Environmental Agreements

Strategic Complements
and Transnational Cooperation

Von der Carl von Ossietzky Universität Oldenburg
- Fakultät II - Informatik, Wirtschafts- und Rechtswissenschaften -

zur Erlangung des Grades
Doktor der Wirtschaftswissenschaften (Dr. rer. pol.)

genehmigte Dissertation

von

Leonhard Kähler

geboren am 14.03.1985 in Halle (Saale)

Referent : Prof. Dr. Carsten Helm
Korreferrent: Prof. Dr. Klaus Eisenack

Tag der Disputation: 03.07.2017

For my child.

Contents

1 Introduction 1

1.1 Climate . . . 1

1.2 Policy . . . 2

1.3 Literature on IEAs . . . 3

1.3.1 General Economic Perspective . . . 3

1.3.2 Adaptation to Climate Change . . . 6

1.3.3 Transnational Environmental Agreements . . . 7

1.3.4 My Contribution . . . 8

1.4 My Main Assumptions . . . 10

1.5 Structure . . . 12

2 Adaptation to climate change can support unilateral emission reductions 14

2.1 Introduction . . . 14

2.2 Basic model . . . 18

2.2.1 Assumptions and notation . . . 18

2.2.2 Basic results and interpretation . . . 21

2.3 Stackelberg games with adaptation . . . 25

2.3.1 Leadership in emissions and adaptation . . . 26

2.3.2 Leadership in adaptation only . . . 29

2.3.3 Leadership in emissions only . . . 31

2.3.4 Mixed leadership . . . 32

2.4 Unilateral action with endogenous sequence of play . . . 33

2.5 Conclusions . . . 38

2.A Proof of Prop. 1 (Extended and optimised damage function) . . . 41

2.B Proof of Prop. 2 (Stackelberg leadership in emissions and adaptation) . . . 41

2.C Proof of Prop. 3 (Stackelberg leadership in adaptation) 43

Interlude I **47**

3 Strategic Complements in International Environmental Agreements: a New Island of Stability **51**

3.1 Introduction 51
3.2 Model Structure 53
  3.2.1 Variables and Assumptions 54
  3.2.2 Game Structure 55
3.3 Game Equilibria 56
  3.3.1 Stage C: Emissions of Non-Members 56
  3.3.2 Benchmark Solution 59
  3.3.3 Stage B: Emissions of Agreement Members 59
  3.3.4 Stage A: Agreement Size 62
3.4 Discussion 64
3.5 Conclusions 66
3.A Proof of Prop. 7 67
3.B Proof of Prop. 8 68
3.C Proof of Prop. 9 71
3.D Proof of Prop. 10 72

Interlude II **75**

4 Transnational Environmental Agreements with Heterogeneous Actors **77**

4.1 Introduction 77
4.2 Current transnational approaches in the global governance literature . . . 80

4.3 Scope and limits of international environmental agreements 84

4.4 Proposals for theoretical analysis 86

4.4.1 Climate clubs 86

4.4.2 City alliances 90

4.4.3 Outlook 94

4.5 Conclusions 96

Interlude III **99**

5 The Effects of a City Alliance on Rural Emissions **102**

5.1 Introduction 102

5.2 Model 105

5.2.1 Basics 105

5.2.2 Complementarity 107

5.2.3 Specifications 109

5.2.4 Solution 109

5.2.5 Independent Freeriding 113

5.2.6 Welfare 115

5.3 Interpretation 116

5.4 Summary 118

5.A Proof of Prop. 11 121

5.B Proof of Lemma 1 122

5.C Proof for Prop. 12 122

6 Overall Conclusion **124**

7 Thanks **127**

1 Introduction

1.1 Climate

The global climate is changing and human-made emissions are largely responsible. This will very likely have catastrophic results like floods, droughts, storms and more. Climate change in general is inevitable by now but its scale is still uncertain and subject to human behaviour.

The exact tipping points in the climate and ecosystems are not known. A target of 2 degrees (Celsius) of global warming is often called acceptable, recently 1.5 degrees are also under discussion. Both numbers seem quite arbitrary. However even if some catastrophic events (like melting of Antarctic glaciers and resulting sea level rise) might take place at more than 2 degrees warming, other dire consequences are starting just now (e.g. severe droughts and increased frequency in storms and other extreme weather phenomena).

Large reductions in greenhouse gas emissions are necessary to limit global warming and avoid more severe disasters. If emissions continue to grow as they do now, large-scale global warming is to be expected.

The main drivers of global warming are carbon dioxide (CO_2) and methane (CH_4) (IPCC, 2013). These gases have very different properties. CO_2 is associated mostly with (heavy) industry, electricity production and transport. These emissions therefore can be interpreted as a byproduct of increasing global industrialisation. CO_2 emissions are large in highly developed countries and growing fast in emerging economies. CO_2 lasts for a long time in the atmosphere so the consequences of contemporary emissions remain relevant for many years (IPCC, 2013).

Attribution of CH_4 emissions is more difficult (IPCC, 2013). Agriculture and mining are large anthropogenic drivers. These emissions result probably from global population growth and increased demand for meat in emerging economies. Wetlands also emit lot of CH_4; this is considered natural but increases as temperatures rise due to global warming. If thermafrost sites thaw due to ongoing climate change, CH_4 can be set free there, too, creating a vicious circle. CH_4 remains in the atmosphere for a much shorter period of time than CO_2. Therefore contemporary methane emissions do not carry such a great weight for the future climate beyond the next few years. This is probably a reason why it is given much less concern in international negotiations about mitigation than CO_2.

1.2 Policy

Climate change depends on the total level of greenhouse gases. These in turn depend on global greenhouse gas emissions. Therefore coordinated action in emission reduction is very important. Individual contributions to mitigation are likely to have little effect on the global climate.

International negotiations on climate change are taking place regularly. Their main focus is on mitigation of CO_2 emissions but other greenhouse gases and adaptation to climate change also play a role. As a result of these negotiations some agreements have been reached with regard to global warming targets and emission reductions. The Kyoto Protocol from 1997 and the Paris Agreement from 2015 are notable examples for treaties with many member countries. Agreements like these are formed between national governments and have the ambition of being unique (i.e. encompassing as many countries as possible), they are usually called International Environmental Agreements (IEAs).

However all IEAs which exist so far are lacking in either ambition or binding force or both. Therefore actual emissions are not falling fast enough to meet a 2 degree target at the moment, much less a 1.5 degree target. In fact global emissions might even continue to grow in the next years.

Negotiations also take place on other levels. Subnational authorities, NGOs and supranational entities negotiate, regional agreements are discussed and alliances between similar countries form something like subgroups[1] in larger negotiations. Such structures aim at international cooperation but do not have a single agreement between national governments as their target and are called Transnational Environmental Agreements (TEAs) here.

Notably cities are joining forces to combat climate change. Some of the city alliances encompass several thousand members already. They are often multi-purpose in scope, including mitigation, adaptation, sustainability in general, technology research and exchange of best practices. Concrete mitigation targets of city alliances have not yet been set in many cases.

1.3 Literature on IEAs

1.3.1 General Economic Perspective

The literature on IEAs is large and growing. Many analyse the so called ‘prisoner’s dilemma’, freeriding, carbon leakage and other perspectives on the topic. There are also simulations of economic pathways for different policy scenarios. The central problem is quite well understood by now: There is little incentive for individual countries to mitigate greenhouse gas emissions because their impact on global climate change is (a) small and (b) probably partly reversed by other countries mitigating even less than they do anyway.

There are also some ideas how to improve the results of climate negotiations. Many

---

[1] Prominent examples are the so called BRICs (Brazil, Russia, India and China) as well as small island states.

works focus on issue linkage with trade deals and carbon border adjustments. Some take more structural approaches like minimum participation rules.

International environmental agreements has been established as a field of economic literature in the 1990s but the roots go back to older literature in environmental economics and cartel economics.[2] These older articles include work by D'Aspremont et al. (1983) which is widely used for the definition of a stable agreement. Bulow et al. (1985) have written about strategic substitutes and complements which is very important in understanding strategic behaviour. Baumol and Bradford (1972) describe effects of non-convexities on strategic behaviour. Further seminal work on economic incentives in climate policy comes from Hoel (1991) who describes the effects of unilateral (environmental) action in two player games.

Perhaps the most influential early work on international environmental agreements in economic literature comes from Barrett (1994). He uses a linear-quadratic model for the payoff of emission policy which has been widely adopted (and modified) since. As a game structure he takes a Stackelberg setup in which the coalition decides on emissions before the non-members do so; this, too, has been adopted by many authors. As a stability concept Barrett (like many other authors) employs the concept of internal and external stability derived from D'Aspremont et al. (1983). Barrett finds that self-enforcing international environmental agreements (if they exist in stable form at all) are either 'broad but shallow' or 'deep and small'. Small agreements (with three countries as members, independent of the number of countries in the world) form if relatively large gains from emissions reductions are possible. Large agreements (up to a 'grand coalition', an agreement of all countries which share the problem) form only if the net benefits are small.

---

[2] Probably due to the roots in cartel literature, agreements are often called coalitions. I use these terms synonymously here.

In another important early article Carraro and Siniscalco (1993) take a slightly more generalized look on coalition size. They use more abstract forms for payoff functions and include an option of transfer payments for stabilising coalitions. Carraro and Siniscalco (1993) employ a Nash game in which the coalition members and the non-members decide on emissions simultaneously in contrast to Barrett (1994) with a sequential game. A different analytic approach comes from Chander and Tulkens (1995). They use the 'core' concept of coalition stability and show a possibility of full cooperation. Nordhaus and Yang (1996) have built a calibrated model called RICE which simulates both climate and economy. They compare multiple equilibrium coalitions in terms of climate consequences and regional net welfare. In their two-country model Finus and Rundshagen (1998) show which equilibria are resilient to renegotiation. They compare different emission control regimes with regard to their stability.

Barrett (2001) models global public good provision with asymmetric countries. He finds that strategy is different and transfer payments are feasible if the asymmetry is large enough. Unilateral emission reductions can lead to self-interested emission reductions by other countries according to Copeland and Taylor (2005). The reason for the possibility of such strategic complements in their model lies in free trade. Carraro et al. (2006) analyse coalition stability in a calibrated model. Their best solution for negotiation design is an ex-post external transfer scheme. Employing another calibrated model, Finus et al. (2006) find potential for large gains to cooperation. However only small coalitions are stable according to their analysis. The review by Stern (2007) gives a very comprehensive view on the economics of climate change. There is a detailed description of mitigation and adaptation costs, potential for cooperation and policy measures. An international carbon market is strongly encouraged. In deviation from standard assumptions Heugues (2012b) analyses effects of strategic complementarity. She finds a larger possible coalition than most other authors. Eichner and Pethig (2013) extend the typical emissions framework with international trade. In this context much larger coalitions are possible but will not lead to significant emission reductions. In a model that includes technology development Helm and Schmidt (2015) analyse the effect of carbon border adjustments and timing of investment in green technology research. Their results show that late research is beneficial to member count but still investment in green technology and abatement are inefficiently low.

The overall consensus seems to be that small coalitions (with about three members) for climate protection are possible even under adverse circumstances. Larger coalitions suffer from freerider effects and therefore usually are not stable. If they can be stabilized either by favourable conditions or policy measures (like carbon border adjustments) they are not expected to lower emissions by much in comparison to a 'business as usual, all singletons' case.

1.3.2 Adaptation to Climate Change

Even if ambitious emission reductions are realized very soon, many countries will have to deal with damage from climate change. If mitigation is postponed more severe damage is to be expected. Therefore adaptation to climate change is going to be necessary.

The effects of adaptation on negotiations about mitigation are not trivial. There are very different views about consequences in the literature. It is likely that the concrete forms and timing of adaptation will play a major role in determining whether adaptation is beneficial or detrimental to mitigation.

