University of Heidelberg

Department of Economics

Discussion Paper Series | No. 584

Strategic Conflicts on the Horizon: R&D Incentives for Environmental Technologies

Daniel Heyen

February 2015

Strategic Conflicts on the Horizon: R&D Incentives for Environmental Technologies *

Daniel Heyen
Heidelberg University
Department of Economics
heyen@eco.uni-heidelberg.de

February 26, 2015

Abstract

Technological innovation is a key strategy for tackling environmental problems. The required R&D expenditures however are substantial and fall on self-interested countries. Thus, the prospects of successful innovation critically depend on innovation incentives. This paper focuses on a specific mechanism for strategic distortions in this R&D game. In this mechanism, the outlook of future conflicts surrounding technology deployment directly impacts on the willingness to undertake R&D. Apart from free-riding, a different deployment conflict with distortive effects on innovation may occur: Low deployment costs and heterogeneous preferences might give rise to 'free-driving'. In this recently considered possibility (Weitzman 2012), the country with the highest preference for technology deployment, the free-driver, may dominate the deployment outcome to the detriment of others. The present paper develops a simple two stage model for analyzing how technology deployment conflicts, free-riding and free-driving, shape R&D incentives of two asymmetric countries. The framework gives rise to rich findings, underpinning the narrative that future deployment conflicts pull forward to the R&D stage. While the outlook of free-riding unambiguously weakens innovation incentives, the findings for free-driving are more complex, including the possibility of super-optimal R&D and incentives for counter-R&D.
Keywords: Environmental Innovation; R&D Game; Innovation Incentives; Externalities; Strategic Conflicts; Climate Engineering.
JEL Codes: Q55; O31; Q54; H41; D62.

1 Introduction

Technological innovation is a key strategy for tackling environmental problems. Important examples include CO_2 abatement technologies (Bosetti et al. 2009; Perino and Requate 2012; Poyago-Theotoky 2007) and 'breakthrough technologies' like no-emission

*I thank conference participants at the World Congress of Environmental and Resource Economists (WCERE) 2014 in Istanbul, the AUROE Young Academics Workshop 2014 in Kiel, seminar participants at the FZU-ZEW Monthly Brown Bag Seminar, the Brown Bag Seminar at the Chair of Environmental Economics, and participants of the climate engineering summer schools at Harvard University 2013 and Heidelberg 2014 for helpful comments.

energy sources (Barrett 2006; Hoffert et al. 2002). A characteristic of novel technologies is that they require costly R&D to become available (Popp 2006; Harstad 2012; Golombek and Hoel 2011). In the absence of a supranational entity undertaking or enforcing the development of these technologies, the main burden for R&D expenditures falls on countries. Domestic interests and strategic considerations, however, often stand in the way of implementing actions that would improve global well-being. The prospects for successful development of potentially welfare enhancing environmental technologies thus critically depend on countries' incentives to engage in R&D.

In this R&D game, incentives can deviate from the global optimal for different reasons. A possible cause for an overall insufficient willingness to develop technologies is the public good nature of knowledge (Stiglitz 1999), leading to free-riding on other countries' R&D efforts (Popp 2010; Hall and Helmers 2013). Another reason for inefficiencies in the technological innovation process is that future deployment of the technologies, usually involving significant externalities, itself is prone to strategic considerations. Such outlook of suboptimal deployment patterns, in turn, may also shape the incentives for technology R&D. This mechanism has been explicitly raised by Popp (2010)[1] and is present, yet less prominently, in a couple of other contributions (Hall and Helmers 2013; Perino and Requate 2012; Barrett 2006; Hoel and De Zeeuw 2010). It is this impact of anticipated technology deployment profiles on R&D incentives that is the focus of the present inquiry.

The mechanism described by Popp (2010) is that R&D incentives are weakened due to the anticipation of free-riding behavior in future use of the technology. But free-riding is not the only deviation from global optimal deployment that can occur in the context of environmental technologies with transboundary effects. To explore the possibility of other deployment conflicts, it is helpful to focus on a specific set of environmental technologies. 'Climate engineering' (CE), or 'geoengineering', is the "deliberate large-scale manipulation of the planetary environment to counteract anthropogenic climate change" (Shepherd 2009). The main categories are so-called *carbon dioxide removal* (CDR) techniques that aim at reducing the stock of greenhouse gases in the atmosphere, for instance by removing CO2 from the ambient air by means of chemicals ('direct air capture', see Keith et al. 2006), and *solar radiation management* (SRM) techniques that would alter the earth's radiation balance, for instance by the release of sulfur particles ('stratospheric aerosol injection', see Caldeira et al. 2013).

CE technologies raise a set of new economical and political questions (Barrett 2008; Finus et al. 2013; Victor 2008). While CDR techniques like direct air capture may, in terms of the surrounding incentive structure, be very similar to the mitigation of greenhouse gases and thus prone to free-riding (Chen and Tavoni 2013), the low deployment costs of SRM are expected to induce a novel strategic conflict, 'free driving' (Weitzman

[1] "Thus, without appropriate policy interventions, the market for technologies that reduce emissions will be limited, reducing incentives to develop such technologies" (Popp 2010).

2012): If preferences for global cooling are different across countries, the country aiming at the strongest temperature reduction, the 'free-driver', will use SRM as to maximize its own payoffs. Due to the low costs and global effects of SRM deployment, the resulting amount of cooling may considerably exceed the level of cooling other countries prefer. The domination of the outcome by one single country jeopardizes the global benefits SRM could provide if deployed as to maximize global welfare.

The present paper asks about the repercussions of these strategic conflicts, free-riding and free-driving, on the incentives for developing such a technology. For this purpose, it develops a simple game-theoretical framework in which two countries non-cooperatively play a threshold R&D game in the first period and, conditional on successful R&D, a deployment game for a transboundary environmental technology in the second period. The general set-up combines well-known building blocks from the literature: The two stage R&D game structure is borrowed from the industrial organization literature (D'Aspremont and Jacquemin 1988; Kamien et al. 1992), the threshold R&D structure stems from Barrett (2006), and the technology deployment game is a standard public good game (Barrett 1994; Finus and Rübbelke 2013; Diamantoudi and Sartzetakis 2006) with heterogeneous players (Barrett 2001; McGinty 2006).

CE technologies are a fitting illustration for this general framework. Different technologies are expected to bring about the different strategic conflicts free-riding and free-driving. Also, CE technologies are not developed yet and thus require sufficient R&D incentives to become available.

The paper makes three main contributions. The first contribution is to demonstrate that free-rider and free-driver equilibria can emerge in the same standard public good framework. In comparison to Weitzman (2012), the smooth (inverse U-shaped) benefit function is more tractable than a kinked benefit function. Also, the cost parameter in the present framework plays an important and informative role while Weitzman (2012) abstracts from technology deployment costs. The second contribution is to provide a simple and tractable framework to analyze the impact of anticipated deployment equilibria on the willingness to undertake R&D. This framework is capable of underpinning the narrative of Popp (2010) of future free-riding weakening today's innovation incentives and able to identify novel effects of free-driving on R&D. The third contribution of this paper is to offer a toolkit for disentangling R&D incentives into a 'non-spillover technology' part and a pure 'technology interaction' part and thus enabling a deeper understanding of R&D.

The framework produces novel and partially surprising findings. The main finding is that strategic conflicts looming on the horizon directly impact on R&D incentives. It is this mechanism at work that weakens innovation incentives due to the anticipation of free-riding (Popp 2010). In contrast to the intuitive repercussions of free-riding, free-driving has more subtle implications for R&D: Not only can it happen that the free-driver's willingness to innovate is *super-optimal* so that technologies get available that ought to remain undeveloped; it is also possible that a country is willing to undertake 'counter-R&D' in order to deprive the free-driver of the technology.

The present study makes contributions to three strands of literature. The first strand of literature revolves around the implications of externalities on strategic interaction. This literature can be applied, as the present paper does, to technology deployment, but the most common field of application in environmental economics is mitigation of greenhouse gas emissions. This literature started from Barrett (1994) and mostly developed into the direction of International Environmental Agreements (for a survey see Wagner 2001). The present paper abstracts from any possible form of treaty, issue-linkage and governance. The topic of asymmetric countries, a crucial feature of the present study, has received attention in Barrett (2001) and McGinty (2006).

The second strand of literature this paper contributes to is environmental innovation. This environmental R&D can either be analyzed in a social planner framework (Goeschl and Perino 2007; Teubal 1978), with regard to an intergenerational dimension (Goeschl et al. 2013), or – most common – the international dimension; the latter literature is closely connected to the industrial organization literature on R&D games (Brander and Spencer 1983; D'Aspremont and Jacquemin 1988; Kamien et al. 1992; Cozzi 1999). In terms of the type of innovation, environmental R&D can either be cost-reducing (Bosetti et al. 2009; Hall and Helmers 2013), pollution-reducing (Perino and Requate 2012; Poyago-Theotoky 2007) or, as in the present paper, technology enabling (Barrett 2006; Hoel and De Zeeuw 2010).

Finally, this paper adds to the literature on the economics of CE with contributions focusing either on intergenerational heterogeneity (Goeschl et al. 2013) or heterogeneity across countries (Moreno-Cruz 2010; Manoussi and Xepapadeas 2014). Most of these papers however are interested the interplay of SRM with abatement from which the present study abstracts. It is closely connected to Weitzman (2012) who, for the fist time, analyzed 'free-driving' behavior in SRM deployment. The new angle the present paper adds to Weitzman (2012) is to put the repercussions on R&D incentives center stage.

The organization of the paper is as follows. Section 2 develops the simple game-theoretical setting in which two asymmetric countries play a non-cooperative R&D and technology deployment game. Section 3 characterizes deployment equilibria for the case of symmetric countries and analyzes the resulting R&D incentives. These findings serve as a useful benchmark for the general case of asymmetric countries that is the focus of section 4, the main section of this paper. Section 5 demonstrates how R&D incentives can be disentangled into two effects that enhance the understanding of the complex findings of section 4. Finally, section 6 concludes.

