GOVERNANCE AND THE EFFICIENCY OF ECONOMIC SYSTEMS
GESY

Discussion Paper No. 366

The Timing of Climate Agreements under Multiple Externalities

Robert C. Schmidt*
Roland Strausz**

*HU Berlin, Germany,
E-Mail: robert.schmidt.1@wiwi.hu-berlin.de
**HU Berlin, Germany,
E-Mail: strauszr@wiwi.hu-berlin.de

November 2011

Financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15 is gratefully acknowledged.

Sonderforschungsbereich/Transregio 15 · www.sfbtr15.de
Universität Mannheim · Freie Universität Berlin · Humboldt-Universität zu Berlin · Ludwig-Maximilians-Universität München
Rheinische Friedrich-Wilhelms-Universität Bonn · Zentrum für Europäische Wirtschaftsforschung Mannheim

Speaker: Prof. Dr. Urs Schweizer. · Department of Economics · University of Bonn · D-53113 Bonn,
Phone: +49(0228)739220 · Fax: +49(0228)739221

The Timing of Climate Agreements under Multiple Externalities*

Robert C. Schmidt† and Roland Strausz‡

November 29, 2011

Abstract

We study the potential of cooperation in global emission abatements with multiple externalities. Using a two-country model without side-payments, we identify the strategic effects under different timing regimes of cooperation. We obtain a positive *complementarity effect* of long-term cooperation in abatement on R&D levels that boosts potential benefits of long-term cooperation and a *redistributive effect* that destabilizes long-term cooperation when countries are asymmetric. We show that whether and what type of cooperation is sustainable, depends crucially on the kind rather than on the magnitude of asymmetries.

Keywords: climate treaty, abatement, long-term commitment, cooperation, low-carbon technology
JEL classification: D62, F53, H23, Q55

*Financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15 is gratefully acknowledged.

†Address: Spandauer Str. 1, 10178 Berlin, e-mail: *robert.schmidt.1@wiwi.hu-berlin.de*

‡Address: Spandauer Str. 1, 10178 Berlin, e-mail: *strauszr@wiwi.hu-berlin.de*

1 Introduction

One of the main global challenges of today is the problem of climate change. At its core lies a public good's problem which, by conventional economic wisdom, could simply be solved by global cooperation that benefits all participants. In practice, however, simple global agreements to curb the emission of greenhouse gases have been proven illusive (e.g., failure of the Copenhagen climate summit, or reluctance of the U.S. to ratify the Kyoto-protocol). This paper contributes to a growing body of literature that tries to identify and understand the difficulties in reaching such agreements.

Our starting point is the notion that multiple externalities are underlying the challenge of climate change. First, countries may resort to abatement efforts and thereby reduce directly those emissions that are held responsible for anthropogenic climate change. Second, countries may invest in R&D activities and thereby develop new production techniques that reduce the emissions intensity of output. Both types of activities exhibit strong externalities globally; countries benefit directly from each other's abatement efforts through a reduction in global emissions, and from each other's R&D efforts indirectly through knowledge spill-overs that reduce their abatement costs. Yet, irrespective of the double public goods problem, international climate negotiations focus primarily on reaching agreements on emission levels.[1]

Given that both abatement and R&D efforts are crucial when tackling the issue of climate change, the goal of this paper is to understand how both types of externalities affect the potential for agreements. Distinguishing between direct abatement efforts and R&D investments, we are able to address also the *timing* of cooperation. Because the implementation of R&D requires more time than the implementation of abatement efforts, countries may cooperate in their choice of future abatement levels either before or after they invest in R&D. We, therefore, distinguish between *early cooperation* where countries commit to long-term abatement targets before they invest in R&D, and *late cooperation* where countries first invest in R&D, and at a later stage sign a short-term climate treaty. Highlighting this temporal dimension in the design of climate agreements, we are able to identify strategic incentives for delaying cooperation.

In order to analyze these issues, we consider a two-country game with environmental externalities and knowledge spill-overs.[2] To capture the practical difficulties of implementing contingent side payments, we exclude them from our analysis. Furthermore, we assume that

[1] An economic explanation for this dichotomy is that agreements on R&D are harder to monitor than agreements on emissions (see Beccherle and Tirole 2011, Golombek and Hoel 2008, Harstad 2010).

[2] Environmental policy design in the presence of innovation externalities is analyzed e.g. by Fischer et al. (2003), Gerlagh et al. (2009), Golombek and Hoel (2004), Heal and Tarui (2010). See also Jaffe et al. (2005).

countries can only cooperate in their level of emissions but not in their R&D efforts. Under these assumptions, we analyze how environmental and R&D externalities affect outcomes and can lead to a delay or a failure of cooperation.

Apart from identifying different strategic effects that can arise in the presence of the two externalities, our main results are as follows. First, there is a positive *complementarity effect* of long-term abatement commitments on R&D levels in the sense that agreements that implement higher levels of abatement, also provide stronger incentives for R&D. In other words, long-term agreements on abatement levels do not crowd out efforts in R&D. Second, the timing of cooperation exhibits a *redistributive effect* that favors late cooperation. More specifically, our results show that late (short-term) cooperation allows, via strategic choices of R&D levels, to reallocate gains from cooperation. Hence, even if late cooperation yields lower aggregated welfare than early (long-term) cooperation, a specific country may, due to this redistributive effect, nevertheless prefer late to early cooperation. As a result, this country would veto any early cooperation. The redistributive effect, therefore, destabilizes long-term cooperation. We point out that this effect is linked to the absence of side payments and occurs only when countries are asymmetric. However, we also demonstrate that already a small asymmetry can sometimes be sufficient to induce a failure of early cooperation. Third, we highlight that it is the type rather than the strength of asymmetries that generally matters for failures in cooperation. In particular, early (long-term) cooperation may sometimes be sustainable even in the presence of *strong* asymmetries. We demonstrate this for an extreme case of unidirectional externalities, where one country exerts only an abatement externality, whereas the other country exerts only an R&D externality. The presence of *both* externalities leads to a mutual dependency in countries' welfare that facilitates early cooperation. Intuitively, the country that benefits from environmental spill-overs, signs a long-term agreement to induce the other country to accept a higher abatement target as well. The country that benefits from knowledge spill-overs, at the same time, is willing to accept this long-term abatement commitment to induce the other country to subsequently invest more in R&D.

Economic literature offers multiple explanations for the difficulties in achieving climate agreements. Barrett (1994) points out that, since climate stabilization is a public good, each country prefers other countries to abate more, but may not benefit from contributing substantially to this public good itself. As a result of this free-rider problem, only a small number of countries may sign a climate treaty and especially so when the potential gains from cooperation are large. Hence, effective climate treaties that must be self-enforcing with respect to participation, are often not sustainable.[3] Because we analyze cooperation between only two players, we can abstract from these concerns. This enables us to highlight the role of multiple externalities and asymmetries between countries as another important reason why cooperation may fail.

A further strand of the literature has identified asymmetries in general as a possible impediment to cooperation. If countries are asymmetric, then side-payments can play an important role in the stabilization of cooperation (e.g., Barrett, 2001; Carraro et al., 2006; Harstad, 2007). For various reasons, however, they may be hard to implement on a large scale, and financial obligations may be difficult to enforce due to the sovereignty of countries and the lack of a higher authority that can impose punishments upon countries. The difficulties of implementing side-payments may, thus, destabilize cooperation among asymmetric countries due to the same fundamental problems that make it difficult to overcome the free-rider problem. In line with these observations, also our analysis highlights the role of asymmetries in achieving cooperation in the absence of side-payments.

Some authors abstract from the various obstacles to cooperation, and instead assume that there is an *exogenous* delay in cooperation. This allows them to identify possible adverse strategic effects that can arise when cooperation cannot be achieved today, but is expected to succeed at some point in the future. Buchholz and Konrad (1994) show that countries may choose technologies with inefficiently high abatement costs for strategic reasons, even when low-emission technologies are available today at no additional cost. The reason is that a country may benefit from committing itself to high marginal abatement costs in the future, if this induces other countries to abate more. Beccherle and Tirole (2011) generalize this approach by introducing costly investments into emission-reducing technologies or other activities that reduce future abatement costs. Similar to Buchholz and Konrad (1994), they show that the anticipation of future cooperation in the choice of abatement targets induces countries to invest less at an earlier stage (before abatement targets are determined). This delay in cooperation can lead to an outcome that, in terms of aggregate welfare, is inferior even to the fully non-cooperative outcome. Because we investigate countries that may agree on future reduction targets already *before* R&D levels are determined, our analysis extends these studies. In particular, we show that, with a long-term commitment, the option to strategically under-invest in R&D vanishes, because in this case countries effectively choose their R&D levels *ex post*, after they are already

[3]Contrasting results are presented by, for instance, Asheim and Holtsmark (2009), Froyn and Hovi (2008), and Heitzig et al. (2011). For an overview, see also Finus (2008). Stressing different punishment strategies, Barrett (1997) and Hoel and Schneider (1997) consider other possible remedies to this problem.

committed to future abatement targets.[4]

The remainder of this paper is organized as follows. Section 2 introduces the basic model. In Section 3, we derive optimality conditions to characterize the outcome in the four different cooperation regimes that we consider: full cooperation, no cooperation, early, and late cooperation. In Section 4, we endogenize the timing of cooperation and identify the effects of different kinds of asymmetries. In Section 5, we introduce a specification of our general model that is especially suited for an application to climate agreements. Finally, Section 6 concludes. The Appendix collects all formal proofs.