McKitrick and Collinge (2002) analyse non-convexities which have their origin in defensive measures (i.e. adaptation). Their formal model is designed to find the optimal regulating policy. In his otherwise very simple model Barrett (2008) distinguishes between rich and poor countries. He finds that the ability of the rich countries to adapt does not harm poor countries. Zehaie (2009) focuses on timing of adaptation. According to this model deciding early on adaptation can lead to a lower emissions abatement, effectively shifting the burden of mitigation to other countries. In their calibrated model de Bruin et al. (2011) check for stable coalitions among twelve regions. They analyse the effect of proactive (i.e. early) adaptation on payoffs and the incentive to leave a coalition. They also find an incentive to over-adapt unilaterally, similar to the theoretical results of Zehaie (2009). Conversely, Benchekroun et al. (2011) find that a higher adaptation effectiveness can reduce freeriding incentives. They use a linear-quadratic model with symmetric countries. Ebert and Welsch (2012) introduce adaptation into a two player game similar to Hoel (1991). A very interesting result of this model is a reaction function (in terms of emissions) which can be upward sloping.

1.3.3 Transnational Environmental Agreements

International environmental agreements (IEAs) have been considered in different strands of scientific literature. My work methodologically builds mostly on the economic literature but also takes input from political science.

We can observe non-state actors like cities, industrial lobby groups and NGOs shaping energy and climate policy. Hagen et al. (2016) for example discuss the influence of lobby groups on international climate negotiations. They have an economic perspective and focus on coalition stability.

Cities in particular have received quite a lot of attention of scientific literature in recent years as actors in international climate policy. The resulting structures have been called City Alliances, Transnational Municipal Networks or Transnational Environmental Agreements. Betsill and Bulkeley (2006) promote a multilevel perspective on the Cities for Climate Protection (CCP) program. They discard the focus on nation states as main actors and a clear distinction between state and non-state actors. In a more concrete look on political action in the US Stewart (2008) examines interdependency of national and subnational cap-and-trade systems. He finds a theoretical advantage for the unitary approach but recommends plural architectures nonetheless. Subnational action can impair federal cap-and trade programs but can also create political pressure for stronger measures. In a combination of a historical and theoretical perspective Bulkeley (2010) examines the role of cities in climate governance. There is a detailed analysis of actors, barriers and institutional structure on the city level. Bansard et al. (2016) examine thirteen transnational municipal networks for geographical membership distribution, commitment of emission reductions and monitoring provisions. Particularly interesting is a comparison of emission reduction pledges on the municipal network and the UNFCCC level.

1.3.4 My Contribution

There are many models already on IEAs that show the prevalence of small coalitions. At best there seems to be an option for 'large but shallow' agreements in which many countries participate but the difference to no cooperation (in terms of both emissions and welfare) is small. I open up two new approaches for finding a better solution.

In my first approach non-convexities lead to strategic complementarity of emissions from the perspective of at least some players which give incentives not to freeride. This is possible by virtue of adaptation to climate change. There is thorough analysis of detrimental effects of adaptation on agreements for mitigation. However the possibility of a beneficial influence is not very well understood.

In 'Adaptation to climate change can support unilateral emission reductions' (Eisenack and Kähler, 2015) we analyse the strategic interaction of adaptation and unilateral action in emission reductions. We find that the possibility of adaptation can change strategic behaviour of a country. It can deter freeriding by the country in question if another actor reduces emissions unilaterally and can even lead to quasi-cooperation. This also holds if the order of emission choices is endogenous among countries.

In 'Strategic Complements in International Environmental Agreements: a New Island of Stability' (Kähler and Eisenack, 2016) we generalise the results from Eisenack and Kähler (2015) to a multi-country setting. We find potential for two different stable agreement sizes, one of them larger than the standard case and with more global emissions abatement.

My second approach focuses on non-state actors as a potentially positive force in climate policy. The public good nature of emissions abatement has lead to the focus on a single IEA between national governments. However this is not necessarily the best option. Multiple agreements, lobby groups, private-public partnership and city alliances offer some hope of succeeding where the UNFCCC negotiations seem to be stuck with mostly good intentions but little hope of sufficient concrete action. Urban initiatives in particular are already growing in size and importance.

In Hagen et al. (2017) we start to bridge the gap between political science and economic literature on transnational environmental agreements. We collect many political ideas on environmental agreements that go beyond the architecture of a single coalition of nation states. Starting from there we ask what kind of contribution an economic perspective can bring. We develop two examples for models on TEAs and develop avenues for further research.

Then I follow up on one of these models (with some modification) separately in 'The

Effects of a City Alliance on Rural Emissions'. There I analyse what effects emission reductions by a city alliance have on the incentives of rural regions. I find that there are conflicting forces. On the one hand rural regions which are politically and economically connected to geographically close cities can benefit from 'mirroring' the mitigation policies of these. On the other hand there is the well-known freerider incentive: if member emissions sink, non-members have an incentive emit more because their marginal damages sink. The overall effect is positive (i.e. leads to quasi-cooperation in the form of emission reductions by non-member regions) if the non-member regions are linked closely enough to the members of the city alliance and if the alliance lowers emissions of its members sufficiently.

1.4 My Main Assumptions

The axioms used in any model define the results to a large degree. It is therefore important to be transparent about their use. Perhaps the best way to justify an assumption is by empirical studies. However since real world phenomena have multiple sources, the attribution of causal relations is sometimes sketchy. Another possible reason to make a certain assumption are literary examples. If other scientists in the field (or related fields) take something for granted, it is probably safe to do the same. This obviously can lead to a continuation of errors. It can also lead to wrong assumptions over time if some facts (particularly in human behaviour or ecosystems) change over time. As a third option assumptions can be derived from other results or assumptions. It is also possible to blend results from different studies into new assumptions. The reliability of such new assumptions of course relies heavily on the correctness of the underlying axioms and on correct derivation.

I explain my assumptions in each model. However there are some assumptions that recur in several or all of my models. Here I give an overview of the most important ones; I explain and motivate them.

Assumption 1. *Each region has an industry which produces emissions as a byproduct. Higher local emissions (c.p.) lead to increasing benefits, but with decreasing marginal benefits. Each region also suffers damages from global emissions. These damages rise if global emissions grow.*

These properties of the (regional) payoff functions are very common in similar models (e.g. Hoel, 1991; Barrett, 1994; Altamirano-Cabrera and Finus, 2006; McGinty, 2007). Negative marginal benefits can be excluded simply because they represent emission choices which are never viable. It is reasonable to assume that the most emission efficient actions are taken 'first', i.e. in any case which allows for emissions at all. If some but not all actions leading to emissions are stopped, the most emission efficient are chosen to continue. However some models assume constant marginal benefits for simplicity reasons.

Positive marginal damages from emissions are a virtually ubiquitous axiom in environmental economics. They are derived directly from natural sciences (Stern, 2007). This simply means that higher emissions always lead to higher damages if nothing else changes. Increasing marginal damages (i.e. convex total damages) are commonly assumed but not universally accepted.

Assumption 2. *Regions are symmetric among each type.*

Of course this simplifies the analysis a lot in comparison to the real world. For a slightly more realistic interpretation it is possible to construe multiple regions in a model together as one larger region when checking for coalition size. Similar simplifications are common in the game theoretic literature on international environmental agreements (e.g. Barrett, 2001; Asheim et al., 2006).

Assumption 3. *Regions which are not members of an agreement decide on their emissions simultaneously with each other (in a Nash subgame) after they know the emissions level of the agreement members.*

It seems sensible to assume that non-member regions play a Nash game among themselves because a leadership position would be very difficult to claim and keep. In the literature both Nash and Stackelberg games between agreement and non-members are common (e.g. Barrett, 1994; Carraro and Siniscalco, 1993); the relationship between non-members of an (existing) agreement is almost always modelled as a Nash (sub-)game.

Even if the agreement can not claim a Stackelberg leader position, the other regions react in the same way, i.e. according to the same reaction function. In this case the agreement members won't be able to profit from it by anticipation and setting their emissions strategically.

1.5 Structure

The rest of this work consists of four main articles, three interludes and an overall conclusion. The first article 'Adaptation to climate change can support unilateral emission reductions' by Klaus Eisenack and myself has been published in Oxford Economic Papers (2015). It deals with a two-player problem of emissions and adaptation. We have both contributed equally to the model calculations and the writing.

The first interlude describes the connection between the first and second article. It also gives a short overview of some GAMS calculations that I have prepared for simulating multi player games of environmental agreements with concave damage functions.

'Strategic Complements in International Environmental Agreements: a New Island of Stability' is a paper by myself and Klaus Eisenack. It is currently under review. Here central results of the first article are developed for a multi player game. I have conceptualised this model and discussed the results. We have equally contributed to the calculation of the model and the proofs.

Interlude II explains the way in which the next part of my work generalises the topic of environmental agreements even further. The third article ’Transnational Environmental Agreements with Heterogeneous Actors’ by Achim Hagen, Klaus Eisenack and me is forthcoming in the book ’Economics of International Environmental Agreements: A Critical Approach’. It describes approaches to the global climate problem that go beyond a single coalition of national governments. We three authors have equally contributed to the concept, structure and writing of the article. The outlook has been formulated by Achim Hagen and me together. The city alliance model is an individual contribution of myself.

The third interlude gives a motivation of the fourth article from the perspective of the third article. It also contains some thoughts on the legal status of cities in an international setting. The fourth article ’The Effects of a City Alliance on Rural Emissions’ (by me alone, currently under review) shows a model of the effects of a city alliance on non-members. I conclude with a summary of the most important results of all four articles.

2 Adaptation to climate change can support unilateral emission reductions

Abstract

Policy advocates frequently call for unilateral action to promote international climate protection. It is still conventional wisdom that unilateral action does not pay off for individual countries due to free-riding incentives for other countries. Does this conclusion change if damage can be reduced by adaptation measures? This paper considers adaptation as an explicit decision variable, and frames unilateral action as Stackelberg game with two countries. The sequence of play is determined endogenously. We show that the Stackelberg leader reduces adaptation expenditures and emissions if the follower's damage function has a specific convexity property where adaptation leads to strategic complements. Then, no country has an incentive to deviate from the sequence of play. Unilateral action in adaptation or in emissions leads to a strict Pareto improvement compared to the non-cooperative Nash solution with lower total emissions and less adaptation.

2.1 Introduction

Solving problems of international pollution involves the reduction of damaging emissions, e.g., greenhouse gases or chlorofluorocarbon. As the reduction of such emissions is a public good, it is difficult to reach an international agreement on emission abatement. If one country were to unilaterally commit to emission reductions, it has to be expected that others would then increase their emissions as a strategic reaction. This is currently exemplified by the slow progress in international climate policy. Due to these difficulties, political discourse in such contexts as the United Nations Framework Convention on Climate

Change (UNFCCC) has focused increasingly on alternative means of addressing climate change. In particular, the idea of promoting adaptation to climate change has gained increasing attention in recent years (see, e.g., Haites, 2011). Adaptation is defined here as "adjustment to actual or expected climate and its effects [...] to moderate or avoid harm or exploit beneficial opportunities" (e.g., IPCC, 2014a)[3]. In contrast to emissions abatement, adaptation considers an externality as given and aims at reducing its damage. How does adaptation change the standard pessimistic analysis of emission abatement? This paper determines conditions under which adaptation leads to incentives for emission-reducing unilateral action.