2 The Model

Two countries play a non-cooperative R&D and technology deployment game. In stage one, countries simultaneously make their individual R&D investments for an environmental technology. The R&D game is a threshold public good game with perfect spillovers: If the sum of R&D contributions exceeds a (commonly known) threshold, the technology is in the second stage available to both countries; otherwise no country can use it. Conditional on successful R&D, countries in the second stage simultaneously choose their deployment levels of the technology. This technology deployment game is a standard public good game with perfect spillovers. Each country bears the full (quadratic) costs of its technology deployment; in contrast to that and as usual in public good games, the other country cannot be excluded form the (inverse U-shaped) benefits. But, due to heterogeneity in the benefit-function, a central feature of the model, this has a flip-side. Non-excludability also implies that a country cannot protect against undesired high technology deployment levels.

The set-up combines different well-established building blocks. The general two-period setting with R&D and successive deployment stage, here transboundary technology use instead of Cournot competition, is borrowed from seminal work in the industrial organization literature (D'Aspremont and Jacquemin 1988; Kamien et al. 1992). The design of the threshold R&D process for a new environmental technology, but without treaty formation, is taken from Barrett (2006).

This public good structure with asymmetric benefits is quite general and capable of embracing a variety of technologies with transboundary effects and heterogeneous preferences. A particular fitting set of technologies however are the above mentioned climate engineering (CE) technologies.[2] What makes CE a good working example is that, first, these technologies are not developed yet so that the relevant R&D question is whether they are made available – in contrast to cost-reducing R&D (for instance Hall and Helmers 2013). Second, CE deployment fits well into the public good structure: Deployment costs are borne by the deploying country alone while the effect of reduced GHG levels and temperatures is inevitably global. Finally, and central in this paper, countries differ in their assessment of an 'optimal climate' (Rosenzweig and Parry 1994; Porter et al. 2014; Manoussi and Xepapadeas 2014; Heyen et al. 2015); as a consequence, the assumption of heterogeneous preferences for the *public gob*[3] is particular reasonable here.

The magnitude of deployment costs will crucially shape the outcome of the game. Direct air capture, a CE technology of the carbon dioxide removal (CDR) class, is a

[2]It should be noted that the simple framework deliberately abstracts from dissimilarities of CE technologies – how quickly they act and how big their unintended side-effects will be – that are crucial in other contexts.

[3]"A pure public gob is a pure public good (more of it is better) for some people under some circumstances and a pure public bad (more of it is worse) for some other people under some other circumstances." (Weitzman 2012)

suited example for an environmental technology with rather high deployment costs; in contrast, stratospheric aerosol injection, the best-known proposal for a solar radiation management (SRM) technology, is a good example of a low deployment cost technology. The rest of this paper will constantly refer to these two CE technologies.

The solution concept of the model is standard subgame perfectness (SPNE). The natural tool to solve the game is thus by backward induction. In the following we will explain the two stages of the model in detail, starting with the technology deployment stage (2.1) and then turning to the R&D stage (2.2).

2.1 Technology deployment stage

In case of successful R&D, countries in the second period choose their technology level $q_i \geq 0$ simultaneously and non-cooperatively. In terms of the working example CE, think of q_i as the reduction in global temperatures that is accomplished by either removing CO2 from the atmosphere (direct air capture) or putting sulfur into the stratosphere (stratospheric aerosol injection). The cost and benefit structure is of the quadratic-quadratic type (Barrett 1994; Finus and Rübbelke 2013; Diamantoudi and Sartzetakis 2006). The cost function, the same for both countries, is

$$C(q_i) = \frac{c}{2} q_i^2 \quad , \; i = 1, 2 \tag{1}$$

with $c > 0$. Costs only depend on the private contribution q_i. In contrast, benefits feature the usual public good structure with perfect spillovers, such that benefits are a function of the total technology level $Q = q_1 + q_2$,

$$B_i(Q) = b \left(a_i Q - \frac{1}{2} Q^2 \right) \tag{2}$$

with $b > 0$ and $a_i > 0$. The marginal benefits dB_i/dQ vanish at $Q = a_i$ which justifies calling a_i *country i's preferred technology level.* A central component of the model is to allow for $a_1 \neq a_2$ and hence heterogeneous technology preferences. Without loss of generality suppose $a_1 \leq a_2$. Instead of a_1 and a_2, it is often more meaningful to focus on the *mean technology optimum* $\bar{a} = (a_1 + a_2)/2$ and the *preference asymmetry* $a_\Delta = (a_2 - a_1)/2$. The condition $a_i > 0$ translates into the restriction $a_\Delta < \bar{a}$.

The countries choose their deployment levels non-cooperatively so that standard Nash equilibrium is the appropriate solution concept. Denote the Nash equilibrium of this game, to be determined in section 3 and 4, by (q_1^*, q_2^*). The social optimal configuration is (q_1^{**}, q_2^{**}). The payoff of country i of this technology deployment game is denoted $\pi_i(q_1, q_2) = B(q_1 + q_2) - C(q_i)$.[4]

[4]The reason to consider only the second stage payoffs, and not total payoffs that would also take into account R&D expenditures, is the following: The main focus of this paper is on R&D incentives, and the second period payoffs crucially shape the willingness to undertake costly R&D.

2.2 R&D stage

In the first period, countries simultaneously choose their R&D levels $r_i \geq 0$ (the possibility of non-negative R&D contributions will be discussed in section 4.3). As in Barrett (2006), R&D is a threshold public good game with perfect knowledge spillovers. If $r_1 + r_2 \geq \bar{R}$, the technology is available to both countries. If $r_1 + r_2 < \bar{R}$, neither country has access to the technology. The threshold $\bar{R}$ is common knowledge. Modeling R&D as a threshold process, and not as cost-reducing R&D (Hoel and De Zeeuw 2010; Hall and Helmers 2013; Bosetti et al. 2009) or emission-reducing (Poyago-Theotoky 2007; Perino and Requate 2012), is a realistic assumption in the context of technologies like direct air capture and stratospheric aerosol injection because CE technologies are at present merely theoretical concepts. The countries choose their individual R&D levels non-cooperatively so that, again, standard Nash equilibrium is the appropriate solution concept.

The analysis of R&D incentives crucially depends on how much the countries are willing to sacrifice to have the environmental technology available.

Definition 1. *Country i's* willingness to pay for the technology (WTP) *is* $R_i = \pi_i(q_1^*, q_2^*) - \pi_i(0,0)$*, where* (q_1^*, q_2^*) *is the Nash equilibrium of the technology deployment game in the second period. The total willingness to pay is* $R = R_1 + R_2$*. Important for comparison is* $R^{**} = (\pi_1(q_1^{**}, q_2^{**}) - \pi_1(0,0)) + (\pi_2(q_1^{**}, q_2^{**}) - \pi_2(0,0))$*, the maximal amount society would be willing to pay for making the technology available.*

Well known (for instance Barrett 2013) is that the Nash equilibria of a threshold public good game do not suffer from underprovision. In our context:

Lemma 1. *The Nash equilibria of the R&D game are all combinations* (r_1, r_2) *with* $r_1 + r_2 = \bar{R}$ *and* $r_i \leq \max(0, R_i)$*.*

The intuition is that every country is willing to fill up R&D investments up to the necessary threshold if the necessary contribution does not exceed its WTP R_i. The reason for the condition $r_i \leq \max(0, R_i)$ in Lemma 1 is that R_1 can be negative (see section 4.3). In this case, Nash equilibria involving zero R&D contributions by country 1, $r_1 = 0$, are still possible.

Lemma 1 demonstrates a further merit of the threshold R&D assumption. In general, due to the public good nature of knowledge, we would expect total R&D contributions in equilibrium to fall short of its optimal level (Stiglitz 1999; Popp 2010). At the same time, and being the central topic of this paper, we also expect strong implications of the anticipated strategic behavior in the deployment stage on R&D incentives. Disentangling both effects would be cumbersome. Due to the favorable equilibrium conditions of the threshold R&D game, it is clear that any effect on total R&D in this model can be fully ascribed to the anticipated strategic outcome of the technology deployment game.

In particular, this paper is not about equilibrium selection. Which of the, in general infinitely many, R&D equilibria is more or less likely to become reality is not the ambition of this inquiry. Rather, the focus is on (i) a comparison of R&D incentives across countries and (ii) the question whether total R&D incentives are strong enough for successful technology development. Both questions are fully determined by analyzing R_i and $R_1 + R_2$.

3 The symmetric benchmark

This section is dedicated to the case in which both countries have homogeneous preferences for the level of technology deployment, $a_1 = a_2$, that is $a_\Delta = 0$. In a first step, section 3.1 derives the deployment equilibrium, reestablishing well-known free-riding results from the literature. Building on this, section 3.2 characterizes the R&D equilibria, demonstrating that the anticipated free-riding deployment profile weakens R&D incentives. Thus, the first contribution of this section is to pin down the narrative of Popp (2010), giving us confidence that the model is adequately designed for making statements about R&D incentives due to anticipated strategic conflicts. The second contribution of this section is to provide benchmark equilibria for the asymmetric case $a_\Delta > 0$ that will be covered in section 4.

3.1 Technology deployment

This section derives the Nash equilibrium (q_1^*, q_2^*) and contrasts it with the social optimal configuration (q_1^{**}, q_2^{**}). The main step for deriving the Nash equilibrium is to determine the reaction function. Given the other country's contribution q_{-i}, country i optimally chooses

$$q_i(q_{-i}) = \max\left\{0\,,\ \frac{b}{b+c}(\bar{a} - q_{-i})\right\} \quad , \ i = 1,2\ . \tag{3}$$

Recall that in the symmetric case $\bar{a}$ is the global technology level that maximizes the countries' benefit function. If the other country's contribution does not exceed $\bar{a}$ (if it does, $q_i = 0$ is the best reply), the optimal response is to deploy some fraction of the remaining amount $\bar{a} - q_{-i}$ that would maximize country i's benefits, and this fraction approaches unity as the deployment costs converge to zero. As usual, the condition for the Nash equilibrium is $q_1(q_2^*) = q_1^*$ and $q_2(q_1^*) = q_2^*$.

