Regional Initiatives and the Cost of Delaying Binding Climate Change Agreements*

Julien Beccherle†
Jean Tirole ‡

December 20, 2010

Abstract

The Kyoto and Copenhagen Protocols on climate change mitigation postponed the specification of binding commitments to a future negotiation. This paper analyzes the strategic implications of delayed negotiations. While, as is well-understood, the incentive to free ride leads to excessive emissions prior to a binding agreement, the cost of delay is magnified by players' attempt to secure a favorable bargaining position in the future negotiation. A "brinkmanship", an "effort substitution", and a "raising rival's cost" effects all concur to generate high post-agreement emissions. The paper applies this general insight to the issuance of forward or bankable permits.

Keywords: International negotiations, climate change, cap and trade, bankable permits.

JEL numbers: D62, F51, H23, Q52.

*The authors are grateful to participants at the World Congress of Environmental and Resource Economics (Montreal, 2010) and at the College de France conference on Managing Global Warming (Paris, 2010), Sören Blomquist, Dominique Bureau, Dominik Grafenhofer, Bill Nordhaus, Jean-Pierre Ponssard, Nick Stern, two anonymous referees and especially Bard Harstad for very helpful comments on an earlier draft.

†Ministère de l'Ecologie, de l'Energie, du Développement Durable, et de la Mer [julien.beccherle@developpement-durable.gouv.fr]

‡Toulouse School of Economics [jean.tirole@tse-fr.eu]

1 Introduction

Climate change is a global issue in need of a global answer. The first attempt at an integrated approach to mitigation, the 1997 Kyoto Protocol, did not specify firm commitments from ratifying countries to cut their emissions. The 2009 Copenhagen conference was meant to define the contours of the post-Kyoto world. Despite widespread agreement on the urgency to act, Copenhagen did not deliver an agreement with binding commitments to emissions cuts, and left the definition of objectives and their verification to future negotiations. A low-aim deal, drafted by a small group of countries led by the US, China, Brazil, South Africa and India, had to be struck in a hurry to avoid returning home empty handed. It was only "taken note of" by the assembly.

The failure of the Copenhagen negotiation to deliver a legally binding commitment to emissions reductions beyond 2012 (the year of expiration of the Kyoto Protocol), has multiple origins. First, the lack of measurement and enforcement protocols and of consensus on instruments made it difficult to even design a sustainable agreement. Second, political will was lacking; indeed no draft had been seriously discussed by heads of states prior to the conference. Finally, the negotiation revealed a high level of distrust among countries.

The paper argues that extending the waiting game until Kyoto 3 (say, 2020) would have serious consequences, that go well beyond the celebrated free-riding incentive. Namely, not only will countries engage in suboptimal efforts to reduce their emissions in the next ten years, but they will also consider how their behavior will impact the outcome of negotiations in 2020.[1]

We consider a two-period framework. In period 1, each region of the world chooses a public policy, anticipating the negotiation of a global agreement at date 2. In the generic version of the model, this policy refers to any instrument that impacts the region's date-2 welfare: It may determine its date-2 technological feasibility set, its installed base of polluting equipments, or, in a key application of our theory, the domestic allocation of property rights on pollution allowances. The key feature of the date-1 policy choice is that it affects the region's marginal cost of date-2 abatement, which in turn implies that the region's date-1 choice of public policy is made with an eye on the future negotiation.

[1]Of course there will be some progress. Carbon permits markets exist or may be created in Europe, the US, Japan and some other developed economies. Emerging countries are taking some action as well. A mixture of collateral damages (the emission of SO_2, a local pollutant, jointly with that of CO_2 by coal plants), the direct impact of own production of CO_2 for large countries like China, and the desire to placate domestic opinion and avoid international pressure will all lead to some carbon control.

The paper's first contribution is to investigate the exact nature of the resulting commitment effect. A natural benchmark is regional optimization or cost minimization. Because the date-1 policy affects the date-2 incentives for abatement, regional optimization aims at minimizing the total (intertemporal) cost to the region. Cost minimization obtains both in the first best, in which a binding agreement is reached at date 1, and in the complete absence of negotiation.

Under delayed negotiation, two new effects concur to push date-2 emissions up. A decrease in one's marginal incentive to abate first implies that the region would pollute more, were the negotiation to fail. The "brinkmanship effect" works through a reduction in the *other* region's payoff when the negotiation fails; it changes the region's threat point in the negotiation and enables it to extract more of the surplus; it is particularly potent when the region has substantial bargaining power in the negotiation. Second, the outcome in the negotiation depends on *one's own* welfare, were the negotiation to fail. Thus a region's date-1 policy is in part guided by its own welfare when negotiations break down at date 2. Because emissions are higher when negotiations break down, the country can afford a lower date-1 investment in pollution control; this "intertemporal substitution effect" by contrast is most potent when the country's bargaining power is weak. These two effects both imply that a delay in negotiating a global agreement increases post-negotiation pollution and not only the pre-negotiation one.

When environmental damage costs are convex, a third strategic effect arises, that reinforces the other two: By committing to a higher pollution level, a region raises the marginal damage cost of all regions and therefore induces others to cut down on their emissions. This is the "raising the other's marginal environmental cost" effect, or "raising rival's cost" effect for short.

We show that delaying negotiation always raises *date-2* emissions compared to the first-best. Indeed, it may even be the case that a delayed negotiation induces more date-2 emissions than in the complete absence of negotiation, and this despite the reduction in emissions brought about by the date-2 agreement. Thus, delayed negotiation may be worse then no negotiation at all.

The paper's second contribution is to apply these generic insights to the issuance of future allowances and to the bankability of pollution allowances (as embedded in the Waxman-Markey bill). In particular, we predict that regions will issue too many forward or bankable allowances.

Proponents of regional cap-and-trade systems, such as the one existing in Europe or those that are/were under consideration in the United States and a number of other developed countries, take the view that regional markets will jump-start climate change mitigation and later lead to convergence to a single, worldwide climate treaty. We investigate the strategic implications of a delayed agreement in the context of the issuance of forward allowances; we also show that similar insights apply to the issuance of bankable permits. In either case, the region puts today into private hands allowances that can be used to cover future emissions.

Focusing on linear environmental damage costs for expositional simplicity, we first show that, following the logic of the Coase conjecture, regions issue forward allowances neither in the complete absence of negotiation nor in the first best (in which negotiation takes place at date 1). By contrast, in the intermediate situation in which negotiations are delayed, the brinkmanship and intertemporal substitution effects both imply that regions issue forward allowances whenever date-2 emissions increase with the number of such allowances. The latter property holds for example if the regions auction off date-2 new (spot) allowances and face a shadow cost of public funds; because regions put some weight on revenue, they have a tendency to "over issue" spot permits at date 2, thereby lowering the price of carbon. Alternatively, the government could internalize only partially the welfare of the holders of forward permits. Or else, the new permits could be distributed for free at date 2, but the government might at date 2 negotiate domestically with a powerful industrial lobby, whose status-quo welfare is stronger, the larger the number of allowances it received at date 1.

Either way, delayed negotiations lead to high future emissions through an excessive issuance of forward or bankable permits, even though they can be retired or fewer spot permits issued. A political implication of our analysis is that if an ambitious climate treaty is impossible today, countries should at least agree to limit banking and forward-selling.[2] Further, we show that markets are merged under symmetrical conditions, but that the agreement may otherwise content itself with a specification of the volume of emissions in maintained regional markets.

The paper is organized as follows: Section 2 sets up the generic model. Section 3 identifies the brinkmanship, intertemporal substitution, and raising rival's cost effects, and derives and illustrates the excess pollution result. Section 4 develops the application to forward allowances and bankable pollution permits. Section 5 concludes.

[2] Of course bankability has the (well-known) benefit of smoothing the carbon price when transitory shocks to economic activity or technological progress would make this price very volatile.