It is not evident, *prima facie*, that adaptation can create incentives for emission reductions. When adaptation is an (imperfect) substitute for emission abatement (both reduce damages), the possibility of adaptation might well reduce incentives for mitigation (Ingham et al., 2007). One political argument put forward in the 1990s was that if too much effort is put into adapting to climate change, this might worsen the prospects for reaching an agreement to abate greenhouse gas emissions (see, e.g., Pielke et al., 2007, for a discussion). There are still at least two considerations suggesting that this argument needs to be qualified. First, the picture might change if multi-stage games are considered. Different game structures (without adaptation) have already been considered in the established literature on international environmental agreements (e.g. Hoel, 1991; Carraro and Siniscalco, 1993). There are models where a coalition takes the role of a Stackelberg leader in reducing emissions (starting with Barrett, 1994). Second, some fundamental considerations on the effects of adaptation show that standard convexity properties of damage functions may break down (e.g. Baumol, 1972). McKitrick and Collinge (2002) have analysed non-convexities that arise from adaptation measures from a policy point of view. If unilateral

---

[3] Beneficial opportunities are well possible, though not explicitly considered in this paper.

emission reductions lead to increasing marginal damages for other countries, they might reduce emissions as well: emissions become strategic complements. With the exception of Heugues (2012a,b), this has not been considered so far in the literature on international environmental agreements and emission abatement games.

This paper addresses the effect of adaptation on the incentives for emission reductions by formulating unilateral action as Stackelberg game. We simplify by concentrating on the case of two countries. Our setting is unique in that it both explicitly represents an adaptation decision with a very general class of damage functions, and considers different types of unilateral action (particularly, in terms of adaptation or mitigation). In a second step, we endogenise the selection of the Stackelberg leader. We find that it is individually rational for the Stackelberg leader to reduce its emissions below the non-cooperative Nash level if the follower has a specific property. We show that the specific property is equivalent to having a concave optimised damage function, and that it leads to strategic complements in emissions. Furthermore, also adaptation and emissions become strategic complements. Then, unilateral adaptation leads to a strict Pareto improvement and reduces total emissions. Moreover, if one country has this property, it endogenously becomes the Stackelberg follower. We thus provide a case for better prospects of unilateral action for climate protection if adaptation is included in the analysis.

Although the effects of adaptation have been mostly neglected in the established environmental economics literature (but see, e.g., Baumol, 1972; Butler and Maher, 1986, for early exceptions where adaptations are called defensive activities, protective measures, or averting behavior), research in the economics of adaptation is currently in an early stage (cf. Heuson et al., 2012). In this paragraph, we briefly summarise existing work on (i) the strategic effects of adaptation, and (ii) unilateral action. Zehaie (2009) addresses questions of timing when adaptation is an option. He shows that adaptation has no effect on emission abatement if it is undertaken after or simultaneously with emission reductions. This is different if both countries decide on adaptation first, so that higher levels of adaptation and emissions result – supporting the political argument that too much adaptation might lead to less ambitious mitigation. His study, however, does not address the question of unilateral action. There is also some first work on adaptation in international environmental agreements that focus on international adaptation finance (e.g., Fankhauser and Martin, 2010; Hof et al., 2011; Eisenack, 2012) or on coalition stability (e.g., Barrett, 2008; Buob and Siegenthaler, 2011; de Bruin et al., 2011; Marrouch and Chaudhuri, 2011). Most of the latter studies work with very specific (mostly linear-quadratic) damage functions. Ebert and Welsch (2011, 2012) investigate a large class of damage functions that explicitly model the effect of adaptation expenditures (when countries decides simultaneously on their adaptation and their emissions). They show that reaction functions in a game with two countries can become upward-sloping. They further determine the Nash equilibrium and the social optimum. Our paper extends this work by determining different Stackelberg equilibria when countries are given the option of adaptation. We can thus address the issue of unilateral action. We stick to a general formulation of damage functions that allows to represent two qualitatively distinct types of strategic behavior.

The effects of unilateral action have been early highlighted by Hoel (1991) – in a setting without adaptation (see also Bosetti and de Chian (2013) for a recent simulation study). He considered the (negative) slope of the reaction functions in a Nash game of emission reductions, and investigated the consequences of one country having preferences for lower total emissions. As a reaction, the other country expands its emissions. In a Stackelberg formulation (without adaptation), unilateral action has recently been analysed by Heugues (2012a). She determines the consequences of strategic complements in the amount of emissions for an endogenous sequence of play. We extend this work by including adaptation as a second decision variable. To our knowledge, our paper is the first that jointly analyses the strategic effects of adaptation in a Stackelberg setting with endogeneous sequence of play. We further prove a crucial link between the damage functions in the literature on international pollution problems with and without adaptation. We show when extended damage models which account for both emissions and adaptation are strategically equivalent to conventional damage models that disregard adaptation. This has implications for integrated assessment modeling, and is helpful for further research into the economics of adaptation.

Section 2.2 presents the basic model setup and collects its properties. We define α and β-type countries, which is an important distinction for the whole paper. We show in Section 2.3 how the solutions for different game structures with adaptation depend on the types of countries involved, and compare their equilibrium payoffs. The subsequent section analyses an extended multi-stage game where the Stackelberg leader is determined endogenously. A discussion section concludes, and the Appendix contains proofs.

2.2 Basic model

This section presents the model setup, introduces some terminology and basic equations that are crucial throughout the paper. It provides a new proposition that sheds light on the interpretation of some of our assumptions, and will simplify technical matters later on. One main achievement of our paper is the analysis of Stackelberg games. It will still be important to compare the Stackelberg equilibrium with the non-cooperative Nash solution, so the latter is summarised in this section.

2.2.1 Assumptions and notation

The basic game considers countries $i \in \{1, \ldots, N\}$ with payoff functions

$$\pi_i(e_i, e_{-i}, a_i) = B_i(e_i) - D_i(e_i + e_{-i}, a_i) - a_i, \tag{1}$$

defined by 'extended damage functions' $D_i(e, a_i)$ and by benefit functions $B_i(e_i)$, with the partial derivatives

$$\begin{aligned}\forall i : \partial_e D_i > 0, \partial_{ee} D_i > 0, \\ \partial_a D_i < 0, \partial_{aa} D_i > 0, \\ \partial_{ea} D_i = \partial_{ae} D_i < 0, \\ B_i' > 0, B_i'' < 0.\end{aligned} \tag{2}$$

Note that the introduction of the extended damage function is the main structural extension to the standard formulation in the literature, where damages only depend on emissions. This formulation defines the marginal adaptation costs to be identical to unity. This avoids the well known problems in defining a common metric of the "amount of adaptation" (see, e.g., Füssel and Klein, 2006). Instead, we consider a_i as expenditures that (*ceteris paribus*) determine the damage. Total emissions are denoted by $e = \sum_{i=1,\ldots,N} e_i$, and $e_{-i} = \sum_{j \neq i} e_j$ is the sum of all countries' emissions except country i. Damage increases convexly with emissions e, and decreases convexly with adaptation expenditures a_i: the most effective adaptations are undertaken first. Moreover, adaptation reduces marginal damage. This is equivalent to the statement that for higher emissions, adaptation does reduce damage more effectively at the margin. The benefit functions $B_i(e_i)$ are assumed to be strictly increasing and concave. This corresponds to countries where emissions are a necessary input for production, which in turn yields a benefit. For greater production benefits, greater emission input is necessary, and there are decreasing returns to scale. We now define

$$\nu_i := \partial_{ee} D_i - \frac{\partial_{ea} D_i^{\,2}}{\partial_{aa} D_i}, \tag{3}$$

and impose the convexity condition

$$\nu_i - B_i'' > 0, \tag{4}$$

for all countries. We call i an

$$\alpha\text{-type country if} \quad \nu_i > 0,$$
$$\beta\text{-type country if} \quad \nu_i < 0.$$

These two cases will be crucial in the remainder of the paper. The existence of β-type countries becomes possible, depending on the parameterisation, because of the indirect effects of adaptation on the emission decision. Note that $\partial_{ea} D_i \neq 0$ is required for β-type countries. Otherwise ν_i would equal $\partial_{ee} D_i$, which is always positive. The interpretation of the country types is further investigated below.

For convenience, we introduce some further notation that helps to analyse the Nash equilibrium and more complex game structures subsequently. The solutions of

$$B_i'(R_i) \equiv \partial_e D_i(R_i + e_{-i}, a_i), \tag{5}$$

define the 'extended reaction functions' $e_i = R_i(e_{-i}, a_i)$, and the solutions of

$$-\partial_a D_i(e, A_i) \equiv 1, \tag{6}$$

the 'optimal adaptation response functions' $a_i = A_i(e)$. The 'optimised reaction functions' $e_i = \tilde{R}_i(e_{-i})$ are defined by

$$B_i'(\tilde{R}_i) \equiv \partial_e D_i(\tilde{R}_i + e_{-i}, A_i(\tilde{R}_i + e_{-i})). \tag{7}$$

Finally, the 'optimised damage functions' are defined as

$$\tilde{D}_i(e) := \min_{a_i} D_i(e, a_i) + a_i. \tag{8}$$

2.2.2 Basic results and interpretation

By generalising the standard results of Hoel (1991), Ebert and Welsch (2011, 2012) established that

$$\tilde{R}_i'(e_{-i}) = \frac{\nu_i}{B_i'' - \nu_i} \in (-1, \infty). \tag{9}$$

If $\nu_i > 0$, the optimised reaction function $\tilde{R}_i$ is downward-sloping as in the case without adaptation. If, however, $\nu_i < 0$, the inclusion of adaptation leads to an upward-sloping optimised reaction function, although the extended reaction functions remain decreasing.

In the Nash equilibrium, each country decides simultaneously on two variables (its emissions and its adaptation expenditures). It is characterised by the first-order conditions $\forall i : \partial_e D_i(e, a_i) = B_i'(e_i)$, $-\partial_a D_i(e, a_i) = 1$. The marginal damage to the individual country equals marginal benefit from its emissions and (negative) marginal costs of adaptation, respectively. Solving these equations jointly for all countries characterises the emission and adaptation decisions. Ebert and Welsch (2011) have shown that this Nash equilibrium can alternatively be characterised by the intersection of the optimised reaction functions,

$$\forall i : e_i = \tilde{R}_i(e_{-i}), \tag{10}$$

$$a_i = A_i(e). \tag{11}$$

We now deduce some new properties from the above. It is straightforward that

$$\partial_e R_i = \frac{\partial_{ee} D_i}{B_i'' - \partial_{ee} D_i} \in (-1, 0), \tag{12}$$

$$\partial_a R_i = \frac{\partial_{ea} D_i}{B_i'' - \partial_{ee} D_i} > 0, \tag{13}$$

$$A_i'(e) = -\frac{\partial_{ea} D_i}{\partial_{aa} D_i} > 0. \tag{14}$$

Furthermore, it is easy to see from Eq. (5) and Eq. (7) that $\tilde{R}_i(e_{-i}) = R_i(e_{-i}, A_i(R_i(e_{-i}) + e_{-i})$ and

$$\begin{aligned} \tilde{D}_i(e) &= D_i(e, A_i(e)) + A_i(e), && (15) \\ \tilde{D}_i'(e) &= \partial_e D_i + \partial_a D_i A_i' + A_i' \\ &= \partial_e D_i(e, A_i(e)). && (16) \end{aligned}$$

As, due to Eq. (6), $(\partial_a D_i + 1)A_i' = 0$, the marginal optimised damage is the same as the marginal damage. The last equation allows for a proposition that links Nash equilibria in our setting to the standard setting without adaptation. It also eases the interpretation and understanding of β-type countries. Until now, the β-type has only been introduced formally, but its meaning is important to get a sense of the results and to judge their relevance.