2 Basic Model

Consider two countries $i \in \{1, 2\}$ that choose emissions targets. The targets are measured in terms of abatement of emissions relative to some "business-as-usual" scenario. The abatement targets are denoted by $a_i \geq 0$. In addition, each country has the possibility to invest in R&D in order to reduce the costs of abatement. Country i's R&D effort is denoted by $r_i \geq 0$. Due to environmental externalities, country i's welfare may depend on overall abatement levels. Furthermore, country i's costs of abatement can, due to knowledge spill-overs, depend also on the other country's R&D effort. More specifically, country i's welfare, which reflects the net benefit of abatement, is denoted by Π_i and given by

$$\Pi_i(a_1, a_2, r_1, r_2) \equiv B_i(a_1, a_2) - C_i(a_i, r_1, r_2), \tag{1}$$

where B_i measures environmental benefits and C_i the abatement costs of country i, including its R&D investment costs. In the absence of R&D spill-overs, C_i depends on r_i but not on r_{-i}.[5] In line with Beccherle and Tirole (2011), we concentrate on benefit functions that are linearly increasing:

$$B_i(a_i, a_{-i}) = a_i + \gamma_{-i} a_{-i},$$

where $\gamma_{-i} \in [0, 1]$. We say that country $-i$ has a positive *abatement externality* on country i if $\gamma_{-i} > 0$. In the extreme $\gamma_{-i} = 1$, so that country $-i$'s abatement has the same effect on

[4]Harstad (2010) endogenizes the length of the commitment period in climate agreements. Using an infinite time horizon, the model allows countries to interact and cooperate repeatedly. Even if countries are *ex ante* symmetric, they can become asymmetric over time if they choose different investments in technology. However, in the Markov perfect equilibrium identified by the author, countries remain symmetric. In contrast to this, we focus more explicitly on the role of asymmetries, and analyze how they affect the stability and the timing of cooperation when side-payments are not feasible.

[5]For the indices, by convention, let $-i \equiv 2$ if $i = 1$, and $-i \equiv 1$ if $i = 2$.

country i as country i's own abatement a_i. The linear specification allows us to cleanly show that direct interactions between abatement and R&D externalities already lead to strategic delays in cooperation.[6]

We consider general cost functions C_i that are convex in (a_i, r_1, r_2), so that the country's welfare function Π_i is concave in (a_1, a_2, r_1, r_2). More specifically, we assume that cost functions satisfy

$$\frac{\partial C_i}{\partial a_i} > 0, \; \frac{\partial^2 C_i}{\partial a_i^2} > 0, \; \frac{\partial^2 C_i}{\partial r_i^2} > 0, \; \frac{\partial C_i}{\partial r_{-i}} \leq 0, \; \frac{\partial^2 C_i}{\partial r_{-i}^2} \geq 0, \; \frac{\partial^2 C_i}{\partial r_i \partial r_{-i}} \geq 0, \; \frac{\partial^2 C_i}{\partial a_i \partial r_i} \leq 0, \; \frac{\partial^2 C_i}{\partial a_i \partial r_{-i}} \leq 0.$$

The intuition behind these conditions is straightforward. The condition $\partial^2 C_i / \partial r_i \partial r_{-i} \geq 0$ for instance captures the standard assumption that R&D efforts are strategic substitutes. The last two conditions imply that R&D reduces (weakly) the marginal cost of abatement. We say that country i has a positive *R&D externality* if $\partial C_{-i} / \partial r_i < 0$. Furthermore, we focus on the natural case where the R&D externalities do not exceed their corresponding direct effects. That is, we assume: $\partial^2 C_i / \partial r_i^2 \geq \partial^2 C_i / \partial r_i \partial r_{-i}$, and $|\partial^2 C_i / \partial a_i \partial r_i| \geq |\partial^2 C_i / \partial a_i \partial r_{-i}|$.

Focusing on the ongoing discussion of climate agreements, we consider countries that have the possibility to cooperate in abatement efforts, but not in R&D (see Beccherle and Tirole, 2011; Golombek and Hoel, 2008; Harstad, 2010). We also follow the literature (e.g. Barrett, 2001; Harstad, 2007) by assuming that if countries agree to cooperate, then the abatement targets are always chosen to maximize the total welfare of *both* countries: $\Pi \equiv \Pi_1 + \Pi_2$. We thereby also abstract from any enforceability issues concerning these agreements. To capture the difficulty of implementing conditional side-payments in practice, we rule out such side-payments altogether. We say that *a country has an incentive to cooperate*, if its payoff from cooperation exceeds its payoff without cooperation. If both counties have an incentive to cooperate, then cooperation succeeds. In other words, *cooperation fails* as soon as one country does not have an incentive to cooperate. This can occur when countries are asymmetric, because the benefits of cooperation are, then, also shared asymmetrically. In the case of global climate agreements these asymmetries are substantial. Hence, we are especially interested in identifying the role of asymmetries in the success and in the failure of achieving cooperation.

To understand how the presence of multiple externalities and the timing of cooperation affect outcomes, our approach is to first analyze the following sub-cases independently:

[6] Concave benefit functions exhibit $\partial^2 B(a_1, a_2)/(\partial a_1 \partial a_2) < 0$, which gives rise to what is known as the "raising rivals' costs effect" in industrial organization. This effect renders the analysis less tractable, while it actually "magnifies the strategic incentive" (Beccherle and Tirole, 2011) of delay.

1. *Full cooperation*: Countries choose $(a_1^f, a_2^f, r_1^f, r_2^f)$ cooperatively to maximize the joint surplus Π.

2. *No cooperation*: Countries first choose R&D levels (r_1^n, r_2^n) non-cooperatively, and subsequently choose the abatement levels (a_1^n, a_2^n) also non-cooperatively.

3. *Early cooperation*: Countries first commit to long-term cooperative abatement levels (a_1^e, a_2^e) and subsequently choose R&D levels (r_1^e, r_2^e) non-cooperatively.

4. *Late cooperation*: Countries first choose R&D levels (r_1^l, r_2^l) non-cooperatively and then choose short-term abatement levels (a_1^l, a_2^l) cooperatively.

The first two cases represent benchmarks which we use in order to evaluate the outcomes under early and late cooperation. Overall, the analysis of these cases enables us to identify and to classify the interactions of the different spill-over effects under different, exogenously-given timing regimes.

In a second step, we then study the cooperation and timing decision by considering an overall game where the choice whether to cooperate early, late, or not arises endogenously and is part of the overall equilibrium outcome. The following figure illustrates the time structure of this overall game:[7]

Figure 1

Intuitively, if countries do not agree to cooperate (neither early nor late), then first the R&D levels are chosen non-cooperatively, and subsequently the abatement levels are chosen non-cooperatively. Underlying this sequence of events is the assumption that R&D is a time-consuming process, so current R&D efforts reduce *future* abatement costs. The abatement levels a_1 and a_2 in our model, thus, refer to some future period of time. Near-term abatement (before R&D levels are chosen) is not explicitly modeled.[8] If countries cooperate, then they transfer their abatement choice to a hypothetical planner who seeks to maximize total welfare. Under early cooperation, the planner takes into account how the assigned long-term abatement targets will affect also countries' non-cooperative choice of R&D efforts. Conversely, if countries

[7]The dotted line at the bottom left of the figure indicates that the implementation of the chosen abatement levels under early cooperation is not a decision.

[8]Underlying this approach is the implicit assumption that near-term abatement efforts do not interact strongly with the variables of the model. See also Beccherle and Tirole (2011).

cooperate late, they fully anticipate in the R&D stage how their choices will affect abatement targets later in the cooperative stage. Cooperation in a certain stage fails, as soon as one country rejects it. A country rejects early cooperation when it expects to gain more from later or no cooperation. A country rejects late cooperation when it expects more from rejecting it.

3 Optimality Conditions

In this section, we derive optimality conditions that characterize an interior solution for each of the four cases: 1. full cooperation, 2. no cooperation, 3. early cooperation, and 4. late cooperation. We show how these optimality conditions capture the different strategic effects that arise in the presence of the two externalities under early, late, and no cooperation. The optimality conditions enable us to characterize inefficiencies that arise in the absence of full cooperation, and to state comparative welfare results.

3.1 Full cooperation

Under full cooperation, countries maximize the joint surplus Π. Hence, they solve the following maximization problem: $\max_{a_1,a_2,r_1,r_2} \Pi(a_1,a_2,r_1,r_2)$. Because the target function is concave, the cooperative solution $(a_1^f, a_2^f, r_1^f, r_2^f)$ must satisfy the following first-order conditions:[9]

$$\frac{\partial C_i}{\partial a_i} = 1 + \gamma_i. \tag{2}$$

$$\frac{\partial C_i}{\partial r_i} + \frac{\partial C_{-i}}{\partial r_i} = 0. \tag{3}$$

These optimality conditions are intuitive. The first condition requires that country i's marginal abatement cost equals its *aggregated* marginal benefit of abatement. The second condition says that the aggregated abatement costs are minimized over r_i. Both abatement and R&D spill-overs are fully internalized. Due to the convexity of C_i, expression (2) indicates that the presence of an abatement externality ($\gamma_i > 0$) implies more abatement. Similarly, expression (3) shows that knowledge spill-overs from R&D ($\partial C_{-i}/\partial r_i < 0$) imply higher levels of R&D.

3.2 No cooperation

Under no cooperation, countries play a sequential game. In the first stage, they simultaneously choose their R&D efforts. In the second stage, they choose their abatement levels. We study

[9]For an ease of notation, functional dependencies are usually suppressed.

the subgame perfect equilibrium of this non-cooperative extensive form game by backward induction.