Proposition 1 (Deployment equilibrium in the symmetric benchmark, $a_\Delta = 0$)**.** *Let $c > 0$. The deployment equilibrium in the symmetric benchmark with $a_1 = a_2 = \bar{a}$ is unique and has the following properties:*

(i) The contributions $q_i^* = \frac{\bar{a}b}{c+2b}$ are positive and monotonically decreasing in c with $\lim_{c\to\infty} q_i^* = 0$.

(ii) *The sum of contributions* $Q^* = q_1^* + q_2^*$ *is smaller than the socially optimal amount* $Q^{**} = \frac{4\bar{a}b}{c+4b}$*, and the fraction* Q^*/Q^{**} *decreases in* c*.*

(iii) *The equilibrium* (q_1^*, q_2^*) *is not Pareto optimal. The social optimal configuration* (q_1^{**}, q_2^{**}) *is Pareto optimal and a Pareto improvement to* (q_1^*, q_2^*)*.*

Proof. See Appendix B. □

These findings are hardly surprising and, for abatement instead of technology deployment choices, widely found in the literature (e.g. Barrett 1994). The key properties of the strategic conflict surrounding the technology use is that free-riding on other countries' contribution exists, giving rise to suboptimal low deployment levels. The social optimal configuration would make both countries better off, but is not stable against unilateral deviations. The higher the deployment costs, the more severe is, in relative terms, the gap between social optimal and actually undertaken technology deployment. In that sense, lower cost technologies are not only beneficial because they boost total net benefits, but also because small deployment costs alleviate the free-riding problem.

Regarding our working example climate engineering technologies, the findings of Proposition 1 imply that, if countries are symmetric and hence regard the same global temperature as optimal, any strategic problem that we can expect, irrespective of the cost structure of the CE technology, is free-riding and thus underprovision of the technology. We would expect this conflict to be stronger for cost-intensive technologies like direct air capture and significantly attenuated for for low cost technologies like stratospheric aerosol injection. We will see in section 4 that heterogeneity in technology preferences substantially changes this favorable picture of low deployment costs.

3.2 R&D

Based on the results from the previous section we can determine the countries' willingness to pay R_i and contrast the total WTP R with the amount R^{**} a global planner would be willing to sacrifice to make the technology available (cf. Definition 1).

Proposition 2 (WTP for R&D in the symmetric benchmark)**.** *The* WTP *in the symmetric benchmark has the following properties:*

(i) *The individual* WTP $R_1 = R_2$*, and hence also the total* WTP $R = R_1 + R_2$*, are positive and decreasing in* c *with* $\lim_{c\to\infty} R = 0$*.*

(ii) *The total* WTP R *falls short of the social optimal amount* R^{**}*.*

Proof. See Appendix B. □

Figure 1 gives a graphical illustration of Proposition 2. What is intuitive in light of the deployment equilibrium in Proposition 1 is the decrease of WTP in the costs

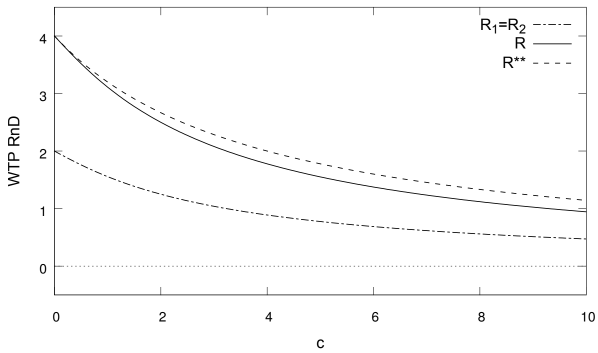


Figure 1: Comparison of the total WTP (solid line) and the social optimal amount (dashed line) as a function of the cost parameter c in the symmetric benchmark ($a_\Delta = 0$). The total WTP R is the sum $R_1 + R_2$ with $R_1 = R_2$. The parameter settings are $\bar{a} = 2$ and $b = 1$.

parameter c: The higher the costs, the lower will be the levels deployed, and thus the less willing are countries to spend money to get the technology.

The interesting feature is that the total WTP R is lower than what a social planner would be ready to pay for the technology's availability, R^{**}. Whether this has implications for the success of technology R&D depends on the threshold $\bar{R}$. If $\bar{R}$ is higher than R^{**}, then the discrepancy of R and R^{**} is inconsequential because R&D should not proceed anyway. Likewise, if $\bar{R}$ is lower than R, the R&D incentives for the (social desirable) technology are strong enough. The divergence of R and R^{**} however implies the existence of threshold values $\bar{R}$ for which development of the technology should proceed, but fails to do so due to weak incentives.

The reason why R falls short of R^{**}, to stress this again, is not the public good nature of R&D. The threshold assumption precludes underprovision due to R&D free-riding. Rather, the reason for $R < R^{**}$ is the anticipated strategic conflict in the technology deployment stage. Foreseeing free-riding in technology deployment, countries' incentives to develop the technology are substantially weakened. The first deliverable of section 3 is thus to underpin the narrative of Popp (2010) in a rigorous framework. The second deliverable is that the findings of Proposition 1 and Proposition 2 serve as a benchmark for the asymmetric case.

4 Asymmetric countries

This central section extends the analysis of R&D incentives to asymmetric countries. In Section 4.1 it will become clear that the cost parameter c plays a crucial role in the analysis. High cost parameter values give rise to deployment and R&D equilibria very similar to the benchmark case. These 'free-riding' equilibria will be covered in Section 4.2. In contrast, and topic of section 4.3, low deployment costs substantially change the strategic set-up, giving rise to 'free-driving' behavior (Weitzman 2012) with far-reaching implications for R&D incentives.

4.1 Two types of deployment equilibria

The first step is again to look at the deployment stage and specifically the reaction functions. The counterpart of 3, now for non-vanishing preference asymmetry $a_\Delta > 0$, is

$$q_1(q_2) = \max\left\{0\,,\,\frac{b}{b+c}(\bar{a} - a_\Delta - q_2)\right\} \quad , \quad q_2(q_1) = \max\left\{0\,,\,\frac{b}{b+c}(\bar{a} + a_\Delta - q_1)\right\}\,. \tag{4}$$

Recall that $\bar{a} - a_\Delta = a_1$ and $\bar{a} + a_\Delta = a_2$ are the preferred technology levels of country 1 and country 2, respectively.

Definition 2. *We call a Nash equilibrium (q_1^*, q_2^*) of the technology deployment game a* free-driver equilibrium *if the country with the lower preference for the technology does not deploy in equilibrium, $q_1^* = 0$. If both countries contribute positive amounts $q_i^* > 0$, we call (q_1^*, q_2^*) a* free-rider *equilibrium.*

In particular, the deployment equilibrium in the symmetric benchmark is of the free-rider type. Weitzman (2012) gives a similar definition for free-driving in an n country setting with a kinked benefit function and without deployment costs.[5]

As a direct consequence of the reaction functions in (4), the following Lemma presents the key role of the cost parameter c in determining the type of deployment equilibrium.

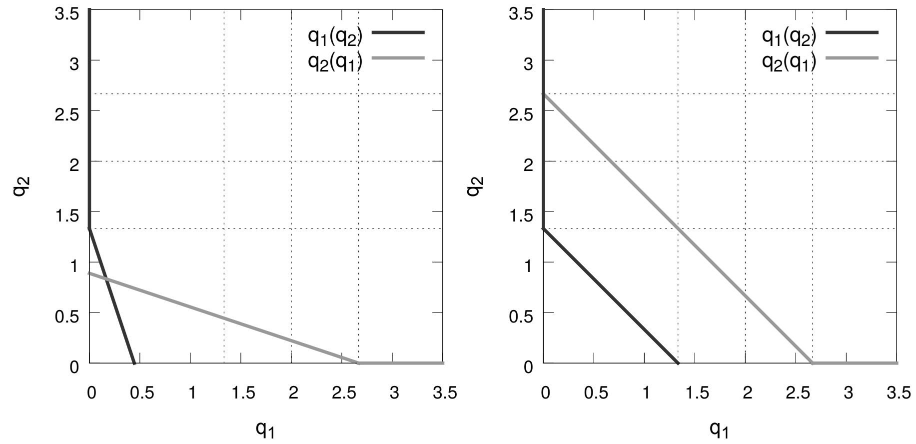
Lemma 2. *The technology deployment game has a unique Nash equilibrium if not $c = 0$ and $a_\Delta = 0$ at the same time. For $a_\Delta > 0$, the cost parameter $\bar{c} := 2ba_\Delta/(\bar{a} - a_\Delta)$ separates the two different strategic outcomes. The equilibrium is of the free-rider type for $c > \bar{c}$ and of the free-driver type for $c < \bar{c}$.*

Proof. See Appendix C. □

The separating cost parameter $\bar{c}$ increases in the preference asymmetry a_Δ and vanishes for $a_\Delta = 0$.

Figure 2 illustrates Lemma 2 by presenting the reaction functions from (4) for two different cost parameter. If $c > \bar{c}$, both countries in equilibrium deploy positive amounts of the technology; for $c \leq \bar{c}$, country 1 would actually prefer a negative deployment level to counteract the high deployment level of country 2; being restricted to non-negative levels country 1's best response is not to deploy. Table 1 summarizes the different outcomes.