Literature review and contribution

This paper builds on several literatures. First, the forward-market application relates to, among others, Allaz and Vila (1993) and Mahenc and Salanié (2004), who investigate the idea that forward markets can be used to influence rivals. Allaz and Vila find that if competing firms set quantities of forward rights over time, then they will sell more than they would in a spot market. Selling one unit today reduces the competitors' future marginal benefits and therefore increases the firm's profit. In much of the paper, we abstract from the Allaz-Vila (raising rival's cost) effect by assuming that marginal pollution damages are linear in total pollution and focus on the impact of negotiation, a question that is moot in Allaz and Vila's oligopoly framework. Laffont and Tirole (1996a,b) study cap-and-trade policies with spot and forward markets and analyze how regulatory commitment and flexibility to news can be made consistent. Like Allaz-Vila, these papers however do not consider negotiations among multiple regulators/countries, which is the focus of the current paper.[3]

Second, by taking the view that negotiations take time, we implicitly study the role of incomplete contracts and the importance of property rights allocations in a context of global externalities. Our paper thereby builds on the incomplete contracts and hold up literature (Grossman and Hart (1986), Hart and Moore (1990), Williamson (1985)). The common thread with that literature is the idea that ex-ante investments influence the outcome of ex-post negotiations. Our paper does not add to the theoretical body of knowledge on incomplete contracting. Rather, it applies it to the environmental negotiations context, and unveils the brinkmanship and intertemporal substitution effects as well as a new raise-your-rival's cost effect. And it contains a novel application to the design of emissions trading systems.

Third, this paper contributes to the growing literature on climate change agreements[4] and casts some light on the effects of delayed negotiations on international climate change agreements. Our contribution is to include dynamics into the analysis and to put forward the potential cost of delayed negotiations. An early paper emphasizing the potential of R&D reduction when investments can be held up in future renegotiations is Buchholz and Konrad (1994). The most closely related paper is Harstad (2009) who develops a very interesting analysis of the dynamics of climate change agreements. Harstad studies an

[3] There is also a growing literature on closed-economy, optimal dynamic multi-instrument policies (for example, carbon price and R&D subsidies, as in Acemoglu et al (2009) and Grimaud and Rouge (2008)).

[4] See e.g., Aldy and Stavins (2007) and Barrett (2005). The paper also builds on the extensive literature on coalition formation in international agreements (e.g., Carraro and Siniscalco (1993)).

economy where sovereign countries repeatedly make investment and emission decisions. The cost of pollution depends on the total stock of emissions, accumulated over time. Harstad demonstrates, as we do, that under some conditions short-lasting agreements lead to higher pollution than no agreement at all. Notably, he establishes this result in an infinite horizon model while we have only two periods.

Under some assumptions,[5] Harstad's state space, which a priori has $(n+1)$ dimensions (the existing stock of pollution, together with the technological state of each the n countries), collapses to a two-dimensional recursive one, in which outputs are fully determined by the existing stock of pollution and the latter's evolution depends on these outputs and the global stock of knowledge. We chose to be more general in terms of technologies and to allow for asymmetries (in particular, in our model, which country is more technology advanced impacts the continuation equilibrium) at the expense of a two-period analysis. This added generality allows us to investigate interesting policy instruments such as forward and bankable allowances, which the Harstad assumptions cannot capture, as well as the consequences of asymmetric preferences and bargaining powers. We also identify the three effects at stake.

2 A generic model

2.1 Timing and utility functions

The model has two periods, $t = 1,\ 2$ and two regions of the world, $i = A,\ B$.[6] For notational simplicity, and without loss of generality, we normalize date-2 payoffs so that regions do not discount the future. In period 1, regional authorities (regulators) non-cooperatively and simultaneously choose single-dimensional[7] policy variables, a_1^i for region i. At this stage $a_1^i \in \mathbb{R}$ may stand for any date-1 instrument available to the regulator that affects the region's date-2 incentive to emit pollutants. We normalize a_1^i such that a higher a_1^i is *less* environment-friendly. For example, a high a_1^i may correspond to a lax pollution standard, a low investment in green technology or a high level of issuance of forward allowances. Thus a_1^i in general is not date-1 pollution, although the case of costly adjustment (in which a_1^i is indeed date-1 pollution and the region incurs an adjustment cost related to the distance between a_1^i and a_2^i) fits our framework.

[5]See Section 3.3 for a statement of these assumptions.

[6]The analysis is generalized to an arbitrary number of regions in Section 3.1.

[7]The results carry over to a multi-dimensional date-1 action space, provided that the standard supermodularity assumptions are satisfied (see e.g., Milgrom and Roberts (1990)).

The negotiation with the other region is delayed until period 2 (see the discussion below). The absence of date-1 agreement implies that the choices of date-1 emissions are driven by the familiar free-riding incentive. Without loss of generality and unless otherwise stated,[8] we omit the corresponding analysis for clarity of exposition. At date 2, the regulators will agree on second-period pollution levels $a_2^i \in \mathbb{R}^+$ in both regions and on a side payment, for example through an allocation of pollution allowances or direct cash transfers. While a_1^i can stand for any (environment-unfriendly) policy that changes region i's date-2 emissions incentives, a_2^i denotes the actual level of date-2 emissions.[9]

Regions' date-1 and date-2 welfares, gross of pollution damages, are denoted $U_1^i(a_1^i)$ and $U_2^i(a_1^i, a_2^i)$. These functions are increasing in a_1^i and a_2^i respectively, and twice differentiable. U_2^i is twice differentiable and strictly concave in a_2^i. Region i's pollution damage depends linearly[10] on the total amount of emissions at date 2; if c^i is region i's marginal damage, then its environmental cost is $c^i a_2$ where $a_2 \equiv a_2^A + a_2^B$. Let $c \equiv c^A + c^B$ denote the total marginal damage cost.

Assumption 1 *(i) The marginal utility of polluting in the second period is decreasing with the first-period pollution control:*

$$\frac{\partial^2 U_2^i}{\partial a_1^i \partial a_2^i} > 0.$$

(ii) Let

$$\Gamma^i(a_1^i) \equiv \frac{dU_1^i}{da_1^i}(a_1^i) + \frac{\partial U_2^i}{\partial a_1^i}\Big(a_1^i, \widehat{a}_2^i(a_1^i)\Big) \quad \textit{for all } a_1^i$$

where

$$\widehat{a}_2^i(a_1^i) \equiv arg\ max\{U_2^i(a_1^i, a_2^i) - ca_2^i\}.$$

There exists a_1^{iFB} such that $\Gamma^i(a_1^{iFB}) = 0$, $\Gamma^i(a_1^i) > 0$ if $a_1^i < a_1^{iFB}$ and $\Gamma^i(a_1^i) < 0$ if $a_1^i > a_1^{iFB}$.

Condition (i) simply states that a lax environmental policy at date 1 (a higher a_1^i) raises region i's marginal cost of pollution abatement, or equivalently raises its marginal utility

[8] Date-1 and date-2 emissions are interdependent in the case of bankable allowances (Section 4.2). There, we will formally introduce date-1 pollution.

[9] As we will later discuss, this is unrestrictive since any date-2 pollution-control policy corresponds to a unique level of emissions.

[10] We make the assumption that the environmental costs depend linearly on the total amount of pollution for simplicity, but we show in Section 3 that the results obtained in the linear case can be extended to a general environmental cost function. The benefit of assuming linear damage functions is that we rule out Allaz-Vila style effects.

of emissions. Delaying pollution abatement then raises the cost of pollution mitigation at a latter stage.[11] This assumption just captures the notion of a "lax environmental policy" in our model.

In the following, "hats" refer to the values of the parameters that emerge from an efficient date-2 agreement, while "star" superscripts correspond to the Nash outcome that would result from a failure to agree. Thus, $\widehat{a}_2^i$ denotes region i's second-period pollution control after the agreement. Condition (ii) is a quasi-concavity condition in which the jointly efficient date-2 reaction to the first period policy, $\widehat{a}_i^2(a_i^1)$, is factored in; this assumption will guarantee the uniqueness of the first-best policy a_1^{iFB} and allow us to sign the bias induced by delayed negotiations. We will check that Assumption 1 is satisfied in all our applications.

Were the negotiation to break down at date 2, regional regulators would choose non-cooperatively second-period pollution levels, a_2^{i*} for region i. This benchmark defines the outside options in the date-2 negotiation. Both $\widehat{a}_2^i$ and a_2^{i*} are functions of region i's prior policy.

Example: investments in green equipments.

Authorities start with a mass 1 of (existing or potential) brown equipments or buildings. In each period, they choose their investment levels: in period 1, region i's regulator makes a fraction X^i of its equipments pollution-free for date 2 (again, we neglect date-1 pollution). Alternatively, X^i might be the fraction of green buildings or equipments built at date 1 and becoming operational at date 2. We assume that the region has some technology for revamping, summarized by a cost function $\phi^i(X^i)$ which is increasing and convex, with $\phi^{i\prime}(0) = 0$ and $\phi^{i\prime}(1) = +\infty$. Similarly, let Y^i denote the fraction of revamped equipment after the date-2 investment (so $1 - Y^i$ is the remaining fraction of "brown" buildings or equipments). The date-2 investment costs $\phi^i(Y^i - X^i)$.

