NBER WORKING PAPER SERIES

ON INTERNATIONAL COOPERATION

Bård Harstad

Working Paper 33161
http://www.nber.org/papers/w33161

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2024

This is the draft of a chapter that will be forthcoming in Handbook of the Economics of Climate Change, Volume 1 (edited by Barrage and Hsiang), Elsevier. Nothing to disclose. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

On International Cooperation
Bård Harstad
NBER Working Paper No. 33161
November 2024
JEL No. C72, F53, H87, Q2, Q5

ABSTRACT

A simple dynamic game is used for analyzing international environmental problems and climate agreements. Different countries are, over time, emitting as well as investing in green technology. In this framework, we can analyze the business-as-usual outcome, short vs. long term agreements, self-enforcing agreements, participation, compliance, alternative designs, and the development from the Kyoto Protocol to the Paris Agreement. The text should be accessible to students at any level.

Bård Harstad
Stanford University
Graduate School of Business
655 Knight Way
Stanford, CA 94305
and NBER
harstad@stanford.edu

Table of contents

1. Introduction and overview
2. Climate policies as dynamic games
    2.1 Greenhouse Gases and Technology Stocks
    2.2 Equilibrium Concept: SPE vs. MPE?
3. What if there is no cooperation?
    3.1 Example 1: Private Windmills are Public Goods
    3.2 Business As Usual
4. Agreements as incomplete contracts
    4.1 Example 2: The Strategic Choice of Technology
    4.2 Short-Term Agreements
    4.3 Example 3: Strategic Emission Caps
    4.4 Long-Term Agreements
    4.5 Climate Contracts Over Many Periods
5. On the duration of agreements
    5.1 Example 4: The Cost of Commitments
    5.2 The Optimal Commitment Period Length
    5.3 Updating and Renegotiation Design
6. Participation and coalition formation
    6.1 Example 5: The Result " m = 3 "
    6.2 Participation – With Fixed Duration
    6.3 Example 6: Endogenous Duration
    6.4 Participation – With Endogenous Duration
    6.5 Further Reading and Farsightedness
7. Compliance with self-enforcing treaties
    7.1 Example 7: A Repeated Game with Trigger Strategies
    7.2 Compliance Technology and Self-Enforcing Agreements
    7.3 Technologies, Thresholds, and Transfers
    7.3 Technologies, Thresholds, and Transfers
8. Deep-but-narrow vs. broad-but-shallow
    8.1 Example 8: Two Types of Countries
    8.2 Pledge-and-Review Bargaining
    8.3 Modesty May Pay
    8.4 Institutional Design
9. From Kyoto to Paris and beyond

1. INTRODUCTION AND OVERVIEW

Amidst a number of tragic events, I find it to be a most fortunate coincidence that "international cooperation" is not only the most important topic to study, but it is also the most interesting one. All the major challenges humanities are facing require international cooperation. In the very long run, the state of the world will hinge little on today's policies, except for how the countries manage to cooperate on climate change and nature conservation. But international cooperation is complicated, and interesting, because it requires a large number of very different nations to interact over time in a caotic situation with brute force and no world government. To understand international cooperation, we must understand laizzes faire, free riding, exernalities, dynamics, compliance, defections, punishments, negotiations, and renegotiations.

But when it comes to climate agreements, we do not have a lot of empirical observations to draw on. The Kyoto Protocol of 1997 was, essentially, the first global climate agreement and the Paris Agreement, of 2015, is another. Two data points are interesting, but not sufficient to deepen our understanding satisfactorily.

For these reasons, it is necessary to draw on theoretical models. An economic model simplifies the complicated world so that there are just one or a few mechanisms at play. The simplicity implies that the model is inaccurate, but that our thinking will be clear. A theoretical model that we can understand is helpful for our understandings of present policies, our exploration of alternatives, and the communication of our ideas.

This chapter employs simple models to shed light on climate policies without and with international cooperation. I include eight examples, or excercises, that can be solved by your undergraduate students, but the full analysis can be challenging for PhD students. Because while it is desirable to keep the models simple, they must still include a few features that are important when it comes to climate policies. First, the model obviously permits multiple countries. Second, policies must include both emissions, causing the problem, and technology investments, necessary for a solution. Third, the dynamics is essential: Greenhouse gases accumulate over time, the green technologies need time to develop, and one country's action might influence what others do in the future.

The three features are already making the models challenging, so I will abstract from many other important aspects (see below).

The basic framework is introduced in Section 2, and I will rely on it throughout the chapter, but the sections can be read independently of one another. After all, some readers might be interested in compliance and not contracts, for example.

Section 3 discusses dynamic common-pool problems and the so-called "business-as-usual" (BAU) scenario. In my view, it is important to begin with BAU to illustrate that the outcome with no cooperation can be much worse than what we would have thought based on static common-pool problems. In a static model, one country will abate too little, and emit too much, because it does not internalize the externality on the rest of the world. In a dynamic setting, a country might emit even more because the larger stock of greenhouse gases will induce other countries to emit less in the future. Similarly, it will invest less in technology, because by investing less it is expected to emit more, and then other countries will find it optimal to emit less and invest more themselves. The fact that the dynamic common-pool problem is especially costly should motivate the development of an international environmental agreement (IEA).

Nevertheless, IEAs can make things even worse. In Section 4, I refer to traditional findings in contract theory, where it is well-known that parties may under-invest prior to negotiations. The intuition is that if a party has already invested in, say, renewable energy, then this party will find it inexpensive to limit emissions and, in a bargaining game, that party will end up with a smaller emission allowance than parties that have invested less. Intuitively, parties that have invested less will find the IEA costly, and thus "hold up" the negotiations unless they obtain more favorable terms. When the hold-up problem is anticipated, investments fall. If investments fall by a lot, and they are important in order to deal with climate change, the parties can end up being worse off – even compared with BAU. The fact that short-term agreements can be harmful should motivate us to study in more detail how agreements ought to be designed.

Long-term agreements, analyzed next, specifies emission caps for such a long commitment period that the parties have time to invest and develop green technology. There might still be underinvestment problems, either because the agreement is finite while the technology is long-lasting, or because of technological spillovers. To motivate countries to invest more, the emission caps should and will be tighter, so that countries will need to invest in renewables in order to be able to consume satisfactory levels of energy.

The mechanisms just discussed suggest that a long-term agreement, or a long-lasting commitment period, is ideal in order to deal with the hold-up problem. The optimal length, discussed in Section 5, trades off the cost of short-termism with the fact that we do not really know what the caps ought to be several decades from now.

Participation is the topic of Section 6. In contrast to districts within a country, the countries themselves are sovereign and free to decide whether they want to participate or free ride. This choice should be a part of the game, so that we can endogenize the size of the coalition. In much of the literature, the trade-off is as follows. The cost of participating is that one will be end up contributing more, or emitting less, than what one could do as a free rider. The cost of free riding, however, is that the other members of the coalition cannot be influenced, and they cannot be expected to internalize the climate change harm experienced by the free rider. This trade-off implies that the equilibrium

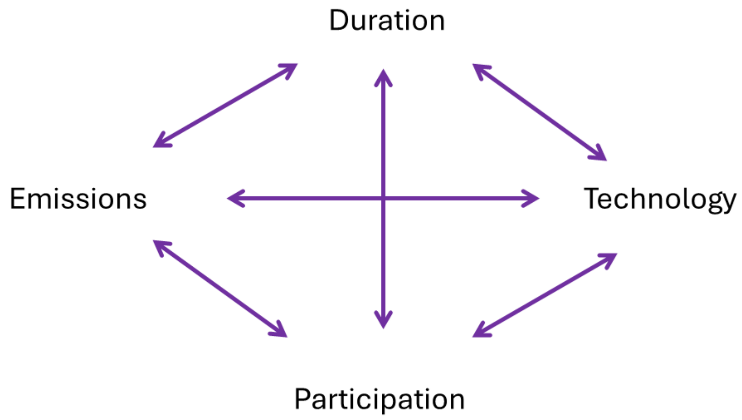

Figure 1: *Section 4 analyzes the two-way interaction between emission caps and investments in technology. Section 5 connects with the commitment period length and Section 6 connects with the coalition size.*

coalition size is relatively small: With linear-quadratic utility functions, for example, at most three countries will participate. When the coalition-formation game is embedded in a dynamic game with both investments and emissions, the results are more nuanced and interesting. In fact, the hold-up problem, reducing everyone's payoff in the earlier sections, can not be taken advantage off, and both the coalition and payoffs can be larger because of it. The intuition for this result is that when a coalition decides on the duration of the agreement, they find it optimal with a long-term agreement iff (i.e., if and only if) the number of members is large. If some countries free ride, however, the coalition might prefer a short-term agreement, hoping that the number of members will be larger in the future. Because the short-term agreements are associated with the hold-up problem, free riding becomes less attractive and, therefore, the coalition can be larger.

Compliance is studied in Section 7. Just as no world government can force a country to participate in an IEA, a participant might not necessarily comply with the pledges or promised that are made. The problem of motivating compliance is related to participation, but there are two major differences. First, while the membership decision is made at the beginning of a game, the decision about whether to comply or defect is made every day. Second, when a country defects, instead of complying, it can reasonably expect that the other members will stick to their promises up to the stage when the defection is observed. For this reason, compliance can be more challenging than participation. To study the limits of what a self-enforcing agreement can do, it is often assumed that the consequences of defection is severe. In many games, the most severe punishment is that everyone returns to BAU as soon as a single defection is observed. From the literature on repeated games, so-called folk theorems tell us that with this type of punishment, the first-best outcome can be sustained as an equilibrium iff the parties' weight on future payoffs (i.e., the discount factor) is sufficiently high. If the discount factor is smaller, the best self-enforcing IEA is inefficient. The inefficiency can be that the emission levels are higher than in the first best (FB), although still smaller than with BAU. This simple insight is derived and presented before we, once again, return to the game in which the parties are, in every period, both investing in technologies and emitting greenhouse gases. In this case, technology will have a strategic role that is different from the other scenaria studied in this chapter. After all, the temptation to defect not only depends on the discount factor, but also on the benefits of emitting. The temptation to emit is weakened if the party can consume energy from renewable energy sources, for example, instead of from fossil fuels. Thus, by investing more (and more than in the FB) in green technology, compliance at the emission stage will be credible. The more difficult it is to motivate countries to comply with the reduced emission levels, the larger are the equilibrium investment levels relative to the first-best levels. If, instead, it is difficult to motivate countries to comply with the high investments, then investments must be reduced relative to the first-best levels. In this case, the best self-enforcing IEA may require countries to be punished if they invest more, and as much as the first-best levels. The intuition for this result is that if countries were allowed to invest more, they would not find it optimal to emit a lot, and thus to punish, if it turns out that another country defects on its pledges.

Section 8 compares deep-but-narrow agreements with those that are broad-but-shallow. The difference can depend on, for instance, the choice between the top-down bargaining procedure associated with the Kyoto Protocol, and the bottom-up pledge-and-review bargaining procedure associated with the Paris Agreement. It is shown that "modesty may pay", as observed by Finus and Maus (2008), in that shallow agreements can attract a larger number of participants. Section 9 discusses how the various models can shed light on the Kyoto Protocol, the Paris Agreement, and the development from one to the other. I conclude by discussing "what's next".

What about everything else? It is reasonable to claim that it is a rather subjective judgement to think that exactly these three features are the most important ones: (1) Multiple countries, (2) green technology, (3) an infinite time horizon. This claim is supported by the fact that much of my own research, which I will draw on below, has investigated these aspects. But, at the least, this focus does make the chapter manageable to write and read.

My focus on feature 2 (technology) and 3 (dynamics) may be what separates this chapter from earlier overviews of international agreements. Those, of course, focus on aspects that I am abstracting from. For earlier overviews, one should consult with Kolstad and Toman (2005) on climate policy, Barrett (2005) on environmental agreements, Calvo and Rubio (2012) on the importance of the dynamics and the stocks, Benchekroun and Long (2012) on the game theory, Caparrós (2016) on the role of bargaining games, Buchholtz and Sandler (2021) for the general public goods aspects, and Bellelli et al. (2023) on the empirical findings.

For other aspects, the readers are anyway better off looking elsewhere. Political economy forces are essential in the real world, of course, but those are emphasized in another book chapter that I am writing (Harstad, 2025). There, I emphasize the importance of heterogeneity, asymmetry, side transfers, elections, and time inconsistency. Thus, I view the two chapters as being complementary to one another. The importance of issue linkages is surveyed by Maggi (2016), disasters by Deryugina (2022), and, in this volume, you can learn more about integrated assessment models (Dietz, 2024), adaptation (Carleton et al., 2024), and domestic climate policies (Kotchen, 2024), for instance. The literate on trade agreements is also related and useful: See Maggi (2014), Grossman (2016), Bagwell and Staiger (2016). The literature on supply-side climate policies is still emerging and must be returned to in another volume.

2. CLIMATE POLICIES AS DYNAMIC GAMES

The motivation to cooperate is that the non-cooperative outcome can be unattractive. Thus, to understand cooperation, we must understand its absence. When every country is free to decide on its policies unilaterally, the countries are playing a game against one another. One policy might influence the future policies of the other countries. When policies are made at multiple points in time, the game is dynamic. Thus, non-cooperative climate policies should be formalized as a dynamic game.

A formalized game will emphasize the strategic nature of the decisions, and it may abstract from a number of behaviors such as altruism and reciprocity preferences, inattention, and irrationality. While these characteristics are extremely important in the daily life of human beings, they are often harder to observe when it comes to hard-core policymaking at the international arena.

Dynamic games can be difficult to solve, however, so we might need to impose several simplifying assumptions. This section introduces a model where both the stock of greenhouse gases and the national stocks of technologies (or capacities to produce renewable energy) develop over time, depending on the choices of the countries. The model will be used throughout the section to illustrate the various aspects that are important to international cooperation.

In the game below, each country is represented by a benevolent decision maker that directly decides on both emission and investments. The idea is that voters might succeed electing a government acting on their behalf, and that this government can introduce national policies that implement the levels that are preferred by the government or the voters. In the chapter by Kotchen (2024), he explains how a government might introduce national climate policies to implement these quantities.