[5] Two comments on the free-driver definition are in order. First, the extension of Definition 2 to an n country setting would involve the choice whether to speak of a free-driver equilibrium when *at least one* country or *all but one* countries do not contribute in equilibrium. Second, the free-driver definition in its current form rests on the impossibility of negative deployment. In the context of Solar Radiation Management, the possibility of counter-geoengineering (Barrett et al. 2014) has been raised. In this case negative contributions would be possible. The analysis of this possibility is left for future research.



(a) $c > \bar{c}$. Free-rider equilibrium. (b) $c < \bar{c}$. Free-driver equilibrium.

Figure 2: The reaction functions in the asymmetric case ($a_\Delta > 0$) for two different cost parameter c. The parameter settings are $b = 1$, $\bar{a} = 2$, $a_\Delta = 2/3$ so that $\bar{c} = 1$. With $c = 2$ (left panel), the equilibrium $(1/6, 5/6)$ is of the free-rider type. With $c = 0$ (right panel), the equilibrium $(0, 8/3)$ is a free-driver equilibrium.

Table 1: Nash outcomes of the deployment game in period 2 for asymmetric countries ($a_\Delta > 0$) specifying technology deployment of country 1 (q_1^*), country 2 (q_2^*), total deployment (Q^*) and social optimal total deployment (Q^{**}).

Type	Free-rider	Free-driver
c	$c > \bar{c}$	$c \leq \bar{c}$
q_1^*	$\frac{b}{c+2b}\bar{a} - \frac{b}{c}a_\Delta$	0
q_2^*	$\frac{b}{c+2b}\bar{a} + \frac{b}{c}a_\Delta$	$\frac{b}{b+c}(\bar{a} + a_\Delta)$
Q^*	$\frac{2b}{c+2b}\bar{a}$	$\frac{b}{b+c}(\bar{a} + a_\Delta)$
Q^{**}	$\frac{4b}{c+4b}\bar{a}$	$\frac{4b}{c+4b}\bar{a}$

In terms of the working example CE technologies the above findings are highly relevant. In light of the sharp difference in deployment costs, we can expect CDR technologies like direct air capture to be prone to very different strategic incentives as the low cost SRM technology stratospheric aerosol injection. Table 1 also demonstrates that the framework outlined in section 2 is capable of reproducing the free-driver behavior, established in Weitzman (2012), in a standard smooth public good setting.

The following two subsections will elaborate and compare the deployment and resulting R&D characteristics for free-rider (4.2) and free-driver (4.3) equilibria.

4.2 Free-rider equilibria, $c > \bar{c}$

The first part of this section specifies the free-rider deployment equilibrium (the second column in Table 1) and is thus the counterpart of Proposition 1 for asymmetric countries $a_\Delta > 0$.

Proposition 3 (Free-rider deployment equilibrium with asymmetric countries)**.** *The unique free-rider equilibrium has the following properties:*

(i) The contributions q_1^* and q_2^* are positive with limits $\lim_{c\to\bar{c}} q_1^* = 0$, $\lim_{c\to\infty} q_1^* = 0$, $\lim_{c\to\bar{c}} q_2^* = a_1$, and $\lim_{c\to\infty} q_2^* = 0$. While q_2^* is monotonically decreasing in c, q_1^* has a maximum in $(\bar{c},\infty)$. While q_1^* increases in the asymmetry parameter a_Δ, q_2^* decreases in a_Δ.

(ii) The sum of contributions Q^* is smaller than the social optimal amount Q^{**}.

(iii) The equilibrium (q_1^*, q_2^*) is not Pareto optimal. Local Pareto improvements consist of $(q_1^* + \delta\alpha, q_2^* + \delta\beta)$, where $\alpha, \beta, \delta > 0$. While the social optimal configuration (q_1^{**}, q_2^{**}) is an improvement to (q_1^*, q_2^*) for country 2, this is in general not true for country 1: For $a_\Delta > 0$, (q_1^{**}, q_2^{**}) is not an improvement to (q_1^*, q_2^*) at least in a neighborhood of $\bar{c}$.

Proof. See Appendix C.1. □

Country 2's technology deployment is very similar to the symmetric benchmark. That increases in a_Δ drive up q_2^* is simply because this implies a higher technology preference, $a_2 = \bar{a} + a_\Delta$.

Quite different is the non-monotonic pattern of country 1. The reason that a decrease in costs close to the separating cost level $\bar{c}$ *reduces* the contribution of country 1 is free-riding of the extremest form: At $\bar{c}$, country 2 contributes $\bar{a} - a_\Delta = a_1$, exactly country 1's preferred level; as a consequence, country 1 does not need to make own costly contributions. Not surprising, see part (ii), is that this leads overall to an underprovision of the technology.

What is interesting is the look at the Pareto improvements in (iii): Small increases in both contributions would make both countries better off, a simple consequence of the underprovision due to free-riding; the social optimal configuration however is often, and definitely when costs are low, *not* an improvement for country 1. The reason is again that, close to $\bar{c}$, country 1 is free-riding in an extreme way on country 2's contribution such that there is no room for better outcomes. As the appendix C.1 shows, a similar situation can also occur for high cost parameter values if the asymmetry a_Δ is high enough.

What are the implications for R&D incentives? The following Proposition, the counterpart of Proposition 2 for asymmetric countries, gives the answer.

Proposition 4 (WTP for R&D with asymmetric countries. Free-rider)**.** *The* WTP *for free-rider technologies fulfills:*

(i) The individual WTP *for R&D R_1, R_2, and hence also the total* WTP *R are positive and decreasing in c. We have $R_1 < R_2$ and $\lim_{c\to\bar{c}} \frac{dR_1}{dc} = 0$.*

(ii) Increases in the preference asymmetry a_Δ decrease R_1 and increase R_2.

(iii) The total WTP *R falls short of the social optimal amount R^{**}. The difference between them increases in the preference asymmetry a_Δ.*

Proof. See Appendix C.1. □

Proposition 4 features strong similarities with the corresponding results in the symmetric benchmark, see Proposition 2. Most importantly, the total WTP falls short of the social optimum (part (iii)), implying the existence of constellations in which a beneficial technology is not developed for strategic reasons. The interpretation for $R < R^{**}$ is, again, that the prospect of an underprovision of the technology directly reduces R&D incentives.

Besides the intuitive finding that an increase in a_Δ drives up country 2's WTP and decreases country 1's WTP (part (ii)), a fact worth mentioning is that country 1's WTP R_1 is strictly positive and flat at $\bar{c}$. Both are direct consequences of the fact that, at $c = \bar{c}$, the technological deployment profile is perfect from the country 1's viewpoint: No private costs, $q_1^* = 0$, but the optimal total technology deployment $Q = a_1$ provided by country 2.

4.3 Free-driver equilibria, $c \leq \bar{c}$

This section demonstrates that free-driver equilibria are very different from their free-rider counterparts, both with regard to the deployment patterns, but also the resulting R&D incentives.

The first part of this section specifies characteristics of the free-driver deployment equilibrium (cf. third column in Table 1).

Proposition 5 (Free-driver deployment equilibrium with asymmetric countries)**.** *The unique free-driver equilibrium has the following properties:*

(i) Country 1 does not contribute, $q_1^* = 0$, while q_2^* is positive, monotonically decreasing in c, and taking on the values $\bar{a} + a_\Delta = a_2$ and $\bar{a} - a_\Delta = a_1$ at the boundaries $c = 0$ and $c = \bar{c}$, respectively.

(ii) The sum of contributions Q^* is higher than the social optimal amount Q^{**} if and only if $c < 4ba_\Delta/(3\bar{a} - a_\Delta) < \bar{c}$.

(iii) The equilibrium (q_1^*, q_2^*) is not Pareto optimal except for $c = \bar{c}$. For $c < \bar{c}$, local Pareto improvements consist of $(q_1^* + \delta\alpha, q_2^* + \delta\beta)$, where $\alpha, \delta > 0$ and $\beta < -\alpha$. At $c = 0$, the social optimal configuration is an improvement for country 1 but not for country 2; at $c = \bar{c}$, it is the other way round. There always exists an inner range where (q_1^{**}, q_2^{**}) is a a Pareto improvement to (q_1^*, q_2^*).

Proof. See Appendix C.2. □

By definition, in the entire free-driver region $c \leq \bar{c}$ country 1 does not deploy the technology, $q_1^* = 0$. The total technology deployment level thus equals country 2's contribution q_2^*. It is not surprising to observe, and according to earlier findings, that this level goes up as costs c decrease. It also makes sense that $q_2^* = a_2$ at $c = 0$: If there are no costs of deployment, country 2 chooses its preferred technology level.

Part (ii) shows that the total deployment level Q^* can exceed the social optimal amount. This finding, unknown in the standard public good literature, is another indication of the highly unusual and fascinating nature of free-driver equilibria. One has to keep in mind however that the comparison of the total values Q^* and Q^{**} is only one dimension of gauging the gap between equilibrium outcomes and social optimal reference points. Due to non-linear deployment costs, the distribution of deployment levels across countries always matters.

Part (iii) describes the possible Pareto improvements. The constellation at $c = \bar{c}$ must be Pareto optimal as country 1, by definition, cannot do any better: The overall deployment level is at the optimal level $Q^* = a_1$ without own costly contributions. Once $c < \bar{c}$, however, the Nash outcome is not Pareto optimal, and local Pareto improvements consist of country 1 contributing positive amounts, which already suffices to make country 2 better off, while country 2, in order to make it an improvement for country 1, would overproportionally decrease its deployment. Interestingly, the social optimal configuration (q_1^{**}, q_2^{**}) can only be a Pareto improvement if costs are in a certain interior range; when costs get extreme, either to $c = 0$ or $c = \bar{c}$, the Nash outcome favors one country too strongly as if the uniform social optimal deployment profile is attractive for both.