In the second stage R&D levels $(\bar{r}_1, \bar{r}_2)$ are given and country i's reaction function follows from maximizing its payoff $\Pi_i(a_1, a_2, \bar{r}_1, \bar{r}_2)$ w.r.t. a_i. The first-order conditions are

$$\frac{\partial C_i}{\partial a_i} = 1.$$

In order to signify that the equilibrium in stage 2 depends on (r_1, r_2), we write $(a_1^n(r_1, r_2), a_2^n(r_1, r_2))$.

In the first stage, countries choose their R&D levels (r_1, r_2) while anticipating the outcome in the second stage. Hence, each country expects the payoff

$$\Pi_i(a_1^n(r_1, r_2), a_2^n(r_1, r_2), r_1, r_2).$$

By the Envelope Theorem, the subgame-perfect Nash equilibrium solves the system

$$\frac{d\Pi_i}{dr_i} = \frac{\partial \Pi_i}{\partial r_i} + \frac{\partial \Pi_i}{\partial a_1}\frac{\partial a_1^n}{\partial r_i} + \frac{\partial \Pi_i}{\partial a_2}\frac{\partial a_2^n}{\partial r_i} = -\frac{\partial C_i}{\partial r_i} + \frac{\partial B_i}{\partial a_{-i}}\frac{\partial a_{-i}^n}{\partial r_i} = 0.$$

To summarize, the non-cooperative outcome $(a_1^n, a_2^n, r_1^n, r_2^n)$ solves for $i = 1, 2$:

$$\frac{\partial C_i}{\partial a_i} = 1, \tag{4}$$
$$\frac{\partial C_i}{\partial r_i} = \gamma_{-i}\frac{\partial a_{-i}^n}{\partial r_i}. \tag{5}$$

Condition (4) indicates that in stage 2 each country chooses its abatement level a_i such that the *individual* (instead of the aggregated) net benefit is maximized. Hence, abatement externalities are neglected. Comparing the left-hand side of (3) to (5) reveals that, without cooperation, countries neglect R&D spill-overs. This negatively affects R&D efforts.

More interesting is the right-hand side of (5), which identifies a *strategic double spill-over effect* that may actually increase R&D incentives. To understand the intuition behind this effect, observe that by raising its R&D effort r_i, the R&D externality will induce the other country to raise its abatement level a_{-i} from which the original country benefits through the abatement spill-over. Hence, this effect occurs only if there are spill-overs in both R&D *and* abatement. The following lemma confirms that this double spill-over effect tends to raise R&D efforts. Hence, it mitigates the aforementioned negative effect on R&D efforts.

Lemma 1 *In the presence of both an abatement* ***and*** *an R&D externality, there is a strategic double spill-over effect that tends to increase R&D incentives.*

To summarize, we identify three qualitatively different effects by which the outcome under no cooperation differs from the outcome under full cooperation: 1. the neglect of abatement externalities, 2. the neglect of R&D externalities, 3. a double externality effect that raises R&D incentives. The first two effects are straightforward and, respectively, lower the incentives for abatement and R&D. The third effect is more subtle and mitigates the second effect.[10]

3.3 Early cooperation

Under early cooperation, countries first commit to long-term abatement choices (a_1^e, a_2^e) cooperatively and subsequently choose their R&D levels (r_1^e, r_2^e) non-cooperatively. In the spirit of subgame perfection, we analyze the cooperative levels (a_1^e, a_2^e) that maximize the joint surplus Π under full anticipation of how the countries react to these abatement levels in stage 2 when choosing their R&D levels non-cooperatively. More specifically, the reaction to abatement levels $(\bar{a}_1, \bar{a}_2)$ is a Nash equilibrium in R&D levels (r_1^e, r_2^e) that solves (for $i = 1, 2$)

$$\frac{\partial \Pi_i(\bar{a}_1, \bar{a}_2, r_1, r_2)}{\partial r_i} = 0. \tag{6}$$

In stage 1, the cooperation levels (a_1^e, a_2^e) therefore solve

$$\max_{a_1, a_2} \Pi(a_1, a_2, r_1^e(a_1, a_2), r_2^e(a_1, a_2)).$$

By the Envelope Theorem, the first-order conditions yield for $i = 1, 2$

$$\frac{\partial \Pi}{\partial a_i} + \frac{\partial \Pi_2}{\partial r_1}\frac{\partial r_1^e}{\partial a_i} + \frac{\partial \Pi_1}{\partial r_2}\frac{\partial r_2^e}{\partial a_i} = 0.$$

To summarize, the solution under early cooperation $(a_1^e, a_2^e, r_1^e, r_2^e)$ satisfies the system

$$\frac{\partial C_i}{\partial r_i} = 0, \tag{7}$$

$$\frac{\partial C_i}{\partial a_i} + \frac{\partial C_{-i}}{\partial r_i}\frac{\partial r_i^e}{\partial a_i} + \frac{\partial C_i}{\partial r_{-i}}\frac{\partial r_{-i}^e}{\partial a_i} = 1 + \gamma_i. \tag{8}$$

Comparing (5) to (7) reveals that also with early cooperation countries fully neglect knowledge spill-overs in their choice of R&D levels in stage 2. Hence, just as in the case without any cooperation each country minimizes its own abatement costs, given its abatement target assigned in stage 1. On the right-hand side of (8), we see the aggregated marginal benefit of abatement, because countries cooperate in stage 1. On the left-hand side, we observe the

[10] Related results are presented by Golombek and Hoel (2004).

marginal abatement cost of country i, plus two *strategic effects* which are both related to the R&D externality.

The first of these effects indicates that the abatement target assigned to country i in the cooperative stage is *raised* in case of positive R&D spill-overs ($\partial C_{-i}/\partial r_i < 0$). Intuitively, by assigning a higher abatement target to country i, additional R&D investments by this country are triggered in the non-cooperative stage. This leads to spill-overs that reduce country $-i$'s abatement costs. Hence, this strategic effect *alleviates* the inefficiency resulting from knowledge spill-overs that are not internalized in stage 2.

The other strategic effect implies that, in the presence of R&D spill-overs ($\partial C_i/\partial r_{-i} < 0$), abatement levels are *reduced* in the cooperative stage. Intuitively, by assigning a higher abatement target to country i, higher R&D investments by this country are triggered. Because R&D efforts are strategic substitutes, they partially crowd-out R&D investments by country $-i$. This reduces the positive spill-overs from country $-i$'s R&D upon country i's abatement costs. To reduce this negative side effect, a_i is reduced in the cooperative stage.

The next lemma shows that, overall, these strategic effects are (weakly) raising total welfare. Moreover, under symmetry, early cooperation in abatement levels induces higher R&D efforts chosen by both countries. Hence, even though there is no direct cooperation in R&D, early cooperation in abatement partially offsets the lack of cooperation in R&D. It, in particular, implies that long-term cooperation in abatement does not lead to a crowding-out in R&D efforts. Hence, under early cooperation, abatement and R&D act as complements rather than substitutes. As a result, we identify the *complementarity effect* of early cooperation.

Lemma 2 *There are two strategic effects under early cooperation, which together tend to increase aggregate welfare. If countries are symmetric, they induce higher R&D levels for each country in the non-cooperative stage, by assigning higher abatement targets in the cooperative stage.*

To summarize, we identify three qualitatively different effects by which the outcome under early cooperation differs from the outcome under full cooperation: 1. the neglect of R&D externalities; 2. a complementarity effect of abatement that raises R&D incentives; 3. a crowding-out effect that lowers R&D incentives, but which does not offset the aforementioned complementarity effect of abatement.

3.4 Late cooperation

Under late cooperation, countries first choose their R&D levels (r_1^l, r_2^l) non-cooperatively and subsequently choose their abatement levels (a_1^l, a_2^l) cooperatively by signing a short-term agreement to maximize the joint surplus Π. Hence, given R&D levels $(\bar{r}_1, \bar{r}_2)$, the abatement levels (a_1^l, a_2^l) solve

$$\frac{\partial \Pi}{\partial a_i} = 0. \tag{9}$$

This yields abatement levels $a_1^l(r_1, r_2)$ and $a_2^l(r_1, r_2)$ as functions of R&D levels.

In the first stage, countries play a Nash equilibrium in R&D levels, anticipating the cooperative abatement levels $\big(a_1^l(r_1, r_2), a_2^l(r_1, r_2)\big)$ in stage 2. Country i's maximization problem is $\max_{r_i} \Pi_i \left(a_1(r_1, r_2), a_2(r_1, r_2), r_1, r_2\right)$. This yields the first-order conditions:

$$\frac{\partial \Pi_i}{\partial r_i} + \frac{\partial \Pi_i}{\partial a_i}\frac{\partial a_i^l}{\partial r_i} + \frac{\partial \Pi_i}{\partial a_{-i}}\frac{\partial a_{-i}^l}{\partial r_i} = 0 \quad \Leftrightarrow \quad \frac{\partial C_i}{\partial r_i} = \left[\frac{\partial B_i}{\partial a_i} - \frac{\partial C_i}{\partial a_i}\right]\frac{\partial a_i^l}{\partial r_i} + \frac{\partial B_i}{\partial a_{-i}}\frac{\partial a_{-i}^l}{\partial r_i}.$$

Using (9), we can simplify this condition so that we can characterize the solution under late cooperation $(a_1^l, a_2^l, r_1^l, r_2^l)$ by the system

$$\frac{\partial C_i}{\partial a_i} = 1 + \gamma_i, \tag{10}$$

$$\frac{\partial C_i}{\partial r_i} = \gamma_{-i}\frac{\partial a_{-i}^l}{\partial r_i} - \gamma_i\frac{\partial a_i^l}{\partial r_i}. \tag{11}$$

Condition (10) coincides with (2) if the fixed R&D levels $\bar{r}_1$ and $\bar{r}_2$ are the same. However, condition (11), which determines the R&D levels, differs from the optimality condition in the full cooperative case (3). As under early cooperation, countries neglect knowledge spill-overs in their choice of R&D levels. As compared to (3), the optimality condition (11), therefore, lacks the derivative $\partial C_{-i}/\partial r_i$ on its left-hand side. On the right-hand side, we observe two strategic effects that depend on the interaction between the abatement and the R&D externality. The first effect, $\gamma_{-i}\partial a_{-i}^l/\partial r_i$, tends to increase R&D incentives. Intuitively, if country i raises its R&D effort in stage 1, this induces country $-i$ to abate more in the cooperative stage due to knowledge spill-overs that reduce its marginal abatement costs. In the presence of abatement externalities, this positively affects country i's own welfare and, hence, increases its R&D incentives. In contrast, the second effect, $-\gamma_i\partial a_i^l/\partial r_i$, lowers R&D incentives. Intuitively, each country has an incentive to enter the cooperative stage with high marginal abatement costs, because this implies that it will be assigned a lower abatement effort. This *strategic commitment effect* reduces a country's R&D incentives. The next lemma shows that, under symmetric abatement externalities, the two effects work in opposite directions, but the overall effect lowers R&D incentives.