2.1 Greenhouse Gases and Technology Stocks

The set of countries is $N := \{1, ..., n\}$. Each country is treated as one player, and I will assume that every $i \in N$ makes two decisions in every time period $t \in \{1, ..., \infty\}$. At the emission stage, all the i's are simultaneously deciding on the $g_{i,t}$'s, where $g_{i,t}$ measures i's emission of greenhouse gases. These emissions accumulate over time to the stock G_t, measuring the additional stock of CO2, for instance, caused by humans. Assuming that the fraction $q_G \in [0, 1]$ depreciates every period, G_t evolves according to:

$$G_t = q_G G_{t-1} + \sum_{i \in N} g_{i,t}. \tag{1}$$

As a substitute to emitting, and as a possible solution to climate change, let every i invest $r_{i,t}$ in "green technology". The investments accumulate to the stock $R_{i,t}$, which might depreciate by the factor $q_R \in [0, 1]$, so that it evolves as follows:

$$R_{i,t} = q_R R_{i,t-1} + r_{i,t}.$$

One interpretation of $R_{i,t}$ is that it represents i's capacity to abate, so that it emits only $g_{i,t} = y_{i,t} - R_{i,t}$, while still benefiting from $y_{i,t}$. Ploeg and de Zeeuw (1991) present one of the first climate-policy games with both emission and abatement decisions, and they formalize abatement in this way. Another interpretation of $R_{i,t}$, which I will stick with below, is that $R_{i,t}$ measures i's capacity to produce renewable energy, so that i's total consumption of energy, $y_{i,t}$, is such that the following inequality is satisfied:

$$y_{i,t} = g_{i,t} + R_{i,t}. \tag{2}$$

It will be useful to study the decisions on $g_{i,t}$ and $r_{i,t}$ separately, and thus to assume that they are made at different stages in period t. A period is defined so that it starts just before one investment stage and it ends just before the next investment stage. With an additive utility function, i enjoys the following utility at time t:

$$u_{i,t} = B(y_{i,t}) - C(G_t) - K(r_{i,t}, R_{i,t-1}). \tag{3}$$

The benefit of energy consumption is $B(y_{i,t})$, the cost of the greenhouse gas stock is

$C(G_t)$, and the investment cost is given by the function $K(r_{i,t}, R_{i,t-1})$, which might depend on the stock developed so far.

The additive form of the utility function is not strictly necessary. If, instead, the utility was given by the Cobb-Douglas function $\widetilde{B}(y_{i,t})^c \widetilde{C}(G_t)^b \widetilde{K}(r_{i,t}, R_{i,t})^\gamma$, for example, we can use logarithmic maximization so that i would still maximize (3) if we define $B(y_{i,t}) = c \ln \widetilde{B}(y_i)$, $C(G_t) = -b \ln \widetilde{C}(G)$, and $K(r_{i,t}, R_{i,t-1}) = -\gamma \ln \widetilde{K}(r_{i,t}, R_{i,t})$.

At the beginning of a period, i would like to maximize its continuation value

$$v_{i,t} = \sum_{t' \geq t}^{\infty} \delta^{t'-t} u_{i,t} = u_{i,t} + \delta v_{i,t+1},$$

where $\delta \in (0,1)$ is the discount factor. In general, the continuation values will depend on the stocks.

When the parties' actions influence the stocks, the game is called a differential game when time is continuous, and a difference game when time is discrete. These games are generally difficult to solve. Thus, it is common to simplify the functional forms. For example, one may assume that the functional forms are linear-quadratic. Engwerda (2005) provides a comprehensive overview of this literature and its methods. While a quadratic investment cost is introduced from Sections 6, Sections 3-5 assume that the investment cost is simply linear:

$$K(r_{i,t}, R_{i,t-1}) = kr_{i,t}. \tag{LK}$$

Evidently, we will abstract from how a country implements the chosen quantities domestically. Kotchen (2024) discusses alternative instruments. Because the countries are assumed to be identical, we also abstract from the importance of side transfers and international emission permit trading.

2.2 Equilibrium Concept: SPE vs. MPE?

The choice of equilibrium concept is a part of the assumptions, and thus the model. In a static game, it would make sense to characterize all Nash equilibria (NEs). In a dynamic game with a finite time horizon, the typical choice would be subgame-perfect equilibrium (SPE). In a repeated game with infinite number of periods, there are typically a large number of SPEs. It is not clear, then, which SPE that will be most reasonable to focus on. At least two approaches seem to be reasonable.

One interpretations of international cooperation is that the talks are allowing the parties to coordinate on a good SPE. Thus, we may focus on what "best" SPE, given that all quantities must be self-enforcing (as an SPE). This is the approach taken in

Section 7.

Alternatively, we can refine the equilibrium concept. When the game includes stocks that change over time, we have a dynamic game (but not a repeated game). To obtain sharp predictions, it is quite common to restrict attention to Markov-perfect equilibrium (MPE). The MPEs allow parties to form strategies that are contingent on stocks that are payoff relevant, but otherwise the strategies do not depend on the history. In lab experiments, subjects often play MPE strategies when the game is dynamic and complex (Vespa, 2011; Battaglini et al, 2015). Since these equilibria are relatively simple to characterize, we will begin by emphasizing the symmetric MPE from the next section (which turns out to be unique). In addition, Section 3 compares the MPE to the so-called "open loop" equilibrium, where the parties are not updating their actions after observing the actions of the others. This is not a reasonable equilibrium, of course, but by comparing that outcome to the MPE we will be able to understand why the dynamics of the climate policy game makes it especially inefficient.

None of the two approaches sheds light on everything, but both of them can be helpful. The choice will influence the type of treaty that will be considered. As mentioned, when we consider SPEs in Section 7, we study self-enforcing treaties. When treaties are discussed in Section 4-6, where we restrict attention to MPEs, we are explicitly ruling out self-enforcement mechanisms and, therefore, we will simply take as given that countries might be able to commit to certain actions (like short-term emission caps) but not to other actions (like technology investments). These analyses are especially relevant for so-called legally binding treaties, where there might be certain political costs associated with noncompliance.

If the countries are bargaining in a symmetric situation, it will be assumed that the outcome is both efficient and symmetric. This would apply if we employ the Nash Bargaining Solution (NBS), for example.)

3. WHAT IF THERE IS NO COOPERATION?

If there is no agreement, we have "business as usual" (BAU). In this scenario, countries make decisions without internalizing the externality on the others. This implies that countries will emit more than in the FB, and the stock of greenhouse gases will be larger than in the FB outcome.

What about the investments? If the emission levels were fixed, the investment levels would be first best when there are no technological spillover associated with the investments. The reason for this claim is simply that one country's investment level causes no externalities on the other countries: They don't care about the investment, per se.

Nevertheless, in the model of Section 2, countries will end up investing too little with

BAU relative to the FB. The explanation for this finding is that emission levels are not fixed in the dynamic game – they are endogenous. If a country invests more, it will find it optimal to emit less later on. When this country emits less, the marginal cost of emission declines for everyone, and other countries will find it optimal to emit more. The larger emission levels by the others are anticipated by the investing country, who thus reduce the investments, relative to the situation in which this effect were not present.

In effect, one country's technology stock ends up being a public good, which benefits all the countries, even if there are no direct technological spillovers in this model.

This logic explains why the BAU outcome is even worse in this dynamic model than it would have been in a static model. To be precise, suppose every country i decided on the sequence of emission levels and investment levels once and for all, at the beginning of the game. When these decisions are made simultaneously, the game is static, and its Nash equilibrium (NE) is typically referred to as the open-loop equilibrium of the dynamic game. The NE is worse than the FB in that countries emit too much, but investments, conditional on the investment levels, are socially optimal. While this outcome is inefficient, the BAU outcome is even worse. With BAU, a country invests less also to induce other countries to emit less and invest more in the future. Furthermore, a country emits more because that induces the other countries to emit less and invest more in the next period. When all countries are adopting these strategies, they end up with more emission and lower payoffs than with the NE. The BAU outcome is thus especially harmful in a climate policy game – because it is dynamic.

This inefficiency should motivate us to study international cooperation in more detail, which we turn to after this section.

3.1 Example 1: Private Windmills are Public Goods

As an example, consider the model of Section 2 but with only one period, period 1, and with quadratic functions for B and C:

$$B\left(y_i\right) = -\frac{b}{2}\left(\overline{y} - y_i\right)^2 \text{ and } C\left(G\right) = \frac{c}{2}G^2 = \frac{c}{2}\left(q_G G_0 + \sum_{j\in N} g_j - R\right), \qquad \text{(Q)}$$

where $R \equiv \sum_{i\in N} R_i$.

What is the SPE of this game?

Deriving the details are left as an exercise. Clearly, the SPE can be found by backward induction. At the emission stage, $B'\left(y_i\right) = C'\left(G\right)$, so every y_i is the same. The intuition is simply that, on the margin, one more unit of consumption leads to another unit of emission, when the renewable capacity is binding. Thus, R_i is, in effect, a public good, even when there are no technological spillovers.

With some algebra, we find that, at the emission stage,

$$y_i = \frac{\overline{y}b - cq_G G_0 + cR}{b + nc}, \text{ so} \tag{4}$$

$$g_i^{BAU}(\mathbf{R}) = \frac{\overline{y}b - cq_G G_0 + cR}{b + nc} - R_i, \tag{5}$$

where $\mathbf{R}$ is the vector of R_i's. Eq. (5) implies that if R_i is larger, i emits less, but country $j \neq i$ emits more. This is intuitive: A country does not need to emit if it can consume renewable energy, but when such a country emits less, C' declines, and other countries emit more. This is referred to as "carbon leakage." While carbon leakage here is the result of the convex $C(\cdot)$ function, carbon leakage can alternatively result from international trade (as in Kotchen, 2024).

All this is anticipated when i invests. Thus, i finds it optimal to invest until marginal benefits equal the marginal investment cost:

$$B'(y_i)(1 + \partial g_i / \partial R_i) - C'(G)\left(\sum_{j \in N} \partial g_j / \partial R_i\right) = k. \tag{6}$$

With (5), and some algebra, we get R:

$$R^{BAU} = n\overline{y} + q_G G_0 - \frac{(b + cn)^2}{cb(b + c)} k.$$

We refer to the above as BAU. The FB, in contrast, is given by the same equations if just c is replaced with nc. For example:

$$g_i^{FB}(\mathbf{R}) = \frac{\overline{y}b - ncq_G G_0 + ncR}{b + n^2 c} - R_i.$$

By comparison, emissions are larger with BAU than in the FB, and investments are lower. The reason is that j benefits if i invests more because with a larger R_i, i will emit less.

Another interesting comparison is with the Nash equilibrium (NE) of the static version of the game. If decisions were made simultaneously, j would *not* emit more if i changed its decisions. So, $\partial g_j / \partial R_i = 0$, and investments would be given by $B'(y_{i,t}) = k$, which equals the socially optimal level, conditional on the level of technologies. Because i's investment does not lead j to emit more, i would find it optimal to invest more in the NE than it does with BAU. Regarding emission levels, (4)-(5) would continue to hold with the NE. By comparing all these conditions, we can conclude:

$$\begin{aligned}
&g_i^{BAU} > g_i^{NE} > g_i^{FB}, \\
&R^{BAU} < R^{NE} < R^{FB}, \text{ so} \\
&u_i^{BAU} < u_i^{NE} < u_i^{FB}, \forall i \in N.
\end{aligned} \tag{7}$$

These inequalities tell us something important. With the NE, investment levels will be just right, conditional on those emission levels. But the NE is inefficient because emission levels will be too high: Country i will not take into account the externality on j, so the level of investments is smaller than in the FB. In BAU, investments are even lower than in the NE, because by investing less, other countries will emit less. When everyone invests less, total emission ends up being higher, and payoffs larger, than in the NE. Thus, the fact that the game is dynamic, and decisions are made sequentially, make the outcome even worse than it would be in the NE outcome, and the NE outcome is, in itself, also inefficient.

3.2 Business As Usual

The basic insights from Example 1 survive in more general settings: all inequalities in (7) will hold, even when there is an infinite number of periods, and even if $B(\cdot)$ and $C(\cdot)$ are general concave and convex functions, respectively (for details, see the theorem in Harstad, 2012:1534).

With the linear investment cost, (LK), introduced in Section 2, the analysis simplifies by a lot, even when we permit an infinite number of periods, and it can be shown that the continuation value will be linear in all the stocks. For this reason, the model is tractable, easy to solve, and there is a unique MPE.

Because $C(\cdot)$ is convex, it is more costly to emit when the stock G_t is large. Thus, every country emits less at time $t+1$ when G_t ends up being large. A country understands this, and thus that if i increases $g_{i,t}$, then, in equilibrium,

$$\frac{\partial g_{j,t+1}}{\partial g_{i,t}} < 0 \text{ and } \frac{\partial r_{j,t+1}}{\partial g_{i,t}} > 0 \ \forall j \in N. \tag{8}$$

Here, the second inequality says that the other countries will also invest more, if i emits more. This is natural, because when j finds it optimal to consume less energy from fossil fuels, it is beneficial to invest more in renewable energy, instead.

From this logic, it follows that if country i decides to invest more, then it will find it optimal to emit less and, because of that, other countries will emit more and invest less in the next period. Formally:

$$\frac{\partial g_{j,t+1}}{\partial r_{i,t}} > 0 \text{ and } \frac{\partial r_{j,t+1}}{\partial r_{i,t}} < 0. \tag{9}$$

All these effects imply that other countries will take advantage of i's "good" behavior. Because i prefers that another country, j, will emit less and invest more, i has an incentive to emit more and invest less than if j could not respond according to (8)-(9).

The situation in which j cannot respond to i's policies is interesting in itself. Because dynamic games are complicated to solve, some scholars simplify by assuming that every party commits to a sequence of decisions at the very beginning of the game. When these commitments are made simultaneously, the game is essentially static, and the Nash equilibrium of the static game corresponds to the so-called open-loop equilibrium of the dynamic game. In this NE, j's decisions will not change if i changes its policy, so (8) and (9) will not apply. Therefore, i finds it optimal to emit less, and invest more.

The NE is not first best, of course: Country i will not take into account the externality on j, when i decides on how much to emit. Because i finds it optimal to emit less if $R_{i,t}$ is large, j would have benefited if i invested more, so i's equilibrium investment level is smaller than the first-best level. To summarize:

The BAU outcome of the dynamic game is worse than the NE outcome of the static game.

In other words, while the static common-pool problem, characterized by the NE, is inefficient and provides lower utilities than in the first best, the BAU of the dynamic climate-change game is even worse.

The fact that dynamic common-pool problems are especially inefficient has been recognized a long time. Levhari and Mirman (1980) developed the insight in their study of fish wars, and Ploeg and de Zeeuw (1991, 1992) showed that the MPE outcome is worse than the NE in climate policy games. Although Dockner and Long (1993) show that MPEs in nonlinear strategies can be better, the simplest MPE, in linear strategies, is highly inefficient.

These inefficiencies should motivate us to search for agreements that can improve on the outcome.

4. AGREEMENTS AS INCOMPLETE CONTRACTS

Because the BAU outcome is so inefficient, the countries may be quite motivated to sign an agreement.

As mentioned in Section 2, when it comes to international agreements, one sometimes distinguish between (legally) binding agreements and self-enforcing agreements. If an agreement is binding, a stakeholder can hold a government accountable if it does not comply. Now, it is not clear which stakeholders that will have this capacity, and it is also unclear what it means to "hold a government accountable": Both aspects may depend on the country in question. In the U.S., for example, ratifying an international treaty implies that it becomes law, so that a state, such as California, can sue the federal government if it does not comply. More generally, IPCC (2014, p. 1020) writes that "a more legally binding commitment ... signals a greater seriousness by states ... These factors increase the costs of violation (through enforcement and sanctions at international and domestic scales, the loss of mutual cooperation by others, and the loss of reputation and credibility in future negotiations)."

With this motivation, we start by considering legally binding agreements, where we abstract from the temptation to defect, but we return to that temptation, and self-enforcing agreements, in Section 7.

Contract theory is a branch of economic theory where binding agreements are analyzed. (For textbook treatments, see Bolton and Dewatripont, 2005.) Two types of contracts can be relevant for climate change agreements. A "complete contract" is an agreement where the parties commit to all variables of interest. In the model of Section 2, the variables include every emission level and every investment level. With this possibility, the countries can and will agree on the first-best outcome, because they are symmetric, by assumption.