Letting $a_2^i \equiv 1 - Y^i$ and $a_1^i \equiv 1 - X^i$, the first-period and second-period utilities U_1^i and U_2^i are equal to (minus) the investment costs: $U_1^i(a_1^i) = -\phi^i(1 - a_1^i)$ and $U_2^i(a_1^i, a_2^i) = -\phi^i\Big(a_1^i - a_2^i\Big)$.

Since ϕ^i is convex, $U_2^i(a_1^i, a_2^i)$ is concave in a_2^i. For given (a_1^i, a_1^j), $a_2^{i*}(a_1^i)$ and $\widehat{a}_2^i(a_1^i)$

[11] Think for instance of electric utilities that invest massively today in coal-fired power plant: this investment choice will have long lasting consequences on the ability to cheaply abate pollution.

are uniquely defined. We can check that Assumption 1 is satisfied.[12]

Discussion (delayed negotiation).

Our approach follows the large literatures on incomplete contracts and on repeated contracting in assuming that no complete long-term contract is signed at date 1. Regions may initially fail to agree for different reasons. First, the compliance environment may not yet be in place; for example, the Kyoto and Copenhagen negotiations were hampered by the absence of a reliable measurement of emissions and by uncertainty as to how these would be measured in the future. Furthermore, there was no consensus on the use of economic instruments and on the nature of enforcement. Second, if a region is governed by a political party opposed to international negotiations on climate change, the other region may choose to wait for this government to be defeated and replaced by a more favorable government before entering real negotiations. Compensating the reluctant government is too expensive and so the environmentally-concerned region prefers to wait until an equally environmentally-concerned government is installed in the other country to negotiate.[13]

Third, another possible explanation is an initial asymmetry of information between regions, leading to a bargaining breakdown in early stages. The simplest such situation consistent with our modelling has the regions hold private information about some fixed

[12]First, differentiating U_2^i with respect to a_1^i and a_2^i

$$\frac{\partial^2 U_2^i}{\partial a_2^i \partial a_1^i} = \phi^{i\prime\prime}\left(a_1^i - a_2^i\right) > 0$$

since ϕ^i is convex. Second, from the first-order condition in the negotiated case

$$\phi^{i\prime}\left(a_1^i - \widehat{a}_2^i(a_1^i)\right) = c.$$

Thus

$$\Gamma^i(a_1^i) = \phi^{i\prime}\left(1 - a_1^i\right) - \phi^{i\prime}\left(a_1^i - \widehat{a}_2^i(a_1^i)\right) = \phi^{i\prime}\left(1 - a_1^i\right) - c.$$

Since $\phi^{i\prime}(0) = 0$ and $\phi^{i\prime}(1) = \infty$, and since ϕ^i is convex, the function Γ^i is decreasing and admits a zero. Thus Assumption 1 is satisfied.

[13]To illustrate the impact of asynchronized political agendas in a simpler environment than ours, take two periods (1 and 2) and two symmetric regions (A and B), and consider a pro-environment policy that creates gross per-region utility u relative to its no-agreement utility. However, the date-1 government of region A has valuation $\widehat{u} < 0$ for this policy, unlike the date-2 government who will value it at u. Region B values it at u at both dates. The discount factor is δ. If no agreement exists by date 2, Nash bargaining then will ensure that each region receives u. Anticipating an agreement at date 2, the date-1 government of region A is willing to take a pro-environment stance at date 1 if it receives compensation T from region 2 such that $[\widehat{u} + T] \geq \delta\widehat{u}$. Similarly, region 2 is willing to accept an agreement if $[u - T] \geq \delta u$. Thus, if $(1-\delta)(u + \widehat{u}) < 0$, i.e., $u + \widehat{u} < 0$, no agreement is signed at date 1.

component of their utility function, such as some infra-marginal cost of implementing a pro-environment policy. Provided that this information becomes common knowledge before date 2, it can create a breakdown of negotiation at date 1 without impacting any of the analysis below. In general, though, private information relates to the marginal cost of abatement or the region's tolerance to emissions; the analysis is then more complex than depicted in the paper, as it involves screening/signaling, but the basic forces unveiled in the paper are robust. Fourth, there may be an agreement at date 1, but some unanticipated loopholes allow regions to deviate from its letter and spirit.[14] Our analysis then refers to the degrees of freedom involuntarily allowed by the date-1 agreement.

2.2 Welfare functions and date-2 bargaining

Letting T^i denote the transfer received by region i as part of the date-2 agreement ($T^A + T^B = 0$), the second-period welfares of the world and of region i, W_2 and W_2^i, respectively, can be written as:

$$W_2(a_2^A, a_2^B, a_1^A, a_1^B) \equiv U_2^A(a_1^A, a_2^A) + U_2^B(a_1^B, a_2^B) - c(a_2^A + a_2^B) \tag{1}$$

and

$$W_2^i(a_2^i, a_2^j, a_1^i) + T^i \equiv U_2^i(a_1^i, a_2^i) - c^i(a_2^i + a_2^j) + T^i. \tag{2}$$

Let, for given date-1 actions,

$$\widehat{W}_2(a_1^A, a_1^B) \equiv W_2(\widehat{a}_2^A(a_1^A), \widehat{a}_2^B(a_1^B), a_1^A, a_1^B),$$

and

$$W_2^{i^*}(a_1^i, a_1^j) \equiv W_2^i(a_2^{i^*}(a_1^i), a_2^{j^*}(a_1^j), a_1^i),$$

where

$$a_2^{i^*}(a_1^i) \equiv \arg\max\{U_2^i(a_1^i, a_2^i) - c^i a_2^i\}$$

and

$$\widehat{a}_2^i(a_1^i) \equiv \arg\max\{U_2^i(a_1^i, a_2^i) - c a_2^i\} < a_2^{i^*}(a_1^i),$$

using revealed preference: A successful negotiation reduces date-2 emissions for any given date-1 policies.[15] Revealed preference also implies that the functions $\widehat{a}_2^i$ and $a_2^{i^*}$ are monotonic in the first-period effort:

[14]See Tirole (2009) for an analysis of pre-contractual cognition and its consequences for contract design and incompleteness.

[15]Interestingly, the European Union ETS price fell by 21% in 2009 as it became more and more unlikely that a satisfactory agreement would be drawn. While some of this decrease may be due to news about the economic recession, the 9% drop in price immediately after the Copenhagen Accord is a clear sign.

Lemma 1. *$\widehat{a}_2^i$ and ${a_2^i}^*$ are non-decreasing (strictly increasing if $c^i > 0$ for all i) functions of a_1^i. Furthermore, for all a_1^i, ${a_2^i}^*(a_1^i) > \widehat{a}_2^i(a_1^i)$.*

Thus, the more stringent the first-period pollution control policy, the lower the second-period pollution. In particular, applied to the case in which negotiations are delayed, by adopting loose pollution control policies in the first-period a region can credibly commit to high date-2 pollution, were the negotiation to break down. This commitment is a key element of our analysis.

Date-2 bargaining: At date 2 the two regulators agree on their pollution levels and on a side payment. We model the bargaining outcome by the Nash bargaining solution.[16] Calling α^A and α^B the bargaining powers of regions A and B (with $\alpha^A + \alpha^B = 1$),[17] region $i \in \{A,\ B\}$'s intertemporal payoff after negotiation, W_{neg}^i, is given by:

$$W_{neg}^i(a_1^i, a_1^j) = U_1^i(a_1^i) + W_2^{i*}(a_1^i, a_1^j) + \alpha^i\left(\widehat{W}_2(a_1^i, a_1^j) - \left(W_2^{A*}(a_1^i, a_1^j) + W_2^{B*}(a_1^i, a_1^j)\right)\right) \quad (3)$$

The term $U_1^i(a_1^i)$ in (3) is the date-1 utility associated with choosing policy a_1^i. This term is fixed (sunk) at date 2. $W_2^{i*}(a_1^i, a_1^j)$ is the outside option of region i, while

$$\alpha^i\left(\widehat{W}_2(a_1^i, a_1^j) - \left(W_2^{A*}(a_1^i, a_1^j) + W_2^{B*}(a_1^i, a_1^j)\right)\right)$$

is the share of the total surplus extracted by region i during the negotiation.