Connecting these findings with the respective part for free-rider equilibria of Proposition 3 gives a comprehensive picture. For country 2, the social optimal configuration is always an improvement except for very small c. For $c > \bar{c}$ this is because it overcomes the free-rider problem, for levels below but close to $\bar{c}$, it saves deployment costs because (q_1^{**}, q_2^{**}) involves positive contributions by country 1. For cost parameter values close to 0, however, the benefits of being able to afford technology levels close to optimal outweigh the costs of being the only contributor. For country 1, things are slightly more complex. Starting at the low cost end with costs close to zero, it is clear that the social optimal configuration would be an improvement; the reason is that here the free-driver behavior of country 2 is extreme, implying a significant divergence between preferred, $Q = a_1$, and actual deployment. Also clear is that for values close to, and on both sides, of the separating cost parameter $\bar{c}$, country 1 is better off under the Nash outcome because the overall deployment level is close to a_1 and own provision costs are low, if not zero. Ambiguous however is the case of higher cost parameters: As explained in 4.2, (q_1^{**}, q_2^{**}) is a Pareto improvement only if the asymmetry a_Δ is not too high.

The main result in this section is concerned with the implications of free-driving behavior for R&D incentives.

Proposition 6 (WTP for R&D with asymmetric countries. Free-driver)**.** *The* WTP *for free-rider technologies fulfills:*

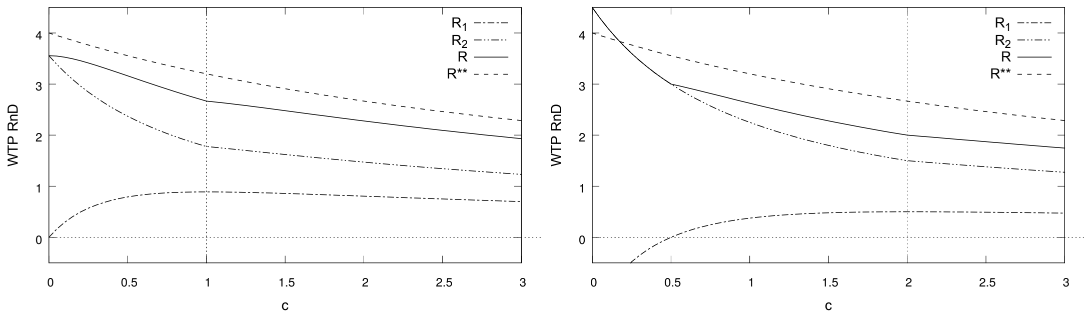
(i) Country 2*'s* WTP R_2 *is positive and decreasing in the cost parameter* c*. Country* 1*'s* WTP R_1 *is increasing in* c *with* $\lim_{c\to\bar{c}} \frac{dR_1}{dc} = 0$*;* R_1 *gets negative for small* c *if* $a_\Delta > \bar{a}/3$*.*

(ii) Increases in the asymmetry a_Δ *decrease* R_1 *and increase* R_2*.*

(iii) The total WTP $R = \max\{R_1, 0\} + \max\{R_2, 0\}$ *is positive and decreasing in* c*. If* $a_\Delta > (\sqrt{2}-1)\bar{a}$*, the total* WTP R *at low values of* c *is higher than the the social optimum* R^{**}*. If this is the case, country* 1*'s* WTP R_1 *is necessarily negative.*

Proof. See Appendix C.2. □

Figure 3 gives a graphical illustration of Proposition 6.



(a) $a_\Delta = 2/3$ (b) $a_\Delta = 1$

Figure 3: The WTP of country 1 (dot-dashed line), country 2 (dot-dot-dashed line), total WTP (solid line) and social optimal WTP (dashed line) as a function of the cost parameter c for different asymmetry levels a_Δ. The vertical dotted line is at $\bar{c}$, separating free-driver equilibria to the left from free-rider equilibria to the right. Note that $R = \max(0, R_1) + \max(0, R_2)$. The parameter settings are $\bar{a} = 2$ and $b = 1$ as before.

It is not surprising that country 2's WTPR_2 is positive and decreasing in costs. Also, that a_Δ drives R_1 up while it decreases R_2 has been found before and follows the same intuition here.

Quite unusual however is the behavior of country 1's WTP R_1. In contrast to the free-riding case, see Proposition 4, and sharply different from country 2, here R_1 *decreases as costs go down.* This new finding however makes sense in light of the perfect constellation country 1 has at $\bar{c}$. Relative to that, lower c values drive country 2 into ever higher deployment levels and thus away from country 1's preferred overall technology level a_1.

Related and also interesting to note is that in the free-driving equilibrium, country 1, although not deploying the technology, is in general ready to sacrifice means to get the technology. The reason for this positive WTP is that, despite excessive technology deployment by country 2, country 1 is often better off with the technology deployment pattern of country 2 than without any deployment.

This, however, can change for low cost levels if the technology preference across countries is high. Then, country 2's deployment strongly exceeds country 1's optimal deployment level so that country 1 would be better off without the technology. The negative WTP R_1 in this case can be interpreted as the willingness to undertake counter-R&D. This demonstrates that low cost technologies like stratospheric aerosol injection that give rise to free-driver behavior not only suffer from strategic conflicts at the deployment stage in some future, but that the anticipation of those future conflicts my directly give rise to conflicts in the present.

Another peculiarity with free-driver equilibria is that R&D incentives can be *too strong.* The total WTP R, when defined based on non-negative R&D contributions $r_i \geq 0$ (for the other case see Remark 1), can exceed R^{**} when the cost parameter c is low and the preference asymmetry between countries a_Δ is high. In Figure 3, the asymmetry a_Δ in the right figure, $a_\Delta = 1$, is large enough to feature $R > R^{**}$ for small c, while this is not the case in the left figure. If $R > R^{**}$, the consequence is that technology R&D takes place (undertaken by country 2 alone) that should, from a societal point, not proceed.

Remark 1. *The analysis in Proposition 6 (iii) is correct if R&D contributions are non-negative, $r_i \geq 0$. Alternatively, one may consider the possibility of counter-R&D. For simplicity assume that one unit of counter-R&D just cancels one unit of R&D. In terms of possible equilibria, note that with $R_1 < 0$ and $R_2 > 0$ there is a unique R&D equilibrium, namely either either $(0, \bar{R})$ or $(0,0)$. The relevant total* WTP *that decides about successful R&D is $R = R_1 + R_2$. It is easy to show (see C.2) that this measure is positive and always falls short of the social optimum; in other words the R&D incentives with counter-R&D are never too strong as in the case discussed before. But also the counter-R&D scenario gives rise to a peculiar outcome: For significant asymmetry between the countries, $a_\Delta > \bar{a}/3$, the total* WTP *R is non-monotonical in c.*

5 Disentangling Effects

The purpose of this final substantive section is to get a deeper understanding of the R&D incentives. For this purpose, divide the R&D process for a transboundary technology in a thought experiment into two steps. We write country i's WTP $R_i = \pi_i(q_1^*, q_2^*) - \pi_i(0,0)$ accordingly as

$$R_1 = \underbrace{\pi_1(q_1^*, q_2^*) - \pi_1(q_1^{\text{Priv}}, 0)}_{=:R_1^{\text{Pub}}} + \underbrace{\pi_1(q_1^{\text{Priv}}, 0) - \pi_1(0,0)}_{=:R_1^{\text{Priv}}} \tag{5}$$

and similar for R_2. The thought experiment consists of splitting R&D artificially into two processes. First, the development of a fully private technology so that the other country can be excluded from any deployment effects. The optimal technology deployment for such a private technology just balances private benefits and private costs with

$$q_i^{\text{Priv}} = \frac{b}{b+c} a_i \quad , \ i = 1, 2 \ . \tag{6}$$

The WTP for such a no-spillover technology is then R_i^{Priv}.

The second process is then to transform the private technology to a transboundary technology with perfect spillovers that can be used by both countries. The willingness to open up a private technology in that way, which will result in the Nash contribution pattern (q_1^*, q_2^*), is R_i^{Pub}.

The first Proposition analyzes the incentives to develop a no-spillover technology.

Proposition 7 (WTP for private technology)**.** *The incentives to develop a fully private technology are* $R_1^{\text{Priv}} = \frac{b^2}{2(b+c)}(\bar{a} - a_\Delta)^2$ *and* $R_2^{\text{Priv}} = \frac{b^2}{2(b+c)}(\bar{a} + a_\Delta)^2$*. Thus:*

(i) For both countries, the WTP *to develop a private technology is positive and decreasing in* c*.*

(ii) Increases in the asymmetry a_Δ *increase* R_1^{Priv} *and decreases* R_2^{Priv}*, and always* $R_1^{\text{Priv}} < R_2^{\text{Priv}}$*.*

Proof. Obvious. □

The findings are hardly surprising. In the absence of any interaction effects with other countries' deployment, the lower the costs and the higher the preferred technology level, the higher the WTP to develop such a private technology. Also, due to any interaction effects, the findings of Proposition 7 are valid for any cost level c.

The simple and obvious nature of the WTP for a private technology cannot explain the rich findings and differences between countries and cost ranges, see Propositions 4 and 6. Before we tackle the asymmetric case, a few words on the symmetric case are helpful to fully appreciate what is to come. In contrast to R_i^{Priv}, which is always monotonic in c, the effect R_i^{Pub} is not monotonic, even in the symmetric case. For $a_\Delta = 0$ we get

$$R_i^{\text{Pub}} = \frac{\bar{a}^2 b^2 c(2c + 3b)}{2(c^3 + 5bc^2 + 8b^2c + 4b^3)} \tag{7}$$

with $\lim_{c \to 0} R_i^{\text{Pub}} = \lim_{c \to \infty} R_i^{\text{Pub}} = 0$. The reason that R_i^{Pub} is not monotonically increasing as costs go down is that opening up a low cost technology has only limited benefits: If costs are low, private deployment levels are already close to optimal; hence there is not much room for benefiting from other's deployment. It is helpful to keep this non-monotonicity in the symmetric case in mind when we now turn to the willingness to open up a technology for interaction for the general, asymmetric case.