Proposition 1. *Denote the equilibrium of the Nash game with payoff functions* $\pi_i = B_i - D_i - a_i, i = 1, 2$ *by* $(a_1^N, e_1^N, a_2^N, e_2^N)$*. Let* $(\tilde{e}_1^N, \tilde{e}_2^N)$ *be the equilibrium of the Nash game with payoff functions* $\tilde{\pi}_i = B_i - \tilde{D}_i, i = 1, 2$*, where* $\tilde{D}_i$ *are the optimised damage functions as defined from the extended damage functions by Eq. (8). Then* $\tilde{D}_i'' = \nu_i$*, and* $\tilde{e}_1^N = e_1^N$*,* $\tilde{e}_2^N = e_2^N$*.*

This proposition clarifies several issues. First, it shows that β-type countries (with negative ν_i) have a concave optimised damage function. Although the extended damage functions were generally assumed to be strictly convex in both arguments, this convexity does not necessarily carry over to the optimised damage function. This does not, however, invalidate the existence of a game equilibrium due to the convexity assumption Eq. (4). This contrasts the standard case in the environmental economics literature where the damage function is assumed to be convex (corresponding to α-type countries with $\nu_i > 0$).

Second, the existence of an interior Nash equilibrium can be easily clarified. By differentiating the first-order condition $\tilde{\pi}_i' = B_i' - \tilde{D}_i' = 0$, we obtain $\tilde{\pi}_i'' = B_i'' - \tilde{D}_i'' = B_i'' - \nu_i$. This second-order condition is strictly negative due to Eq. (4), so that sufficiency and uniqueness are established. Further assumptions are required to show the existence, for example Inada-conditions for B_i.

Third, Prop. 1 sheds light on what characterises β-type countries, and helps making plausible that such countries might exist. We have assumed that the extended damage function of both α and β-type countries is convex in the amount of emissions. In other words, if adaptation expenditures are kept constant, there are always increasing marginal damages (as in the standard literature). If adaptation is variable, however, β-type countries suffer a decreasing marginal optimised damage from emissions. Prop. 1 thus shows that β-type countries are exactly those that are able to 'concavify' their damages when they adapt in an optimal way. Why this might be the case can be seen by inferring from Eq. (3) that $\nu_i < 0$ (β-type) if and only if

$$\partial_{ee} D_i \cdot \partial_{aa} D_i < \partial_{ea} D_i^2. \tag{17}$$

This shows that β-type countries are characterised by a non-trivial configuration of properties. We can identify (*ceteris paribus*) three possibilities from this inequality: a low value of the second derivatives $\partial_{ee} D_i, \partial_{aa} D_i$, or a high absolute value of the cross derivative $\partial_{ea} D_i$. Countries with very small $\partial_{ee} D_i$ are those that suffer a damage close to being linear in global emissions when they do not adapt (almost constant marginal damage). In this case, even simple adaptation measures can induce the optimised damage function to become concave. Countries with very small $\partial_{aa} D_i$ are those where the effectiveness of additional adaptation expenditures is not decreasing very strong. They dispose of a portfolio of adaptations where the difference between the effectiveness of inframarginal and marginal adaptation expenditures is not large. Adaptation can be scaled up without significant diminishing returns.

For countries with a large value of $|\partial_{ea} D_i| = |\partial_{ae} D_i|$, on the one hand additional adaptations become significantly more cost-effective when emissions increase, and on the other hand adaptations are available in the country that substantially reduce also marginal damages (in addition to absolute damages). This is conceivable, for example, if the geographical conditions of the country are such that climate change primarily manifests in more frequent extreme events (like floods or droughts), but not so much in a stronger intensity of extreme events. Note that climate change patterns are expected to be quite different from region to region (see IPCC, 2013). In this case some adaptation measures (e.g. education that helps people to respond to hydro-meteorological disasters in a better way, irrigation systems, desalination plants or flood protection infrastructure) introduce economies of scale as they can be used more frequently with increasing climate change (cf. Eisenack, 2014). There are several other plausible arguments for such a case, including those made by Baumol (1972). He considers an externality where affected residents can substantially reduce (absolute and marginal) damages by migration. If additional adaptations become increasingly cost-effective with rising pollution, a convex extended damage function can lead to a concave optimised damage function. It is yet not straightforward to generally assign a set of countries to the two types, as different economic and geographical conditions will influence all second derivatives of the extended damage function in different ways[4] . Some of these conditions may be changed by specific policies, while others

---

[4] Countries of the β-type need not necessarily be those with a low vulnerability, as was suggested by Ebert and Welsch (2012). For example, $\partial_{ee} D_i$ might become quite small for high levels of emissions (as implied by the argument of Baumol (1972)): Although damages increase with emissions, they will increase less when damages come close to a maximum level. Thus, countries that already suffer high damages (compared to a maximum damage, probably small island developing states), might qualify as β-type countries. The interpretation of Ebert and Welsch (2012) rests on the formalisation of climate sensitivity as the second derivative $\partial_{ee} D_i$, and adaptive capacity as $\partial_{ea} D_i^2 / \partial_{aa} D_i$. We would argue, however, that sensitivity is more

may not be changed.

Fourth, Prop. 1 shows that the Nash equilibrium of an emissions game with adaptation can be determined by reducing the game to a simpler emissions game that neglects adaptation. One only needs to replace the extended damage function by the optimised damage function, and can proceed with the established standard analysis. This is claimed to be done implicitly by some established models on the integrated assessment of climate change (e.g., Nordhaus and Boyer, 2000). Game equilibria yet change if sequential play is considered. The proposition thus provides a bridge between the established literature on damages without adaptation, and the more recent literature that includes adaptation in the analysis.

2.3 Stackelberg games with adaptation

What are the consequences for unilateral action when there are β-type countries with concave optimised damage functions? Are there differences between unilateral adaptation and unilateral mitigation? Can unilateral action become individually rational in this context? We thus model different kinds unilateral action with adaptation as Stackelberg games. The game structures considered in the subsections make different assumptions about whether there is leadership in emissions, or adaptation, or both. These structures are assumed to be exogenously given in each subsection. The choice of game structure will be endogenised in Section 4. To bring the main mechanisms to the fore, we simplify our analysis to the case of two countries (as is common in the literature, e.g. Hoel, 1991; Zehaie, 2009; Ebert

---

appropriately expressed by the first derivatives. This can be illustrated in the context of barriers to adaptation (cf. Moser and Ekstrom, 2010; Eisenack et al., 2014). Barriers have the consequence that damages are above the optimised damage function, e.g. due to inefficiencies in implementing adaptations. Then, resolving barriers - in the literature frequently synonymous with reducing vulnerability - does not affect the second derivatives of the extended damage function.

and Welsch, 2011).

2.3.1 Leadership in emissions and adaptation

We consider two countries $i \in \{L, F\}$, and assume that country L is the Stackelberg leader in both emissions and adaptation. Country L announces its adaptation expenditures a_L and its emissions e_L in the first stage of the game. In a second stage, country F reacts by determining its adaptation expenditures a_F and its level of emissions e_F. To compare this Stackelberg solution with the non-cooperative Nash equilibrium, we denote the latter by $(a_L^N, e_L^N, a_F^N, e_F^N)$, and the former by $(a_L^S, e_L^S, a_F^S, e_F^S)$.

By backward induction, we start with the second stage. The follower maximises π_F with respect to a_F, e_F and takes the first-stage decisions a_L, e_L as given parameters. This problem is formally the same as F's problem in the non-cooperative Nash solution (Eq. 10, Eq. 11), so that $e_F = \tilde{R}_F(e_L)$ and $a_F = A_F(e_L + e_F)$, with the properties as laid out in Eq. (9) and Eq. (14).

In the first stage, by anticipating the follower's reaction, the leader determines

$$\max_{a_L, e_L} \pi_L = B_L(e_L) - D_L(e_L + \tilde{R}_F(e_L), a_L) - a_L. \tag{18}$$

Due to Prop. 1, this problem can be slightly reformulated, so that the outcome of the decision depends on the properties of the follower as summarised in the following proposition (see Appendix for the proof):

Proposition 2. *The existence of an e_L^S that satisfies the condition*

$$B_L'(e_L^S) = (1 + \tilde{R}_F'(e_L^S))\, \tilde{D}_L'(e_L^S + \tilde{R}_F(e_L^S)), \tag{19}$$

is necessary and sufficient for a Stackelberg equilibrium with leadership in emissions and adaptation $(a_L^S, e_L^S, a_F^S, e_F^S)$, with $e_F^S = \tilde{R}_F(e_L^S), e^S = e_L^S + \tilde{R}_F(e_L^S), a_F^S = A_F(e^S)$ and

$$a_L^S = A_L(e_L^S + \tilde{R}_F(e_L^S)). \tag{20}$$

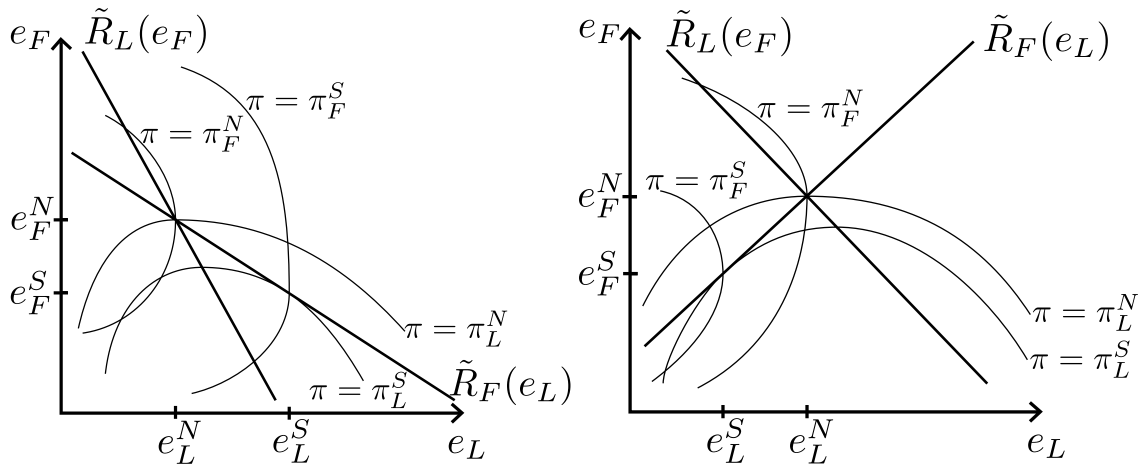

Figure 1: Nash and Stackelberg equilibrium with α-type country L. In the first case (left) country F is of the α type as well, in the second case (right) country F is of the β type.

It holds that $e_F^S < e_F^N$, $\pi_L^S > \pi_L^N$. *There are two cases:*

1. *For an* α*-type country F,* $e_L^S > e_L^N$, $e^S > e^N$, $a_L^S > a_L^N$, $a_F^S > a_F^N$, $\pi_F^S < \pi_F^N$.
2. *For a* β*-type country F,* $e_L^S < e_L^N$, $e^S < e^N$, $a_L^S < a_L^N$, $a_F^S < a_F^N$, $\pi_F^S > \pi_F^N$.

The interesting case is that with a β-type country as follower. As the follower has an upward-sloping optimised reaction function, it reacts with an expansion of emissions if the leader emits more. Since the leading country anticipates this, it can unilaterally reduce emissions, which the follower will then do as well. This joint reduction of emissions lowers the leader's damage more than the foregone benefits. The intuition behind this works as follows: the leading country lowers its emissions from the point where its marginal (foregone) benefits equal its individual marginal (avoided) damages (Nash equilibrium) to the point where the leading country puts a modified weight on the marginal damages that accounts for the anticipated reaction of the follower.