2.3 Benchmarks: first best and autarky

Before tackling the case of delayed negotiations, let us look at the two polar cases in which negotiations take place at both dates (first best)[18] or never take place (autarky).

First best: The pollution levels agreed upon during the negotiation $(\widehat{a}_2^A,\ \widehat{a}_2^B)$ are optimal given the first-period choices $(a_1^A,\ a_1^B)$. From the envelope theorem, the first best can be obtained by differentiating

$$U_1^A(a_1^A) + U_1^B(a_1^B) + U_2^A(a_1^A, a_2^A) + U_2^B(a_1^B, a_2^B) - c(a_2^A + a_2^B)$$

[16] They are other approaches to modeling bargaining. It seems reasonable to assume that the outcome of a negotiation depends at least partially on the outside options available to the parties.

[17] Bargaining theory has so far had relatively little to say about the determinants of bargaining power, and so it is reasonable to allow arbitrary sharing coefficients in our theory.

[18] It does not matter whether the two regions negotiate at date 1 a long-term agreement for the two periods, or negotiate at each date t an agreement specifying the countries' short-term policies.

with respect to $\{a_1^A, a_1^B\}$, and evaluating it at $\{\widehat{a}_2^A,\ \widehat{a}_2^B\}$. This gives, for $i \in \{A, B\}$:

$$\Gamma^i(a_1^i) \equiv \frac{dU_1^i}{da_1^i}(a_1^i) + \frac{\partial U_2^i}{\partial a_1^i}\left(a_1^i, \widehat{a}_2^i(a_1^i)\right) = 0 \tag{4}$$

As announced in Assumption 1, let a_1^{iFB} denote the unique value of a_1^i satisfying this equation.

Autarky: The outcome when there is never any negotiation is given by a similar regional optimization equation:

$$\frac{dU_1^i}{da_1^i}(a_1^i) + \frac{\partial U_2^i}{\partial a_1^i}\left(a_1^i, {a_2^i}^*(a_1^i)\right) = \Gamma^i(a_1^i) + \left(\frac{\partial U_2^i}{\partial a_1^i}(a_1^i, {a_2^i}^*(a_1^i)) - \frac{\partial U_2^i}{\partial a_1^i}(a_1^i, \widehat{a}_2^i(a_1^i))\right) = 0. \tag{5}$$

Let ${a_1^i}^* > a_1^{iFB}$ denote the solution to equation (5), which we will assume is unique.

3 The cost of delaying negotiations

3.1 The intertemporal substitution and brinkmanship effects

Suppose now that regions negotiate only at date 2. We assume that the game with payoff functions $\{W_{neg}^i(a_1^i, a_1^j),\ W_{neg}^j(a_1^i, a_1^j)\}$ admits a unique Nash equilibrium, $\{a_1^{iDN},\ a_1^{jDN}\}$, where "$DN$" stands for "Delayed Negotiation". Before proving our main result, we identify more precisely the effects mentioned above. Region i chooses a_1^i so as to maximize W_{neg}^i as defined by equation (3). Note that, while W_{neg}^i depends on a_1^j, its derivative with respect to a_1^i does not. From the envelope theorem, the first-order condition is:

$$\begin{aligned}\frac{\partial W_{neg}^i}{\partial a_1^i}(a_1^i) = &\left(\frac{dU_1^i}{da_1^i}(a_1^i) + \frac{\partial U_2^i}{\partial a_1^i}(a_1^i, \widehat{a}_2^i(a_1^i))\right) \\ &+ (1-\alpha^i)\left(\frac{\partial U_2^i}{\partial a_1^i}(a_1^i, {a_2^i}^*(a_1^i)) - \frac{\partial U_2^i}{\partial a_1^i}(a_1^i, \widehat{a}_2^i(a_1^i))\right) \\ &+ \alpha^i c^j \frac{d{a_2^i}^*}{da_1^i}(a_1^i) = 0.\end{aligned} \tag{6}$$

The regional optimization term is

$$\left(\frac{dU_1^i(a_1^i)}{da_1^i} + \frac{\partial U_2^i}{\partial a_1^i}(a_1^i, \widehat{a}_2^i(a_1^i))\right)$$

and is nil at $a_1^i = a_1^{iFB}$.

The intertemporal substitution term is given by

$$(1-\alpha^i)\left(\frac{\partial U_2^i}{\partial a_1^i}(a_1^i, a_2^{i*}(a_1^i)) - \frac{\partial U_2^i}{\partial a_1^i}(a_1^i, \widehat{a}_2^i(a_1^i))\right).$$

It is positive since $a_2^{i*}(a_1^i) > \widehat{a}_2^i(a_1^i)$ for all a_1^i (the negotiation leads to less pollution) and $\partial^2 U_2^i(a_1^i, a_2^i)/\partial a_1^i \partial a_2^i > 0$. Furthermore, it is larger, the smaller is α^i. A region with a low bargaining power behaves as if it anticipated autarky, and therefore a high date-2 emissions level.

The brinkmanship term is given by

$$\alpha^i c^j \frac{da_2^{i*}}{da_1^i}(a_1^i).$$

Because a_2^{i*} increases with a_1^i from Lemma 1, this term is always positive. It is larger when α^i is large: A country with substantial bargaining power is able to extract most of the surplus created by an agreement and therefore benefits from lowering the other party's outside option. Also the brinkmanship effect is more important, the larger the other region's environmental cost. Indeed, if region j is sensitive to pollution (c^j is large), then region i's threat is all the more effective.

And so

$$\Gamma^i(a_1^{iDN}) < 0$$

Assumption 1 (ii) then implies that $a_1^{iDN} > a_1^{iFB}$. In equilibrium there is an overprovision of a_1^i relative to the first-best level:

Proposition 1. *Delayed negotiations raise the post-negotiation pollution by encouraging lax environmental policies a_1^{iDN} prior to the negotiation: For all i, $a_1^{iDN} > a_1^{iFB}$ and so $a_2^{iDN} \equiv \widehat{a}_2^i(a_1^{iDN}) > \widehat{a}_2^i(a_1^{iFB})$.*

Comparison with autarky

Comparing conditions (5) and (6), we note that under autarky, the negotiation-related brinkmanship effect disappears, while the intertemporal substitution effect has full strength.

Further assumptions are needed in order to compare the autarky level a_1^{i*} with the equilibrium value $a_1^{i\,DN}$ under period-2-only negotiation.

For example, in the investment-in-green-equipments illustration introduced earlier, the date-1 investment is the same : $a_1^{iDN} = a_1^{i*}$ (and so there is more date-2 pollution under autarky). But one can also find examples in which the brinkmanship effect is particularly strong relative to the intertemporal substitution one and so, not only $a_1^{iDN} > a_1^{i*}$, but also date-2 pollution is higher under delayed negotiation then under autarky (see our discussion paper, Beccherle and Tirole (2010)).

Proposition 2. *The date-2 pollution under delayed negotiation may be higher or smaller than under autarky.*

n-region version. The analysis generalizes straightforwardly to n regions. Letting α^i denote region i's bargaining power ($\Sigma_i \alpha^i = 1$), condition (6) remains valid provided that c^j be replaced by $(\Sigma_{j \neq i} c^j)$ in the brinkmanship term. A small region is likely to have little bargaining power and so the intertemporal substitution effect has full strength. The brinkmanship effect by contrast is likely to vanish provided that the externality $(\Sigma_{j \neq i} c^j)$ does not increase fast when the country becomes small. For example, an increase in the number of regions stemming from the breakup of alliances reduces the brinkmanship effect and reinforces the intertemporal substitution one.

3.2 Non-linear pollution damages: raising the rival's cost

The linear environmental-costs assumption allowed us to cleanly focus on the effect of the date-2 negotiation on the date-1 incentive. Yet the effects of greenhouse gases (GHG) seem to be convex rather than linear. The nonlinearity in the environmental damage function actually magnifies the strategic incentive. The intuition resembles that developed in Allaz and Vila (1993) in a different context.

Let $C^i(a_2)$ denote the environmental cost function of region i. Let $C(a_2) \equiv C^A(a_2) + C^B(a_2)$ denote the total environmental cost. We assume that C^i is twice differentiable and is increasing and convex (that is $\frac{dC^i}{da_2}(a_2) > 0$, and $\frac{d^2C^i}{da_2{}^2}(a_2) > 0$).