Lemma 3 *In stage 1 of the late cooperation game, there are two strategic effects. If abatement externalities are symmetric ($\gamma_1 = \gamma_2$), they work in opposite directions. The overall effect, however, leads to **lower** investments in R&D, and, thus, reduces total welfare.*

To summarize, we identify three qualitatively different effects by which the outcome under late cooperation differs from the outcome under full cooperation: 1. the neglect of R&D externalities; 2. a strategic commitment effect that lowers R&D incentives; 3. a double externality effect that raises R&D incentives but does not offset the second effect when abatement externalities are not too asymmetric.[11]

3.5 Comparisons

In the following, we use our previous findings to derive some comparative welfare results.

Proposition 1 *Without R&D externalities, early cooperation leads to the full cooperative outcome.*

To understand this result note that, in the absence of R&D externalities, there are no potential gains from cooperation in terms of R&D. Hence, also under full cooperation, given the assigned abatement target, each country minimizes its *own* abatement costs. But this is also achieved under early cooperation.

Proposition 2 *Without abatement externalities, the outcome under late cooperation coincides with the outcome under no cooperation so that early cooperation is welfare superior to late cooperation.*

In the absence of abatement externalities, there are for any *fixed* R&D levels no potential gains from cooperation. Hence, late cooperation has no effect upon the final outcome when compared to no cooperation. Early cooperation, however, can achieve welfare gains because the hypothetical planner in stage 1 can, by assigning higher abatement targets to each country in stage 1, trigger additional R&D spill-overs in stage 2.

Proposition 3 *If abatement externalities are symmetric, early cooperation induces a higher total welfare than late cooperation.*

[11] By continuity, the assumption of symmetric abatement spill-overs ($\gamma_1 = \gamma_2$) in Lemma 3 can be replaced by the requirement that abatement externalities are not too asymmetric.

Proposition 3 extends the results from Propositions 1 and 2 to situations where there exist both abatement *and* R&D externalities. At first sight, one may think that early cooperation should *always* dominate late cooperation in terms of aggregate welfare. Yet, the proposition has the qualifier that abatement externalities are symmetric. This begs the question whether we can actually dispense with this qualifier. Clearly, by continuity the proposition will hold also for small asymmetries, but our next example shows that, with large asymmetries, early cooperation can yield a lower aggregate welfare than late cooperation.

The example considers an extreme asymmetry in abatement and R&D externalities. In particular, suppose country 1 benefits only from abatement by country 2, whereas country 2 does not benefit from any abatement at all: $B_1(a_1, a_2) = a_2$ and $B_2(a_1, a_2) = 0$. Moreover, suppose country 1 does not benefit from R&D spill-overs, while country 2 benefits only from the R&D spill-overs from country 1 and not from its own R&D effort: $C_1(a_1, r_1, r_2) = r_1^2 + a_1$ and $C_2(a_2, r_1, r_2) = a_2^3/(1 + 2r_1) + r_2$.[12] In this example, country 1's abatement induces only costs but no benefits. Independent of cooperation, its optimal level is therefore zero: $a_1 = 0$. Similarly, R&D of country 2 induces only costs. Hence, country 2 never exerts any positive R&D effort: $r_2 = 0$. Using $a_1 = r_2 = 0$, the payoff functions of the two countries simplify to: $\Pi_1 = a_2 - r_1^2$ and $\Pi_2 = -a_2^3/(1 + 2r_1)$. We next show how the abatement level a_2 and R&D level r_1 depend on the timing of cooperation.

Under early cooperation, the abatement levels are fixed at the R&D stage, and because R&D is only costly to country 1, it optimally chooses $r_1 = 0$. Given that country 1 exerts no R&D effort, the optimal level of abatement under early cooperation maximizes $a_2 - a_2^3$, which yields $a_2 = \sqrt{3}/3 \approx 0.58$. This outcome yields aggregate welfare of approximately 0.38.

Under late cooperation, the abatement levels are chosen after the R&D efforts and, in contrast to early cooperation, country 1 now has an incentive to invest in R&D in order to reduce country 2's costs of abatement and, thereby, trigger a higher abatement level at the cooperative stage. Indeed, given some R&D effort r_1, late cooperation leads to $a_2 = \sqrt{3 + 6r_1}/3$. Anticipating this level of abatement, it is optimal for country 1 to choose an R&D level of $r_1 \approx 0.42$, resulting in a total welfare of approximately $0.43 > 0.38$.

This counter-intuitive example, where late cooperation outperforms early cooperation, is best understood by recalling the different strategic effects of early versus late cooperation. The example is constructed in such a way that *only* the double externality effect that raises R&D

[12] Alternatively, think of country 1 as a country that benefits also from its own abatement but faces prohibitively high marginal abatement costs. Hence, this country never chooses a positive a_1. Similarly, country 2 can be interpreted as one with prohibitively high marginal R&D costs.

incentives under late cooperation is active, while under early cooperation, the planner is unable to trigger any positive R&D effort by country 1. Hence, only with late cooperation incentives for R&D are present and this results in higher aggregate welfare.

Finally, we also point out that there is no unambiguous ranking in total welfare of late cooperation versus no cooperation. Indeed, late cooperation can sometimes lead to a *lower* aggregated welfare than no cooperation. This is the case if the strategic incentive to enter the cooperative stage with high marginal abatement costs is sufficiently strong and induces more distortions than in the non-cooperative case, where neither environmental externalities nor knowledge spill-over effects are internalized. In a setup with side-payments, Beccherle and Tirole (2011) obtain a similar result.

4 The Endogenous Timing of Cooperation

In this section, we address the countries' incentives concerning the different stages of cooperation. In order to endogenize the timing of cooperation, we can no longer analyze the different timing games separately. It may be, for instance, that one of the countries prefers not to cooperate in the early stage, because it expects a better deal under late cooperation. Consequently, we must perform an integrated analysis of both early and late cooperation.

Because one of the paper's main interests lies in the interaction of abatement and R&D spill-overs, we focus in particular on the effects of asymmetries in spill-overs. Asymmetric spill-overs may arise for different reasons. For instance, countries may be asymmetric with respect to their technological state in that one country is a technological leader, while the other is a follower. In this case, knowledge spill-overs will mostly be unidirectional, from the leader to the follower. Similarly, asymmetries in abatement spill-overs may exist when some countries are, due to their specific location, particularly vulnerable to climate change, while others are less affected, or when one country attaches a higher value to this problem. A classical example of such asymmetries would be two countries that share a river, with one country being located upstream and the other one downstream. Under unilateral environmental externalities, one country benefits from increased abatement efforts by the other country, but not vice versa.

4.1 Unilateral externalities

In this subsection, we study extreme asymmetries in spill-over effects in the form of *unilateral externalities*, where only one country generates a given type of spill-over. We first address the stability and the timing of cooperation under such unilateral externalities:

Proposition 4 *Cooperation always fails when there is only a unilateral abatement externality but no R&D externality, or when there is a unilateral R&D externality but no abatement externality.*

Intuitively, when there is only a unilateral externality in either abatement or R&D, then the welfare of the country that generates the externality does neither depend on the abatement nor on the R&D choice of the other country. It therefore can attain its maximum welfare without cooperation. Hence, when cooperation does change the final outcome, then this country loses from cooperation and, therefore, has an incentive to veto it.

Proposition 5 *When there is both a unilateral abatement and a unilateral R&D externality, cooperation fails if both externalities positively affect the* ***same*** *country. If each of the externalities affects a different country, then late cooperation fails, while early cooperation can succeed if both externalities are of comparable strength.*

Proposition 5 demonstrates that for early cooperation to work, both unilateral effects must be sufficiently strong. To understand the intuition behind this result observe that in the absence of an R&D externality, the country that generates the abatement externality, say country 1, always rejects to cooperate, because it would be assigned a higher abatement level under cooperation, which reduces its welfare. Conversely, in the absence of an abatement externality, country 2 rejects to cooperate, because it is assigned a higher abatement level by the planner in order to induce this country to invest more in R&D. By continuity, one can find intermediate cases where both effects keep each other in balance so that both countries benefit from early cooperation.