The dependence on the properties of the follower is illustrated in Fig. 1. It is possible to represent the situation in this game in the (e_L, e_F) plane, if we consider the optimised damage and reaction functions. The left graph shows the case where the follower country F is an α-type country (with decreasing optimised reaction function $\tilde{R}_F$), while the other graph depicts a β-type follower country L (with increasing optimised reaction function $\tilde{R}_F$). The non-cooperative Nash solution is at the intersection of both reaction functions. Due to the convexity assumptions, iso-payoff curves are U-shaped, and the curves $\pi_L = \pi_L^N, \pi_F = \pi_F^N$ go through the Nash equilibrium. Payoffs for country L are higher below the $\pi_L = \pi_L^N$ curve, while the payoffs for country F increase to the left of $\pi_F = \pi_F^N$. Thus, the points between both curves to the lower left of (e_L^N, e_F^N) represent Pareto improvements compared to the Nash equilibrium.

Consider the left graph in Fig. 1 first, where country F is of the α type. The leader then selects e_L under the assumption that the follower reacts with emissions on the optimised reaction function $\tilde{R}_F$. The leader's maximum payoff is thus reached where an iso-payoff curve $\pi_L = \pi_L^S$ is tangent to $\tilde{R}_F$. This is only possible if the leader expands its emissions, $e_L^S > e_L^N$, and improves the leader's payoff in comparison to the Nash equilibrium to $\pi_L^S > \pi_L^N$. As the slope of $\tilde{R}_F$ is less than unity, total emissions increase $e^S > e^N$. As a consequence, both countries need to expand their adaptation expenditures, and the follower's payoff is reduced to $\pi_F^S < \pi_F^N$.

This is different in the case illustrated by the right graph in Fig. 1, where the follower is a β-type country. Again, emissions e_L are selected such that the iso-payoff curve $\pi_L = \pi_L^S$ is tangent to the optimised reaction function $\tilde{R}_F$, which leads to the emission reduction $e_L^S < e_L^N$. Due to its β-type, the follower also reacts with emission reductions. In sum, total emissions decrease to $e^S < e^N$, both countries can reduce their adaptation expenditures, and a Pareto improvement is achieved as both countries increase their payoff.

It is not straightforward to establish the existence of the solution to the first-order conditions. By further differentiating Eq. (37) from the Appendix,

$$\frac{d^2}{de_L^2}\pi_L(e_L, \tilde{R}_F(e_L), A_L(e_L + \tilde{R}_F(e_L))) = B_L'' - (1 + \tilde{R}_F')^2\nu_L - \tilde{R}_F''\tilde{D}_L', \qquad (21)$$

is obtained. If this is strictly negative, uniqueness of the Stackelberg equilibrium would be implied. The proof of Prop. 2 would then show that the existence of an interior Nash equilibrium would also imply the existence of the Stackelberg equilibrium. Yet, strict negativity cannot be established in general. The sign of $\tilde{R}_F''$ involves the third derivatives of the extended damage function, on which no assumptions have been imposed[5].

To sum up this section, there is a case for unilateral action contributing to the solution of global environmental problems. If the follower in this game is of the β-type, unilateral action achieves emission reductions compared to the non-cooperative Nash solution. There are also lower adaptation expenditures in both countries. When there are both α-type and β-type countries, the result depends on who takes the lead. This is investigated in a later section.

2.3.2 Leadership in adaptation only

We now investigate another type of unilateral action. Suppose that one country can commit to Stackelberg leadership in terms of adaptation, but not in terms of emissions. To compare the game equilibrium of this section with results from other game structures, we will denote it by $(a_L^A, e_L^A, a_F^A, e_F^A)$. The leader country L choses a_L in the first stage of the game. In the second stage, the leader's emissions e_L, follower's emissions e_F and adaptation expenditures a_F are chosen simultaneously.

---

[5] Strategic complements are also studied in the international trade policy literature (e.g. Neary, 1994; Bandyopadhyay, 1997), that encounters similar difficulties with establishing the existence of a maximum in general.

Consider the second stage first. Country F optimises its payoff π_F as in the non-cooperative Nash solution according to $e_F = \tilde{R}_F(e_L)$ and $a_F = A_F(e_F + e_L)$. The optimised reaction function and the adaptation response function have the same properties as laid out in Eq. (14) and Eq. (9). Country L optimises its payoff for a given decision of the other country and its own adaptation decision from the first stage, i.e.

$$\max_{e_L} B_L(e_L) - D_L(e_L + e_F, a_L) - a_L. \tag{22}$$

The first order condition is the same as in the Nash game, so that $e_L = R_L(e_F, a_L)$. The second stage equilibrium (e_L^A, a_F^A, e_F^A) can thus be characterised by the implicit equations

$$e_L^A = R_L(\tilde{R}_F(e_L^A), a_L), \tag{23}$$

$$a_F^A = A_F(e_L^A + e_F^A), \tag{24}$$

$$e_F^A = \tilde{R}_F(R_L(e_F^A, a_L)). \tag{25}$$

The proof below will show the following: If the Stackelberg leader would expand its adaptation expenditures, total emissions would increase. Consequently, the follower would expand its adaptation expenditures as well. The follower would also expand its emissions if it is of the β-type, but would reduce emissions otherwise.

In the first stage, the leader selects a_L, and anticipates the resulting effects on the second stage equilibrium

$$\max_{a_L} \pi_L = B_L(e_L^A(a_L)) - D_L(e^A(a_L), a_L) - a_L, \tag{26}$$

yielding the following result (see Appendix for the proof).

Proposition 3. *A necessary condition for a Stackelberg equilibrium with leadership in adaptation* $(a_L^A, e_L^A, a_F^A, e_F^A)$ *is*

$$1 + \partial_a D_L(e_L^A + \tilde{R}_F(e_L^A), a_L^A) = -\partial_e D_L(e_L^A + \tilde{R}_F(e_L^A), a_L^A)) \frac{de_F^A}{da_L}, \tag{27}$$

$$e_L^A = R_L(\tilde{R}_F(e_L^A), a_L^A). \tag{28}$$

Then $e_F^A = \tilde{R}_F(e_L^A)$, $e^A = e_L^A + e_F^A$ *and* $a_F^A = A_F(e^A)$. *It holds that* $\pi_L^S > \pi_L^A > \pi_L^N$, *and additionally*

1. *if country F is of* α*-type, then* $e_L^A > e_L^N$, $e^A > e^N$, $a_L^A > a_L^N$, $a_F^A > a_F^N$, $\pi_F^A < \pi_F^N$,
2. *if country F is of* β*-type, and country L of* α*-type, then* $e_L^A < e_L^N$, $e^A < e^N$, $a_L^A < a_L^N$, $a_F^A < a_F^N$, $\pi_F^A > \pi_F^N$.

The proposition shows that leadership only in adaptation leads to different results than leadership in both emissions and adaptation. The leader in adaptation can always improve its payoff compared to the non-cooperative Nash solution, but not so much as if the country has an additional degree of freedom in terms of emissions. We have further established in this subsection that strategic complementarity is also possible between adaptation and emissions. If the country L is of α-type, and country F of β-type, the leader unilaterally selects lower adaptation expenditures than in the Nash solution, since it anticipates that this would lead to less total emissions. A Pareto improvement is the consequence.

2.3.3 Leadership in emissions only

We next consider the situation where the Stackelberg leader can only commit to chose its emissions in the first stage, but not its adaptation expenditures. In the second stage, the countries determine a_L, a_F, e_F, depending on e_L. As above, the reaction functions in the second stage are identical to the Nash setting, so that the solution of the second stage is characterised by $e_F = \tilde{R}_F(e_L)$, $a_L = A_L(e_L + e_F)$, $a_F = A_F(e_L + e_F)$. In the first stage, the leader consequently determines

$$\max_{e_L} \pi_L = B_L(e_L) - \tilde{D}_L(e_L + \tilde{R}_F(e_L)). \tag{29}$$

This yields the first-order condition $0 = B'_L - (1 + \tilde{R}'_F)\tilde{D}'_L$. We see, by comparing with Prop. 2, that the equilibrium is characterised by the same equations as in the case with leadership in both adaptation and emissions. We conclude that leadership only in emissions leads to the same outcome as if there is leadership in both adaptation and emissions. This is plausible since the decisions of one country do not depend on the adaptation decision of the other country. Only if adaptation is chosen first, the results differ.

2.3.4 Mixed leadership

The equivalence to leadership in both adaptation and emissions is *prima facie* not as simple for a mixed case. Suppose that one country can commit to unilateral action in terms of emissions, while the other to unilateral action in terms of adaptation. We need to be careful with our notation here. Although it is a mixed case, it will be convenient to denote the decision variables of the leader in emissions by (a_L, e_L), and the decision variables of the leader in adaptation by (a_F, e_F). So, in the first stage (e_L, a_F) are determined, and (a_L, e_F) in the second stage.

The second stage result is, again, characterised by the same reaction functions as in the Nash game, i.e. $e_F = \tilde{R}_F(e_L), a_L = A_L(e_L + e_F)$.

In the first stage, in contrast to the previous games structures, both countries need to make a decision. The leader in adaptation maximises its payoff subject to the anticipated equilibrium in the second stage, and by taking the emissions of the other country e_L as given:

$$\max_{a_F} \pi_F = B_F(\tilde{R}_F(e_L)) - D_F(e_L + \tilde{R}_F(e_L), a_F) - a_F. \tag{30}$$

The first-order condition simply yields $-\partial_a D_F - 1 = 0$, so that $a_F = A_F(e_L + e_F)$: the leader in adaptation choses the same adaptation expenditures as if it is a follower in adaptation.

The leader in emissions maximises its payoff subject to the anticipated equilibrium in the second stage, i.e.

$$\max_{e_L} \pi_L = B_L(e_L) - \tilde{D}_L(e_L + \tilde{R}_F(e_L)). \tag{31}$$

This decision is independent from the other country's adaptation decision, so that the problem is – again – formally identical to that of a leader in both adaptation and emissions in Prop. 2.

So, the outcome of the mixed case is not different from the other cases of unilateral action, except from the case where one country only leads in adaptation, while the other follows in both adaptation and emissions. The following section thus investigates the consequences of comparing three possible outcomes: $(\pi_i^N, \pi_j^N), (\pi_L^S, \pi_F^S), (\pi_L^A, \pi_F^A)$. The results depend on the country types.

2.4 Unilateral action with endogenous sequence of play

The previous analysis takes Stackelberg leadership in one or multiple variables as given. There might be historical or other reasons that define leadership at a given time. In this section, however, we solve a 3-stage game where Stackelberg leadership is determined endogenously in stage 1. The game with endogenous sequence of play is modelled in an 'observable delay' fashion as Hamilton and Slutsky (1990) have done for duopoly games.

In detail, the sequence of play is endogenised as follows. In stage 1, countries determine whether one of them commits to unilateral action in adaptation and/or emissions. This stage is played simultaneously in a Nash-fashion, and its outcome determines the game structure of the following stages. In the subgame of the following stages, the emissions and adaptation expenditures are selected in the determined sequence (stage 2, and possibly 3). We determine the equilibria of the complete 3-stage game, depending on the country types.