A complete contract is not realistic, of course. In the real world, there are a large number of political variables that influence the climate change outcome, and there is no hope that an international agreement can specify everything that is potentially relevant. The Kyoto Protocol, for example, emphasized emission levels, but left the countries' investment levels to be nationally determined. This approach has been confirmed by later agreements. According to article 114 of the 2010 Cancun Agreement, which was confirmed in Durban in 2011, "technology needs must be nationally determined, based on national circumstance and priorities."

There might be good reasons for why investment levels are left out. As argued by Golombek and Hoel (2006, p. 2), "it would hardly be feasible for a country (or some international agency) to verify all aspects of other countries' R&D policies".

Thus, a more realistic concept is "incomplete contract", where it is recognized that certain variables remain outside of the agreement. The literature on incomplete contracts literature was originally developed to study interaction between firms (see Hart, 1995). The literature recognizes that the variables that are not contracted on will be strategically chosen so as to influence the next round of (re)negotiations, but also that these variables may be influenced by the decisions that are committed to with the contract. In the climate policy game, this implies that there is a two-way interaction between the investment levels and the emission caps.

On the one hand, the agreed-upon emission levels are not exogenous, but the outcome of bargaining. At the time when the emission caps are negotiated, the countries that are ill-prepared for a treaty can "hold up" the countries that have already invested in green technology, and the laggards can demand to be granted larger emission quotas than the leaders who have invested more. When this hold-up problem is anticipated, countries might be reluctant to invest. Thus, investments levels might be low, in equilibrium, when it is expected that the more one invest, then less one will be able to emit. If the investments are socially important, the hold-up problem associated with so-called short-term agreements can make them worse than BAU, it is argued in Section 4.1. The idea that technology investments can fall before climate negotiations go back to Buchholtz and Konrad (1994)

There is evidence that the hold-up problem is important in reality. For example, The New York Times (17 October 2008, p. A4) reported that "Leaders of countries that want concessions say that nations like Denmark have a built-in advantage because they already depend more heavily on renewable energy."

On the other hand, because investing in renewables is a substitute to emission, a country will find it optimal to invest more if the emission cap is tight. By reducing the caps, the countries will be induced to invest more. Several papers have documented that regulation can motivate technological change in this way (Jaffe et al., 2003; Newell et al., 2006; Dugoua, 2023). This effect is likely to be important in so-called long-term agreements, where countries do have time to develop new technology between the negotiation stage and the emission stage.

By combining the two forces, we can arrive at the following insight. If an incomplete contract specifies emission levels for a large number of future periods, then countries will be less worried about the hold-up problem that might arise sometime in the future, and then investments will be first best iff the emission caps are first best. (After all, we have assumed away direct technological spillovers on the other countries.) When the current commitment period is about to end, and one thinks about the next round of bargaining, then the hold-up problem will begin to discourage investments in technology. To maintain the incentive to invest, also in this situation, the emission caps must be tighter, and more demanding. Consequently, the shorter is the duration of the commitment period, or the shorter it is until it expires, the more ambitious must the abatement levels be for investments to remain at the efficient level: See Sections 4.2 and 4.3.

4.1 Example 2: The Strategic Choice of Technology

Buchholtz and Konrad (1994) showed that technology investments might be strategically small, both in a situation without negotiations, and also if the negotiations are coming up. The title of this subsection is named after their article, although the formalization of the model is as in Harstad (2016).

To capture the basic insight, consider the one-period Example 1 in Section 3.1, but suppose the timing of a period is as given in Figure 2, and that the g_i's are determined by the NBS. With this, the SPE of the game can be determined by backward induction. The details are left as an exercise.

As discussed after eq. (Q), differences in the R_i's are irrelevant at the emission stage. And when the technology stocks are given, the countries are, in effect, negotiating the y_i's. Countries have identical benefits and costs when it comes to the y_i's, and the threat point, given by y_i^{BAU}, is also the same for every i. With the NBS, therefore, all the y_i's will be the same, and they will be efficient, conditional on the technology level: $g_i^{ST}\left(\mathbf{R}^{ST}\right) = g_i^{FB}\left(\mathbf{R}^{ST}\right)$. The investment levels are non-cooperative, and given by (6), as with BAU. However, in this equation, G will be different because of the negotiations: $g_i^{ST}(\mathbf{R}) < g_i^{BAU}(\mathbf{R})$. Thus, the marginal cost, C', will also be smaller, and the benefit from emitting even less, by investing, is reduced. Hence, equilibrium R is smaller when short-term agreements are expected, than with BAU: $\mathbf{R}^{ST} < \mathbf{R}^{BAU}$. The reduced technology level implies $g_i^{ST} < g_i^{FB}$, and that payoffs can be smaller too. With some algebra, we find:

$$
\begin{aligned}
& g_i^{ST} < g_i^{BAU}, \\
& R^{ST} < R^{BAU} < R^{NR} < R^{FB}, \\
& u_i^{ST} < u_i^{BAU} < u_i^{NE} < u_i^{FB} \textit{ iff } k > n/\left(n-1\right).
\end{aligned}
$$

4.2 Short-Term Agreements

Even if there is an infinite number of periods, it turns out that all the inequalities from Example 2 hold in this model, except that the condition is weakened as follows. (See Harstad, 2016, for details.)

$$
u_{i,t}^{ST} < u_{i,t}^{BAU} \textit{ iff } k > \left(1-\delta q_R\right) n/\left(n-1\right). \tag{10}
$$

Why can the countries be worse off when they negotiate emission caps, compared to with BAU?

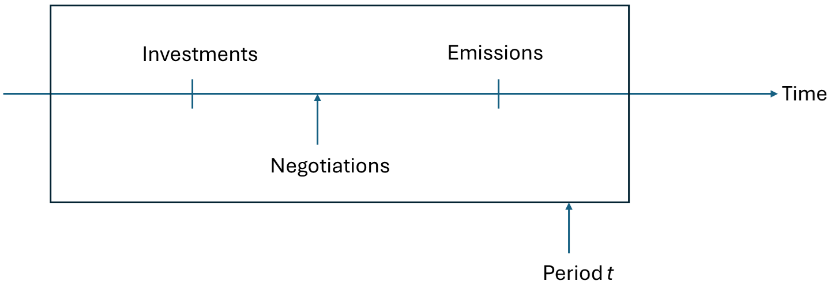

Figure 2: *With short-term agreements, the relevant technology has already been invested in.*

To explain the intuition, remember that differences in the technologies (i.e., the $R_{i,t}$'s) turn out to be irrelevant when it comes to the benefits and costs of consumption levels (the $y_{i,t}$'s). These differences are also irrelevant in the BAU scenario, as observed after eq. (Q), and BAU serves as the threat point when the countries negotiate and decide on the quantities or, equivalently, the $y_{i,t}$'s. Thus, the countries will sign an agreement which makes every $y_{i,t}$ the same, no matter differences in the technology stocks. While it is simple to show this mathematically, the intuition can be related to the hold-up problem: If a country invests a lot, then other countries, who invest less, can "hold up" the technology leader and demand that it takes on the lion's share of the total emission cuts. A country with a small $R_{i,t}$, in contrast, will find it costly to reduce $g_{i,t}$, and it can demand a larger $g_{i,t}$ in return for agreeing.

When this hold-up problem is anticipated at the investment stage, a country realizes that it will not be able to capture the full surplus of its investments. Thus, it invests less than the socially optimal level.

The investments before upcoming negotiations can be lower than with BAU.

The country also invested less than the socially optimal level in the BAU scenario. But the two situations are not identical. When party i expects that the countries will negotiate the emission levels, and thus to solve the climate change problem, to some extent, then it expects that the marginal cost, $C'(G_t)$, will be reduced. The reduced $C'(G_t)$ implies that it is less important that i invests in green technology as a way to solve the climate-change problem. The countries will, collectively, solve the problem anyway. For this reason, i might end up investing less when i expects that the countries will negotiate the $g_{i,t}$'s than i would have done in the BAU scenario.

The smaller investments are harmful to everyone, of course, because, even with BAU, investments are lower than the first-best levels. When the investments are further reduced, because of the reduced $C'(G_t)$, it is possible that the anticipation of the short-term agreements make everyone worse off, compared to BAU, because the cost of reduced investments might outweigh the benefit from reduced emission levels. As inequality (10) states, this will be the case when, for example, n is large, k is large, and q_R is large.

The payoffs with short-term agreements can be lower than with BAU.

The intuition for why the inequality is weakened to (10) with many periods, so that it is more likely that $u_{i,t}^{ST} < u_{i,t}^{BAU}$, is that when some of the technology survives to another period ($q_R > 0$), then the under-investment problem, associated with the short-term agreements, is even costlier than in the one-period model (especially when the weight on the future, δ, is large).

Condition (10) is weakened further, so that it is more likely that $u_{i,t}^{ST} < u_{i,t}^{BAU}$, if there are technological spillovers associated with the investments. Thus, if intellectual property rights are weak or poorly enforced at the international level, it is more plausible that short-term agreements make everyone worse off, relative with BAU, because countries will be further discouraged from investing when they anticipate future negotiations.

It is worth returning to the possibility that payoffs are lower with agreements than without. In Section 2, we already concluded that the payoff in the BAU scenario was lower than in the static common-pool situation, which in turn was lower than the first-best payoffs. When payoffs in BAU are so low, it is worrisome that payoffs can be even lower when the countries are doing their best, negotiating caps on the emission levels. In this situation, the countries would have been better off if they committed to remain with BAU, before the investment decisions were made. After the investments are sunk, however, they all benefit from negotiating the emission levels. Thus, the countries are facing a time inconsistency problem, where they will find it optimal to negotiate later in the game, even though they could have been better off if they committed to not negotiate short-term agreements.

The details of this analysis are available in Harstad (2016), but, as mentioned, the idea goes back to Buchholtz and Konrad (1994). In their model, the effect of the technology is multiplicative, and it reduces the per-unit cost of abatement effort. In the above model, in contrast, the effect of the technology is additive: see eq. (2). Furthermore, Buchholtz and Konrad permit a quite general utility function, although they consider only one period. Beccherle and Tirole (2011) also discuss the hold-up problem that arises when there is a delay in reaching an international agreement. Despite the differences between these models, the under-investment problem is present in all of them. This suggests that the result is quite robust.

The possibility that short-term agreements can be worse than BAU should motivate us to continue the search for a better design.

4.3 Example 3: Strategic Emission Caps

To introduce the benefits of a long-term agreement, return to Example 1 and 2, with only one period and quadratic $B(\cdot)$ and $C(\cdot)$. Assume, however, that the parties decide on the $g_{i,t}$'s at the beginning, and that the bargaining outcome is characterized by the NBS. What are the equilibrium emission caps and investment levels in this game?

The details are left as an exercise, but, of course, the SPE can be derived by backward induction. After g_i has been determined, country i can expect to benefit from consuming $g_i + R_i$ by investing. Thus, i invests as much as:

$$B'(g_i + q_R R_{i,0} + r_i) = k \Leftrightarrow$$
$$r_i = \overline{y} - k/b - g_i - q_R R_{i,0}.$$

The simple equation has three implications:

(1) Country i invests less if $g_{i,t}$ is large. Conversely, if the emission caps are tight, so that the $g_{i,t}$'s are small, then the parties find it optimal to invest more. In this way, the incomplete contract, which only specifies $g_{i,t}$, induce the parties to invest.

(2) Country i invests the socially optimal amount, conditional on $g_{i,t}$, no matter what $g_{i,t}$ is. This follows simply because there are, by assumption, no technological spillovers. Thus, as soon as the countries have pinned down the $g_{i,t}$'s, there is no need to negotiate the $r_{i,t}$'s, because countries will make those choices efficiently.

(3) Country i invests less if i starts out with a larger technology stock, so that $q_R R_{i,t-1}$ is large. At the negotiation stage, the countries anticipate all this. They know that every country will end up consuming $y_{i,t} = \overline{y} - k/b$ by investing $r_{i,t}$, that $g_{i,t}$ pins down r_i, and thus that negotiating the g_i's is equivalent to negotiating the induced r_i's. With the NBS, they agree on g_i's that are such that their costs, and thus the r_i's, will be the same. With simple algebra, we find:

$$g_i = (k - cqG_0)/nc + (R_0/n - R_{i,0})\, q_R/n.$$

With this, the first-best outcome is implemented. The g_i's are first best, by negotiation, and the r_i's are first best, because there are no spillovers associated with them.

4.4 Long-Term Agreements

With short-term agreements, the time distance between the negotiations and the compliance stage is too short for new technology to be developed. For a long-term agreement, instead, there will be time to develop new technology. The simplest way of capturing this situation is to assume that the emission caps are decided on at the very beginning of the period. The choice of emission caps will influence how much the countries will find it worthwhile to invest, and this influence can be taken into account by the countries when they decide on the emission caps.

Unfortunately, the efficient outcome in Example 3 does not survive when there is an infinite number of periods if each contract lasts only one period. The reason is, as in Section 4.1, that investments are inefficiently low when some of the technology survive to the next bargaining stage. To motivate countries to invest, even in this situation, it will be optimal to reduce the emission caps, below the ex post optimal level, because tight caps will motivate countries to invest more. (This is in line with point (1) in Example 3.)

Formally, suppose the parties fix the $g_{i,t}$'s at the beginning of the period. Then, i can expect to consume one more unit of energy at time t by investing one more unit in technology. Furthermore, next period starts with q_R more units of technology. If the future $g_{i,t+1}$'s are not yet decided on, however, $R_{i,t}$ is, essentially, a public good, and all the n countries will invest less, as a consequence (just as in the previous subsection). When this is anticipated, i invests as much as:

$$B'\left(g_{i,t}+R_{i,t}\right)=\left(1-\delta q_R/n\right)k \Leftrightarrow r_{i,t}=B'^{-1}\left(1-\delta q_R/n\right)-g_{i,t}-q_R R_{i,t-1}.$$

Because the first-best investment level is

$$r_{i,t}^{FB}=B'^{-1}\left(1-\delta q_R\right)-g_{i,t}-q_R R_{i,t-1},$$

we can make a few observations:

(1) As in Example 3, the country invests more if its emission allowance ($g_{i,t}$) is small. By playing around with the $g_{i,t}$'s, we can motivate i to invest any level that we want.

(2) If $\delta q_R > 0$, and $n > 1$, i invests less than the socially optimal level. The reason is that the other countries benefit when period $t+1$ starts with a larger $R_{i,t+1}$, because it will then be expected that this country will emit less.

Trade-off. Combining the two points, we realize that the parties will benefit from making the following trade-off. By reducing $g_{i,t}$ below the level that is ex post first best, the countries experience an inefficiency regarding the suboptimally low emission levels, but countries are induced to invest more, and that is beneficial in a situation in which the countries invest too little because of the hold-up problem. The optimal $g_{i,t}$'s will thus be smaller than the levels that are socially optimal, given the technology levels, because of the need to motivate additional investments. It can be shown that:

$$g_{i,t}\left(\mathbf{R}_t^{LT}\right)<g_{i,t}^{FB}\left(\mathbf{R}_t^{LT}\right),$$

where $\mathbf{R}_t^{LT}$ is the equilibrium technology vector under these types of long-term agreements. (See Harstad, 2012, for details.)