Proposition 8 (WTP for opening up a private technology)**.** *The incentives to open up an existing private technology to a fully spillover technology are as follows:*

(i) Country 1. In the free-rider region, R_1^{Pub} is positive and non-monotonic in c. In the free-driver region it increases with c and is negative for $c < \frac{4ba_\Delta^2}{\bar{a}^2+2a_\Delta\bar{a}-3a_\Delta^2}$.

(ii) Country 2. In the free-rider region, R_2^{Pub} is positive and non-monotonic in c with $lim_{c\to\bar{c}} R_2^{\mathrm{Pub}} = 0$. In the free-driver region, $R_2^{\mathrm{Pub}} = 0$.

(iii) The WTP *to open the private technology is always higher for country 1, $R_1^{\mathrm{Pub}} > R_2^{\mathrm{Pub}}$, except for small c when $R_1^{\mathrm{Pub}} < 0$, see (i).*

Proof. See Appendix D. □

Figure 4 gives a graphical representation of both effects.

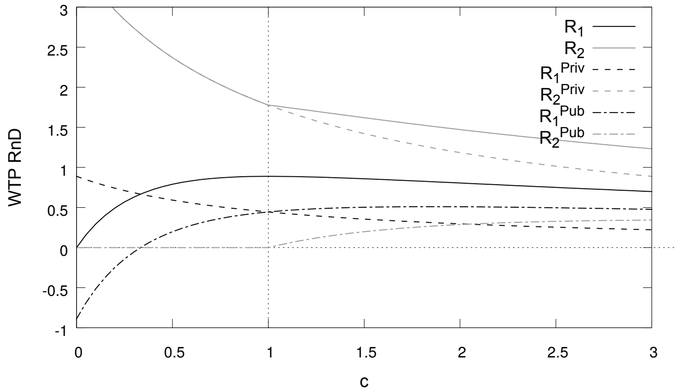


Figure 4: Disentangling two drivers of the WTP for country 1 (black lines) and country 2 (gray lines). The vertical dotted line is at $\bar{c}$, separating free-driver equilibria to the left from free-rider equilibria to the right. The functions depicted are the total WTP R_i (solid lines), the WTP to develop a no-spillover technology R_i^{Priv} (dashed lines), and the WTP to transform a private technology to a perfect spillover technology R_i^{Pub} (dot-dashed lines). The parameter settings are $\bar{a} = 2$, $a_\Delta = 2/3$ and $b = 1$.

Not surprising is the non-montonic pattern of the willingness to open up a technology that we find for both countries in the free-rider region. This is essentially the same effect that we already isolated in the symmetric case, see (7).

An interesting feature is that $R_2^{\mathrm{Pub}} = 0$ over the entire free-driver region. The reason for that is that country 2, with country 1 in the Nash equilibrium not deploying at all, is already using the technology privately.

In terms of interpretation, (iii) shows that over the entire free-rider region $R_1^{\mathrm{Pub}} > R_2^{\mathrm{Pub}}$, and the gap increases with the preference asymmetry a_Δ (see appendix D). This surprising finding can be interpreted as that country 1 is benefiting more from the pubic good characteristic of the technology as country 2 does; country 1 is free-riding heavily on country 2's contribution.

Disentangling the different drivers also helps us to develop a more nuanced interpretation of the free-driving zone. At first sight, country 1 could be interpreted as a loser from the technology interaction because it is in the hand of the 'free-driver' and would actually prefer negative own contributions. However, if c is not too small, $R_1^{\text{Pub}} > 0$ (and in particular larger than R_2^{Pub}) because country 1 still benefits from country 2's technology deployment. So the story is more subtle. Even though we are in a 'free-driver' equilibrium in which country 1 does not deploy, country 1 is still free-riding on country 2's technology deployment. Only if the cost parameter c gets very small, country 1 is actually a loser from the spillover character of the technology. The reason that Weitzman (2012) only covers the latter case is that in his model $c = 0$. Due to the general public good structure with continuous cost parameter values c, the present model is capable of generating more complex and subtle effects.

6 Conclusions

Technologies constitute a central component in the portfolio of measures against environmental problems. An important example is climate engineering (CE), a set of environmental technologies that has recently received increasing attention. Novel environmental technologies, and CE in particular, however require substantial R&D expenditures to become available. In a decentralized world largely shaped by domestic interests, the incentives for innovation are crucial for successful R&D and thus deserve a thorough analysis.

The present paper focused on a specific problem surrounding R&D incentives. The anticipation of strategic conflicts in their future use, for instance free-riding, can be expected to have repercussions on the willingness to develop these technologies. In fact, the rigorous, yet parsimonious, framework developed in this paper has proved capable of underpinning a narrative of (Popp 2010) holding that the anticipation of free-riding weakens R&D incentives.

Starting from there, the paper turned to a different strategic effect. In sharp contrast to free-riding, *free-driving* occurs when the country with the highest preference for deployment dominates the outcome (Weitzman 2012). This novel strategic effect materializes – and the present paper has provided further support for this in a standard smooth public good setting – if preferences for the technology deployment are heterogeneous and deployment costs for the technology are low. Stratospheric aerosol injection, and in general SRM, are expected to exhibit these characteristics and may thus be prone to free-driving outcomes.

This paper demonstrated that the outlook of free-driving has novel and rich impacts on innovation incentives. Most notably, the anticipated future deployment conflict may 'pull forward' to the R&D stage: The free-driver, keen to get the technology, may push technology development even if this is against the global best. Accordingly, the other country, foreseeing the free-driver's extreme deployment level, may be willing to costly counteract the free-driver's innovation efforts, giving rise to an R&D conflict.

The present paper provides only the first step into the formal and rigorous analysis of how deployment conflicts impact on R&D incentives. There are a couple of valuable extensions that future research should envisage. The first possible extension is to generalize the two-country setting to n countries. It is far from clear how to generalize the definitions of 'free-riding' and 'free-driving' to the general case as it involves non-trivial and meaningful choices. This is particularly true in light of the subtleties surrounding free-driving that already emerged in the two country setting: The 'dominated' country, at first sight an obvious loser from the technology interaction, can for many parameter settings be expected to substantially benefit from the technological interaction. The question about winners and losers from the technology in a general setting is a fascinating research question.

A second possible line of research is to modify the framework to cost-reducing R&D. This type of innovation is pervasive in environmental economics (Bosetti et al. 2009; Hall and Helmers 2013), emphasizing the welfare-improving role of low cost technologies. The findings in the present paper suggest that free-driver technologies may be characterized by opposition against cost reducing R&D. The country with the lower preference for the technology, concerned of being worse-off under the free-driver's increasing deployment, may be willing to prevent cost-reducing innovation.

Another important extension of the existing framework is to consider governance structures and treaty formation. The present paper deliberately refrained from discussing these issues. Accordingly, the focus was on non-cooperative behavior and pure Nash outcomes as the relevant solution concept. The motivation to do so was mainly to develop a framework for making positive statements about the R&D incentives – in particular its deviations from global optimal – that we may expect without any governance regime in place. The present framework is thus the ideal starting point to see which, if any, governance structures or treaty options can help to overcome the strategic incentive problems.

Finally, future research should also incorporate an angle from the CE literature by focusing on the interplay of technology and abatement. In particular, Moreno-Cruz (2010) and Goeschl et al. (2013) demonstrated, in very different settings, that abatement can serve as a tool to attenuate strategic conflicts about the use of a technology by shifting their deployment incentives. A promising research question is whether abatement can also play this role in the context of innovation incentives for free-driver technologies.

References

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

Barrett, S. (2001). International cooperation for sale, *European Economic Review* **45**(10): 1835–1850.

Barrett, S. (2006). Climate treaties and "breakthrough" technologies, *The American Economic Review* **96**(2): 22–25.

Barrett, S. (2008). The incredible economics of geoengineering, *Environmental and Resource Economics* **39**(1): 45–54.

Barrett, S. (2013). Climate treaties and approaching catastrophes, *Journal of Environmental Economics and Management* **66**(2): 235–250.

Barrett, S., Lenton, T. M., Millner, A., Tavoni, A., Carpenter, S., Anderies, J. M., Chapin III, F. S., Crpin, A.-S., Daily, G., Ehrlich, P., Folke, C., Galaz, V., Hughes, T., Kautsky, N., Lambin, E. F., Naylor, R., Nyborg, K., Polasky, S., Scheffer, M., Wilen, J., Xepapadeas, A. and de Zeeuw, A. (2014). Climate engineering reconsidered, *Nature Climate Change* **4**(7): 527–529.

Bosetti, V., Carraro, C., Duval, R., Sgobbi, A. and Tavoni, M. (2009). The role of R&D and technology diffusion in climate change mitigation: new perspectives using the WITCH model, *FEEM Working Paper 14*.

Brander, J. A. and Spencer, B. J. (1983). Strategic commitment with R&D: the symmetric case, *The Bell Journal of Economics* **14**(1): 225–235.

Caldeira, K., Bala, G. and Cao, L. (2013). The science of geoengineering, *Annual Review of Earth and Planetary Sciences* **41**(1): 231–256.

Chen, C. and Tavoni, M. (2013). Direct air capture of CO2 and climate stabilization: A model based assessment, *Climatic Change* **118**(1): 59–72.

Cozzi, G. (1999). R&D cooperation and growth, *Journal of Economic Theory* **86**(1): 17–49.

D'Aspremont, C. and Jacquemin, A. (1988). Cooperative and noncooperative R & D in duopoly with spillovers, *The American Economic Review* **78**(5): 1133–1137.

Diamantoudi, E. and Sartzetakis, E. S. (2006). Stable international environmental agreements: An analytical approach, *Journal of Public Economic Theory* **8**(2): 247–263.

Finus, M. and Rübbelke, D. T. G. (2013). Public good provision and ancillary benefits: The case of climate agreements, *Environmental and Resource Economics* **56**(2): 211–226.