		Country 2			
		Aa/Ae	Aa/Te	Ta/Ae	Ta/Te
Country 1	Aa/Ae	(a_1, e_1, a_2, e_2) $\pi_1^N, \quad \pi_2^N$	$(e_2), (a_1, e_1, a_2)$ $\pi_{1,F}^S, \quad \pi_{2,L}^S$	$(a_2), (a_1, e_1, e_2)$ $\pi_{1,F}^A, \quad \pi_{2,L}^A$	$(a_2, e_2), (a_1, e_1)$ $\pi_{1,F}^S, \quad \pi_{2,L}^S$
	Aa/Te	$(e_1), (a_1, a_2, e_2)$ $\pi_{1,L}^S, \quad \pi_{2,F}^S$	(a_1, e_1, a_2, e_2) $\pi_1^N, \quad \pi_2^N$	$(e_1), (a_1, a_2, e_2)$ $\pi_{1,L}^S, \quad \pi_{2,F}^S$	$(a_2), (a_1, e_1, e_2)$ $\pi_{1,F}^A, \quad \pi_{2,L}^A$
	Ta/Ae	$(a_1), (e_1, a_2, e_2)$ $\pi_{1,L}^A, \quad \pi_{2,F}^A$	$(a_1, e_2), (e_1, a_2)$ $\pi_{1,F}^S, \quad \pi_{2,L}^S$	(a_1, e_1, a_2, e_2) $\pi_1^N, \quad \pi_2^N$	$(e_2), (a_1, e_1, a_2)$ $\pi_{1,F}^S, \quad \pi_{2,L}^S$
	Ta/Te	$(a_1, e_1), (a_2, e_2)$ $\pi_{1,L}^S, \quad \pi_{2,F}^S$	$(a_1), (e_1, a_2, e_2)$ $\pi_{1,L}^A, \quad \pi_{2,F}^A$	$(e_1), (a_1, a_2, e_2)$ $\pi_{1,L}^S, \quad \pi_{2,F}^S$	(a_1, e_1, a_2, e_2) $\pi_1^N, \quad \pi_2^N$

Table 1: Payoff matrix of stage 1 and resulting game structure for stage 2 and stage 3. Variable(s) in the first brackets are chosen in stage 2, those in the second brackets (if applicable) in stage 3 .

Stage 1 is represented as follows. A country selects the strategy Ta if it is willing to take unilateral action in terms of adaptation (Take adaptation), and Te if it is willing to take unilateral action in terms of emissions (Take emissions). If a country is willing to accept unilateral action by the other country in terms of adaptation (Accept adaptation), it selects Aa. It selects Ae if it accepts unilateral action in terms of emissions (Accept emissions). The strategies for emissions and adaptation can be combined in all possible ways, so that each country i selects its strategy u_i from a set with four options. If the strategy profile is such that one country takes unilateral action in a specific type of variable (adaptation or emissions), and the other country accepts this, the first becomes the Stackelberg leader with respect to this variable in the subsequent subgame. If no country is willing to accept unilateral action in one variable type, or if no country wants to take unilateral action for one variable type, both countries fall back to select this variable simultaneously in the subsequent stage (see Tab. 1). The complete game can be solved by backward induction, where the solution of stages 2 and 3 are already determined in the previous section. Note that this analysis requires additional indices in order to distinguish cases where a specific country $i \in \{1, 2\}$ is a follower or leader, respectively. Thus $\pi^S_{i,L}$ denotes the payoff for country i if it is the Stackelberg leader in both adaptation and emissions, $\pi^A_{i,F}$ the payoff for country i if it is the follower in the game with leadership in adaptation only, etc.

The game can be solved for different settings. We first concentrate on the most interesting case with heterogeneous country types. We assume (without loss of generality) that country 1 is of α-type, and country 2 of β-type. Then, Prop. 2 and Prop. 3 show that $\pi^S_{1,L} > \pi^A_{1,L} > \pi^N_1 > \pi^S_{1,F}, \pi^A_{1,F}$ and $\pi^S_{2,F}, \pi^A_{2,F} > \pi^N_2$. Note that some ambiguity remains, in particular on $\pi^S_{2,F} \lesseqgtr \pi^A_{2,F}$, which can shown to be equivalent to $e^S \gtreqless e^A$. These relations support the following equilibria.

Proposition 4. *Let country 1 be of α-type, and country 2 of β-type. The equilibria of the 3-stage game are as follows*

1. *If $e^S < e^A$, then the stage 1 Nash equilibria are*

$$(u_1, u_2) \in \{(Aa/Te, Aa/Ae), (Aa/Te, Ta/Ae),$$
$$(Ta/Te, Aa/Ae), (Ta/Te, Ta/Ae)\}.$$

These equilibria have identical payoffs $\pi^S_{1,L}, \pi^S_{2,F}$, with Stackelberg leadership in emissions, and possibly also in adaptation.

2. *If $e^A < e^S$ and $\pi^S_{2,F} < \pi^A_{2,L}$, then the stage 1 Nash equilibrium is $(u_1, u_2) = (Ta/Te, Aa/Te)$. Country 1 is the leader in adaptation only, and the payoffs are $\pi^A_{1,L}, \pi^A_{2,F}$.*

3. *If* $e^A < e^S$ *and* $\pi^S_{2,F} > \pi^A_{2,L}$*, then the stage 1 Nash equilibria are* $(u_1, u_2) \in \{(Aa/Te, Aa/Ae), (Aa/Te, Ta/Ae), (Ta/Te, Aa/Te)\}$*. Country 1 is either the leader in adaptation only or in emissions. The payoffs are either* $\pi^A_{1,L}, \pi^A_{2,F}$ *or* $\pi^S_{1,L}, \pi^S_{2,F}$*.*

In all possible equilibria the α*-type country becomes the Stackelberg leader in at least one variable, while the* β*-Type country is always the follower. In comparison to the non-cooperative Nash solution, all equilibria are Pareto-superior with less adaptation expenditures and less emissions.*

This result rests on how the Stackelberg equilibria (in stage 2 and 3) change payoffs in comparison to the non-cooperative Nash equilibrium (in stage 2). For leadership in both adaptation and emissions, Prop. 2 has shown that the outcome depends on the type of the Stackelberg follower. Due to Prop. 3, this basically also holds for leadership in adaptation only. All other game structures can be reduced to one of these cases. When the follower is an α-type country, the leader expands adaptation expenditures or emissions to improve its payoff at the expense of the follower. If, in contrast, the follower is of β-type, the leader reduces adaptation expenditures or emissions, such that the follower reduces emissions as well. So, both countries improve their payoff compared to the Nash equilibrium. In this case the β-type country is willing to accept that the α-type country takes the lead (in adaptation or emissions). If, in contrast, the β-type country would lead, it can anticipate that the following α-type country would react to reduced adaptation expenditures or emissions with higher emissions.

This is one positive result of this paper. In the 3-stage game equilibrium, adaptation expenditures and total emissions come closer to the social optimum. This improves the situation for both countries. If it is not determined from the outset whether the “right” country takes the lead in adaptation or emissions, the equilibrium of the multi-stage game will fortunately lead to a configuration with Pareto improvement. When there are heterogeneous countries, the α-type country undertakes unilateral action to the benefit of all.

Now turn to the case where both countries are of α-type. Prop. 2 and Prop. 3 show that $\pi_L^S > \pi_L^A > \pi_L^N$, $\pi_F^S, \pi_F^A < \pi_F^N$. Inspection of Tab. 1 then yields:

Proposition 5. *If both countries are of α-type, then both play (Ta/Te) in the game equilibrium. This leads to the non-cooperative Nash solution.*

This shows that in a game without β-type countries, no country wants to submit to a follower position, neither in terms of emissions, nor in terms of adaptation. If there is a Stackelberg leader, it would always exploit its advantage by increasing its emissions or by increasing its adaptation expenditures – at the expense of the other country. Thus, every country wants to become the Stackelberg leader, leading to a kind of "draw" in stage 1. So, the "lowest common denominator" is the non-cooperative Nash solution. This proposition also shows that the political fear that unilateral adaptation might impede international mitigation efforts needs to be qualified. If there are only α-type countries, this is only a theoretical option that would not evolve as a game equilibrium. If there are countries of both types, unilateral adaptation would even improve global cooperation.

To complete, turn to the case of both countries being of β-type. Prop. 2 and Prop. 3 show that $\pi_L^S > \pi_L^A > \pi_L^N$, and $\pi_F^S > \pi_F^N$, so that inspection of Tab. 1 yields:

Proposition 6. *If both countries are of β-type, then the non-cooperative Nash-solution cannot be the game equilibrium. If, additionally, at least one of the conditions (i) $\pi_{1,F}^S > \pi_{1,F}^A$, $\pi_{2,F}^S > \pi_{2,F}^A$, or (ii) $\pi_{1,F}^S > \pi_{1,L}^A$, $\pi_{2,F}^S > \pi_{2,L}^A$, holds, all equilibria are determined by a Stackelberg game with leadership in emissions (and possibly also adaptation). These equilibria lead to a Pareto improvement compared to the non-cooperative Nash solution.*

This case is less conclusive than the others since Eq. (43) has an ambiguous sign. Without the additional conditions, either Stackelberg leadership in emissions or in adaptation can be the outcome in stage 2. It is also not possible to say which country will take unilateral action. Every country knows that the other will react to unilateral action in a way that benefits all countries. So, both countries prefer any kind of leadership to none. Under the additional conditions, Prop. 2 further implies that any equilibrium is associated with less total emissions and adaptation expenditures than the non-cooperative Nash solution.

2.5 Conclusions

This paper has analysed whether the option of adaptation – in addition to mitigation – improves the prospects for contributing to a global public good. Our model depicts unilateral action as a Stackelberg game with two countries. We assume a quite general class of extended damage functions that explicitly represent adaptation, and compare different Stackelberg equilibria (unilateral adaptation and/or mitigaton) with the non-cooperative Nash solution. In a further step, we determine the type of Stackelberg leadership endogenously by adding a further game stage before the emission and adaptation decisions.

We find that the Stackelberg leader unilaterally reduces adaptation expenditures and emissions below its non-cooperative Nash level if the extended damage function of the follower has a specific convexity property (denoted as β-type country). We show that this property is equivalent to a concave optimised damage function. Thus, strategic complementary of emissions or of adaptation can be brought about by countries having the option of adaptation. While their marginal damage is increasing if they do not adapt, marginal damage is decreasing if countries adapt optimally. Then, all Stackelberg equilibria are Pareto-superior to the non-cooperative Nash solution. Total emissions and adaptation expenditures are reduced. This raises the question whether β-type countries would indeed be the followers. The result is a positive one: If the role of the leader and the follower is determined endogenously, α-type countries (being equivalent to a country with convex optimised damage function) are willing to take leadership in adaptation or emissions. Then, β-type countries have no incentive to unilaterally deviate from their role as a follower. Thus, if there is at least one β-type country, Stackelberg leadership is endogenously determined such that emissions and adaptation expenditures are reduced to the benefit of all. In sum, we find cases where unilateral action in adaptation or mitigation (i) is individually rational, (ii) reduces emissions and adaptation expenditures (below the non-cooperative Nash solution), and (iii) is Pareto-improving.

Some climate policy implications can be derived from our results. First, unilateral adaptation and unilateral emission reductions can be rational and supportive to climate protection at the same time. Second, it is not necessarily beneficial if all countries commit to emission reductions and adaptation at the same time. The outcome of climate negotiations can improve, if negotiations prior to the countries' announcement of mitigation and adaptation decisions do stipulate the sequence of these announcements. If a first group of countries is willing to announce its adaptation or mitigation strategy first, while a second group of countries is willing to wait and accept the announcements of the first group, then a Pareto improvement may be obtained. Third, in contrast to some policy advocates and earlier work (Zehaie, 2009), there seem to be little problems with unilateral adaptation. Either, unilateral adaptation leads to emission reductions (if there are countries of both types), or unilateral adaptation would not emerge as a game equilibrium (if all countries have convex optimised damage functions).

The results of this paper are theoretical in nature, and thus only provide a first step in the direction of further analyses. We are not aware of empirical studies that indicate the existence of β-type countries for international pollution problems. Yet, in the field of climate change, the empirical base for damage functions is still very much evolving (e.g., Tol, 2005; Watkiss, 2011), so that it is currently difficult to make a case for or against β-type countries. In general, however, (partially) concave damage functions are not im-plausible, e.g., when damages come close to a maximum level (cf. Baumol, 1972). Further explanations for β-type countries to exist are provided above. It is interesting to observe that some (implicitly optimised) damage functions used in the literature on the integrated assessment of climate change (e.g., Nordhaus, 1993; Nordhaus and Boyer, 2000; Warren et al., 2006; Ortiz et al., 2011; Marrouch and Chaudhuri, 2011) are concave in parts of their domain, although the authors do not relate this to adaptation, or do not exploit this property explicitly. We suspect that this does not lead to problems with these models when their computed equilibria are in the convex parts of the domains of their damage functions. To our knowledge, only the analysis of Heugues (2012a,b) explicitly starts from the assumption of strategic complements in global pollution games. Finally, we build a bridge between analyses that take regard of adaptation on the one hand, and the established literature without adaptation on the other hand. The proposition that links the extended to the optimised damage function and the Nash equilibria is crucial for integrated assessment modeling of climate change and will be helpful for further research into the economics of adaptation.