The second-period welfare functions are now:

$$W_2(a_2^A, a_2^B, a_1^A, a_1^B) = U_2^A(a_1^A, a_2^A) + U_2^B(a_1^B, a_2^B) - C(a_2) \tag{7}$$

$$W_2^i(a_2^i, a_2^j, a_1^i) = U_2^i(a_1^i, a_2^i) - C^i(a_2) \tag{8}$$

Under *autarky*, a region's second-period emissions impact the other region's marginal damage cost. Regions play a Nash equilibrium of the game with payoff functions

$$W_2^i(a_2^i, a_2^j, a_1^i).$$

Let $\{{a_2^i}^*(a_1^i, a_1^j)\}_{i\in\{A,B\}}$ denote the Nash equilibrium of the game which, we assume, is unique and stable. Similarly, $\{\widehat{a}_2^i(a_1^i, a_1^j)\}_{i\in\{A,B\}}$ maximizes $W_2(a_2^A, a_2^B, a_1^A, a_1^B)$ and is assumed to be unique.

For given date-1 actions, let

$$\widehat{W}_2(a_1^A, a_1^B) \equiv W_2(\widehat{a}_2^A(a_1^A, a_1^B), \widehat{a}_2^B(a_1^A, a_1^B), a_1^A, a_1^B),$$

and

$$W_2^{i*}(a_1^i, a_1^j) \equiv W_2^i({a_2^i}^*(a_1^A, a_1^B), {a_2^j}^*(a_1^A, a_1^B), a_1^i).$$

Region i's intertemporal payoff after negotiation, W_{neg}^i, is:

$$W_{neg}^i(a_1^i, a_1^j) = U_1^i(a_1^i) + W_2^{i*}(a_1^i, a_1^j) + \alpha^i\left(\widehat{W}_2(a_1^A, a_1^B) - \left(W_2^{i*}(a_1^i, a_1^j) + W_2^{j*}(a_1^i, a_1^j)\right)\right).$$

From the envelope theorem, the first-order condition is:

$$\begin{aligned}
\frac{\partial W_{neg}^i}{\partial a_1^i}(a_1^i, a_1^j) = &\left(\frac{dU_1^i(a_1^i)}{da_1^i} + \frac{\partial U_2^i}{\partial a_1^i}\big(a_1^i, \widehat{a}_2^i(a_1^i, a_1^j)\big)\right) \\
&+ (1-\alpha^i)\left(\frac{\partial U_2^i}{\partial a_1^i}\big(a_1^i, {a_2^i}^*(a_1^i, a_1^j)\big) - \frac{\partial U_2^i}{\partial a_1^i}\big(a_1^i, \widehat{a}_2^i(a_1^i, a_1^j)\big)\right) \\
&+ \alpha^i C^{j\prime}\big(a_2^*(a_1^A, a_1^B)\big)\frac{\partial {a_2^i}^*}{\partial a_1^i}\big(a_1^i, a_1^j\big) \\
&+ (1-\alpha^i)C^{i\prime}\big(a_2^*(a_1^A, a_1^B)\big)\Big(-\frac{\partial {a_2^j}^*}{\partial a_1^i}(a_1^i, a_1^j)\Big) = 0
\end{aligned} \tag{9}$$

Condition (9) is very similar to equation (6); the three effects identified earlier (regional optimization, intertemporal substitution and brinkmanship) are still at work. But introducing non-linear environmental costs adds a fourth term:

$$(1-\alpha^i)C^{i\prime}\big(a_2^*(a_1^A, a_1^B)\big)\Big(-\frac{\partial {a_2^j}^*}{\partial a_1^i}(a_1^i, a_1^j)\Big).$$

An increase in a_1^i makes date-2 abatement more costly to region i and so commits that region to high emissions in the absence of agreement. This raises region j's marginal damage cost and thus induces region j to reduce its own emissions under autarky. Region j faces a higher marginal cost of pollution and reduces its second-period pollution: Because we assumed that the Nash equilibrium is stable, $a_2^{j^*}$ is decreasing in a_1^i (and increasing in a_1^j). This is the *raising rival's cost effect.*

This effect is larger, the larger is $1-\alpha^i$, that is, the smaller the bargaining power of the region. The new effect reinforces the intertemporal substitution and brinkmanship effects. We thus conclude that Proposition 1 holds a fortiori.[19]

3.3 The symmetric additive specification

Harstad (2009) assumes that region i derives an increasing and concave benefit $B(y_2^i)$ from date-2 consumption y_2^i.[20] Symmetry obtains, as $B(\cdot)$ as well as the damage function $C(\cdot)$ are the same for both regions. Finally, emissions take an additive form:

$$a_2^i = y_2^i + a_1^i.$$

Thus

$$B^{'}\left(a_2^{i*} - a_1^i\right) = C^{'}\left(a_2^*\right), \text{ where } a_2^* \equiv \Sigma_i a_2^{i*}$$

and

$$B^{'}\left(\widehat{a}_2^i - a_1^i\right) = 2C^{'}\left(\widehat{a}_2\right), \text{ where } \widehat{a}_2 \equiv \Sigma_i \widehat{a}_2^i.$$

And so, letting $a_1 \equiv \Sigma_i a_1^i$, $\psi(x) \equiv (B^{'})^{-1}\big(C^{'}(x)\big)$, and $\widehat{\psi}(x) \equiv 2(B^{'})^{-1}\big(2C^{'}(x)\big)$ (ψ and $\widehat{\psi}$ are decreasing),

$$a_2^* = a_1 + \psi(a_2^*)$$

and

$$\widehat{a}_2 = a_1 + \widehat{\psi}(\widehat{a}_2).$$

[19]This analysis again generalizes to n regions. The additional term associated with the convexity of the damage functions is then:

$$(1-\alpha^i)C^{i\prime}\big(a_2^*(\vec{a_1})\big)\big[-\sum_{j\neq i}\frac{\partial a_2^{j^*}}{\partial a_1^i}(\vec{a_1})\big] + \alpha^i \sum_{j\neq i}\sum_{k\neq j,i}\left[C^{j\prime}\big(a_2^*(\vec{a_1})\big)\right]\frac{\partial a_2^{k^*}}{\partial a_1^i}(\vec{a_1})$$

where $\vec{a_1} = (a_1^1, a_1^2, \ldots a_1^n)$.

[20]Recall that his model has an infinite horizon. We recast it in our two-period framework for the sake of comparison.

The linear additive assumption of Harstad's model implies that knowledge is a pure public good, on par with environmental quality. The outcomes in the presence or absence of negotiation are independent of who made the prior investments in green technologies (only a_1 matters).

We can push the analysis a bit further and allow for asymmetric bargaining weights in the symmetric additive specification. The intertemporal substitution, brinkmanship and raising rival's costs effects are given by the following three terms, respectively:

$$\begin{aligned}\Delta(a_1) \equiv (1-\alpha^i)\Big[2C'(\widehat{a}_2(a_1)) - C'(a_2^*(a_1))\Big] \\ + \alpha^i C'(a_2^*(a_1))\left(\frac{1-\psi'\big(a_2^*(a_1)\big)}{1-2\psi'\big(a_2^*(a_1)\big)}\right) \\ + (1-\alpha^i)C'(a_2^*(a_1))\left(\frac{-\psi'\big(a_2^*(a_1)\big)}{1-2\psi'\big(a_2^*(a_1)\big)}\right).\end{aligned}$$

For symmetric bargaining powers ($\alpha^i = 1/2$ as is assumed in Harstad), the total strategic effect is $\Delta = C'(\widehat{a}_2(a_1))$.

Another simple case arises when damage costs are linear ($\psi' = 0$). Then Δ is independent of the bargaining powers and $da_2^{i*}/da_1^i = d\widehat{a}_2^i/da_1^i = 1$.

The following result is a direct application of these formulas:

Proposition 3. *Consider the symmetric additive specification and assume that the damage function $C(a_2)$ is linear. Then date-1 efforts are the same whether or not negotiation occurs at date 2. Put differently, the brinkmanship and intertemporal substitution effects cancel out: $a_1^{iDN} = {a_1^i}^*$. And so $a_2^{iDN} < {a_2^i}^*$ for all i.*

An application of the symmetric additive specification is to the choice at date 1 of a pollution standard[21] that defines the environmental quality of equipments that will not be revamped at date 2. Letting a_1^i denote the resulting *date-2* pollution and y_2^i denote the new pollution at date 2, then $a_2^i = y_2^i + a_1^i$. This is but a reinterpretation of our example with investments in green equipments.