Finus, M., Kotsogiannis, C. and McCorriston, S. (2013). The international dimension of climate change policy, *Environmental and Resource Economics* **56**(2): 151–160.

Goeschl, T. and Perino, G. (2007). Innovation without magic bullets: Stock pollution and R&D sequences, *Journal of Environmental Economics and Management* **54**(2): 146–161.

Goeschl, T., Heyen, D. and Moreno-Cruz, J. (2013). The intergenerational transfer of solar radiation management capabilities and atmospheric carbon stocks, *Environmental and Resource Economics* **56**(1): 85–104.

Golombek, R. and Hoel, M. (2011). International cooperation on climate-friendly technologies, *Environmental and Resource Economics* **49**(4): 473–490.

Hall, B. H. and Helmers, C. (2013). Innovation and diffusion of clean/green technology: Can patent commons help?, *Journal of Environmental Economics and Management* **66**(1): 33–51.

Harstad, B. (2012). Climate contracts: A game of emissions, investments, negotiations, and renegotiations, *The Review of Economic Studies* **79**(4): 1527–1557.

Heyen, D., Wiertz, T. and Irvine, P. (2015). Regional disparities in solar radiation management impacts: Limitations to simple assessments and the role of diverging preferences, *IASS Working Paper*.

Hoel, M. and De Zeeuw, A. (2010). Can a focus on breakthrough technologies improve the performance of international environmental agreements?, *Environmental and Resource Economics* **47**(3): 395–406.

Hoffert, M. I., Caldeira, K., Benford, G., Criswell, D. R., Green, C., Herzog, H., Jain, A. K., Kheshgi, H. S., Lackner, K. S., Lewis, J. S., Lightfoot, H. D., Manheimer, W., Mankins, J. C., Mauel, M. E., Perkins, L. J., Schlesinger, M. E., Volk, T. and Wigley, T. M. L. (2002). Advanced technology paths to global climate stability: Energy for a greenhouse planet, *Science* **298**(5595): 981–987.

Kamien, M. I., Muller, E. and Zang, I. (1992). Research joint ventures and R&D cartels, *The American Economic Review* **82**(5): 1293–1306.

Keith, D. W., Ha-Duong, M. and Stolaroff, J. K. (2006). Climate strategy with CO2 capture from the air, *Climatic Change* **74**(1-3): 17–45.

Manoussi, V. and Xepapadeas, A. (2014). Cooperation and competition in climate change policies: Mitigation and climate engineering when countries are asymmetric, *SSRN Scholarly Paper ID 2535720*, Social Science Research Network, Rochester, NY.

McGinty, M. (2006). International environmental agreements among asymmetric nations, *Oxford Economic Papers*.

Moreno-Cruz, J. (2010). Mitigation and the geoengineering threat, *Unpublished Manuscript*. Available at: http://works.bepress.com/morenocruz/3.

Perino, G. and Requate, T. (2012). Does more stringent environmental regulation induce or reduce technology adoption? When the rate of technology adoption is inverted U-shaped, *Journal of Environmental Economics and Management* **64**(3): 456–467.

Popp, D. (2006). R&D subsidies and climate policy: is there a free lunch?, *Climatic Change* **77**(3-4): 311–341.

Popp, D. (2010). Innovation and climate policy, *NBER Working Paper 15673*, National Bureau of Economic Research.

Porter, J. R., Xie, L., Challinor, A. J., Cochrane, K., Howden, S. M., Iqbal, M. M., Lobell, D. B. and Travasso, M. I. (2014). Food security and food production systems, *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel of Climate Change* pp. 485–533.

Poyago-Theotoky, J. (2007). The organization of R&D and environmental policy, *Journal of Economic Behavior & Organization* **62**(1): 63–75.

Rosenzweig, C. and Parry, M. L. (1994). Potential impact of climate change on world food supply, *Nature* **367**(6459): 133–138.

Shepherd, J. (2009). *Geoengineering the climate: science, governance and uncertainty*, Royal Society.

Stiglitz, J. E. (1999). Knowledge as a global public good, *Global public goods* **1**(9): 308–326.

Teubal, M. (1978). Threshold R&D levels in sectors of advanced technology, *European Economic Review* **7**(4): 395–402.

Victor, D. (2008). On the regulation of geoengineering, *Oxford Review of Economic Policy* **24**(2): 322–336.

Wagner, U. J. (2001). The design of stable international environmental agreements: Economic theory and political economy, *Journal of Economic Surveys* **15**(3): 377–411.

Weitzman, M. L. (2012). A voting architecture for the governance of free-driver externalities, with application to geoengineering, *NBER Working Paper 18622*, National Bureau of Economic Research.

Appendix A Toolkit for proving inequalities

There is often the need to prove inequalities of the type $P < 0$ or $P > 0$ where P is a polynomial in c. In the former case we want to find upper bounds of negative terms to get sufficient conditions for $P < 0$. For free-rider technologies, $c > \bar{c}$, this is possible by replacing c by $\bar{c}$. If we have to prove $P > 0$ for free-driver technologies, $c \leq \bar{c}$, replacing c in negative terms by $\bar{c}$ gives the needed lower bound.

Remark 2. *For the proof of inequalities it makes sense to deviate from expressing a_1 and a_2 in terms of $\bar{a}$ and a_Δ; rather, a_1 and δ where $a_2 = a_1 + \delta$ is the best choice. With that, $\bar{c} = b\delta/a_1$*

Example If we have to prove $-3a_1bc^2 + 2\delta b^2 c < 0$ for $c < \bar{c}$, a sufficient condition is $-3a_1bc\bar{c} + 2\delta b^2 c < 0$, which is true because $-3a_1bc\bar{c} = -3\delta b^2 c$. In more complex situations, we will indicate the terms that are combined in that way.

Appendix B On section 3

Proof of Proposition 1 The social optimal deployment is $q_1^{**} = q_2^{**} = \frac{2\bar{a}b}{c+4b}$, so that $Q^{**} = \frac{4\bar{a}b}{c+4b}$.

(i) This part of the Proposition is obvious.

(ii) The only fact that needs clarification is the behavior of the fraction $Q^*/Q^{**} = (c+4b)/(2c+4b)$: It decreases monotonically from 1 at $c = 0$ to $1/2$ in the limit $c \to \infty$.

(iii) The condition $\pi_i(q_1^*, q_2^*) < \pi_i(q_1^{**}, q_2^{**})$ is equivalent with $\frac{a^2b^2c^2}{c^3+8bc^2+20b^2c+16b^3} > 0$ so that (q_1^{**}, q_2^{**}) is a Pareto improvement to (q_1^*, q_2^*). The social optimum must be Pareto optimal because otherwise there would exist a constellation that has higher total welfare.

Proof of Proposition 2

(i) We have $R_1 = R_2 = \frac{a^2b^2(3c+4b)}{2(c+2b)^2)} > 0$ with $\frac{dR_1}{dc} = -\frac{a^2b^2(3c+2b)}{2(c+2b)^3)} < 0$.

(ii) The social optimal reference point is $R^{**} = \frac{4a^2b^2}{(c+4b)}$, which obviously declines in c. The condition $R < R^{**}$ is equivalent with $\frac{a^2b^2c^2}{c^3+8bc^2+20b^2c+16b^3} > 0$.

Appendix C On section 4

Proof of Lemma 2 The reaction functions in (4) do not have an interior intersect iff $c \leq \bar{c}$. The only statement that remains to be shown is $\frac{d\bar{c}}{da_\Delta} = \frac{2b\bar{a}}{(\bar{a}-a_\Delta)^2} > 0$.

C.1 Free-rider

Proof of Proposition 3

(i) We have $q_1^* = \frac{b}{c+2b}\bar{a} - \frac{b}{c}a_\Delta > 0$ because $c > \bar{c}$ and $q_2^* = \frac{b}{c+2b}\bar{a} + \frac{b}{c}a_\Delta > 0$. In particular, $q_1^* = 0$ and $q_2^* = \bar{a} - a_\Delta = a_1$ at $c = \bar{c}$. In terms of the derivatives,

$$\frac{dq_1^*}{dc} = -\frac{b}{(c+2b)^2}\bar{a} + \frac{b}{c^2}a_\Delta \quad , \quad \frac{dq_2^*}{dc} = -\frac{b}{(c+2b)^2}\bar{a} - \frac{b}{c^2}a_\Delta < 0 \ .$$

Here, dq_1^*/dc has a root at $\bar{c} + 2b\sqrt{\bar{a}a_\Delta}/(\bar{a} - a_\Delta) > \bar{c}$.

(ii) From (i) we get $Q^* = \frac{2b}{c+2b}\bar{a}$. Easy to verify is $Q^{**} = \frac{4b}{c+4b}\bar{a}$. Thus, $Q^* < Q^{**}$ is obvious.