4.2 Asymmetric spill-overs and the timing of cooperation

In order to illustrate how asymmetries in spill-overs affect the success and timing of cooperation, we now introduce a specific version of our general model. This specification allows us to derive closed-form solutions for all cases, but is nevertheless rich enough to provide some general insights regarding the timing of cooperation.

Let the cost function of country i have the following form

$$C_i(a_i, r_i, r_{-i}) = a_i^2 + r_i^2 - (1 + a_i)r_i - \lambda_{-i}r_{-i},$$

where $\lambda_{-i} \in [0,1]$ measures the R&D externality of country $-i$ on country i. Country i's payoff under this specific cost function is

$$\Pi_i(a_1, a_2, r_1, r_2) = a_i - a_i^2 + r_i - r_i^2 + a_i r_i + \gamma_{-i} a_{-i} + \lambda_{-i} r_{-i}.$$

This function is symmetric with respect to a_i and r_i. Hence, the interpretation of these variables as abatement and R&D is, at this general level, arbitrary. In principle, they could reflect any type of activities that cause externalities. What distinguishes these two variables is only the assumption that countries can cooperate in their choice of (a_1, a_2) but not in their choice of (r_1, r_2).

As an aggregate measure of the externalities caused by country i, it is useful to define

$$\mu_i \equiv (\gamma_i + \lambda_i)^2 + 2\gamma_i \lambda_i.$$

If country i does not exert an abatement externality ($\gamma_i = 0$), it follows that $\mu_i = \lambda_i^2$. Conversely, when $\lambda_i = 0$ then $\mu_i = \gamma_i^2$. Under the above assumptions, the following results are obtained in a straightforward manner:

1. Full cooperation:

$$a_i^f = \frac{3 + 2\gamma_i + \lambda_i}{3} \text{ and } r_i^f = \frac{3 + \gamma_i + 2\lambda_i}{3};$$

$$\Pi_i^f = 1 + \gamma_{-i} + \lambda_{-i} + \frac{\mu_{-i} + (\gamma_{-i} - \lambda_{-i})^2}{3} - \frac{\mu_i + (\gamma_i - \lambda_i)^2}{6}; \tag{12}$$

$$\Pi^f = 2 + \gamma_1 + \gamma_2 + \lambda_1 + \lambda_2 + \frac{\mu_1 + (\gamma_1 - \lambda_1)^2}{6} + \frac{\mu_2 + (\gamma_2 - \lambda_2)^2}{6}.$$

2. No cooperation:

$$a_i^n = 1 \text{ and } r_i^n = 1;$$

$$\Pi_i^n = 1 + \gamma_{-i} + \lambda_{-i}; \tag{13}$$

$$\Pi^n = 2 + \gamma_1 + \gamma_2 + \lambda_1 + \lambda_2.$$

3. Early cooperation:

$$a_i^e = \frac{3 + 2\gamma_i + \lambda_i}{3} \text{ and } r_i^e = \frac{3 + \gamma_i + \lambda_i/2}{3} < r_i^f;$$

$$\Pi_i^e = 1 + \gamma_{-i} + \lambda_{-i} + \frac{(2\gamma_{-i} + \lambda_{-i})^2}{6} - \frac{(2\gamma_i + \lambda_i)^2}{12}; \tag{14}$$

$$\Pi^e = 2 + \gamma_1 + \gamma_2 + \lambda_1 + \lambda_2 + \frac{(2\gamma_1 + \lambda_1)^2}{12} + \frac{(2\gamma_2 + \lambda_2)^2}{12}.$$

4. Late cooperation:

$$a_i^l = 1 + \gamma_i/2 \text{ and } r_i^l = 1 = r_i^n;$$

$$\Pi_i^l = 1 + \gamma_{-i} + \lambda_{-i} + \gamma_{-i}^2/2 - \gamma_i^2/4; \tag{15}$$

$$\Pi^l = 2 + \gamma_1 + \gamma_2 + \lambda_1 + \lambda_2 + \gamma_1^2/4 + \gamma_2^2/4.$$

Using these results, we obtain the following ranking of aggregated payoffs:

$$\Pi^f \geq \Pi^e \geq \Pi^l \geq \Pi^n.$$

Using (14) and (15), we find that country i prefers early cooperation over the late cooperation outcome if $\Pi_i^e - \Pi_i^l = (2\mu_{-i} - \mu_i)/12 > 0$, hence, if $\mu_i < 2\mu_{-i}$. In words, a country prefers early over late cooperation if the positive externalities it causes (μ_i) are not more than twice the size of the externalities caused by the other country (μ_{-i}). Furthermore, using (13) and (15), we find that country i prefers late cooperation over the non-cooperative outcome if $\gamma_i < \sqrt{2}\gamma_{-i}$, hence, if the abatement externality it causes is not more than $\sqrt{2}$ times the size of the abatement externality caused by the other country. Otherwise, late cooperation fails. However, early cooperation may still succeed, but the reference case is then the non-cooperative rather than the late cooperation outcome.[13]

Let us now apply the example to a case with two unilateral externalities. Suppose, country 1 benefits only from R&D spill-overs ($\gamma_1 > \gamma_2 = 0$), and country 2 benefits only from an abatement externality caused by country 1 ($\lambda_2 > \lambda_1 = 0$). Using the above results, we find:

$$\Pi_1^n = 1 + \lambda_2 \ , \ \Pi_2^n = 1 + \gamma_1;$$

$$\Pi_1^e = 1 + \lambda_2 + \frac{1}{6}\lambda_2^2 - \frac{1}{3}\gamma_1^2 \ , \ \Pi_2^e = 1 + \gamma_1 + \frac{2}{3}\gamma_1^2 - \frac{1}{12}\lambda_2^2;$$

$$\Pi_1^l = 1 + \lambda_2 - \frac{1}{4}\gamma_1^2 \ , \ \Pi_2^l = 1 + \gamma_1 + \frac{1}{2}\gamma_1^2.$$

Clearly, $\Pi_1^n > \Pi_1^l$, so that country 1 always rejects late cooperation. Hence, either early cooperation succeeds, or the non-cooperative outcome is obtained. For early cooperation to

[13] Under other specifications of benefits and costs, a country can manipulate its R&D choice upwards when early cooperation fails, in order to *induce* the other country to agree to late cooperation. This prevents some of the efficiency losses under no cooperation. In the above example, however, this type of strategy is ineffective, because (for any fixed r_{-i}) the effect of a change in r_i upon Π_{-i}^l and Π_{-i}^n is identical.

succeed, it must hold that $\Pi_1^e \geq \Pi_1^n$, which yields the condition $\sqrt{2}\gamma_1 \leq \lambda_2$. Similarly, the condition $\Pi_2^e \geq \Pi_2^n$ yields $\lambda_2 \leq 2\sqrt{2}\gamma_1$. Overall, we find that early cooperation succeeds if

$$\sqrt{2}\gamma_1 \leq \lambda_2 \leq 2\sqrt{2}\gamma_1.$$

Hence, as Proposition 5 indicates, given two unilateral externalities that go in opposite directions, both externalities must be of comparable size for early cooperation to succeed. Otherwise, the country that causes the stronger externality rejects early cooperation.

5 An Application to Global Warming

In the specific version of the model introduced above, many of the strategic effects identified in Section 3 are actually zero. In this section, we introduce an alternative specification of benefit and cost functions that seems suitable especially for an application of the model to the issue of global warming.

For global warming, it is appropriate to set the abatement externality equal to one ($\gamma_1 = \gamma_2 = 1$). Hence, the benefits of abatement for each country depend only on the *aggregate* level of abatement: $a \equiv a_1 + a_2$. Furthermore, we will assume that also R&D is a pure public good, highlighting the problem of knowledge spill-overs across countries (e.g. Jaffe et al. 2005). In particular, we use the following cost function:

$$C_i(a_i, r_i, r_{-i}) = \frac{a_i^2}{r} + r_i^2,$$

where $r \equiv r_1 + r_2$. Under this specification, country i's investment cost in R&D is a quadratic function of r_i. Similarly, the abatement cost is quadratic in a_i and declining in the aggregate level of R&D. Contrary to the previous example, r_{-i} has an effect not only on country i's total cost, but also on its *marginal* abatement cost $\partial C_i / \partial a_i$. As a result, several strategic effects will affect the outcome that were absent in the previous example.

The asymmetry we consider in this example is one where countries differ in their appreciation of abatement. In particular, we assume the benefit functions

$$B_1(a) = (1 + \delta)a \quad \text{and} \quad B_2(a) = (1 - \delta)a,$$

where $\delta \in [0, 1/2]$ assures an interior solution. If $\delta > 0$, country 1 values abatement more than country 2. Note, however, that the aggregated benefit $B_1(a) + B_2(a)$ is always $2a$. This property allows us to derive closed-form solutions for all four cases:

1. Full cooperation:

$$a_i^n = 2\ ,\ r_i^n = 1;$$

$$\Pi_1^f = 1 + 4\delta\ ,\ \Pi_2^f = 1 - 4\delta; \tag{16}$$

$$\Pi^f = 2.$$

2. No cooperation:

$$a_1^n = \frac{(1+\delta)(3-\delta^2)}{8}, \quad a_2^n = \frac{(1-\delta)(3-\delta^2)}{8}, \quad r_1^n = \frac{3+2\delta-\delta^2}{8}, \quad r_2^n = \frac{3-2\delta-\delta^2}{8};$$

$$\Pi_1^n = \frac{27+12\delta-22\delta^2-4\delta^3+3\delta^4}{64}, \quad \Pi_2^n = \frac{27-12\delta-22\delta^2+4\delta^3+3\delta^4}{64}; \tag{17}$$

$$\Pi^n = \frac{27-22\delta^2+3\delta^4}{32}.$$

3. Early cooperation:

$$a_i^e = 216/125 = 1.728, \quad r_i^e = 18/25 = 0.72;$$

$$\Pi_1^e = 108(1+4\delta)/125 = 108\Pi_1^f/125, \quad \Pi_2^e = 108(1-4\delta)/125 = 108\Pi_2^f/125; \tag{18}$$

$$\Pi^e = 216/125 = 108\Pi^f/125.$$

4. Late cooperation[14]:

$$a_i^l = 1, \quad r_1^l = 1/2 + \delta, \quad r_2^l = 1/2 - \delta;$$

$$\Pi_1^l = 3/4 + \delta - \delta^2, \quad \Pi_2^l = 3/4 - \delta - \delta^2; \tag{19}$$

$$\Pi^l = 3/2 - 2\delta^2.$$

Using these results, we obtain the following ranking of aggregated payoffs:

$$\Pi^f > \Pi^e > \Pi^l > \Pi^n.$$

Endogenization of cooperation:
In order to obtain an endogenous outcome for the timing of cooperation, we consider the overall game and solve it by backwards induction. The following lemma specifies the conditions under which each individual country benefits from late cooperation rather than from rejecting it.