[21] Examples include CO_2 emission standards for automobiles in Europe, and in the United States, a minimum mileage legislation for cars and trucks (miles traveled per gallon of gasoline). In France, an environmental law package called "Grenelle de l'Environnement" introduced upper limits on housing consumption (by 2012 every new building will have to consume less than 50 kWh/m^2/year).

4 Market consolidation under forward or bankable allowances

4.1 Forward allowances

4.1.1 Overview

As the United States, the emerging countries and actually most of the world resist binding agreements, many experts place their hopes in the emergence of regional pollution permit markets such as the already existing European Union Emission Trading System (ETS) for CO_2 emission permits.

Proponents of regional ETS initiatives take the view that regional markets will jump-start climate change mitigation and will pave the way for an international binding agreement over region's emissions. However a deal will not happen overnight. We investigate the strategic implications of a delayed agreement. As we previously did, we ensure strategic independence under autarky by assuming that the damages depend linearly on total emissions. We first focus on strategic choices regarding forward allowances; we later show that similar insights apply to the issuance of bankable permits. In either case, the region puts today into private hands allowances that can be used to cover future emissions. We also assume, as we did before, that the date-2 agreement defines both regions' emissions. The resulting outcome is equivalent to a merger of the two ETS systems only in a symmetric configuration.

The brinkmanship effect applies provided that country i's date-2 pollution under autarky increases with the number of forward allowances distributed or sold at date 1. The intertemporal substitution effect further requires that domestic pollution be reduced by an international agreement. We posit that the government faces a shadow cost of public funds and auctions off date-2 allowances,[22] and show that these properties are indeed satisfied. But the properties may hold even if permits are distributed for free at date 2. For example, the government might at date 2 negotiate domestically with a powerful industrial lobby, whose status-quo welfare is stronger, the larger the number of allowances it received at date 1. Let us preview our results.

We first consider a situation in which the two regions sell forward allowances at date 1 in a non-coordinated way. In the absence of future negotiation, we show that it is suboptimal for a region to sell forward. This *regional optimization effect* is a reinterpretation of

[22] Existing or planned ETS schemes, while distributing most allowances for free, plan to move to a full auctioning approach (of course there is some uncertainty as to the credibility of such commitments).

the Coase conjecture. If a region sells some allowances at date 1, at date 2 it will not fully internalize the decrease in value of these forward allowances when issuing spot allowances. Anticipating this incentive, the buyers of forward allowances buy the allowances below the price that would prevail if only a spot market existed. A region that chooses to sell allowances forward ends up selling more allowances than it would in a date-2 spot market.

By contrast, it becomes profitable for a region to sell some allowances forward when regions negotiate over emissions at date 2. As earlier, the outside options in the date-2 negotiation consist in signing no agreement and choosing noncooperatively the number of spot allowances.

To understand the two effects at work it is useful to consider the polar case in which one region has all the bargaining power. Because fewer allowances are issued than in the absence of an agreement, the date-2 price of allowances (P^i) is higher than in the noncooperative case (p^i). The buyers of forward allowances, rationally anticipating the negotiation outcome, are willing to purchase the forward allowances at a price P^i greater than p^i. For a region with no bargaining power, issuing some allowances forward increases its immediate profit by P^i per unit, while only decreasing its outside option by p^i. So the region has an incentive to free-ride on the negotiated outcome by selling at date 1 forward allowances at the post-agreement price. The *intertemporal substitution effect* thus creates an incentive for the region without bargaining power to sell forward allowances.

The region with full bargaining power also has an incentive to sell forward allowances albeit for a different reason: It benefits from lowering the other region's outside option. The lower the other region's outside option, the larger the surplus that it will be able to extract in the negotiation. By selling allowances forward, the region credibly commits to increase *ex post* the total number of allowances it will sell in the absence of agreement. This *brinkmanship effect* gives an incentive for the region with bargaining power to sell allowances forward.

4.1.2 Description of the forward allowance game

In the case of forward trading, the first-period strategic action a_1^i is the number of forward allowances region i sells at date 1. Similarly, the second-period strategic action a_2^i, that is the level of emissions produced by region i in period 2, is the total number of allowances issued in region i (a_2^i is equal to a_1^i *plus* the number of spot allowances issued at date 2).

Firms are captive within regions.[23] In each region, the firms take the price of allowances

[23] That is, we assume that outsourcing costs are important, so that firms are fixed and therefore obliged

in their region as given when choosing their output. We call $p^i(a_2^i)$ region i's inverse demand function for allowances. p^i is non-increasing in a_2^i.

The region's second-period utility, U_2^i, is made of two terms. First, the regulator values the pollution of firms producing on its soil to the extent that this economizes on abatement costs, generating more profits, employment or taxes. We stay as general as possible and assume that the value associated with domestic firms emitting a_2^i units of pollution is $V_2^i(a_2^i)$. The function V_2^i is increasing in a_2^i and concave. In the absence of taxes or other externalities on the rest of society, one can assume that $V_2^i(a_2^i)$ is equal to the domestic firms' profit $\pi^i(a_2^i)$, so that $V_2^{i\prime}(a_2^i) = \pi_2^{i\prime}(a_2^i) = p^i(a_2^i)$ but we do not impose such a restriction for the moment.

Second, the regulator internalizes part of the cash generated by the sale of allowances; indeed we assume that the regions face a shadow cost of public funds λ, that is that raising \$1 of public money costs society \$$(1+\lambda)$. And so:

$$U_2^i(a_2^i, a_1^i) \equiv \lambda[a_2^i - a_1^i]p^i(a_2^i) + V_2^i(a_2^i) \tag{10}$$

Since $V_2^i(a_2^i)$ is concave, then for either small λ or decreasing marginal revenue, $U_2^i(a_2^i, a_1^i)$ is concave in a_2^i. For given (a_1^i, a_1^j), ${a_2^i}^*(a_1^i)$ and $\widehat{a}_2^i(a_1^i)$ are uniquely defined.

The first-period utility associated with the sale of forward allowances is λ times the revenues from the sale.[24] More precisely, the forward allowances a_1^i are sold at price $p^i(\widehat{a}_2^i(a_1^i))$ as agents rationally anticipate that the date-2 agreement will lead to a price $p^i(\widehat{a}_2^i)$. Let

$$U_1^i(a_1^i) \equiv \lambda a_1^i p^i(\widehat{a}_2^i(a_1^i)) \tag{11}$$

Checking Assumption 1

to buy permits from the region where they were initially located. This is typically the case for utilities: since electricity cannot be transported over long distances, electricity producers must have power plants located relatively close to their end consumers.

[24]Whether forward allowances are sold or distributed for free at date 1 is irrelevant for the argument.

These two functions satisfy the assumptions of the generic model.[25] In particular

$$\Gamma^i(a_1^i) = \frac{dU_1^i}{da_1^i}(a_1^i) + \frac{\partial U_2^i}{\partial a_1^i}\Big(a_1^i, \widehat{a}_2^i(a_1^i)\Big) = \lambda p^{i\prime}(\widehat{a}_2^i(a_1^i))\frac{d\widehat{a}_2^i}{da_1^i}a_1^i$$

has the same sign as $(-a_1^i)$. Assumption 1(ii) is satisfied with $a_1^{iFB} = 0$, and Proposition 1 applies.

The first-order condition takes the specific form:

$$\frac{\partial W_{neg}^i}{\partial a_1^i}(a_1^i) = \Big[\lambda p^{i\prime}(\widehat{a}_2^i(a_1^i))\,\frac{d\widehat{a}_2^i}{da_1^i}a_1^i\Big] + (1-\alpha^i)\lambda\Big[p^i(\widehat{a}_2^i(a_1^i)) - p^i(a_2^{i*}(a_1^i))\Big] + \alpha^i c^j\frac{da_2^{i*}}{da_1^i}(a_1^i).$$

Policy implications

Applied to forward trading, Lemma 1 implies that the level of date-2 emissions in region i (under autarky or negotiation) is an increasing function of the number of its forward allowances. Thus, issuing forward allowances is a credible commitment to emit more in the future. Proposition 1 in turn implies that forward allowances are issued while none would be in the first best, and that delayed negotiation results in more pollution at date 2.