We can do comparative static w.r.t. the length of the period. Suppose the *annual* depreciation rate is q_a, the annual discount rate is δ_a, and the number of years in a period is l. Then, $q_R = q_a^l$ and $\delta = \delta_a^l$, which both decrease in l. Thus, a more long-lasting agreement, where l is large, will motivate countries to make investments that are closer to the first-best levels, and thus it is optimal to negotiate emission caps that are close to the socially optimal level, as well. If, however, l is small, so that the agreement has a shorter duration, then δq_R is smaller, the investments are less than the first-best level, and to encourage more investments it is optimal to reduce the emission allowances relative to the level that is ex-post first best (given the equilibrium investment levels). In other words, an IEA that has a shorter duration, or commitment period, should be more ambitious and demanding, so that it will still motivate the parties to invest.

The shorter the duration, the deeper the cuts must be.

4.5 Climate Contracts Over Many Periods

The situations above are quite stylized, of course. In reality, countries can negotiate emission caps for more than one period at the time. By allowing for this reasonable scenario, we will see how the insight from Sections 4.1 and 4.2 can be combined.

Suppose that the countries pin down the $g_{i,t}$'s for $t \in \{1, ..., T\}$ at the beginning of period 1. Consider a period $t < T$. Party i knows that if i invests one unit more at time t, and q_R units less at time $t+1$, R_{t+1}, as well as the rest of the game, will remain unchanged. Thus, the cost of raising $R_{i,t}$ by one unit, in this way, is $(1 - \delta q_R) k$, so i invests to that:

$$B' (g_{i,t} + q_R R_{i,t-1} + r_{i,t}) = (1 - \delta q_R) k \Leftrightarrow$$
$$r_{i,t} = B'^{-1} ((1 - \delta q_R) k) - g_{i,t} - q_R R_{i,t-1}.$$

Because there are no technological spillovers, i's investment is first best, conditional on $g_{i,t}$, regardless of what the $g_{i,t}$'s will be. Thus, the countries will negotiate and set the $g_{i,t}$'s equal to the optimal levels, and rest assured that the countries will invest the socially optimal amount. With this, the incomplete contract implements the first best outcome, just as in Example 3, even if countries decide on the investments independently.

This efficiency result is no longer true when we arrive at time T, however. Because the emission caps for the next period are not yet negotiated, the investments are as in the case with the one-period contract, and, thus, the optimal emission caps are less than the ex post socially optimal levels (conditional on the technology) in order to motivate countries to invest more. Consequently, the caps will be more demanding to comply with at the end.

With some algebra, we can compare the outcome with long-term (LT) agreements with the FB:

$$\text{For } t < T : g_{i,t}^{LT}\left(\mathbf{R}_t^{LT}\right) = g_{i,t}^{FB}\left(\mathbf{R}_t^{LT}\right), \text{ and } R_t^{LT} = R_t^{FB}.$$
$$\text{For } t = T : g_{i,t}^{LT}\left(\mathbf{R}_t^{LT}\right) < g_{i,t}^{FB}\left(\mathbf{R}_t^{LT}\right), \text{ and } R_t^{LT} < R_t^{FB}.$$

Technological spillovers. Based on the discussion so far, it may not be difficult to guess the consequences of introducing technological spillovers. If the spillovers are positive and large, countries are investing too little both because of the spillover, and because of the hold-up problem before new commitments are negotiated. The lower are the equilibrium investment levels relative to the socially optimal investment levels, the smaller should the $g_{i,t}$'s be, relative to the socially optimal $g_{i,t}$'s (conditional on the technology), because the tight emission caps will motivate the countries to invest more and thus closer to the socially optimal levels. In other words, when there are large technological spillovers, the caps should and will be more tighter so that they induce the countries to invest more. Technological spillovers are permitted in Harstad (2016) but the idea goes back to Golombek and Hoel (2005) who developed this insight in a finite game.

5. ON THE DURATION OF AGREEMENTS

An important aspect of an IEA is the commitment period length. This aspect defines the years for which the commitments are relevant, before they must be negotiated, once again. The first commitment period for the Kyoto Protocol was five years, 2008-2012. The second commitment period was eight years, 2013-2020. As observed by Bodansky et al. (2017, p. 203): "Parties disagreed on several issues including: the length of the commitment period—whether it should be five years (like the first commitment period) or eight years (to coincide with the scheduled launch of the 2015 agreement)." He continues: "the eight-year duration of the second commitment period was chosen so as to end when the Paris Agreement's NDCs were expected to take effect, and thus to avoid a commitment gap" (p. 205). The Paris Agreement has returned to the original commitment period length, and stipulates NDCs that are to be updated every five years.

There might be several important trade-offs involved when the optimal length is determined, but, unfortunately, we do not have a large literature on the optimal duration of a commitment period.

The title of this section is borrowed from Harris and Holmstrom (1987). They studied the optimal contract length in a situation in which they traded off the cost of uncertainty with a fixed cost of rewriting the contract. The larger is the uncertainty, the shorter is the optimal contract length. The larger is the cost of rewriting the agreement, or the cost of acquiring information, the longer is the optimal length. The authors did not have climate agreements in mind when analyzing their model, but their insight is relevant also in our context.

The "fixed" cost of negotiating new commitments might be related to the hold-up problem when it comes to international climate policies – as discussed above. That discussion emphasized the benefit of long-term commitments, because the parties will underinvest, due to the hold-up problem, when they approach a new round of negotiations. The benefit of increasing the commitment period from T to $T+1$ is that the hold-up problem is postponed. In Harris and Holmstrom (1987), the benefit is that the fixed cost can be postponed, but the shocks accumulate over time, so the uncertainty about the state of the world for time $T+1$ is greater than the uncertainty about the state for time T.

Uncertainty is crucial also when it comes to climate agreements. Thus, there are good reasons for why the commitment period is finite in reality. We do not really know what the emission caps ought to be 500 years from now. It would be rather stupid to tie our hands to certain actions so far in advance, because the world is more uncertain than what we have assumed in the simple model above. One uncertainty regards technology: If we ends up being positively surprised by technological breakthroughs, we might want to reduce the emission caps. Another uncertainty regards the cost of climate change. If the cost turns out to be larger than what we expect, the emission caps should be tightened. Either of these uncertainties is sufficient for the infinite-length conjecture to fail. The larger is the uncertainty, the smaller is the optimal T.

In our context, it is natural to assume that we are uncertain about the future amount of greenhouse gases and, thus, the marginal cost of emitting. This can be formalized by assuming:

$$G_t = q_G G_{t-1} + \sum_{i \in N} g_{i,t} + \theta_t. \tag{11}$$

The shock θ_t may be arbitrarily distributed with mean 0 and variance σ^2. The shock can reflect Nature's influence on G_t, and thus on the marginal cost of emission, $C'(G_t)$.

5.1 Example 4: The Cost of Commitments

Consider the model and timing of Example 3, with only one period, period $t=1$, and with (11). What are the optimal g_i's if θ_t is known from the beginning, if it is not known, and what is the cost of deciding on the g_i's before learning θ_t?

The details are left as an exercise, but if θ_t is large, it will naturally be optimal to emit less, and invest more. Thanks to the linear investment cost, there is no cost associated with the stochastic θ_t, if just the realization is known before emission caps and investment levels are determined. If θ_t is not know, however, the additional cost will be $c\sigma^2/2$. Thus, $c\sigma^2/2$ represents the cost of committing to the g_i's before θ_t is realized.

5.2 The Optimal Commitment Period Length

Although the shocks are i.i.d. across periods, they have long-lasting impacts through their effect on G_t. Over time, the accumulated effects of the shocks can have large impacts on the marginal cost of emitting.

If the θ_t's are verifiable, so that the IEA can make the emission caps contingent on them, then the uncertainty does not matter and it remains optimal to agree on an infinite long commitment period, where the emission caps are functions of the history of shocks.

If the θ_t's are observable, but not verifiable, then it is difficult to write an IEA in which the caps are functions of the shocks. In this case, it might be simpler to negotiate new caps, once we have learned about the shocks and thus the cost of emitting. We are then left with a trade-off: It the length of the commitment period, T, is small, we benefit because we can update the caps once we have learned the shocks. By selecting a large T, however, incentives to invest will be larger. The optimal T, denoted T^*, trades off the two benefits. If the hold-up problem is more severe, because q_R is large (or because there are technological spillovers, making larger investments socially desirable), then T^* is larger. If the variance σ^2 is large, T^* is smaller.

With (LK) and (Q), we can calculate both the cost of the hold-up problem and also the cost of $g_{i,t}$'s that aren't reflecting the state of the world. It can then be shown that T^* is finite iff:

$$\sigma^2 > (q_R k)^2 \cdot \left[\delta\left(1-q_G^2\right)\left(1-\delta q_G^2\right)(1-1/n)^2 / bcq_G^2\right],$$

and, under this condition, T^* is smaller if the variance σ^2 is large, but T^* is larger if the hold-up problem is very costly (e.g., if $q_R k$ is large). The proof also allows for technological spillovers and show that, when these are larger, the hold-up problem is more costly, and T^* is larger. This comparative static implies that if the world suffers from weak protection of intellectual property rights (because the TRIPS agreement is not enforced, for example), then the optimal IEA should not only be more ambitious, but the commitment period should also be longer.

T^* is reduced when the uncertainty is large and the hold-up problem inexpensive.

While the proof of the above equation is in Harstad (2016), Schmidt and Strausz (2015) derive related results in a setting where two countries are asymmetric.

There are other concerns that determine T^*, of course. For example, if it turns out that many countries are free riding, then it might be wise to let the agreement expire relatively soon, so that it's easier to negotiate with a larger set of countries. We will return to this possibility in Section 6, where the possibility to free ride, instead of participating, is introduced and analyzed.

5.3 Updating and Renegotiation Design

Letting the commitments expire is actually not the best way of allowing the parties to update the agreement. Alternatively, the commitments can stay in place, as the default or threat point when the parties negotiate new commitments. In this way, the parties will take new information into account, but they will continue to have incentives to invest in new technology, because of this default outcome. The lower are the emission caps in the default, the larger are the incentives to invest, because being comfortable with the default is what gives a party bargaining power.

Chung (1991) and Aghion et al. (1994) have shown that "renegotiation design" can be beneficial when contracts are incomplete. The idea is that the terms will anyway be optimal, ex post, when the parties renegotiate. Thus, the initial, default agreement can be used to provide incentives to invest.

Edlin and Reichelstein (1996) showed that the first-best is achieved by initially contracting on the expected optimal quantity and thereafter allowing for renegotiation. Segal and Whinston (2002) introduced technological spillovers and found that the larger the spillover associated with the investments, the larger is the optimal default quantity, compared with the expected ex post first-best quantity. To ensure ex post efficiency, it is then necessary to renegotiate, also on the equilibrium path.

Applied to our dynamic climate policy game, it would work as follows. Suppose the agreement is only one period, as in Section 4.1. The countries start by agreeing on low default emission levels, but, after the investment stage, they renegotiate and permit the caps to be larger (and equal to the ex post optimal levels). The default levels should be low because, then, the countries that do not invest will be uncomfortable with the default outcome, they will desperately need to renegotiate, and this gives them a poor bargaining position, with an unattractive outcome. To obtain a better bargaining position, and a better outcome, a country will be incentivized to invest more if the default caps are low.

It is useful to return to the trade-off discussed in Section 4.2, and in the final period in Section 4.3. There, the emission caps traded off the benefits of larger investments, when g_i^{LT} was reduced, with the cost that the caps would be inefficiently small, ex post.

When this cost vanishes, thanks to the renegotiation, then the "default caps", agreed to initially, can be further reduced. The optimal emission caps in the default (DE), and also the equilibrium levels, are characterized by tight caps, so that they motivate countries to invest the optimal levels. When $\delta q_R > 0$, then:

$$g_{i,t}^{DE}\left(\mathbf{R}_t^{LT}\right) < g_{i,t}^{LT}\left(\mathbf{R}_t^{LT}\right) < g_{i,t}^{FB}\left(\mathbf{R}_t^{LT}\right), \text{ and } R_t^{DE} = R_t^{FB}.$$

In fact, any level of investment can be induced by selecting the appropriate default caps. If, for example, there are technological spillovers, and these are large, then it is beneficial to motivate more investments, and the default caps should be even smaller (see Harstad, 2012, for details).

By observing the procedure empirically, one might mistake it for being a time inconsistency problem. In every period, the countries pledge to limit emission by a lot but, later on, they renegotiate and allow themselves to emit more. Although this procedure is, empirically, similar to a time inconsistency problem, it can implement the first best.

6. PARTICIPATION AND COALITION FORMATION

There is no world government. Every country is free to decide whether or not to participate in an international agreement. Freedom to choose is generally a good thing, but this freedom also implies that it can be challenging to motivate appropriate actions when there are international externalities.

In reality, the level of participation varies quite a lot between agreements. The Kyoto Protocol, of 1997, distinguished between Annex I and Annex II countries, where only Annex II countries were expected to commit to reductions in their emission levels. In the end, only 37 countries committed to reduce emissions and one of them, Canada, withdrew in 2011. The Paris Agreement, in contrast, includes nearly every country in the world. The level of participation also varies for other types of international agreements, ranging from defence to international trade.

There is a large literature in economics that endogenizes the size and composition of coalitions. Much of this literature draws on how collusions among subset of firms have been modelled in the literature on industrial organization. Seminal papers include those of D'Aspermont et al. (1983) and Palfrey and Rosenthal (1984). A major result in this literature is that equilibrium agreements are predicted to include a very small number of countries (Hoel 1992, Carraro and Siniscalco 1993, Carraro et al. 2006, Barrett 1994, Dixit and Olson 2000). With linear-quadratic utility functions, the coalition size is, at most, three! The contrast between the theoretical prediction and actual coalitions, which are often much larger, has been referred to as the "Paradox of International Agreements" (Kolstad and Toman, 2005).

In this section, I first illustrate the small-coalition result based on a simple two-stage participation game that is often used in the literature. The game is as follows:

(1) Every $i \in N$ decides whether to free ride or be a member.

(2) (a) The set of members, $M \subseteq N$, decides the terms of the agreements while (b) the free riders act noncooperatively.

When the members have symmetric preferences, it is typically, and reasonably, assumed that they select contributions to maximize the sum of the members' payoffs.

In this game, M is an equilibrium coalition if there is a subgame-perfect equilibrium in which (i) every $i \in M$ prefers to be a member, while (ii) every $i \notin M$ decides to free ride, anticipating (13). Part (i) is sometimes referred to as "internal consistency", because it requires that every that are inside the coalition is satisfied with their choice. Part (ii) is often referred to as "external consistency", because it requires that those external to the coalition are satisfied with their choice.

In Section 6.2, we embed the participation game in the dynamic game that we are used to by now, where countries both emit and invest in green technologies. If the coalition can also decide on the duration of the agreement (as in Section 5), then it turns out that the coalition can be larger.

If contracts were complete, so that countries could make agreements on both emission levels and investment levels, the small-coalition result would emerge for the same reasons as before.