[14] The parameter restriction $\delta \in [0, 1/2]$ ensures that all non-negativity constraints are automatically satisfied.

Lemma 4 *Given the above specification of benefit and cost functions, late cooperation succeeds iff* $\delta < 3 - 2\sqrt{2} \approx 0.17$.

The lemma shows that late cooperation succeeds if countries are not too asymmetric. If $\delta \leq 0.17$, the equilibrium payoffs are given by (19), and if $\delta > 0.17$, they are given by (17). When δ changes in a comparative statics sense, then at the critical value for δ, there is a discontinuity. Note that when $\delta \geq 0.17$, late cooperation fails even if both countries would prefer it to the non-cooperative outcome. The reason is that for $\delta \geq 0.17$, countries anticipate that for any *given* R&D levels, late cooperation fails and, in anticipation of this, countries choose the non-cooperative R&D levels according to expression (17).

Given these results, we can solve for the equilibrium outcome of the overall game. The next proposition endogenizes the timing of cooperation, depending on the degree of asymmetry δ.

Proposition 6 *Given the above specification of benefit and cost functions, early cooperation succeeds iff* $\delta < 0.047$*. Hence, if* $\delta \in (0.047, 0.17)$*, late cooperation succeeds, while early cooperation fails. If* $\delta > 0.17$*, the non-cooperative outcome is obtained.*

The result of Proposition 6 is best understood by considering Figure 2. It shows country 2's payoff when early cooperation succeeds and when it fails (using (17), (18), and (19)):

Figure 2

The aforementioned discontinuity at the location $\delta = 0.17$ reflects - given that early cooperation fails - the transition from the late cooperation outcome to the non-cooperative outcome. The figure illustrates that only when δ is sufficiently small ($\delta < 0.047$), early cooperation succeeds, while late cooperation succeeds for a larger set of parameter values.

Hence, although early cooperation leads to a higher aggregated welfare than late cooperation, country 2, for strategic reasons, rejects early cooperation already under a small degree of asymmetry between countries. Under early cooperation, the hypothetical social planner who fixes the countries' future abatement targets assigns an equal abatement level to both countries (due to the symmetry of the abatement cost functions). Under late cooperation, in contrast, the assigned abatement targets depend on the countries' previous investments in R&D. Therefore, under late cooperation, the total costs of abatement (including R&D costs) can effectively be redistributed between countries via their choice of R&D levels, which makes cooperation more sustainable also for higher degrees of asymmetry (*redistributive effect*). This is not possible under early cooperation. This suggests that an asymmetry in the countries' appreciation of abatement makes early cooperation less likely to occur, because it does not allow countries to redistribute the efforts of abatement via the strategic choice of R&D levels. The redistributive effect, thus, favors late cooperation and destabilizes early cooperation when countries are asymmetric.

6 Conclusion

The Tinbergen rule on public policy (e.g., Tinbergen 1952) tells us that in order to implement multiple policy targets, at least one policy tool is required for each target. Applying this rule to cooperation, one cannot expect a climate treaty that addresses only abatement targets to attain the first-best outcome in the presence of multiple market failures such as environmental externalities and knowledge spill-overs. Going beyond this classical Tinbergen rule, our analysis shows that things can actually be much worse: (long-term) treaties that address only one problem may fail to materialize at all and, therefore, yield no improvements whatsoever. We demonstrated this feature in a stylized model of cooperation between two countries and two externalities that act asymmetrically.

A pessimistic insight from our analysis is that even small asymmetries between countries can cause failures in cooperation. Due to the redistributive effect, especially early (long-term) cooperation is vulnerable to small asymmetries. The reason for this is that a delay in cooperation can put one of the countries in a strategically more favorable position. A country that benefits relatively less from climate protection can, via its strategic choice of R&D levels, shift the burden of abatement costs towards the other country. This favors late (short-term) cooperation, and makes it sustainable also for higher degrees of asymmetry.

On a more optimistic note, we identified a positive complementarity effect of long-term abatement, so that long-term agreements on abatement levels do not crowd-out efforts in R&D. Moreover, even strong asymmetries do not necessarily destabilize early cooperation. With primarily unilateral externalities, for instance, where one country has (relatively) higher benefits of abatement, while the other country benefits mostly from knowledge spill-overs, long-term treaties are sustainable. Intuitively, early cooperation requires both countries to adopt more ambitious abatement targets. Therefore, a country with higher benefits of abatement, such as for instance an industrialized country with a relatively high willingness-to-pay for climate stability, may agree to cooperate early in order to induce the other country to abate more as well. Conversely, a country that benefits mostly from knowledge spill-overs, such as for instance a developing country with a high potential for abatement, may agree to a long-term abatement target in order to induce the other country to subsequently invest more in R&D. This interdependency in countries' welfare in the presence of multiple market failures can actually facilitate an early agreement on long-term emission targets. The general conclusion whether asymmetries foster or impede cooperation is therefore that, in the context of asymmetries, type rather than size matters.

Appendix

This appendix collects the formal proofs of the lemmas and propositions.

Proof of Lemma 1: Due to the convexity of C_i, it is to show that the right-hand side of (5) is non-negative. In the presence of an abatement externality, we have $\gamma_{-i} > 0$ so that it remains to be shown that $\partial a_{-i}^n / \partial r_i \geq 0$. To determine the sign of $\partial a_{-i}^n / \partial r_i$, we use the first-order conditions (4) of stage 2 and insert the functions $a_1^n(r_1, r_2)$ and $a_2^n(r_1, r_2)$ that describe the equilibrium outcome:

$$\frac{\partial C_i(a_i^n(r_1, r_2), r_1, r_2)}{\partial a_i} = 1.$$

These conditions are, by the definition of $a_i^n(r_1, r_2)$, fulfilled for *any* (r_1, r_2). Therefore, the derivative w.r.t. r_1 and r_2 is zero. This yields in particular for r_{-i}:

$$\frac{\partial^2 C_i}{\partial a_i^2} \frac{\partial a_i^n}{\partial r_{-i}} + \frac{\partial^2 C_i}{\partial a_i \partial r_{-i}} = 0.$$

Swapping the indices i and $-i$, we, thus, obtain:

$$\frac{\partial a_{-i}^n}{\partial r_i} = -\left(\frac{\partial^2 C_{-i}}{\partial a_{-i} \partial r_i}\right) / \left(\frac{\partial^2 C_{-i}}{\partial a_{-i}^2}\right).$$

Because by assumption $\partial^2 C_i / \partial a_i \partial r_{-i} \leq 0$, the numerator is non-positive, and due to $\partial^2 C_i / \partial a_i^2 > 0$ the denominator is positive. Hence, $\partial a_{-i}^n / \partial r_i$ is non-negative. Q.E.D.

Proof of Lemma 2: To show that the overall strategic effect in stage 1 of the early cooperation game increases total welfare, note that the hypothetical social planner who chooses the abatement targets in stage 1 could neglect this effect. The outcome would then be defined by the following system:

$$\frac{\partial C_i}{\partial r_i} = 0 \quad \text{and} \quad \frac{\partial C_i}{\partial a_i} = 1 + \gamma_i.$$

However, the planner maximizes total welfare in stage 1, in full anticipation of the outcome in stage 2. Therefore, total welfare is strictly higher whenever the strategic effect is included in the planner's decision-making and affects the final outcome.