Giving consideration to adaptation thus leads to a case for strategic complements and the importance of unilateral action. We think that strategic complements in the provision of public goods and international environmental agreements are a research direction that deserves further attention. The current analysis focuses on the case with two countries. The natural next step it to extend the analysis to N heterogeneous countries (in particular, when there are countries of both types as defined and analysed in this paper). Then, more realistic settings and matters of coalition stability can be studied. Nevertheless, this paper has already shown at least one new reason that gives cause for greater optimism about solving international environmental problems.

Appendix to 'Adaptation to climate change can support unilateral emission reductions'

2.A Proof of Prop. 1 (Extended and optimised damage function)

Proof. The equilibrium $(a_1^N, e_1^N, a_2^N, e_2^N)$ is characterised by the first-order conditions $\forall i : B_i'(e_i^N) = \partial_e D_i(e^N, a_i^N)$, $a_i^N = A_i(e^N)$. This is equivalent, by Eq. (16), to $\forall i : B_i' = \tilde{D}_i'$. The latter, however, is the first-order condition for the Nash equilibrium $(\tilde{e}_1^N, \tilde{e}_2^N)$ in the standard case without adaptation. The solutions of both games are identical. Differentiating Eq. (16) further implies that

$$\tilde{D}_i'' = \partial_{ee} D_i + \partial_{ea} D_i A_i'.$$

Substituing Eq. (14) and comparing with Eq. (3) then yields $\tilde{D}_i'' = \nu_i$. □

2.B Proof of Prop. 2 (Stackelberg leadership in emissions and adaptation)

Proof. We first characterise the necessary conditions for an (interior) Stackelberg equilibrium. This also establishes the ordinal relations between the Nash and Stackelberg equilibrium. We then turn to the sufficiency of Eq. (19).

Country F reacts in the same way as in the Nash game with $e_F = \tilde{R}_F(e_L), a_F = A_F(e_L + e_F)$. Country L anticipates this reaction of the follower. The derivative of Eq. (18) with respect to a_L yields the first-order condition $-\partial_a D_L - 1 = 0$, so that the same optimal adaptation response function $a_L = A_L(e)$ as in the Nash game is implied, namely Eq. (20).

Substituting this into Eq. (18) leads to

$$\pi_L = B_L(e_L) - D_L(e_L + \tilde{R}_F(e_L), A_L(e_L + \tilde{R}_F(e_L))) + A_L(e_L + \tilde{R}_F(e_L)) \tag{32}$$

$$= B_L(e_L) - \tilde{D}_L(e_L + \tilde{R}_F(e_L)), \tag{33}$$

and differentiating with respect to e_L leads to the first-order condition Eq. (19), which needs to hold with necessity for an interior solution:

$$B'_L = (1 + \tilde{R}'_F)\, \tilde{D}'_L. \tag{34}$$

To prepare the comparison of the equilibria, note that Prop. 1 entails that $\pi_F = B_F(e_F) - \tilde{D}_F(e_F + e_L)$, which yields

$$\frac{d\pi_F}{de_L} = -\tilde{D}'_F < 0. \tag{35}$$

Also recall from Prop. 1 that the non-cooperative Nash equilibrium is characterised by

$$B'_L(e_L) = \tilde{D}'_L. \tag{36}$$

Now turn to the two cases. If the follower is an α-type country, $\tilde{R}'_F < 0$, so that the right-hand side of Eq. (34) is smaller than that of Eq. (36). Thus, since B'_L is strictly decreasing, it must hold that $e_L^S > e_L^N$. As the follower country has a downward-sloping reaction function, it reduces its emissions compared to the Nash equilibrium. As also $-1 < \tilde{R}'_F$, these reductions are smaller than the leaders' additional emissions, so the total emissions are higher for the Stackelberg equilibrium. As A_L, A_F are upward sloping, this leads to higher adaptation expenditures for both countries in the Stackelberg equilibrium. Finally, $\pi_F^S < \pi_F^N$ due to Eq. (35) and $e_L^S > e_L^N$.

In contrast, if the follower is a β-type country, $\tilde{R}'_F > 0$ implies that the right-hand side of Eq. (34) becomes larger than that of Eq. (36). Thus, in contrast to the other case, $e_L^S < e_L^N$, so that Eq. (35) leads to $\pi_F^S > \pi_F^N$. Now, the follower country has an upward-sloping reaction function, so that it reduces emissions below the Nash equilibrium likewise (which leads to the same outcome as in the first case). Total emissions are consequently lower in the Stackelberg equilibrium, and so are the adaptation expenditures in both countries (since $A'_L, A'_F > 0$).

Finally, turn to the sufficiency of the first-order conditions. First note that Eq. (20) indeed optimises payoffs for any given level of total emissions e, since $\frac{d^2}{da_L^2}\pi_L = -\partial_{aa} D_L < 0$. Second, evaluate Eq. (34) at the Nash equilibrium (with $B'_L = \tilde{D}'_L$) to obtain

$$\frac{d}{de_L}\pi_L = B'_L - \tilde{D}'_L - \tilde{R}'_F \tilde{D}'_L = -\tilde{R}'_F \tilde{D}'_L. \tag{37}$$

Since $\tilde{D}'_L > 0$, the local change of payoff in the Nash equilibrium has the opposite sign of $\tilde{R}'_F$. If the follower is an α-type country (such that $e_L^S > e_L^N$), increasing emissions improves the leader's payoff at least locally. This guarantees the existence of an optimum, as–due to continuity–payoff only ceases to increase further if ultimately e_L^S is reached.

If there are multiple vectors that satisfy the first-order conditions, then one of them must describe the optimum. If the follower is a β-type country, the analogue argument can be made if emissions e_L are reduced below e_L^N. Thus, a decision a_L^S, e_L^S from the first-order conditions indeed optimises (and strictly improves) payoff for the Stackelberg leader in both cases, $\pi_L^S > \pi_L^N$. □

2.C Proof of Prop. 3 (Stackelberg leadership in adaptation)

Proof. By backward induction, we first determine how the equilibrium of the second stage depends on the adaptation decision in the first stage. The total differential of Eq. (23) with respect to e_L^A and a_L yields

$$\frac{de_L^A}{da_L} = \frac{\partial_a R_L}{1 - \partial_e R_L \tilde{R}'_F} > 0. \tag{38}$$

The denominator is always positive: If the follower is a β-country, $\tilde{R}'_F > 0$ by Eq. (9), and generally $\partial_e R_L < 0$ due to Eq. (12). If the follower is an α-country, then $\tilde{R}'_F \in (-1, 0)$ according to Eq. (9), and generally $\partial_e R_L \in (-1, 0)$ due to Eq. (12). The numerator is positive due to Eq. (13). Thus, since $e_F = \tilde{R}_F(e_L)$,

$$\frac{de_F^A}{da_L} = \frac{\partial_a R_L \tilde{R}'_F}{1 - \partial_e R_L \tilde{R}'_F} \begin{cases} > 0 & \text{if } \beta\text{-type follower,} \\ < 0 & \text{if } \alpha\text{-type follower.} \end{cases} \tag{39}$$

Then, Eq. (38) together with Eq. (39) imply that

$$\frac{de^A}{da_L} = \frac{\partial_a R_L(\tilde{R}'_F + 1)}{1 - \partial_e R_L \tilde{R}'_F} > 0. \tag{40}$$

The numerator is positive due to Eq. (13) and Eq. (9).

Now turn to the first stage. Note that the leader's optimised damage function $\tilde{D}_L$ plays no role in the leader's optimisation problem Eq. (26). The first-order condition evaluates to

$$0 = (B'_L - \partial_e D_L)\frac{de_L^A}{da_L} - \partial_e D_L \frac{de_F^A}{da_L} - (\partial_a D_L + 1). \tag{41}$$

The first term vanishes due to $e_L^A = R_L$, so that an interior solution of the first stage is characterised by

$$1 + \partial_a D_L = -\partial_e D_L \frac{de_F^A}{da_L}. \tag{42}$$

This characterisation allows for comparing the game equilibrium $(a_L^A, e_L^A, a_F^A, e_F^A)$ with the non-cooperative Nash solution $(a_L^N, e_L^N, a_F^N, e_F^N)$. Note that the right-hand side of Eq. (42) is positive if the follower is an α-type country (due to Eq. 39), and negative for a β-type follower. How about the left-hand side?

We know from Eq. (11), that the left-hand side vanishes in the Nash-equilibrium

(a_L^N, e_L^N, e_F^N). It can further be verified that

$$\frac{d}{da_L}\partial_a D_L(e^A(a_L), a_L) = \partial_{aa} D_L + \partial_{ae} D_L \cdot \left(\frac{(\tilde{R}_F' + 1)\partial_a R_L}{1 - \partial_e R_L \tilde{R}_F'} \right)$$
$$= \frac{\partial_{aa} D_L (B_L'' - (1 + \tilde{R}_F') \cdot \nu_L)}{B_L'' - (1 + \tilde{R}_F')\partial_{ee} D_L}. \tag{43}$$

This expression is positive if there is a least one α-type country due to the following argument. The denominator is always negative since $1 + \tilde{R}_F' > 0$. In the numerator, $\partial_{aa} D_L > 0$ by assumption. Now consider the second term in the numerator. If the leader is of α-type, then $(1 + \tilde{R}_F') \cdot \nu_L > 0$, so that the whole expression becomes positive. If the leader is of β-type, and the follower of α-type, note that in $B_L'' - (1 + \tilde{R}_F') \cdot \nu_L = B_L'' - \nu_L - \tilde{R}_F' \nu_L$, the first difference is always negative due to the convexity assumption Eq. (4). As the last term is positive for these country types, the whole expression also becomes positive.

The sign of Eq. (43), together with a vanishing $1 + \partial_a D_L$ in the Nash equilibrium then implies the following. If the follower is an α-type country, Eq. (42) can only be positive if $a_L^A > a_L^N$. If the follower is a β-type country and the leader an α-type country, then $a_L^A < a_L^N$. If both the follower and the leader are of β-type, nothing can be said here.

Emissions then compare between the Nash solution and Stackelberg leadership in adaptation only according to Eq. (40) and Eq. (14): If the follower is an α-type country, then $e^A > e^N$ and $a_F^A > a_F^N$. If the follower is a β-type country and the leader an α-type country, then $e^A < e^N$ and $a_F^A < a_F^N$.

We can further conclude that $\pi_L^A > \pi_L^N$ strictly holds. It is clear that the leaders' payoff can not decrease below the Nash solution, as the leader has one additional degree of freedom in the first stage of the game. By evaluating how the leaders' payoff in the Stackelberg game changes at a_L^N, we obtain (cf. Eq. 42):

$$\frac{d\pi_L}{da_L} = -\partial_e D_L \frac{de_F^A}{da_L} \neq 0. \tag{44}$$

This implies that a_L^N cannot optimise the payoff, so that $\pi_L^A = \pi_L^N$ is ruled out.