With incomplete contracts, however, the coalition can be much larger. With incomplete contracts, the members cooperate on emission levels but investments are decided on noncooperatively. In fact, incomplete contracts turn out to be beneficial because of the large coalitions, and because of the hold-up problem (which will not be present on the equilibrium path). This result emerges if the countries can also decide on the duration of the agreement. When they can, they prefer a long-term agreement because it avoids the hold-up problem, but a short-term agreement if one of the members deviates by not participating in a given period. When it is expected that this free rider will return to the equilibrium and participate at the next round, the coalition prefers to have a short-term agreement instead of locking in a smaller coalition. When the short-term agreement comes along with the costly hold-up problem, the temptation to deviate is weakened, and a larger coalition can be sustained. In other words, the hold-up problem, which was problematic in the previous sections, is here beneficial. The hold-up problem arises only off the equilibrium path, if one of the members deviates by free riding, so that the coalition opt for a short-term agreement until the country returns. The cost of the hold-up problem implies that it is more attractive to participate, so that the coalition will be large enough for it to sign a long-term agreement.

At the end of this section, we will discuss other aspects that influence the coalition size. In particular, if we move away from the two-stage game, and countries are farsighted, they may cooperate because they would be worried that if they do not, then other countries will also opt out of the coalition later on.

6.1 Example 5: The Result " m = 3 "

Suppose there is one period, no investments, and country i's payoff from emitting g_i is:

$$-\frac{b}{2}(\overline{y}-g_i)^2 - c\sum_{j\in N} g_j,$$

What are the emission levels and the payoffs for the free riders and the coalition members? What is the equilibrium coalition size?

With BAU, g_i is $g_i^{BAU} = \overline{y} - c/b$. The FB is given by the same equation if just c is replaced with nc, while M's best policy requires c to be replaced with mc.

If we define i's abatement or contribution as $x_i := g_i^{BAU} - g_i$, we can write

$$u_i = -bx_i^2/2 + c\sum_{j\in N\setminus i} x_j, \tag{12}$$

plus a constant. With this, we find:

$$\begin{aligned} x_i &= (m-1)\, c/b,\ i \in M, \\ x_i &= 0,\ i \notin M. \end{aligned} \tag{13}$$

Thus, by free riding, a member reduces x_i to 0, but each of the other members reduces x_i by c/b. Thus, a member is better off participating iff:

$$-b\,[x_i\,(m)]^2/2 + c\sum_{j\in N\setminus i} x_i\,(m) \geq c\sum_{j\in N\setminus i} x_i\,(m-1) \Leftrightarrow$$
$$m \leq 3.$$

Formally, M^* is an equilibrium coalition if and only if $m^* = |M^*| \in \{0,2,3\}$.

The coalition cannot include more than three countries. For every larger coalition, the contributions would be so costly for a member that its payoff would be larger by free riding. If $m = 3$, each member is exactly indifferent between participating and free riding, and thus $m = 2$ is also an equilibrium. In addition, $m = 0$ is an equilibrium, because every country will set exactly the same policy noncooperatively as if it enters and creates a coalition alone. Note that $m = 1$ is *not* an equilibrium, because a free rider would be better off by entering the coalition with one other member.

In Example 5, at most three countries participate in the coalition.

Two aspects are striking about this result. First, the parameters (c and b) do not influence the equilibrium. The reason is that each parameter, such as c, will both raise the cost of participation, and also the benefit of participating: With the equilibrium x_i, both terms in (12) will be quadratic in c. (Similarly, both terms will be proportional to $1/b$.) Thus, the parameter cancels out when the payoffs are compared.

Second, the equilibrium coalition is very small. When m is large, both the benefit of dropping out, and the cost when the other $m-1$ members respond, are large. In this model, where the cost of contributing is convex while the benefit of the others' contribution is large, a large m will imply that the benefit of dropping out increases faster in m than does the cost. This comparison limits how large m can be, in equilibrium. In the two-stage participation game, the benefit of free riding is large also because it is assumed that the remaining coalition, with one less member, will simply continue as normal (although with lower x_i's).

6.2 Participation – With Fixed Duration

To generalize the model in Section 6.1, consider the dynamic game in Section 2 but with the following simplifying functional forms:

$$B\left(y_{i,t}\right) = -\frac{b}{2}(\overline{y} - y_{i,t})^2, \tag{14}$$

$$C\left(G_t\right) = \underline{c}G_t, \text{ and}$$

$$K(R_{i,t}, R_{i,t-1}) = \frac{\underline{\kappa}}{2}\left(R_{i,t}^2 - q_R R_{i,t-1}^2\right), \tag{15}$$

where $K(\cdot)$ is the investment cost. The linear $C(\cdot)$ implies that the present-discounted cost of emitting one more unit at time t is $c \equiv \sum_{\tau=t}^{\infty}(\delta q_G)^{\tau-t}\underline{c} = \underline{c}/(1-\delta q_G)$. With the linear $C(\cdot)$, we get interior solutions only if the investment cost is convex. With (15), the cost is simply quadratic if $q_R = 0$, but the cost of reaching $R_{i,t}$ is smaller if $R_{i,t-1}$ and q_R are large. The net marginal investment cost is, say, $\kappa R_{i,t}$, iff $\kappa \equiv (1-\delta q_R)\underline{\kappa}$. With these functional forms, first-best investment, conditional on $g_{i,t}$, is

$$R_i^{FB}\left(g_{i,t}\right) = \left(\overline{y} - g_{i,t}\right)b/\left(b+\kappa\right). \tag{16}$$

Because the first-best emission level is $g_i^{FB}\left(R_{i,t}\right) = \overline{y} - R_{i,t} - nc/\kappa$, we have

$$R_{i,t} = nc/\kappa,\ y_{i,t} = \overline{y} - nc/b, \text{ and } g_{i,t} = \overline{y} - nc/b - nc/\kappa.$$

The coalitional decisions are the same if just n is replaced by m, while a free-rider's decisions are the same if just n is replaced by 1.

As a start, suppose the countries play the two-stage participation game before the emission stage in period 1. With this model, M's policies are "payoff-irrelevant" for the nonparticipant, and vice versa. That is, every free-riding country has a dominant strategy, so that its optimal policies are independent of what M will decide. For these reasons, the outcome of the game is the same whether M's decisions are before, after, or simultaneous with the decisions of the free riders (i.e., (a) and (b) at stage 2 in the participation game, described above).

The coalition members negotiate the $g_{i,t}$'s for period $t \in \{1, ..., T\}$, while later emission caps may be negotiated just before the emission stage in period $T+1$. The $r_{i,t}$'s are decided on noncooperatively, and i will find it optimal to set $\kappa R_{i,t} = b(\overline{y} - y_{i,t})$, which is socially optimal, in every $t \in \{2, ..., T\}$, but $\kappa R_{i,t} = c$ at $t = 1$, and $t = T+1$, etc, i.e., just before the next emission stage. As before, the intuition is that before future caps are to be negotiated, i's technology is, essentially, a public good, because it motivates i to emit less, and i captures only a fraction of this benefit.

With these investments, as a function of the $g_{i,t}$'s, the coalition will negotiate the first-best $g_{i,t}$'s. The details are left as an exercise, but, by comparing the payoff for member i with what i can obtain from opting out, we learn that i is willing to participate iff the coalition includes at most three countries. Thus, the three-member result holds, also in this setting: M^* is an equilibrium coalition if and only if $m^* = |M^*| \in \{0, 2, 3\}$.

As discussed in Section 5, the duration of the commitment period may also be endogenous. When the coalition size is endogenous, it may depend on the coalition size.

6.3 Example 6: Endogenous Duration

Consider the same payoffs as in Section 6.1, so that we abstract from investment levels, but let there be many periods. After M is formed, M can not only negotiate the caps, but also the duration. Then, the optimal duration depends on number of members at the time being, m, and how this compares to the equilibrium number, m^*. In particular, let M^* denote an equilibrium coalition of size $m^* \equiv |M^*|$ and consider a realized coalition $M \subseteq M^*$, with $m = |M| \leq m^*$. M finds it optimal to contract for $T(m)$ periods, where:

$$T(m) = \begin{cases} 1 & \text{if } m < m^* \\ \{1, ..., \infty\} & \text{if } m = m^* \end{cases}.$$

That is, if few countries participate at time t, they prefer to sign a one-period agreement so that the coalition can return to the larger, equilibrium size in the next period. If, in contrast, $M = M^*$, which will be the case on the equilibrium path, every $T \geq 1$ is equally good to the coalition. The reason is that after T periods, the same coalition will form again, in equilibrium, and they will contribute exactly the same when it comes to emission cuts and investments. The choice of T is thus irrelevant.

6.4 Participation – With Endogenous Duration

When we return to the model with investments, it is rather different.

Duration. When deciding on T, there is, as in Example 6, an incentive to reduce T when the actual number of participants, m, is small relative to the equilibrium number, m^*, so that the larger coalition can be formed soon. However, the reduced T is now coming along with a cost. The cost is that every member invests less before the next bargaining stage because of the hold-up problem. When investments are lower, emission levels will be larger at the next bargaining stage. Therefore, when $m = m^*$, the members strictly prefer a long-lasting agreement. (With complete contracts, they were indifferent.) If it turned out that a member deviated, so that the coalition right now is just $m^* - 1$, then the remaining coalition faces a dilemma. Should it "lock in" the coalition and sign a long-term agreement with the smaller coalition? Or should it sign a short-term agreement, where the benefit is that the coalition size can be expected to be m^* next period, but the cost is that investments will be small, in the meantime, because of the hold-up problem associated with short-term agreement?

Naturally, the answers also depend on how much it matters to include just one more country. If m^* is large, it may not be worthwhile that $m^* - 1$ countries should invest little, just to get another country on board. If m^* is small, one more member is relatively important. The threshold is given by:

$$m^* \leq m_M(\chi) \equiv 1 + \frac{1}{1 - \sqrt{(\chi + \delta) / (\chi + 1)}}, \text{ where } \chi \equiv \frac{\kappa}{b}. \tag{17}$$

That is, M finds it worthwhile to sign a one-period contract is a single member deviates by free riding iff $m^* \leq m_M(\chi)$. (For details, see Battaglini and Harstad, 2016.)

The answers to the two questions also depend on how costly the hold-up problem is. In this model, the hold-up problem is larger if κ/b is small. The reason is that if κ is small, or b is large, then the climate change problem is, most of all, relying on new technology, rather than reduced energy consumption. So, if κ/b is small, it is costly to sign a short-term agreement, and threshold $m_M(\chi)$ is small.

Participation. The cost of the hold-up problem is faced also by members considering to free ride. If (17) holds, a supposed-to-be member that deviates and free rides will cause the remaining coalition to sign a one-period agreement. Then, free riding comes along with the cost that the members will invest exactly as in BAU.

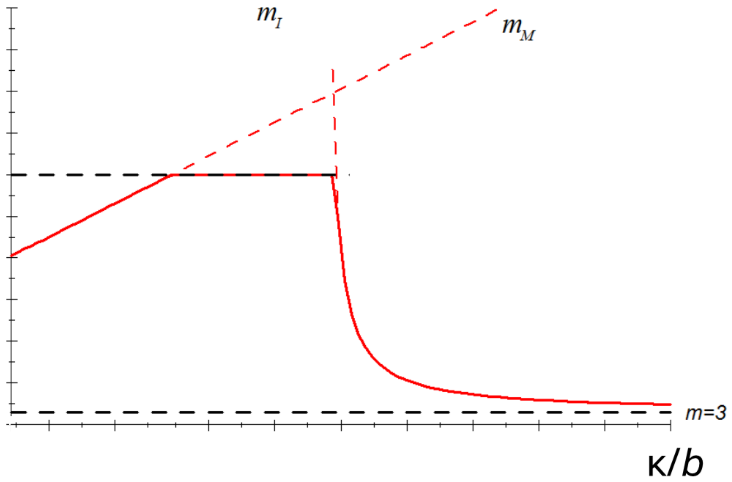

Figure 3: *The coalition size can be large, depending on the cost of the hold-up problem.*

The additional cost makes free riding less attractive, and a member benefits from staying in the coalition for a larger set of parameters than when contracts are complete. Here, is $m_I(\chi)$ the largest coalition size such that a member is better off participating rather than free riding for one period, given that investments will be as in BAU during that period. Because the cost of participating is convex in m^*, $m_I(\chi)$ is limited, especially when the hold-up problem is not that expensive for a member considering to free ride. By comparing one's alternative payoffs, we can show that a member is better off participating iff:

$$m^* \leq m_I(\chi) \equiv 3 + \frac{2\delta}{\chi - \delta}$$

When $\kappa \to \infty$, and therefore $\chi \to \infty$, $m_I(\chi) \to 3$, as in the case without any investment. Thus, for a large κ and, thus, χ, m^* is severely constrained by $m_I(\chi)$. By comparison, for a small χ, we had that m^* was severely constrained by $m_M(\chi)$. In combination, m^* is an equilibrium coalition size iff $m^* \leq \min\{m_I(\chi), m_M(\chi)\}$ .Thus, m^* will be limited by $m_M(\chi)$ when χ is small, and by $m_I(\chi)$ if χ is large. If χ is moderate, m^* can be quite large, as illustrated in Figure 3.

In other words, the hold-up problem associated with short-term incomplete contracts disciplines and motivates coalition members to participate.

The hold-up problem, as an out-of-equilibrium threat, can motivate a larger coalition to form.

6.5 Further Reading and Farsightedness

On the one hand, Eichner and Kollenbach (2023) showed that the possibility for the above hold-up problem is indeed important to generate large coalitions. They consider a timing of the game that leads to no hold-up problem, even for short-term agreements, and they find that, in this case, the coalition is small, just as when contracts are complete. Participation can also be low if investments are made prior to the participation stage (Helm and Schmidt, 2015).

On the other hand, there are other reasons for why the coalition can be large, of course. Kovac and Schmidt (2021) show that if the remaining coalition cannot sign one-period agreements, the coalition will be large, even without the hold-up problem. The result that the small-coalition result gives the number 3 in the static model is, moreover, because of the linear-quadratic functional forms. Karp and Simon (2013) develop a non-parametric approach to the problem, and they identify conditions under which the equilibrium coalition size can be large, depending on the functional forms.

Altruistic preferences at the participation stage can lead to larger coalitions, but altruism at the emission-setting stage can reduce the equilibrium coalition size, because free riders can take advantage of the coalition's benevolent behavior (Schopf, 2024).

Heterogeneous preferences often reduce the equilibrium coalition size, but (therefore) increasing the effectiveness of adaptation in highly vulnerable countries can help supporting an IEA, when this reduces the heterogeneity (Li and Rus, 2019). When countries remain heterogeneous, transfers might be helpful to motivate some of the countries to participate (Dutta and Radner, 2020; Kotchen, 2020; Okada, 2023). Barrett (1997) analyzed how the coalition can be larger if the members could impose sanctions on the free riders. While Hoel and Schneider (1997) show that transfers can be harmful, and lead to more emissions, Barrett (2001) showed that if the countries are very asymmetric, transfers can be necessary for a large coalition to form. When the heterogeneity is private information, transfers are a natural part of the optimal mechanism (Martimort and Sand-Zantman, 2016). Besides explicit side payments, there can be various forms of favours or issue linkages (Maggi, 2016), including trade liberalization (Nordhaus, 2015), that can also motivate participation.