To show the second claim, note that the functions $r_1^e(a_1, a_2)$ and $r_2^e(a_1, a_2)$ that describe the outcome of R&D competition in stage 2, are implicitly defined by the system $\partial C_i(a_i, r_1^e(a_1, a_2), r_2^e(a_1, a_2))/\partial r_i = 0$ for $i = 1, 2$ (see (7)). Differentiation w.r.t. a_i yields

$$\frac{\partial^2 C_i}{\partial r_i \partial a_i} + \frac{\partial^2 C_i}{\partial r_i^2}\frac{\partial r_i^e}{\partial a_i} + \frac{\partial^2 C_i}{\partial r_i \partial r_{-i}}\frac{\partial r_{-i}^e}{\partial a_i} = 0 \quad \text{and} \quad \frac{\partial^2 C_{-i}}{\partial r_{-i}^2}\frac{\partial r_{-i}^e}{\partial a_i} + \frac{\partial^2 C_{-i}}{\partial r_i \partial r_{-i}}\frac{\partial r_i^e}{\partial a_i} = 0.$$

From which follows:

$$\frac{\partial r_i^e}{\partial a_i} = \left(-\frac{\partial^2 C_i}{\partial r_i \partial a_i}\frac{\partial^2 C_{-i}}{\partial r_{-i}^2}\right) \Big/ \left(\frac{\partial^2 C_i}{\partial r_i^2}\frac{\partial^2 C_{-i}}{\partial r_{-i}^2} - \frac{\partial^2 C_i}{\partial r_i \partial r_{-i}}\frac{\partial^2 C_{-i}}{\partial r_i \partial r_{-i}}\right);$$
$$\frac{\partial r_{-i}^e}{\partial a_i} = \left(\frac{\partial^2 C_i}{\partial r_i \partial a_i}\frac{\partial^2 C_{-i}}{\partial r_i \partial r_{-i}}\right) \Big/ \left(\frac{\partial^2 C_i}{\partial r_i^2}\frac{\partial^2 C_{-i}}{\partial r_{-i}^2} - \frac{\partial^2 C_i}{\partial r_i \partial r_{-i}}\frac{\partial^2 C_{-i}}{\partial r_i \partial r_{-i}}\right).$$

Given the assumptions $\partial^2 C_i/\partial r_i^2 > 0$, $\partial^2 C_i/\partial r_i \partial r_{-i} \geq 0$, and $\partial^2 C_i/\partial r_i^2 \geq \partial^2 C_i/\partial r_i \partial r_{-i}$, the denominator is always positive. Using $\partial^2 C_i/\partial a_i \partial r_i \leq 0$ and $\partial^2 C_i/\partial r_i \partial r_{-i} \leq 0$, we find that $\partial r_i^e/\partial a_i \geq 0$ and $\partial r_{-i}^e/\partial a_i \leq 0$. Using the condition $\partial^2 C_i/\partial r_i^2 \geq \partial^2 C_i/\partial r_i \partial r_{-i}$, we find $\partial r_i^e/\partial a_i \geq \left|\partial r_{-i}^e/\partial a_i\right|$. If countries are symmetric, then $\partial C_{-i}/\partial r_i = \partial C_i/\partial r_{-i}$. The overall strategic effect is, then, always non-positive, which (by (7) and (8)) implies higher abatement and higher R&D levels than in the absence of the strategic effect. Q.E.D.

Proof of Lemma 3: Using $\gamma = \gamma_1 = \gamma_2$ and inserting the function $a_i^l(r_1, r_2)$ into (10) yields

$$\frac{\partial C_i\left(a_i^l(r_1, r_2), r_1, r_2\right)}{\partial a_i} = 1 + \gamma.$$

Taking the derivative w.r.t. r_i and r_{-i}, we obtain the following conditions:

$$\frac{\partial^2 C_i}{\partial a_i^2}\frac{\partial a_i^l}{\partial r_i} + \frac{\partial^2 C_i}{\partial a_i \partial r_i} = 0 \quad \text{and} \quad \frac{\partial^2 C_i}{\partial a_i^2}\frac{\partial a_i^l}{\partial r_{-i}} + \frac{\partial^2 C_i}{\partial a_i \partial r_{-i}} = 0.$$

After rearranging terms we find

$$\frac{\partial a_i^l}{\partial r_i} = -\left(\frac{\partial^2 C_i}{\partial a_i \partial r_i}\right) \Big/ \left(\frac{\partial^2 C_i}{\partial a_i^2}\right) \quad \text{and} \quad \frac{\partial a_i^l}{\partial r_{-i}} = -\left(\frac{\partial^2 C_i}{\partial a_i \partial r_{-i}}\right) \Big/ \left(\frac{\partial^2 C_i}{\partial a_i^2}\right).$$

Using the regularity conditions $\partial^2 C_i/\partial a_i^2 > 0$, $\partial^2 C_i/\partial a_i \partial r_i \leq 0$, $\partial^2 C_i/\partial a_i \partial r_{-i} \leq 0$, and $|\partial^2 C_i/\partial a_i \partial r_i| \geq |\partial^2 C_i/\partial a_i \partial r_{-i}|$, we find that $\partial a_i^l/\partial r_i$ and $\partial a_i^l/\partial r_{-i}$ are both non-negative, and $\partial a_i^l/\partial r_i \geq \partial a_i^l/\partial r_{-i}$. Hence, we obtain for the overall strategic effect $\gamma\left(\partial a_{-i}^l/\partial r_i - \partial a_i^l/\partial r_i\right) \leq 0$, which (by (11)) implies lower R&D efforts than in the absence of strategic effects. The reduction in R&D-levels aggravates the problem of under-investments in R&D (relative to the full cooperation case) and, therefore, unambiguously reduces total welfare. Q.E.D.

Proof of Proposition 1: Without R&D externalities we have $\partial C_{-i}/\partial r_i = 0$. Using (2) and (3), the cooperative solution $(a_1^f, a_2^f, r_1^f, r_2^f)$ is with $i = 1, 2$ characterized by

$$\frac{\partial C_i}{\partial a_i} = 1 + \gamma_i \quad \text{and} \quad \frac{\partial C_i}{\partial r_i} = 0.$$

According to (7) and (8), early cooperation $(a_1^e, a_2^e, r_1^e, r_2^e)$ is characterized by the system

$$\frac{\partial C_i}{\partial r_i} = 0 \quad \text{and} \quad \frac{\partial C_i}{\partial a_i} = 1 + \gamma_i.$$

The result follows immediately because the two systems of optimality conditions coincide. Q.E.D.

Proof of Proposition 2: Without abatement externalities we have $\gamma_1 = \gamma_2 = 0$. Using (4) and (5), the solution without cooperation, $(a_1^n, a_2^n, r_1^n, r_2^n)$, solves for $i = 1, 2$

$$\frac{\partial C_i}{\partial a_i} = 1 \quad \text{and} \quad \frac{\partial C_i}{\partial r_i} = 0.$$

Using (10) and (11), we find that the late cooperation outcome $(a_1^l, a_2^l, r_1^l, r_2^l)$ is characterized by the same system of optimality conditions. Hence, the outcome without cooperation and with late cooperation are identical.

Early cooperation satisfies the system (7) and (8), which differs from the other system of equations due to the presence of strategic effects. By ignoring them, the hypothetical planner in stage 1 can assure a total welfare at least as large as under late or no cooperation. Q.E.D.

Proof of Proposition 3: The outcome under late cooperation is characterized by the system (10) and (11). According the Lemma 3, the overall strategic effect in stage 1 (see (11)) is welfare-reducing if $\gamma_1 = \gamma_2$. The outcome under early cooperation satisfies the system (7) and (8), which differs from (10) and (11). According to Lemma 2, the overall strategic effect in stage 1 (see (8)) is welfare-enhancing. The claim, thus, follows immediately from Lemma 2 and Lemma 3. Q.E.D.

Proof of Proposition 4: To show that cooperation in the countries' choices of abatement targets fails when there is a unilateral abatement externality but no R&D externality, note that the country that generates the abatement externality can achieve its own welfare maximum without any cooperation. Hence, this country can never gain from cooperating. Although by Proposition 1 early cooperation leads to a higher *total* welfare, the country that generates the externality loses from cooperation as it is assigned an abatement effort above its individual optimum.

To show that cooperation fails when there is a unilateral R&D externality but no abatement externality, note first that by Proposition 2 late cooperation in abatement yields no welfare gains. Hence, for any *fixed* R&D levels, the option not to cooperate weakly dominates the option to cooperate because by assumption countries only cooperate if at least one country gains from cooperation. Furthermore, early cooperation fails, because the country that generates the R&D externality has no gains from it. The planner assigns a higher abatement level to this country in the cooperative stage (stage 1) in order to trigger additional R&D investments by this country in stage 2. This reduces the welfare of this country that can achieve its own welfare maximum without any cooperation. Q.E.D.

Proof of Proposition 5: To show the first claim, note that the country that generates both externalities achieves its individual welfare maximum without any cooperation. Hence, whenever cooperation affects the final outcome, it negatively affects this country's welfare.

To show that *late cooperation* fails if each of the unilateral externalities affects a different country, note that for fixed R&D levels, the country that generates the abatement externality suffers from cooperation in stage 2. For any *given* R&D levels, it attains its individual welfare maximum when ignoring the abatement externality. Hence, it rejects late cooperation.