Comparison with autarky

Under autarky, rational agents anticipate second-period price of allowances $p^i({a_2^i}^*(a_1^i))$. Therefore ${U_1^i}^*(a_1^i) = \lambda a_1^i p^i({a_2^i}^*(a_1^i))$[26]. The intertemporal welfare function of region i in the autarky situation is

$$\begin{aligned} W_{aut}^i(a_1^i, a_1^j) &= {U_1^i}^*(a_1^i) + U_2^i({a_2^i}^*(a_1^i), a_1^i) - c^i[{a_2^i}^*(a_1^i) + {a_2^j}^*(a_1^j)] \\ &= \lambda p^i({a_2^i}^*(a_1^i)){a_2^i}^*(a_1^i) + V_2^i({a_2^i}^*(a_1^i)) - c^i[{a_2^i}^*(a_1^i) + {a_2^j}^*(a_1^j)]. \end{aligned} \tag{12}$$

Since ${a_2^i}^*(a_1^i)$ is non-decreasing, then ${a_1^i}^* = 0$. The effect of introducing a date-2 negotiation is to induce regions to sell allowances forward, while in autarky they would have refrained from selling forward. The following proposition is proved in the Appendix.

[25]First, we have

$$\frac{\partial^2 U_2^i}{\partial a_1^i \partial a_2^i}(a_2^i, a_1^i) \;=\; -\lambda p^{i\prime}(a_2^i) > 0.$$

Second,

$$\frac{dU_1^i}{da_1^i}(a_1^i) + \frac{\partial U_2^i}{\partial a_1^i}(a_1^i, \widehat{a}_2^i(a_1^i)) = \lambda p^{i\prime}(\widehat{a}_2^i(a_1^i))\frac{d\widehat{a}_2^i(a_1^i)}{da_1^i}a_1^i.$$

[26]${U_1^i}^*(a_1^i)$ differs from the first-period utility under delayed negotiation: Under autarky forward allowances sell at price $p^i({a_2^i}^*(a_1^i))$, while they sell at price $p^i(\widehat{a}_2^i(a_1^i)) > p^i({a_2^i}^*(a_1^i))$ under delayed negotiation.

Proposition 4. *For linear demand functions and when V_2^i is equal to profit, the strategic incentive to be in a stronger bargaining position and the reduction of pollution achieved through date-2 negotiation cancel out: $a_2^{DN} = a_2^*$.*

4.1.3 Merger of emission trading systems

Without loss of generality, we have assumed that at date 2 the two regions agree on a vector of emissions, one for each region. When regions are completely symmetric ($c^A = c^B = c/2$, $\alpha^A = \alpha^B = 1/2$, $V^A(x) = V^B(x)$ and $p^A(x) = p^B(x)$ for all x) bargaining over regional allowances amounts (on the equilibrium path) to merging the regional markets and setting a total number of allowances.

The optimality of merging markets requires that the optimal agreement leads to equal prices of CO_2 in both regions. This condition however is not in general satisfied. It does not hold in the symmetric case off the equilibrium path (that is, when one of the regions deviates from the equilibrium number of forward permits) or in the asymmetric case. Intuitively, the date-2 carbon price determines the rent enjoyed by those private actors who acquired forward permits at date 1. Because such rents are costly to the regions, the latter at date 2 cooperatively set regional allowances and prices with an eye on limiting them. This incentive for price discrimination implies that setting equal prices for the two regions – a corollary of market merger – is not optimal in general.

With a few adaptations, though, our model can deal with the case where the negotiation is constrained to focus on merging the two systems and on setting the total allowances world-wide. We here content ourselves with a sketch of the main insights.

Under carbon-price equalization, the regional optimization and the brinkmanship effects remain. The intertemporal substitution effect however can go in the opposite direction. This effect is driven by the price difference between the autarky outcome (the threat point) and the negotiated outcome. If regions differ sufficiently in their industry structures (and thus abatement costs), the autarky price in one region may be larger than the negotiated price. In that case, the intertemporal substitution effect becomes negative for this region. By contrast, if both autarky prices are below the negotiated price, the analysis of the general model remains valid. The following proposition is proved in the Appendix.

Proposition 5. *For linear inverse demand functions, and when the negotiation bears on merging the two allowance markets:*

(i) When regions have the same technology (i.e. $p^i(x) = p^j(x)$ for all x), then both regions

sell forward.

(ii) When regions have different technologies, then provided that technologies are differentiated enough (alternatively that environmental costs are low enough), only the region with the most performing technology (region i such that $p^i(x) < p^j(x)$ for all x) sells forward.

4.2 Banking of pollution allowances

Another policy through which the authorities can put future permits into private hands is the banking clause in Emission Trading Systems. For example, the Waxman-Markey bill which passed the House of Representatives in June 2009 had built in such a scheme. The bill adopted a cap-and-trade approach, set the quantity of permits to be issued over the next 20 years, and allowed permit holders to bank their permits for future use.

Strategic actions and utility functions

Let n_1^i denote region i's number of allowances issued at date 1. These allowances can be used in period 1 or banked for use at time 2. The amount of allowances banked by the private sector is a_1^i. So $n_1^i \equiv a_1^i + q_1^i$ where q_1^i is the number of allowances used in period 1. As earlier, a_2^i denote emissions in region i in period 2.

For notational convenience, we assume that the polluting firms' payoff function is the same in both periods. Firms' net benefit function in periods 1 and 2, $\pi^i(q_1^i)$ and $\pi^i(a_2^i)$, respectively, is increasing and convex. They anticipate a post-negotiation second-period allowance price $p^i(\widehat{a}_2^i(a_1^i))$. So, assuming that there is banking in equilibrium, they will bank until their marginal benefit to produce is equal to that price:

$$\pi^{i\prime}(n_1^i - a_1^i) = p^i(\widehat{a}_2^i(a_1^i)) = \pi^{i\prime}(\widehat{a}_2^i(a_1^i))$$

and so

$$n_1^i - a_1^i = \widehat{a}_2^i(a_1^i) \tag{13}$$

Because the benefit is convex and the inverse demand function of region i is decreasing, condition (13) defines an increasing function $a_1^i(n_1^i)$. Regions therefore maximize indifferently over a_1^i or over n_1^i. We choose a_1^i as the first-period strategic action.

The pollution at date 1 has a direct cost $c_1^i q_1^i$ to society.[27] The regulator values a share λ of the proceedings from the sale of the allowances. We therefore have

[27] If all effects of pollution are delayed and there is no regeneration, then $c_1^i = c^i$ (the second-period cost). In general c_1^i can be greater or smaller than c^i.

$$U_1^i(a_1^i, a_1^j) = \lambda p^i(\widehat{a}_2^i(a_1^i))(\widehat{a}_2^i(a_1^i) + a_1^i) - c_1^i(\widehat{a}_2^i(a_1^i) + \widehat{a}_2^j(a_1^j)) + \pi^i(\widehat{a}_2^i(a_1^i)).$$

Region i's second-period utility is:

$$U_2^i(a_1^i, a_2^i) = \lambda p^i(\widehat{a}_2^i(a_1^i))(\widehat{a}_2^i(a_1^i) - a_1^i) + \pi^i(\widehat{a}_2^i(a_1^i)).$$

Checking Assumption 1.

We verify that $\dfrac{\partial^2 U_2^i}{\partial a_1^i \partial a_2^i} = -\lambda p^{i\prime}(\widehat{a}_2^i(a_1^i)) > 0$. So assumption 1 (i) is satisfied. We will assume that Assumption 1 (ii) is also satisfied. We have checked that this is indeed the case for linear demand functions. Although U_1^i depends on both a_1^i and a_1^j, the general result applies since $U_1^i(a_1^i, a_1^j)$ is separable in a_1^i and a_1^j.

Proposition 6. *A delayed negotiation increases the quantity of banked allowances and yields an overprovision of allowances in the first period.*

5 Alleys for future research

The introduction has already summarized the main insights of the analysis. These concluding notes rather focus on how it could be enriched. Besides the obvious extension to a longer horizon, several alleys for research seem particularly promising.

First, while our model already covers a wide range of instruments, its generality could be further enhanced. Consider for instance the clean development mechanism (CDM) set up in the aftermath of Kyoto. This mechanism allows countries that have committed to emission abatement targets to implement these in part through projects in countries that have ratified the Kyoto protocol but are not subject to such targets. The developed countries' willingness to go along with the CDM impacts not only their own effort prior to the negotiation of a binding agreement (through the earned credits), but also the effort of the emerging countries' region. The basic model could be enriched to account for such interdependencies.