The two-stage participation game gives a lot of power to free riders. By opting out, the coalition is reduced by one party, even if it turns out that the remaining parties might also get second thoughts. It might be more realistic to allow other players to also reach and change their decision.

To capture this situation, it is sometimes assumed that other parties can also leave. When this is anticipated, and a potential free rider is "farsighted", then opting out might be less attractive. When free riding is less attractive, the coalition can be larger.

To illustrate, suppose that countries can only leave the coalition, and that free riders can not change their mind and join later on. Based on the above analysis, we know that, in the simple example of Section 6.1, a coalition of size 3 is farsightedly stable: A third member does not strictly want to leave, no matter what the other countries are doing. Therefore, a coalition of size 4 is not stable, because one member can leave without triggering other countries to leave. But because the size 4 is not stable, a coalition of size 5 will be stable: A member would anticipate that if it opts out, then the coalition will not end up being of size 4, which is unstable, but of size 3. That reduction in the coalition size will motivate all five countries to participate. Given this, a coalition of size 6 is not stable, because if a single country deviates, the remaining coalition, with five countries, is stable. Thus, the next larger stable coalition consists of eight participants, then 13, and so 20. There can be many coalition sizes that are farsightedly stable, and some of them can be quite large. The numbers from the example with linear-quadratic utility functions are from Ray and Vohra (2001). (But see also Chew, 1994; Ray and Vohra, 1999; Ray, 2007; Osmani and Tol, 2009; Diamantoudi and Sartzetakis, 2018; Vosooghi et al., 2024.)

Instead of a game where countries can only leave the coalition, one could study a game where countries can only join. These games are studied in the literature on ratification (see, for instance, Wagner, 2016) and when there are minimum participation thresholds (Black et al., 1993; Harstad, 2006; Carraro et al., 2009; Weikard et al., 2015), ignored in this chapter.

Another strong assumption that is imposed in most of the literature is that there is, at most, one environmental coalition. The assumption does not match well with the real world, however, because we observe a large number of coalitions in many other settings, including the environment, and international trade. Theoretically, this is also an unreasonable assumption: In Example 5, for instance, there might be only three countries in a coalition, but there can be a large number of coalitions that are active, side by side. Thus, Asheim et al. (2006) show that two agreements can sustain more: They find that a regime with two agreements can Pareto dominate a regime based on a single global treaty.

7. COMPLIANCE WITH SELF-ENFORCING TREATIES

What happens if a country does not comply with its pledges?

There is no world government that can force a sovereign country to comply. Thus, an IEA must be self-enforcing, and designed in a way so that countries will comply voluntarily. The IPCC (2014:1015) states that: "*From a rationalist perspective, compliance will occur if the discounted net benefits from cooperation (including direct climate benefits, co-benefits, reputation, transfers, and other elements) exceed the discounted net benefits* *of defection.*"

This motivation, however, may not always be sufficiently strong. During the first commitment period of the Kyoto Protocol, for example, Canada found it costly to comply and, in 2011, it simply withdrew.

It is arguably more difficult to motivate compliance than to induce countries to participate in the first place: If a country does not participate, the other coalition members will contribute less; if a country participates but defects at the emission stage, the other coalition members will, in equilibrium, contribute just as expected. By comparison, the temptation to participate and defect might be stronger than the temptation to free ride from the very beginning.

This section studies the incentive to comply and derives conditions under which an IEA is self-enforcing. To check the boundaries of what self-enforcement can accomplish, it is often assumed that the punishment following non-compliance is as severe as possible. The harshest punishment when there is a single defecting country is that the other countries take actions that minimizes the maximal payoff for the defecting country, when this country maximizes its own payoff, given the punishment that it faces: These are the min-max strategies, often referred to in the literature on repeated games (for a textbook treatment, see Mailath and Samuelson, 2006).

A simpler punishment is that the cooperative arrangement breaks down, and that the countries return to the business-as-usual outcome. This punishment is also symmetric, and thus achievable if, and even if, it cannot be observed which country that is responsible to the higher-than-agreed-upon pollution level. This type of punishment is typically considered in the literature on self-enforcing agreements: see Barrett (1994; 2006), Dutta and Radner (2004; 2006), and Kerr, Lippert, and Lou (2024), among others.

In some cases, the BAU strategies coincides with the min-max strategies. That will be the case in the simple benchmark model that I start out with below. This model is useful to illustrate the key factors that determine whether an agreement is self-enforcing: The discount factor must be large, and contributions must be modest. Thereafter, we return to the dynamic game, also used in the previous section, where countries are both emitting and investing in green technologies. In this model, the technology will take on a new, strategic role. When the temptation to defect at the emission stage is large, the temptation can be reduced if a country is endowed with green technology that reduces the need to emit. For this reason, the best self-enforcing treaty might require countries to invest more than in the first best, only to reduce the temptation to defect. This situation arises when first-best climate policies place a larger emphasis on the reduction of energy consumption rather than the development of new, green technology. In the reverse case, where investments are the most important ingredient, then countries might be more tempted to defect at the investment stage, instead of at the emission stage. In this situation, the best self-enforcing treaty permits and, in fact, requires the countries to invest less, even though they are expected to reduce the energy consumption by a lot. In equilibrium, a country would be tempted to invest more at the investment stage, but higher investments must be punished by the other members. The intuition for this result is that if a country did indeed invest more, it would find it optimal to emit less in the future, also in the event in which another country has defected. That is, it is not credible that defection will be punished sufficiently by the other countries, unless investments are kept low.

7.1 Example 7: A Repeated Game with Trigger Strategies

As a start, consider the payoffs in Example 6, given by (12), and assume that, if any country contributes another amount than x^*, all contributions will forever after be 0 (i.e., BAU). With this as the threat point, the most tempting defection is to set $x_i = 0$, but compliance with x^* is better than defection if and only if the following compliance constraint holds:

$$
\begin{aligned}
c(m-1)x^* - \frac{b}{2}(x^*)^2 \geq (1-\delta)\,c(m-1)x^* \Leftrightarrow \\
\delta \geq 1 - \frac{c(m-1)x^* - \frac{b}{2}(x^*)^2}{c(m-1)x^*} = \frac{bx^*}{2c(m-1)} \Leftrightarrow \\
x^* \leq 2\delta c(m-1)/b.
\end{aligned}
$$

Thus, an IEA with contributions x^* can be self-enforcing iff x^* is small, δ is large, and m is large.

This inequality has several interpretations. First, it is defining a lower boundary for what the discount factor can be for countries to be willing to comply with x^*. Second, for any give discount factor, the inequality defines the largest contribution level that a coalition of size m is willing to comply with. The larger is the coalition, the larger is the threshold for the contribution x^*. Third, for any given δ and x^*, the inequality is requesting m to be large for compliance to be worthwhile. Hence, free riding can be discouraged if just the coalition is sufficiently large. This contrasts the findings in the previous section, where the incentive to free ride was stronger when m was large.

Part of the explanation for this difference is that, in the previous section, x^* was endogenous. We know that the coalition prefers countries to set $x^* = c(m-1)/b$. With these contributions, the inequality simplifies to:

$$
\delta \geq \frac{bx^*}{2c(m-1)} = \frac{1}{2}.
$$

If this condition fails, the contributions must be reduced below the first-best level for the agreement to be self-enforcing.

This simple analysis can be extended in several directions. For example, suppose that if a country defects, it will be observed and punished only with probability $q \in [0,1]$. Furthermore, suppose the punishment phase will last only $T \leq \infty$ periods, before the coalition will renegotiate and re-start its cooperation phase. With these two generalizations, compliance is attractive only if:

$$v = c(m-1)x^* - \frac{b}{2}(x^*)^2 \geq (1-\delta)c(m-1)x^* + (1-q)\delta v + q\delta^{T+1}v.$$

If q or T is small, the right-hand side is larger, and thus x^* might need to be reduced for compliance to remain attractive. Vice versa: If the coalition can improve on its ability to monitor, so that q increases, then also contributions can increase, without violating the countries's compliance constraints.

With better monitoring technology, the cuts can be deeper.

7.2 Compliance Technology and Self-Enforcing Agreements

Consider now the model in Section 6.2. A symmetric stationary self-enforcing agreement is now a pair (g_i, R_i), so that the countries prefer to comply and stick with (g_i, R_i) forever rather than to defect. Because the two variables are determined at different points in time, there are two types of compliance constraints that must be satisfied: One for the investment stage, and another for the emission stage. Each compliance constraint will require that the payoff is larger if one selects the agreed-upon level than if one defects. If a country defects, it will find it optimal to select the BAU level. For simplicity, it will be assumed that from the subsequent stage, all countries revert to the BAU outcome. With this, the two compliance constraints require that a country's payoff from complying forever is larger than if a country returns to its BAU strategy, before the other countries return to BAU at the next stage.

In general, for every pair (g_i, R_i), the compliance constraint at the investment stage will hold if $\delta \geq \delta^{R_i}(\mathbf{g}, R)$, while the compliance constraint at the emission stage holds iff $\delta \geq \delta^{g}(\mathbf{g}, R)$. With the first-best pair $\left(g_i^{FB}, R_i^{FB}\right)$, the two compliance constraints can be written as follows. (For details, see Harstad et al., 2022.)

$$\delta \geq \overline{\delta}^{R_i} \equiv \delta^{R_i}(\mathbf{g}^{\mathbf{FB}}, \mathbf{R}^{FB}) = \frac{b-\kappa}{2b}, \tag{CC-r}$$

$$\delta \geq \overline{\delta}^{g} \equiv \delta^{g}(\mathbf{g}^{\mathbf{FB}}, \mathbf{R}^{FB}) = \frac{\kappa}{b+2\kappa}. \tag{CC-g}$$

If δ is larger than both thresholds, the first best is attainable. If δ is smaller, the type of distortion depends on which compliance constraint that is binding first.

If investments are relatively costly (so that κ/b is large), investment levels are low, even in the first best. The low investment levels reduce the temptation to defect at the investment stage so, when δ is smaller, the first compliance constraint to bind is the one at the emission stage (i.e., $\delta \geq \overline{\delta}^{R_i}$). When (CC-g) binds, the emission level must be limited, and smaller than the first best, in order to motivate compliance. This is replicating the insight developed in Section 7.1.

When δ falls further, the best self-enforcing agreement moves along the horizontal arrow in Figure 4: g_i will be permitted to increase, but there is no reason to reduce the investment levels, so the best self-enforcing agreement continues to require $R_i = R_i^{FB}$. Consequently, R_i is larger than the ex post first-best investment level, given g_i. When g_i increases, it is ex post optimal that R_i falls, in line with (16). The best agreement, however, requires countries to invest more, and too much, relative to the ex post optimal level. The intuition for this result is that if countries were allowed to invest less, and as little as $R_i^{FB}(g_i)$, the temptation to defect at the emission stage would be greater. To motivate compliance, g_i would need to fall even further. The inefficiently high investment levels reduces the need to consume fossil fuel, and thus they reduce the temptation to emit more. For this reason, the best self-enforcing treaty requires countries to invest more than what is socially optimal, given the emission level that is associated with the agreement. If the country implements the decision with taxes and subsidies, the outcome is that if δ falls, the country must reduce the tax on emission (below the Pigouvian level) but raise the subsidy on the investments, even if there are no technological spillover associated with the technology.

If κ/b is small, so that investments are inexpensive and thus important, then the first compliance constraint that binds when δ becomes smaller is the one at the investment stage. When the first best can no longer be supported as a self-enforcing agreement, the distortion that is necessary in the best self-enforcing agreement is that countries must be allowed to invest less. There is no reason to distort the energy consumption level, $y_{i,t}$, however, so $y_{i,t}$, will remain at the level that is first best. Because $y_{i,t}$ is constant, while R_i is smaller, g_i must be larger. If the discount factor falls, the best pair (g_i, R_i) moves along the downward-sloping line in Figure 4. This implies that R_i is not optimal conditional on the g_i that will be chosen. Instead, $R_i < R_i^{FB}(g_i)$. This may be surprising, because countries would volunteer to invest as much as $R_i^{FB}(g_i)$ when they know that they must or will, in any case, emit g_i. (After all, there are no technological spillovers associated with the investments.) Thus, the best agreement requires that countries invest less than what they would prefer, given g_i. The intuition for this policy is that if countries

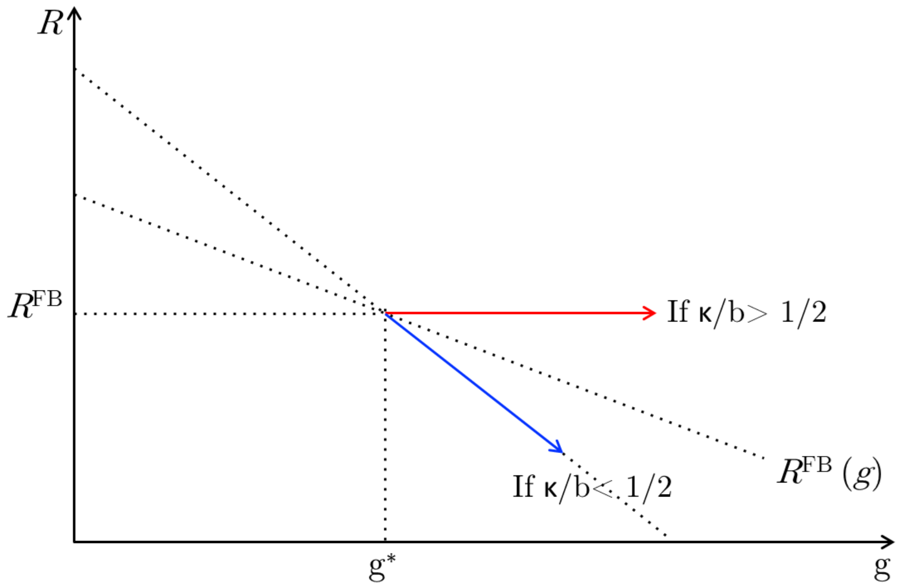

Figure 4: *The best self-enforcing treaty might require investments that are too large, or too small, relative to the first-best level.*

were allowed to invest more, it would be too tempting to defect at the investment stage because, for those high investment levels, the other countries would be unable or unwilling to emit a lot as a punishment in this period. Thus, the large investment levels would reduce the credibility of punishment, and countries will find it beneficial to defect at the investment stage. If the country implements the decision with taxes and subsidies, the outcome is now that if δ falls, the country must tax investments in green technology.

7.3 Technologies, Thresholds, and Transfers

While green technology can help to motivate compliance, "brown technology", which can be strategic complements to emissions, can raise the temptation to emit. Thus, for an agreement to be self-enforcing, it might be necessary to limit how much countries can be permitted to invest in brown technologies. (These limits can justify taxes on the investments in drilling technologies, even when there are taxes also on the emission levels.) In Harstad et al. (2019), brown technology is permitted, and it is shown that even investments in adaptation technology might need to be regulated. If the temptation to defect is strong, then compliance will be credible only for countries that invest less in adaptation.