To show that *early cooperation* succeeds if both externalities are sufficiently strong and each affects a different country, rewrite the optimality conditions for early cooperation ((7) and (8)) for this case. Without loss of generality, suppose country 1 exerts an abatement externality on country 2 but not vice versa ($\gamma_1 > \gamma_2 = 0$) and country 2 exerts an R&D externality on country 1 but not the other way around ($\partial C_1/\partial r_2 < \partial C_2/\partial r_1 = 0$). Hence, we obtain:

$$\frac{\partial C_i}{\partial r_i} = 0, \quad \frac{\partial C_1}{\partial a_1} + \frac{\partial C_1}{\partial r_2}\frac{\partial r_2^e}{\partial a_1} = 1 + \gamma_1 \quad \text{and} \quad \frac{\partial C_2}{\partial a_2} + \frac{\partial C_1}{\partial r_2}\frac{\partial r_2^e}{\partial a_2} = 1.$$

From the proof of Lemma 2, we obtain the following expressions:

$$\frac{\partial r_2^e}{\partial a_2} = \left(-\frac{\partial^2 C_2}{\partial r_2 \partial a_2}\frac{\partial^2 C_1}{\partial r_1^2}\right) \Big/ \left(\frac{\partial^2 C_2}{\partial r_2^2}\frac{\partial^2 C_1}{\partial r_1^2} - \frac{\partial^2 C_2}{\partial r_2 \partial r_1}\frac{\partial^2 C_1}{\partial r_2 \partial r_1}\right);$$

$$\frac{\partial r_2^e}{\partial a_1} = \left(\frac{\partial^2 C_1}{\partial r_1 \partial a_1}\frac{\partial^2 C_2}{\partial r_1 \partial r_2}\right) \Big/ \left(\frac{\partial^2 C_1}{\partial r_1^2}\frac{\partial^2 C_2}{\partial r_2^2} - \frac{\partial^2 C_1}{\partial r_1 \partial r_2}\frac{\partial^2 C_2}{\partial r_1 \partial r_2}\right).$$

Since C_2 is not a function of r_1, these expressions simplify to

$$\frac{\partial r_2^e}{\partial a_2} = \left(-\frac{\partial^2 C_2}{\partial r_2 \partial a_2}\frac{\partial^2 C_1}{\partial r_1^2}\right) \Big/ \left(\frac{\partial^2 C_2}{\partial r_2^2}\frac{\partial^2 C_1}{\partial r_1^2}\right) \quad \text{and} \quad \frac{\partial r_2^e}{\partial a_1} = 0.$$

Using the following regularity assumptions: $\partial^2 C_i/\partial r_i^2 > 0$ and $\partial^2 C_i/\partial a_i \partial r_i \leq 0$, we find that $\partial r_2^e/\partial a_2 \geq 0$. Hence, the product $\partial C_1/\partial r_2 \cdot \partial r_2^e/\partial a_2$ is non-positive, which implies a higher abatement level a_2 than in the absence of this strategic effect, and by $\partial C_2/\partial r_2 = 0$ also a higher R&D level r_2. To summarize: under early cooperation and unilateral externalities, the following conditions are fulfilled:

$$\frac{\partial C_i}{\partial r_i} = 0, \quad \frac{\partial C_1}{\partial a_1} = 1 + \gamma_1, \qquad \text{and} \quad \frac{\partial C_2}{\partial a_2} - \left(\frac{\partial C_1}{\partial r_2}\frac{\partial^2 C_2}{\partial r_2 \partial a_2}\frac{\partial^2 C_1}{\partial r_1^2}\right) / \left(\frac{\partial^2 C_2}{\partial r_2^2}\frac{\partial^2 C_1}{\partial r_1^2}\right) = 1.$$

The expression γ_1 on the right-hand side of the second equation induces a higher abatement level a_1 (as compared to no cooperation), and by $\partial C_1/\partial r_1 = 0$ also leads to a higher R&D level r_1. The higher abatement levels (a_1^e, a_2^e) increase country 2's welfare that benefits from the abatement externality, and the higher R&D levels (r_1^e, r_2^e) are beneficial to country 1 that benefits from knowledge spill-overs. Hence, both countries can benefit from these effects if both externalities are sufficiently strong. Q.E.D.

Proof of Lemma 4: For any *given* R&D levels (r_1, r_2), late cooperation succeeds if both countries achieve a higher welfare than in the absence of cooperation. Hence, late cooperation succeeds if $\Delta\Pi_1^{ln}(r_1, r_2) \equiv \Pi_1^l(r_1, r_2) - \Pi_1^n(r_1, r_2) \geq 0$ and $\Delta\Pi_2^{ln}(r_1, r_2) \equiv \Pi_2^l(r_1, r_2) - \Pi_2^n(r_1, r_2) \geq 0$. Using the results from Section 3, it is straight-forward to derive the following expressions for welfare (for fixed R&D levels (r_1, r_2)):

$$\Pi_1^l(r_1, r_2) = (1 + 2\delta)(r_1 + r_2) - r_1^2, \quad \Pi_2^l(r_1, r_2) = (1 - 2\delta)(r_1 + r_2) - r_2^2;$$

$$\Pi_1^n(r_1, r_2) = (3 + 2\delta - \delta^2)(r_1 + r_2)/4 - r_1^2, \quad \Pi_2^n(r_1, r_2) = (3 - 2\delta - \delta^2)(r_1 + r_2)/4 - r_2^2.$$

This implies

$$\Delta\Pi_1^{ln} = (1 + 6\delta + \delta^2)(r_1 + r_2)/4 > 0 \quad \text{and} \quad \Delta\Pi_2^{ln} = (1 - 6\delta + \delta^2)(r_1 + r_2)/4.$$

Hence, country 1 is always willing to cooperate. Country 2 cooperates if $1 - 6\delta + \delta^2 \geq 0$, which, for the relevant interval $\delta \in [0, 1/2]$, yields the critical value $3 - 2\sqrt{2} \approx 0.17$. If δ is below this value, it holds that $\Delta\Pi_2^{ln} > 0$. Note, that the critical value for δ is *independent* of r_1 and r_2. Therefore, late cooperation succeeds iff $\delta \leq 3 - 2\sqrt{2}$. Q.E.D.

Proof of Proposition 6: The result follows directly from a comparison of (17), (18), and (19). Q.E.D.

References

Asheim, G.B., Holtsmark, B. (2009) Renegotiation-proof climate agreements with full participation: conditions for Pareto-efficiency, *Environmental and Resource Economics*, 43, p. 519-533

Barrett, S. (1994) Self-enforcing international environmental agreements, *Oxford Economic Papers*, 46, p. 878-894

Barrett, S. (1997) The strategy of trade sanctions in international environmental agreements, *Resource and Energy Economics*, 19, p. 345-361

Barrett, S. (2001) International cooperation for sale, *European Economic Review*, 45, p. 1835-1850

Beccherle, J., Tirole, J. (2011) Regional Initiatives and the Cost of Delaying Binding Climate Change Agreements, forthcoming in *Journal of Public Economics*, doi:10.1016/j.jpubeco.2011.04.007

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

Carraro, C., Eyckmans, J., Finus, M. (2006) Optimal transfers and participation decisions in international environmental agreements, *Review of International Organizations*, 1, p. 379-396

Finus, M. (2008) Game theoretic research on the design of international environmental agreements: insights, critical remarks, and future challenges, *International Review of Environmental and Resource Economics*, 2, p. 29-67

Fischer, C., Parry, I.W.H., Pizer, W.A. (2003) Instrument choice for environmental protection when technological innovation is endogenous, *Journal of Environmental Economics and Management*, 45, p. 523-545

Froyn, C.B., Hovi, J. (2008) A climate agreement with full participation, *Economics Letters*, 99, p. 317-319

Gerlagh, R., Kverndokk, S., Rosendahl, K.E. (2009) Optimal Timing of Climate Change Policy: Interaction Between Carbon Taxes and Innovation Externalities, *Environmental and Resource Economics*, 43, p. 369-390

Golombek, R., Hoel, M. (2004) Unilateral Emission Reductions and Cross-Country Technology Spillovers, *B.E. Journal of Economic Analysis & Policy*, vol. advances 4, issue 2, article 3

Golombek, R., Hoel, M. (2008) Endogenous technology and tradable emission quotas, *Resource and Energy Economics*, 30, p. 197-208

Harstad, B. (2007) Harmonization and Side Payments in Political Cooperation, *American Economic Review*, 97, p. 871-889

Harstad, B. (2010) The Dynamics of Climate Agreements, working paper, No 2962, CESifo Working Paper Series from CESifo Group Munich

Heal, G., Tarui, N. (2010) Investment and emission control under technology and pollution externalities, *Resource and Energy Economics*, 32, p. 1-14

Heitzig, J., Lessmann, K., Zou, Y. (2011) Self-enforcing strategies to deter free-riding in the climate change mitigation game and other repeated public good games, forthcoming in *Proceedings of the National Academy of Sciences*, doi:10.1073/pnas.1106265108

Hoel, M., Schneider, K. (1997) Incentives to Participate in an International Environmental Agreement, *Environmental and Resource Economics*, 9, p. 153-170

Jaffe, A.B., Newell, R.G., Stavins, N. (2005) A tale of two market failures: Technology and environmental policy, *Ecological Economics*, 54, p. 164-174

Tinbergen, J. (1952), *On the Theory of Economic Policy.* Amsterdam: North-Holland.

Figures:

Figure 1: Game tree

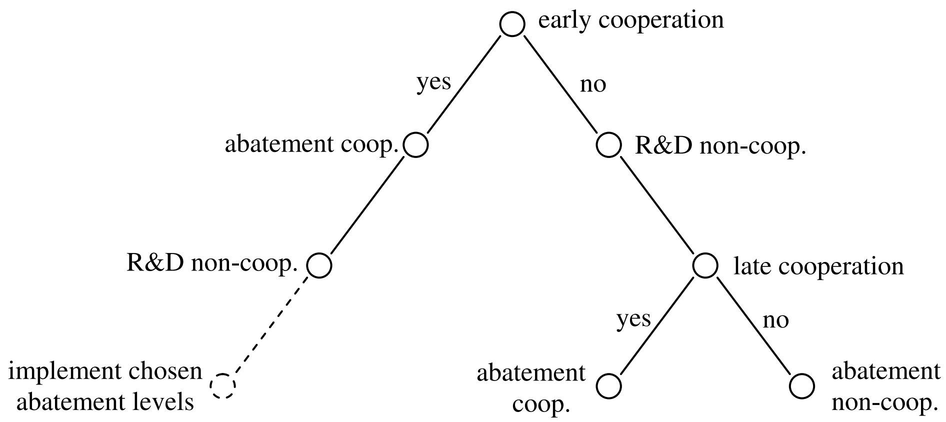

Figure 2: Endogenization of cooperation