A second promising alley is to consider asymmetric information between regions regarding, say, their political resolve to combat climate change. As we noted, such asymmetries may be one of the causes of delay, as regions are engaged in a form of "war of attrition". Furthermore, this extension could generate interesting insights regarding signaling strategies. For example, before negotiating in Copenhagen, Europe made a commitment to a 20% emission reduction relative to 1990 (30% in case of a "satisfactory agreement") and several countries added a carbon tax for those economic agents who are not covered by the ETS system. Some observers argued that European politicians were thereby putting themselves in a weak bargaining position because they made concessions before negotiating and over-signalled their eagerness to reach an agreement; others disagreed and viewed this move as a way to signal good intentions and to jump-start real negotiations.

We have considered only a global agreement between the regions. While we feel that negotiating a global agreement is the most reasonable way to go, many advocate a sectoral agreement approach. The study of multiple negotiations could be fascinating, as interest group politics would then play a major role.

Finally, a common political argument in favor of tough pollution control at home is that strict anti-pollution policies give the country's industry a technological edge by stimulating R&D in green technologies and help create tomorrow's "green growth". This strategic trade effect might alleviate the impact of the strategic effects of delayed negotiation.

We leave these open questions and other exciting issues on the climate negotiation research agenda to future work.

References

Acemoglu, D., Aghion, P., Bursztyn, L., and D. Hemous (2009) "The Environment and Directed Technical Change," NBER Working Paper 15451 (October).

Aldy, J., and R. Stavins (ed.) (2007) *Architectures for Agreement: Addressing Global Climate Change in the Post-Kyoto World*. Cambridge University Press.

Allaz, B., and J.L. Vila (1993) "Cournot Competition, Forward Markets and Efficiency," *Journal of Economic Theory*, 59(1): 1–16.

Barrett, S. (2005) "The Theory of International Environmental Agreements," in K.-G. Mäler and J.R. Vincent, eds., *Handbook of Environmental Economics*, 3, Elsevier.

Beccherle, J. and J. Tirole, (2010) "Regional Initiatives and the Cost of Delaying Binding Climate Change Agreements," Working Paper, Toulouse School of Economics.

Buchholz, W., and K. Konrad (1994) "Global Environmental Problems and the Strategic Choice of Technology," *Journal of Economics*, 60(3): 299–321.

Carraro, C., and D. Siniscalco (1993) "Strategies for the International Protection of the Environment," *Journal of Public Economics*, 52: 309–328.

Grimaud, A., and L. Rouge (2008) "Environment, Directed Technical Change and Economic Policy," *Environmental Resource Economics*, 41: 439–463.

Grossman, S., and O. Hart (1986) "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, 94(4): 691–719.

Harstad, B. (2009) "The Dynamics of Climate Change Agreements," Working Paper, Northwestern University, Center for Mathematical Studies in Economics and Management Science.

Hart, O., and J. Moore (1990) "Property Rights and the Nature of the Firm," *Journal of Political Economy*, 98(6): 1119–1158.

Laffont, J.J., and J. Tirole (1996a) "Pollution Permits and Compliance Strategies," *Journal of Public Economics*, 62(1-2): 85–125.

—— (1996b) "Pollution Permits and Environmental Innovation," *Journal of Public Economics*, 62(1-2): 127–140.

Mahenc, P., and F. Salanié (2004) "Softening Competition Through Forward Trading," *Journal of Economic Theory*, 116(2): 282–293.

Milgrom P., and J. Roberts (1990) "Rationalizability, Learning and Equilibrium in Games with Strategic Complementarities," *Econometrica*, 58: 1255–1278.

Salop, S., and D. Sheffman (1983) "Raising Rivals' Costs," *American Economic Review, Papers & Proceedings*, 73: 267–271.

Tirole, J. (2009) "Cognition and Incomplete Contracts," *American Economic Review*, 99(1): 265–294.

Willamson, O. (1985) *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*. New York: Free Press.

Appendix

Proof of Proposition 4

Define $p^i(a_2^i) = \beta^i - \gamma^i a_2^i$. Then

$$a_2^{i*}(a_1^i) = \frac{(1+\lambda)\beta^i - c^i + \lambda\gamma^i a_1^i}{(1+2\lambda)\gamma^i}.$$

Since $a_1^{i*} = 0$, the autarky level of emission is $\frac{(1+\lambda)\beta^i - c^i}{(1+2\lambda)\gamma^i}$.

Similarly

$$\widehat{a}_2^i(a_1^i) = \frac{(1+\lambda)\beta^i - c + \lambda\gamma^i a_1^i}{(1+2\lambda)\gamma^i}.$$

Using the first-order condition with respect to a_1^i, $a_1^{iDN} = \frac{c^j}{\lambda\gamma^i}$, and so $\widehat{a}_2^i(a_1^{iDN}) = a_2^{i*}(a_1^{i*})$. $\square$

Proof of Proposition 5

Take again $p^i(a_2^i) = \beta^i - \gamma^i a_2^i$. We define the merged-market price $P(a_2)$ by the implicit equation: $p^{A^{-1}}(P(a_2)) + p^{B^{-1}}(P(a_2)) = a_2$.

So $P(a_2) = \dfrac{\beta^i\gamma^j + \beta^j\gamma^i - \gamma^i\gamma^j a_2}{\gamma^i + \gamma^j}$. Define $\beta \equiv \frac{\beta^i\gamma^j + \beta^j\gamma^i}{\gamma^i+\gamma^j}$ and $\gamma \equiv \frac{\gamma^i\gamma^j}{\gamma^i+\gamma^j}$. Then $P(a_2) = \beta - \gamma a_2$.

The first-order condition with respect to a_1^i writes

$$\frac{\partial W_{neg}^i}{\partial a_1^i}(a_1^i, a_1^j) = \lambda P'(\widehat{a}_2(a_1))\frac{d\widehat{a}_2}{da_1^i}a_1^i + (1-\alpha^i)\lambda\Big(P(\widehat{a}_2(a_1)) - p^i(a_2^{i*}(a_1^i))\Big) + \alpha^i c^j \frac{da_2^{i*}}{da_1^i}(a_1^i) = 0$$

Hence:

$$\frac{\partial W_{neg}^i}{\partial a_1^i}(a_1^i, a_1^j) = \frac{\lambda}{1+2\lambda}\Big[(1-\alpha^i)\lambda(\beta-\beta^i) + c^j - (1-\alpha^i)\lambda\gamma a_1^j - \lambda\Big(\gamma - (1-\alpha^i)(\gamma^i - \gamma)\Big)a_1^i\Big],$$

and:

$$\frac{\partial W_{neg}^i}{\partial a_1^i}(0,0) = \frac{\lambda}{1+2\lambda}\Big(\lambda(1-\alpha^i)\frac{\gamma^i}{\gamma^i+\gamma^j}(\beta^j - \beta^i) + c^j\Big) \equiv \Delta^i.$$

First note that for $c^i = c^j = 0$ then $\Delta^i > 0$ if and only if $\beta^j > \beta^i$. For simplicity we will assume that $\gamma^i = \gamma^j = 2\gamma$. We verify that

$$\frac{\partial^2 W^i_{neg}}{\partial a^i_1 \partial a^j_1} \leq 0$$

Furthermore, since $\gamma^i = \gamma^j = 2\gamma$ then $\gamma - (1 - \alpha^i)(\gamma^i - \gamma) \geq 0$ and so W^i_{neg} is concave in a^i_1 for any a^j_1.

We conclude that for $(1-\alpha^i)\lambda\frac{\gamma^i}{\gamma^i+\gamma^j}(\beta^j-\beta^i)+c^j \geq 0$ and $(1-\alpha^j)\lambda\frac{\gamma^j}{\gamma^j+\gamma^i}(\beta^i-\beta^j)+c^i \leq 0$ then $a^i_1 > 0$ and $a^j_1 = 0$.

Indeed, if the conditions above hold: $\frac{\partial W^j_{neg}}{\partial a^j_1}\left(0, a^i_1\right) < 0$ for all $a^i_1 \geq 0$ and so by quasi-concavity of W^j_{neg} in a^j_1, $a^j_1 = 0$ for all $a^i_1 \geq 0$. Then the quasi-concavity of W^i_{neg}, together with $\frac{\partial W^i_{neg}}{\partial a^i_1}(0, 0) > 0$, implies that $a^i_1 > 0$. $\square$