In general, our insight on self-enforcement draws on the literature on repeated games (for a textbook-treatment, see Mailath and Postlewaite, 2006). Dutta and Radner (2009) allow strategies to be conditioned on the stock of greenhouse gases (even thought it is not "payoff relevant"), so that all countries emit little as long as the threshold is not crossed, but the countries revert to BAU when the threshold is crossed. These types of trigger strategies can motivate the countries to emit less, under some conditions. If countries are asymmetric, Kerr et al. (2024) show how transfers, and technology transfers, can help the agreement to be self-enforcing.

8. NARROW-BUT-DEEP VS. BROAD-BUT-SHALLOW

The emission cuts for the Annex II signatories in the Kyoto Protocol (KP) faced substantial legally binding emission cuts. However, only 37 countries ended up with such commitments. The Paris Agreement (PA), in contrast, include 195 signatories, but the pledges are nationally determined, and they are not legally binding. By comparison, one may argue that the KP was narrow but deep, while the PA is broad but shallow.

Ideally, an international environmental agreement (IEA) should be both broad and deep, but this can be difficult because of the problems discussed in the earlier sections. When only second-best IEAs are possible, the debate regards whether it is worse to give up on depth vs. breadth. Aldy et al. (2003) and Aldy and Stavins (2007; 2009) include several chapters studying and comparing alternative climate agreement designs, and the trade-off arises in many of them.

The trade-off is especially natural when we take into account that countries are heterogeneous, as in Example 8. If the IEA is ambitious, only the most eager countries might be willing to pay the (high) cost of contributing. If it is important to get the most reluctant countries to cooperate, as well, it may be necessary to give up on ambition, and accept that contributions are less impressive.

Now, there is not always such a trade-off. A treaty can be broad because it is good, or it can be deep because it is broad. In Sections 4-6, the endogenous emission cuts were larger when the number of participants was larger, because the externalities on a larger number of countries were internalized. In Section 7, the incentive to defect, rather than to comply, was stronger when the cuts were deep, but weaker when the number of participants was large. Thus, a larger coalition can make it possible to deepen the emission cuts, without making it tempting for the countries to defect and emit more. When it is possible to make an IEA both broader and deeper, the members would to advantage of this possibility. After these possibilities have been taken advantage of, one might be left with a trade-off.

8.1 Example 8: Two Types of Countries

As a simple illustration, suppose n_S countries are willing to contribute $\underline{x}$, while n_N would be willing to contribute $\overline{x}$. With the utility functions in Example 6 and 7, the n_N

countries are better off with the broad but shallow agreement iff:

$$(n_N - 1)\, c\overline{x} - b\overline{x}^2/2 < (n_N + n_S - 1)\, c\underline{x} - b\underline{x}^2/2 \Leftrightarrow$$
$$n_S > \frac{(n_N - 1)\, c\,(\overline{x} - \underline{x}) + b\,(\underline{x}^2 - \overline{x}^2)/2}{c\underline{x}},$$

which holds if n_S is large (relative to n_N) and if $(\overline{x} - \underline{x})$ is small, for example. The optimal design thus depends on how many more countries one can persuade, and what the cost of persuasion would be. (A better design can be that n_S contributes $\underline{x}$ while n_N contributes $\overline{x}$. This is referred to as common but differentiated responsibilities.)

8.2 Pledge-and-Review Bargaining

Even without heterogeneity, this trade-off can be important. To shed light on it, we must understand what makes the cuts deep vs. shallow, and that brings us to the choice of bargaining procedure.

In applications, the outcome of a bargaining game is often approximated by the NBS. One reason for this is that the NBS satisfies reasonable axioms. In addition, the NBS is the outcome of standard noncooperative games. When the parties are symmetric, the NBS predicts that the outcome is first best for the parties involved. With the linear-quadratic utility function (12), from Section 6.1, for example, the NBS predicts $x_i = (m-1)\, c/b$, so that all externalities from the other members are fully taken into account.

In reality, negotiations might be less efficient than what is predicted by the NBS.

First, transaction costs typically arise when the parties are not fully informed about the preferences of one another. There is a large literature on bargaining under incomplete information, and it typically predicts that the negotiations are delayed or the outcome is inefficient.

Second, the negotiations might be inefficient by design. The bargaining procedure used in the PA was very different from that used for the KP. The PA relies on "nationally determined contributions" which are decided "bottom-up" rather than "top-down", one may argue.

With nationally determined contributions, one may fear that a country will just propose the contribution it prefers anyway. But if it is uncertain whether the other participants will bother to accept, ratify, or comply with the treaty, then a country might be willing to raise its contribution a little, just to reduce the risk. When all countries are doing so, the agreement is more valuable, and it makes sense to further raise one's contribution since the risks are costly when the agreement is valuable. The equilibrium of such a game can imply that the contributions are larger than in BAU, but lower than with the NBS, so that each contribution maximizes a weighed sum of payoffs, where the relative weight placed on others' payoffs is positive but less than 1. (See Harstad, 2023a, for details.)

For such reasons, we may expect that the relative weight on others' payoffs, w, is less than 1, when the countries employ the P&R procedure:

$$u_i + w \sum_{j \in M \setminus i} u_j,$$

With the NBS, and symmetric utility functions, $w = 1$.

With the utility, (12), described in Section 6.1, the equilibrium contribution becomes:

$$\begin{aligned} x_i &= w(m-1)c/b, \, i \in M, \\ x_i &= 0, \, i \notin M, \end{aligned} \tag{18}$$

The immediate effect of $w < 1$ is that contributions will be less ambitious. Thus, an agreement that is "shallow", rather than "deep", may be associated with a smaller w. By adapting the P&R bargaining procedure, the world community might have moved to a game in which w was smaller for the PA than it was in the KP.

Naturally, the P&R procedure has been criticized by policymakers and scholars. Keohane and Oppenheimer (2016, p. 142) speculated that: "Many governments will be tempted to use the vagueness of the Paris Agreement, the discretion that it permits, to limit the scope or intensity of their proposed actions", while Tirole (2017, p. 209) wrote that: "The strategy of voluntary commitments has several significant defects, and is an inadequate response to the climate change challenge."

8.3 Modesty May Pay

Suppose the members are deciding on the contributions using a bargaining procedure associated with a smaller w. For a given coalition size, m, the effect of a smaller w is that countries will emit more, and be worse off. The coalition size might change with w, however.

In fact, the effect of w on the equilibrium coalition size might be more important than the effect on the contribution level. Because the cost of participating is less when w is small, a member may be satisfied being a member even if m is large. By comparing the utility of a member to the utility this member could obtain from free riding, we get that it is better to participate if and only if:

$$m \leq 1 + 2/w. \tag{19}$$

Thus, if w is small, m can be large. This is shown by Finus and Maus (2008), which motivated the title of this subsection. (See also Colombo et al., 2022; Barrett, 2002.)

The prediction that m is larger when w is small fits well with the fact that many more countries contributed in a meaningful way to the Paris Agreement than in the Kyoto Protocol. The choice of bargaining procedure can thus lead to the trade-off studied in this section. By adopting the P&R procedure, the concern is that the cuts will be less deep, but the consequence is (also) that the coalition will be broader.

8.4 Institutional Design

To understand the trade-off, we must analyze the welfare effects of both the level of the cuts as well as the size of the coalition.

With the utility function (12), it turns out that if m decreases in w as in eq. (19), then also the payoff of every member decreases in w, so that all members would prefer a bargaining procedure characterized by a smaller w, where the coalition would be larger. In fact, when w declines, the increase in m is, in itself, motivating the countries to contribute more. This effect is balancing the effect that the contributions are reduced when w is low, for a fixed m. (For details, see Harstad, 2023b.)

In reality, however, m might not decrease in w exactly as predicted by (19). First, we might have a constraint $m \leq \overline{m}$, because, after all, there is a finite number of countries that are relevant when it comes to global climate policies. Second, countries are more heterogeneous than what the model above assumes. The heterogeneity implies, for example, that some countries might participate even if others strictly prefer not to: They might be "committed". The countries in the EU, for example, cannot easily opt out from a treaty in which the other EU countries participate. To recognize this situation, let $\underline{m}$ measure the number of committed countries, which participate even if w is large.

With boundaries on m, so that $m \in [\underline{m}, \overline{m}]$, a bargaining game associated with a small $\underline{w}$ gives a larger total payoff to the world community than does a game associated with $\overline{w} > \underline{w}$ if the constraints on m are not too severe. The condition for when the members prefer $\underline{w}$ (e.g., P&R) is given by

$$\frac{\min\{\overline{m}-1, 2/\underline{w}\}}{\max\{\underline{m}-1, 2/\overline{w}\}} > \Omega \equiv \sqrt{\frac{\overline{w}\,(1-\overline{w}/2)}{\underline{w}\,(1-\underline{w}/2)}} \in \left(1, \frac{\overline{w}}{\underline{w}}\right). \tag{20}$$

The inequality is derived and better explained elsewhere (Harstad, 2023b), and it can be illustrated easily in Figure 5.

Thus, when we take into account that the constraints on m might bind, we are facing a trade-off. A broad but shallow agreement (associated with $\underline{w} < \overline{w}$) is better than a narrow but deep agreement (associated with $\overline{w} > \underline{w}$) iff there is a large number of

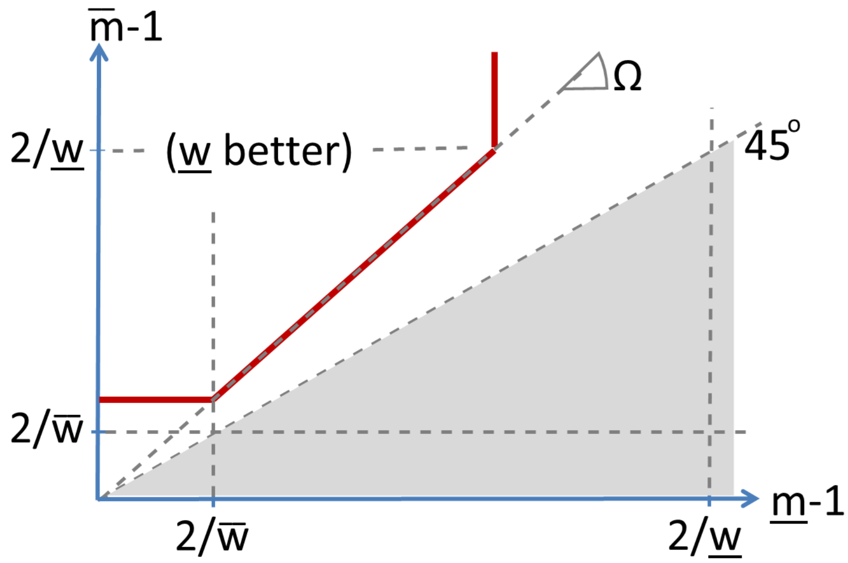

Figure 5: *A broad-but-shallow bargaining procudure is better top-left in the figure.*

potential members, and if the heterogeneity is limited, so that few of them are committed to participate regardless of the design.

Regarding compliance, a reduced w, and thus a reduced x^* or larger m, will also reduce the temptation to defect. From Section 7.1, we have that $x^* \leq 2\delta c\left(m-1\right)/b$, and, with (18), the compliance constraint becomes

$$w\left(m-1\right)c/b \leq \delta 2c\left(m-1\right)/b,$$

so, whether or not we endogenize m by (19), the compliance constraint becomes:

$$w \leq 2\delta,$$

which is more likely to hold when w is small. Thus, emission cuts are more likely to be self-enforcing (so that they do not need to be "legally" binding) if they are the outcome of P&R bargaining rather than the top-down approach associated with the KP (were, we have argued, w might be larger).

The model above is simple, but the results hold in more general settings. With the models in Section 6.2, countries bargain contributions for a number of periods before they decide on investments or investment policies noncooperatively. The investments will be functions of the contributions, and thus a country's continuation value can be written as a function of the x_i's. These continuation values are linear-quadratic, and can be written as (12), if just b and c are appropriately defined.

9. FROM KYOTO TO PARIS AND BEYOND

United Nation's approach to climate change has evolved quite a lot over the last few decades. In 1997, only 37 countries were committed by the Kyoto Protocol to limit emissions while, since 2015, nearly all countries in the world are participating in the Paris Agreement.

The two agreements have a lot in common. In both, the focus is on emission reduction, and while new technology is recognized as being crucial, the investment decisions are left to the individual countries and their market participants. The treaties also emphasize the need for transparency, so that it possible to observe how other countries are doing regarding their commitments. There is no world governments, or effective sanction in place if countries do not comply, and the participation is voluntary and not yet linked to explicit sanctions.

But there are several differences between the agreements. Not only is the number of committed countries quite different, but the bargaining procedure is also fundamentally different. The procedure under the KP has been referred to as "top-down", while the pledge-and-review procedure, associated with the PA, is described as "bottom-up". With P&R, we have nationally determined contributions. The commitments are also weaker in the sense that they are not legally binding, while they were legally binding under the KP. Finally, the two alternative designs were preferred at different moments in time. In the 1990s, the world community opted for the top-down approach but twenty years later, they opted for the bottom-up approach. That fact is, in itself, a puzzle that one must explain.

Empirical research on climate agreement is extremely important, but also challenging given the few data points that we have. Fortunately, economic theory also has a lot to contribute. In fact, several of the observation can be rationalized with game theory and contract theory. When international climate policies are decided on noncooperatively, we have a dynamic common-pool problem (formalized in Section 2), and the inefficiencies are even worse that in the game's static counterpart (Section 3). When we adopt the assumption that countries negotiate emission caps but not investment levels, then we can explain why we might see insufficient breakthroughs on the technology front, why countries might not invest enough, and how the investment levels can be influenced by the emission caps (Section 4). The so-called hold-up problem sheds light on why the commitment period is not annual, and how the optimal length trades off different inefficiencies when it is hard to predict how much one ought to emit in the future (Section 5).

The fact that the United Nation's approach does not ask countries to specify investment levels might be a puzzle in itself, given the costly hold-up problem, but this is consistent with the finding in Section 6 where we observed that the equilibrium coalition size can be larger when emission levels, but not as well investment levels, are negotiated. The need to motivate compliance, at a time without a world government, explains why peer pressure and transparent compliance are important for the treaties.

Economic theory can also shed light on the differences between the treaties. The benefit of P&R is that it is less expensive for a country to participate in the IEA. The lower expense explains why many more countries will participate with this type of procedure, and also why the emission cuts do not need to be legally binding in order to be credible. With these benefits, we can understand why the world community might be better off with P&R, rather than with the KP's top-down procedure. However, they are better off only if there is a large number of countries that are important to attract, and if few countries are sufficiently committed so that they will participate no matter what the procedure is. Both these conditions are more likely to be true in the 2010s, than they were in the 1990s. In the 1990s, many countries were less developed but some of them have since become emerging economies. This transition implies that the number of important potential members has grown. At the same time, some of the countries that were committed in the KP's first commitment period were not willing to remain committed to this type of treaty. Both changes suggest that the world might have moved from being located bottop-right in Figure 5, to top-left in the same figure. If that is the case, economic theory can rationalize why the top-down procedure was optimal in the 1990s, while the bottom-up P&R procedure was more attractive in the 2010s.