(iii) The general directional derivative is as follows: Define $f_i(t) := \pi_i(q_1^* + t\alpha, q_2^* + t\beta)$. Then $f_i'(0)$ is the (α, β)-directional derivative of π_i at the Nash equilibrium. In this specific case we get

$$f_1'(0) = \frac{\beta b(\bar{a} - a_\Delta)}{c + 2b}(c - \bar{c}) \quad , \quad f_2'(0) = \frac{\alpha b}{c + 2b}((\bar{a} + a_\Delta)c + 2ba_\Delta)$$

and both expressions are positive for $\alpha, \beta > 0$. Whether the social optimal configuration is a Pareto Improvement for country 1 depends on the sign of

$$\pi_1(q_1^{**}, q_2^{**}) - \pi_1(q_1^*, q_2^*) = \frac{\left(\bar{a}^2 - 6a_\Delta\bar{a} + {a_\Delta}^2\right) b^2c^3 + \left(8{a_\Delta}^2 - 20a_\Delta\bar{a}\right) b^3c^2 + \left(20{a_\Delta}^2 - 16a_\Delta\bar{a}\right) b^4c + 16{a_\Delta}^2b^5}{2c^4 + 16bc^3 + 40b^2c^2 + 32b^3c}$$

At $c = \bar{c}$, this expression reads $-16b^3a_\Delta^2\bar{a}^2(\bar{a}+a_\Delta)/(\bar{a}-a_\Delta)^3$ and is negative when $a_\Delta > 0$. By continuity, this extends to a full neighborhood. For large a_Δ the c^3-coefficient in the numerator gets negative, implying that (q_1^{**}, q_2^{**}) is not a Pareto improvement for large c. The situation is different for country 2. Here,

$$\pi_2(q_1^{**}, q_2^{**}) - \pi_2(q_1^*, q_2^*) = \frac{\left(\bar{a}^2 + 6a_\Delta\bar{a} + {a_\Delta}^2\right) b^2c^3 + \left(20a_\Delta\bar{a} + 8{a_\Delta}^2\right) b^3c^2 + \left(16a_\Delta\bar{a} + 20{a_\Delta}^2\right) b^4c + 16{a_\Delta}^2b^5}{2c^4 + 16bc^3 + 40b^2c^2 + 32b^3c}$$

and this is clearly positive.

Proof of Proposition 4 For the notation used in proofs of inequality, see Appendix A.

(i) $\frac{dR_1}{dc} = \frac{b^2}{2c^2(c+2b)^3}\Big(\underbrace{-2a_1\delta c^3}_{(B)} \underbrace{-3a_1^2c^3}_{(A)} \underbrace{+2\delta^2bc^2}_{(B)} \underbrace{-2a_1^2bc^2}_{(C)} \underbrace{+3b^2\delta^2c}_{(A)} \underbrace{+2b^3\delta^2}_{(C)}\Big) < 0$ with equality for $c = \bar{c}$. $\frac{dR_2}{dc} = \frac{b^2}{2c^2(c+2b)^3}\Big(-\delta^2c^3 - 4a_1\delta c^3 \underbrace{-3{a_1}^2c^3}_{(B)} - 4a_1b\delta c^2 \underbrace{-2{a_1}^2bc^2}_{(A)} \underbrace{+3b^2\delta^2c}_{(B)} \underbrace{+2b^3\delta^2}_{(A)}\Big)$ which is strictly negative at $c = \bar{c}$.

(ii) $\frac{dR_1}{da_\Delta} = -\frac{(\bar{a}+a_\Delta)b^2c+2a_\Delta b^3}{c^2+2bc} < 0$ and $\frac{dR_2}{da_\Delta} = \frac{b^2((\bar{a}-a_\Delta)c-2a_\Delta b)}{c^2+2bc} > 0$ (because $c > \bar{c}$).

(iii) $R^{**} - R = \frac{b^2c^2}{c^3+8bc^2+20b^2c+16b^3}\bar{a}^2 + \frac{b^2}{c}a_\Delta^2$

C.2 Free-driver

Proof of Proposition 5

(i) Obvious from $q_2^* = \frac{b}{b+c}(\bar{a} + a_\Delta)$.

(ii) Simple algebra.

(iii) The directional derivatives are

$$f_1'(0) = \frac{(\beta + \alpha)b(\bar{a} - a_\Delta)}{c + b}(c - \bar{c}) \quad , \quad f_2'(0) = \frac{\alpha bc}{c + b}(\bar{a} + a_\Delta)$$

and both expressions are positive when $\alpha > 0$ and $\beta < -\alpha$. Simple algebra shows that $\pi_2(q_1^{**}, q_2^{**}) - \pi_2(q_1^*, q_2^*) < 0$ iff $c < 4a_\Delta^2b/(3\bar{a}^2 + 6a_\Delta\bar{a} - a_\Delta^2)$. Evaluating $\pi_1(q_1^{**}, q_2^{**}) - \pi_1(q_1^*, q_2^*)$ at this point gives $ba_\Delta^2\frac{12\bar{a}^3+30a_\Delta\bar{a}^2+8a_\Delta^2\bar{a}-2a_\Delta^3}{9\bar{a}^3+36a_\Delta\bar{a}^2+45a_\Delta^2\bar{a}+18a_\Delta^3}$. The numerator is positive due to $\bar{a} \geq a_\Delta$.

Proof of Proposition 6

(i) $R_1 = \frac{b^2}{2(c+b)^2}\left((2c+b)\bar{a}^2 - 2ba_\Delta\bar{a} - (2c+3b)a_\Delta^2\right)$ with a root at $c_{R_1=0} = \frac{(3a_\Delta - \bar{a})b}{2(\bar{a}-a_\Delta)}$

$R_2 = \frac{b^2}{2(c+b)}\left(\bar{a}+a_\Delta\right)^2$

$\frac{dR_1}{dc} = \frac{b^2}{(c+b)^3}\Big(\underbrace{\delta^2 b}_{(A)}\underbrace{+a_1\delta b}_{(B)}\underbrace{-a_1\delta c}_{(A)}\underbrace{-a_1^2 c}_{(B)}\Big) > 0$ with equality for $c = \bar{c}$.

(ii) See (i).

(iii) Helpful notations: $R_2 = R^{**}$ at $c_{R_2=R^{**}} = -\frac{4b(\bar{a}^2-2a_\Delta\bar{a}-a_\Delta^2)}{7\bar{a}^2-2a_\Delta\bar{a}-a_\Delta^2}$. This is positive iff $a_\Delta > (\sqrt{2}-1)\bar{a}$. For $R = R_1 + R_2$, $\frac{dR}{dc} = 0$ at $c_{R'=0} = \frac{(3a_\Delta-\bar{a})b}{3\bar{a}-a_\Delta}$. This is positive iff $a_\Delta > \bar{a}/3$ (same condition as "R_1 has a root"). **If counter-R&D is not possible**, the statement that R is positive and monotonical is justified with two arguments: For all $c > c_{R_1=0}$, where $R = R_1 + R_2$ anyways, this results from $c_{R'=0} < c_{R_1=0}$ (obvious because these expressions only differ in the denominator). For all other c we have $R = R_2$ and thus R inherits the characteristics 'positive' and 'decreasing in c'. That at the crossing point $R = R^{**}$ necessarily $\max\{R_1, 0\} = 0$ results from the proof $R < R^{**}$ if $R = R_1 + R_2$ below. **If counter-R&D is possible**, $R = \frac{b^2}{2(c+b)^2}\left((\bar{a}^2 - a_\Delta^2)(c+2b) + 2\bar{a}(\bar{a}+a_\Delta)\right) > 0$ with $\frac{dR}{dc} = -\frac{b^2}{2(c+b)^3}\left((3c+b)\bar{a}^2 - 2(b-c)a_\Delta\bar{a} - (c+3b)a_\Delta^2\right)$. We have $R^{**} - R = \frac{b^2}{2c^3+12bc^2+18b^2c+8b^3}\Big((5a_1^2 + 4a_1\delta + \delta^2)c^2 + 2a_1^2 bc\underbrace{-2a_1\delta bc}_{(A)}\underbrace{+2\delta^2 b^2}_{(A)}\Big) > 0$.

Appendix D On section 5

Proof of Proposition 8 In the free-rider region: With $C_1 = \frac{b^2}{2c(c^3+5bc^2+8b^2c+4b^3)}$,

$$R_1^{\text{Pub}} = C_1\Big(\underbrace{2a_1\delta c^3}_{(A)} + 2a_1^2c^3 \underbrace{+4a_1\delta bc^2}_{(B)}\underbrace{+3a_1^2bc^2}_{(C)} + 2a_1\delta b^2 c\underbrace{-\delta^2 bc^2}_{(A)}\underbrace{-2\delta^2b^2c}_{(B)}\underbrace{-\delta^2b^3}_{(C)}\Big) > 0$$

The derivative $\frac{dR_1^{\text{Pub}}}{dc}$ at $\bar{c}$ is $a_1^4/(2(a_1+\delta)^2) > 0$. Thus, with $\lim_{c\to infty} R_1^{\text{Priv}} = 0$, R_1^{Priv} cannot be monotone. With $C_2 = \frac{b^2}{2c(c^3+5bc^2+8b^2c+4b^3)}$,

$$R_2^{\text{Pub}} = C_2\Big(\underbrace{2a_1\delta c^3}_{(A)}\underbrace{+2a_1^2c^3 + 2a_1\delta bc^2}_{(B)}\underbrace{+3a_1^2bc^2}_{(C)}\underbrace{-2\delta^2bc^2}_{(A)}\underbrace{-4\delta^2b^2c}_{(B)}\underbrace{-2a_1\delta b^2c - \delta^2b^3}_{(C)}\Big) > 0\ ,$$

getting zero at $\bar{c}$. We have $R_1^{\text{Pub}} - R_2^{\text{Pub}} = \frac{2b^3\bar{a}a_\Delta}{c^2+3bc+2b^2} > 0$.

In the free-driver region:

$$R_1^{\text{Pub}} = \frac{b^2}{2(c+b)^2}\left((\bar{a}^2 + 2a_\Delta\bar{a} - 3a_\Delta^2)c - 4a_\Delta^2 b\right)\ ,$$

which is zero at $0 < c = \frac{4ba_\Delta^2}{\bar{a}^2+2a_\Delta\bar{a}-3a_\Delta^2} < \bar{c}$. The c-derivative is, with $a_2 = a_1 + \delta$,

$$\frac{dR_1^{\text{Pub}}}{dc} = \frac{b^2}{2(b+c)^3}\Big(\underbrace{2\delta^2 b}_{(A)}\underbrace{+2a_1\delta b}_{(B)} + a_1^2 b\underbrace{-2a_1\delta c}_{(A)}\underbrace{-a_1^2 c}_{(B)}\Big) > 0\ .$$

Clearly, $R_2^{\text{Pub}} = 0$.