Unfortunately, both the deep-but-narrow and the broad-but-shallow designs are second best. To deal with climate change successfully, the world needs to find a way to implement treaties that are both broad and deep. To achieve this, the cost of free riding and defecting on the treaty must be larger than what we have assumed above. Alternatively, the benefits of participating and complying need to be larger. The challenge is that there are not a lot of sticks or carrots that are available in world politics today. Economic sanctions tend to be ineffective, costly to impose, and thus not credible to sustain in the long run. Explicit monetary transfers are challenging for political reasons, and also because governments rarely have the necessary liquidity available. At the end of the day, perhaps market access and trade liberalization stand out among the few or only carrots that are possible to use. The idea to connect trade and environmental agreements is old and has been analyzed extensively by environmental economists (Barrett, 1997), trade economists (Copeland and Taylor, 2005), and Nobel laureates (Nordhaus, 2015), for example, and. Kotchen (2024) discuss carbon border adjustments elsewhere in this volume.

What about political economics? The normative approach to environmental policy and international agreements is challenged by the positive political economics approach. Political economics, as a field, sheds light on the political forces that exist and that can rationalize the inefficient policies that we actually observe in the world. At the least, policies and agreements must take these forces into account, for the policies to be feasible and effective. My companion chapter (Harstad, 2025) discusses the political economics of international public goods, how domestic political forces must be taken into account, and how the forces can actually strengthen international cooperation if the agreements are designed in the right way.

References

Aghion, Philippe, Dewatripont, Mathias, and Rey, Patrick. (1994): "Renegotiation Design with Unverifiable Information," Econometrica 62(2): 257-282.

Aldy, Joseph E, and Robert N Stavins (2007): "Architectures for Agreement" Cambridge University Press

Aldy, Joseph E, and Robert N Stavins (2009): "Post-Kyoto International Climate Polic" Cambridge University Press

Aldy, Joseph E.; Barrett, Scott and Stavins, Robert N. (2003): "Thirteen plus one.."

Asheim, Geir B., Froyn, Camilla Brettville, Hovi, John, and Menz, Fredric C. (2006): "Regional Versus Global Cooperation for Climate Control," Journal of Environmental Economics and Management 51: 93-109.

Bagwell, Kyle, and Staiger, Robert W. (2016): "The design of trade agreements," Handbook of Commercial Policy 1: 435-529.

Barrett, Scott (1994): "Self-Enforcing International Environmental Agreements" Oxford Economic Papers 46, 878-894.

Barrett, Scott (1997): "The strategy of trade sanctions in international environmental agreements," Resource and Energy Economics 19(4): 345-61.

Barrett, Scott (2001): "International Cooperation for Sale," European Economic Review45 (10): 1835-50.

Barrett, Scott (2002): "Consensus Treaties," Journal of Institutional and Theoretical Economics, 158 (4): 529-547.

Barrett, Scott (2005): "The Theory of International Environmental Agreements," Handbook of Environmental Economics 3: 1458-93.

Barrett, Scott (2006): "Climate Treaties and "Breakthrough" Technologies," American Economic Review, papers and proceedings 96 (2): 22-5.

Battaglini, Marco, and Harstad, Bard (2016): "Participation and Duration of Environmental Agreements" Journal of Political Economy 124(1), 160-204.

Bellelli, Francesco S. , Ashar Aftab, and Riccardo Scarpa (2023): "The Participation Dilemma: A Survey of the Empirical Literature on International Environmental Agreement Ratification" , Review of Environmental Economics and Policy Volume 17, Number 1, Winter 2023.

Benchekroun, Hassan and v. Long, Ngo (2012): "Collaborative environmental management: A review of the literature", International Game Theory Review, Volume 14, Issue 04.

Black, J., Levi, M. D. and de Meza, D. (1993), Creating a Good Atmosphere: MinimumParticipation for Tackling the “Greenhouse Effect”, Economica 60, 281–293

Bolton, Patrick, and Dewatripont, Mathias (2005): "Contract Theory" The MIT Press.

Buchholz, Wolfgang, and Konrad, Kai (1994): "Global Environmental Problems and the Strategic Choice of Technology," Journal of Economics 60 (3): 299-321.

Buchholz, Wolfgang, and Sandler, Todd (2021): "Global public goods: A Survey", Journal of Economic Literature, 59(2), 488–545

Calvo, E., and SJ Rubio (2012): " Dynamic models of international environmental agreements: A differential game approach", International Review of Environmental and Resource Economics.

Caparrós, Alejandro (2016): "Bargaining and international environmental agreements", Environmental and Resource Economics Volume 65 Pages 5-31.

Carleton, Tamma; Duflo, Esther; Jack, Kelsey and Zappala, Guglielmo (2024): "Adaptation to Climate Change", in Handbook of the Economics of Climate Change, Volume 1, Elsevier B.V.

Carraro, Carlo and Siniscalco, Domenico (1993): "Strategies for the international protection of the environment," Journal of Public Economics 52 (3): 309-28.

Carraro, Carlo, Eyckmans, Johan, and Finus, Michael. (2006): "Optimal Transfers and Participation Decisions in International Environmental Agreements," The Review of International Organizations 1(4): 379-96.

Carraro, Carlo; Marchiori, Carmen and Oreffice, Sonia (2009): "Endogenous Minimum Participation in International Environmental Treaties," Environmental and Resource Economics42(3): 411-25.

Chung, T.-Y. (1991), "Incomplete Contracts, Specific Investment, and Risk Sharing", Review of Economic Studies, 58, 1031–1042.

Chwe, Michael Suk-Young (1994): "Farsighted Coalitional Stability" Journal of Economic Theory 63 (2), 299-325.

Colombo, Luca; Labrecciosa, Paola, and Van Long, Ngo (2022): "A dynamic analysis of international environmental agreements under partial cooperation" European Economic Review 143: 104036.

d'Aspremont, Claude; Jacquemin, Alexis; Gabszewicz, Jean, and Weymark, John (1983): "On the Stability of Collusive Price Leadership" Canadian Journal of Economics 16 (1), 17-25.

Deryugina, Tatyana (2022): "The Fiscal consequences of natural disasters" in: Skidmore, M. (eds) Handbook on the Economics of Disasters, Edward Elgar Publishing 208-228.

Diamantoudi, Effrosyni & Eftichios S. Sartzetakis (2018): "International Environmental Agreements—The Role of Foresight", Environmental and Resource Economics Volume 71, pages 241–257.

Dietz, Simon (2024): "Introduction to Integrated Climate-Economic Modeling" in Handbook of the Economics of Climate Change, Volume 1, Elsevier B.V.

Dixit, Avinash, and Olson, Mancur (2000): "Does voluntary participation undermine the Coase Theorem?" Journal of Public Economics 76 (3): 309-35.

Dockner, Engelbert J., and v. Long, Ngo (1993): "International Pollution Control: Cooperative versus Noncooperative Strategies", Journal of Environmental Economics and Management, Volume 25, Issue 1: 13-29

Dugoua, Eugenie (2023): "Induced Innovation and International Environmental Agreements: Evidence from the Ozone Regime" Review of Economics and Statistics 1-45.

Dutta, Prajit K. and Radner, Roy (2004): "Self-enforcing climate-change treaties," Proc. Nat. Acad. Sci. U.S., 101, 4746-4751.

Dutta, Prajit K. and Radner, Roy (2006): "A Game-Theoretic Approach to Global Warming," Advances in Mathematical Economics 8: 135-53.

Dutta, Prajit K. and Radner, Roy (2009): "A Strategic Analysis of Global Warming: Theory and Some Numbers," Journal of Economic Behavior & Organization 71 (2):

187-209.
Dutta, Prajit K., and Radner, Roy (2020): "The Paris Accord Can Be Effective if the Green Climate Fund is Effective," mimeo, New York University.
Edlin, Aaron S., and Reichelstein, Stefan (1996): "Holdups, Standard Breach Remedies, and Optimal Investment," American Economic Review86 (3): 478-501.
Eichner, Thomas, and Kollenbach, Gilbert (2023): "Participation and duration of environmental agreements: investment lags matter", Journal of the Association of Environmental and Resource Economists 10.1: 33-53.
Engwerda, Jacob. (2005): LQ Dynamic Optimization and Differential Games, Wiley.
Finus, Michael and Maus, Stefan (2008): "Modesty May Pay!" JPET 10(5): 801-26.
Golombek, Rolf, and Hoel, Michael (2005): "Climate Policy under Technology Spillovers," Environmental and Resource Economics 31 (2): 201-27.
Golombek, Rolf, and Hoel, Michael (2006): "Second-Best Climate Agreements and Technology Policy," Advances in Economic Analysis & Policy 6 (1), Article 1.
Grossman, Gene M. (2016): "The purpose of trade agreements," in K. Bagwell and R. Staiger (Eds.), Handbook of Commercial Policy 1: 379-434.
Harris, Milton, and Holmstrom, Bengt (1987): "On The Duration of Agreements," International Economic Review 28(2): 389-
Harstad, Bård (2006): "Flexible Integration? Mandatory and Minimum Participation Rules," Scandinavian Journal of Economics 108(4): 683-702.
Harstad, Bård (2012): "Climate Contracts: A Game of Emissions, Investments, Negotiations, and Renegotiations" Review of Economic Studies 79, 1527-1555.
Harstad, Bård (2016): "The Dynamics of Climate Agreements" Journal of the European Economic Association 14(3): 719–752.
Harstad, Bård (2023a): "Pledge-and-Review Bargaining," Journal of Economic Theory 207, 105574.
Harstad, Bård (2023b): "Pledge-and-Review Bargaining: from Kyoto to Paris" The Economic Journal 133 (651), 1181-1216.
Harstad, Bård (2025): "The Politics of Global Public Goods", forthcoming in *Handbook of Political Economy*, Elsevier.
Harstad, Bard; Lancia, Francesco, and Russo, Alessia (2019): "Compliance Technology and Self-enforcing Agreements" Journal of the European Economic Association 9 17(1):1–29.
Harstad, Bard; Lancia, Francesco, and Russo, Alessia (2022): "Prices vs. quantities for self-enforcing agreements" Journal of Environmental Economics and Management 111, 102595.
Hart, Oliver (1995): "Firms, Contracts, and Financial Structure" Clarendon Press.
Helm, Carsten, and Schmidt, Robert C. (2015): "Climate cooperation with technology investments and border carbon adjustment", European Economic Review 75, 112-130.
Hoel, Michael (1992): "International environment conventions: The case of uniform reductions of emissions," Environmental and Resource Economics 2 (2): 141-59.
Hoel, Michael, and Kerstin Schneider (1997): “Incentives to Participate in an International Environmental Agreement.” Environmental and Resource Econ. 9: 153–70.
IPCC. (2014). Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge: Cambridge University Press.

Jaffe, Adam B., Newell, Richard G., and Stavins, Robert N. (2003): "Technological Change and the Environment," Handbook of Environmental Economics Volume 1 (Edited by K.-G. Mäler and J.R. Vincent), Elsevier.

Karp, Larry, and Simon, Leo (2013): "Participation games and international environmental agreements: A non-parametric model" Journal of Environmental Economics and Management 65 (2), 326-344.

Keohane, Robert, and Oppenheimer, Michael (2016): "Paris: Beyond the Climate Dead End through Pledge and Review?" Politics and Governance 4(3), pp. 42–51.

Kerr, Suzi; Lippert, Steffen, and Lou, Edmund (2024): "Transfers in climate action teams", Economic Theory

Kolstad, Charles D. and Toman, Michael. (2005): The Economics of Climate Policy, Handbook of Environmental Economics 3: 1562-93.

Kotchen, Matthew (2020): "On the scope of climate finance to facilitate international agreement on climate change" Economics Letters 190, 109070.

Kotchen, Matthew (2024): "Climate Policy Options" in Handbook of the Economics of Climate Change, Volume 1, Elsevier B.V.

Kováč, Eugen, and Robert C. Schmidt (2012): "A simple dynamic climate cooperation model", Journal of Public Economics Volume 194, February, 104329.

Levhari, David and Mirman, Leonard J. (1980): "The Great Fish War: An Example Using a Dynamic Cournot-Nash Solution," Bell Journal of Economics 11 (1): 322-34.

Li, Hongxiu, and Rus, Horatiu A. (2019): "Climate change adaptation and international mitigation agreements with heterogeneous countries" Journal of the Association of Environmental and Resource Economists 6.3: 503-530.

Maggi, Giovanni (2014): "International trade agreements," in G. Gopinath, E. Helpman, and K. Rogoff (eds.), Handbook of International Economics 4: 317-90. Elsevier.

Maggi, Giovanni (2016): "Issue linkage," in K. Bagwell and R. Staiger (Eds.), Handbook of commercial policy Vol. 1, pp. 513-64.

Mailath, George J., and Samuelson, Larry (2006): "Repeated games and reputations: long-run relationships," Oxford University Press

Martimort, David and Wilfried Sand-Zantman (2016): "A mechanism design approach to climate-change agreements", Journal of the European Economic Association 14 (3): 669-718.

Newell, Richard G., Jaffe, Adam B., and Stavins, Robert N. (2006): "The Effects of Economic and Policy Incentives on Carbon Mitigation Technologies," Energy Economics 28: 563-78.

Nordhaus, William (2015): "Climate clubs: Overcoming free-riding in international climate policy," American Economic Review 105(4): 1339-70.

Okada, Akira (2023): "A dynamic climate negotiation game achieving full cooperation" Journal of Economic Behavior and Organization 214: 657-669.

Osmani, Dritan, and Tol, Richard (2009): "Towards farsightedly stable international environmental agreements" Journal of Public Economic Theory 11(3), 455–492.

Palfrey, Thomas R., and Rosenthal, Howard (1984): "Participation and the provision of discrete public goods: a strategic analysis," Journal of Public Economics 24 (2): 171-93.

Ploeg, Frederick van der, and de Zeeuw, Aart (1991): "A differential game of international pollution control," Systems & Control Letters 17 (6): 409-14.

Ploeg, Frederick van der, and de Zeeuw, Aart (1992): "International aspects of pollution control," Environmental and Resource Economics 2 (2): 117-39.

Ray, Debraj (2007): "A game-theoretic perspective on coalition formation" Oxford University Press

Ray, Debraj, and Vohra, Rajiv (1999): " A theory of endogenous coalition structures" Games and Economic Behavior 26 (2), 286-336.

Ray, Debraj and Vohra, Rajiv (2001): "Coalitional Power and Public Goods," Journal of Political Economy 109 (6): 1355-1384.

Schmidt, Robert C., and Strausz, Roland (2015): "On the Timing of Climate Agreements" Environmental and Resource Economics

Schopf, Mark (2024): "Self-enforcing international environmental agreements and altruistic preferences" Environmental and Resource Economics 1-51.

Segal, Ilya, Whinston, Michael D. (2002):"The Mirrless Approach to Mechanism Design With Renegotiation." Econometrica 70 (1): 1-45.

Tirole, Jean (2017): "Economics for the Common Good" Princeton University Press

Vosooghi, Sareh; Arvaniti, Maria, and van der Ploeg, Frederick (2024): "Climate Coalitions with Sophisticated Policy Makers" CESifo Working Paper No. 9768.

Wagner, Ulrich (2016): "Estimating strategic models of international treaty formation," The Review of Economic Studies 83 (4), 1741-1778.

Weikard, H.-P., Wangler, L., and Freytag, A. (2015). Minimum participation rules with heterogeneous countries. Environ. Resour. Econ. 62, 711–727. doi: 10.1007/s10640-014-9861-1.