Now, compare π_F^A with π_F^N. Differentiation shows that

$$\frac{d\pi_F}{da_L} = \frac{d}{da_L} B_F(e_F^A(a_L)) - \tilde{D}_F(e_L^A(a_L) + e_F^A(a_L)) \tag{45}$$

$$= (B_F' - \tilde{D}_F')\frac{de_F^A}{da_L^A} - \tilde{D}_F' \frac{de_L^A}{da_L} \tag{46}$$

$$= -\tilde{D}_F' \frac{de_L^A}{da_L} < 0, \tag{47}$$

due to Prop. 1 and Eq. (38). Thus, if the follower is an α-type country, $a_L^A > a_L^N$ implies $\pi_F^A < \pi_F^N$. If the follower is a β-type country, and the leader of α-type, $\pi_F^A > \pi_F^N$.

We finally compare country L's payoff with leadership in adaptation only to that of Stackelberg leadership in both emissions and adaptation. Obviously, country L has one degree of freedom less to chose from in the former setting, so that $\pi_L^A \leq \pi_L^S$. Furthermore, with leadership in adaptation only, $B_L'(e_L^A) = \partial_e D_L(e_L^A + e_F^A, a_L)$ holds. If country L can also dispose over emissions in the first stage, the first-order condition for this choice is $0 = B_L'(e_L) - (1 + \tilde{R}_F')\partial_e D_L(e_L + e_F, a_L)$. Since $\tilde{R}_F' \neq 0$, it is clear that (a_L^A, e_L^A, e_F^A) cannot maximise π_L with respect to e_L, so that strictly $\pi_L^A < \pi_L^S$. □

Interlude I

So far we have shown that adaptation to climate change can offer strategic options which are universally beneficial in a two player game. Specifically there is the possibility of a concave effective damage function. A country with this property has a positively-sloped reaction function (in emissions). It will therefore not freeride on unilateral emissions reductions by the other country but instead quasi-cooperate (i.e. react with emission reductions of its own).

If this is realised, the country or region in question will endogenously choose to be a Stackelberg follower rather than risk getting stuck in a Nash equilibrium. This is a great improvement by itself (compared to the standard case presented by Hoel (1991)) because usually no player wants to give the other the advantage of Stackelberg leadership. With no player conceding the first mover advantage to the other the usual outcome is a Nash equilibrium. We show that for a country with a concave effective damage function the Stackelberg follower position is preferable to an 'equalised' position in a Nash game (though not necessarily to a Stackelberg leader position). Therefore a Stackelberg game is chosen endogenously. The resulting emissions are lower than in a Nash game and welfare is higher for both countries.

Adaptation is an explicit decision variable in this first model with adaptation costs as a unit for measurement. The order of decisions between the two players is endogenous in both emissions and adaptation. Therefore the results are very general. The analysis shows that perfectly rational adaptation can be integrated into the damage function (both the costs of adaptation and the damage-reducing effect at once) which leads to an effective damage function which simplifies further analysis.

These results of the first model are very promising but they do not account for strategic choices of more than two players so far. I have built several models in GAMS to study coalition stability in a world where some countries have concave damage functions (so called β countries). Without transfer payments between countries I have found no stable agreements including β countries. Depending on parameters however the existence of β countries can lead to the stability of a coalition that contains all α countries.

If transfer payments are included a coalition that contains both α countries and β countries can be stable, up to and including a grand coalition of all countries. I have found that coalitions with very little or very many α countries can contain β countries and still be stable, but that coalitions with about half the existing α countries are not stable if they include β countries even allowing for transfer payments.

I have also studied the reaction function of an aggregate of countries (that play a Nash subgame among themselves). In particular I have analysed an aggregate of both α and β countries. The idea is that such a group of countries acts non-cooperatively among themselves but is seen as a quasi-homogeneous entity by a coalition that may act as a Stackelberg leader. The agreement members are only interested in the aggregated response of the non-members to any emissions changes of the coalition.

I have kept the functional forms very abstract in order to achieve maximum generality. The slope of the reaction function of such a mixed aggregate depends on the slopes of the reaction functions of the individual countries, of course. If all countries in the aggregate are of the same type, the slope of the reaction function of the aggregate has the same algebraic sign as the slopes of the individual reaction functions. This has been well-known for α countries (i.e. conventional countries) but has not been tested for β countries before to my knowledge.

If one of the countries in question changes its individual reaction function towards a more positive slope, the slope of the reaction function of the aggregate (in a Nash subgame equilibrium) becomes more positive, too. (And vice versa: if an individual country changes its reactions function to be more negatively sloped, the slope of the reaction function of the aggregate becomes more negative.)

The absolute value of the reaction function of the aggregate is also important. If the reaction function of one country changes in absolute terms, the reaction function of the aggregate changes, too. The change in the aggregate is larger than the original change if the aggregate of all countries except the one with the original shift has a negatively sloped-reaction function. The increase of the aggregate reaction function is larger than the original increase if the reaction function of the aggregate of the other countries is positively-sloped.

Of course this shall not imply that changing the slope of the reaction function is something that can be achieved intentionally by human action. It rather means that the discovery of regional features that make emissions more strategically complementing for an individual country shifts the perception of the slope of the reaction function of the aggregate in the same direction.

Obviously coalitions stability heavily depends on parameter values. My GAMS models were not calibrated but I checked for a wide range of parameter options. The analytical studies of reaction functions have given me a good insight into Nash subgame equilibria of mixed aggregates. Nonetheless I wanted to understand the analytic requirements for coalition stability.

This is the main idea of the next part: The 'Island of Stability' model takes the concave effective damage functions of the first part and the Stackelberg follower choice of the players that have this (so called β) property and analyses the results in a coalition game.

The model is more specific and simplified in comparison to the first one. It has a linear benefit function and determines the Stackelberg follower role of β countries exogenously. The model is then sufficiently tractable to be solved for coalition stability. The interesting result shows that the existence of β countries can lead to dual equilibria. One of these equilibria is a small coalition well known in literature (e.g. Barrett, 1994). The other equilibrium however is larger and leads to lower global emissions and higher global welfare. This is particularly interesting because larger coalitions (if possible at all) in the standard literature have been shown to be ’shallow’, i.e. yield very little advantage over the non-cooperative outcome in terms of emission reductions and welfare gains.

3 Strategic Complements in International Environmental Agreements: a New Island of Stability

Abstract

International environmental agreements have had varying success in the past; the theoretical literature on international environmental agreements (IEAs) explains why freeriding is so common. This paper allows for two strategically different types of countries. Damage functions are concave for some countries (contrary to the standard convexity assumption). This leads to strategic substitutes and complements in emissions reduction within the same model. The interaction of both country types can lead to a stable agreement that is larger than in the standard case, and to more global abatement. Such a stable agreement constitutes an island of stability in addition to the small standard agreement.

3.1 Introduction

International environmental agreements (IEAs) suffer from the well-known freerider problem: Countries which are not committed to membership of an agreement have very low incentives to reduce emissions if the members of an agreement do so. In light of the fundamental implications of climate change, this pessimistic analysis is yet of little help. What alternatives can be offered to improve prospects for reaching an agreement on global emission reductions?

The root of this pessimism lies in countries' emissions being strategic substitutes, i.e. if one country reduces emissions, the other countries respond by expanding them. Established models of global emissions games assume strategic substitutes in emissions, both in the theoretical (e.g. in the seminal work of Hoel, 1991; Carraro and Siniscalco, 1993;

Barrett, 1994) and in the simulation literature (e.g. Bosetti et al., 2006; Nagashima et al., 2009; Lessmann et al., 2009). This assumption partially drives the common trade-off between broad-but-shallow and deep-but-small IEAs. However, some authors have shown that there are good theoretical or empirical reasons that emissions can also be strategic complements for some countries, e.g. due to technological spillovers, trade, or adaptation to climate change (Fredriksson and Millimet, 2002; Copeland and Taylor, 2005; Ebert and Welsch, 2009; Eisenack and Kähler, 2015). It is thus important to know whether strategic complementarity might ease the provision of a global public good in a self-enforcing IEA. Only few studies up to date analyse the implications of strategic complements in global emissions games. Ebert and Welsch (2009); Eisenack and Kähler (2015); Heugues (2012a) consider the case of two countries with quite general classes of damage functions. While the latter two endogenise the sequence of play in a Stackelberg setting, the former two consider adaptation to damages as a further decision variable. For the n country case, Heugues (2012b) determines stable agreements for specifically parametrized damage functions. In her setting, all countries' emissions are strategic complements. Our paper determines stable agreements for the n country case with heterogeneous country types and quite general damage functions: while some countries' emissions are strategic complements, others' are substitutes. We also explore how game equilibria depend on the number of countries of each type.

We analyse a three-stage game with emissions of countries as decision variables. Countries can have either convex or concave damage functions, which is tied to strategic substitutes or complements (Heugues, 2012a; Eisenack and Kähler, 2015). In the first stage, countries decide about being members of an agreement. In the second stage, the agreement jointly acts as a Stackelberg leader by maximizing their sum of payoffs. In the third stage, followers of all country types play a simultaneous Nash game for the given agreement structure and the the agreement's emissions. The paper analyses how the agreement's maximization problem depends on the number and type of the non-members.

It is shown that non-members with a convex damage function react as free-riders, while non-members with a concave damage function emit less in the equilibrium if the agreement reduces emissions sufficiently. We find that, independently of the number of countries with strategic complements, the usual small agreements remain stable. However, we also find than an additional range of stable agreement sizes exists under reasonable conditions. These agreements are larger, have lower total emissions, and are Pareto-superior to the usual stable agreements. They are yet not much larger than the usual agreements. We call such agreements 'islands of stability' since their size range can be disconnected from the usual range of stable agreement sizes, and since the range is small.

We first introduce the game structure with n countries of two types in section 3.2. In section 3.3 the model is solved by backward induction. Finally we discuss the findings with a view on parameter influence in section 3.4 and conclude with a summary of results and an outline of further steps to understand agreement stability. The appendix contains the proofs.

3.2 Model Structure

This paper determines stable international environmental agreements of multiple countries that deal with a public bad. In the absence of a supranational agency that can enforce a first-best level of mitigation, the agreement has to be self-enforcing and will typically not include all countries. In this section, the variables, basic assumptions and the game structure are introduced.

3.2.1 Variables and Assumptions

The model considers n countries, each denoted by subscript i. Countries choose their own emissions $e_i \in [0, 1]$, i.e. we assume that per country emissions have an upper bound due to capacity constraints. Aggregate emissions by all countries except i are denoted by e_{-i}, so that total emissions are $e = e_i + e_{-i}$.

Emissions are assumed to be a substitutable input for production, that at the same time generate increasing damages D_i, depending on the global emissions level. Therefore mitigation of emissions is a public good. We assume countries' payoff-functions of the form $\pi_i = b \cdot e_i - D_i(e)$ with $D_i(0) = 0, D_i' > 0$. Damages are non-linear to account for strategic substitutes and complements as will become clear below. While our assumptions about damages are rather general, benefits $b \cdot e_i$ are restricted to the linear case in order to keep the analysis tractable (cf. Asheim et al., 2006; Barrett, 1999, 2001). More generalization requires future work.

There are two types of countries, α and β, which differ in the properties of their damage functions. All countries of the same type are identical:

$$\alpha \text{ countries: } D_i'' > 0, D_i' < b, \tag{48}$$

$$\beta \text{ countries: } D_i'' < 0. \tag{49}$$

This means that α countries have convex damage functions whereas β countries have concave damage functions. The former are those countries that are conventionally considered in the literature on international environmental agreements and the integrated assessment of climate change. The latter type of countries, being less conventional, lead to strategic complements, as has been investigated for other settings by Ebert and Welsch (2012); Heugues (2012a); Eisenack and Kähler (2015), who also discuss possible reasons for β countries to exist.[6] Eq. (48) further implies that there is no incentive for a single

---

[6] There are at least two possible reasons why a country could be of the β type. First it could be that the

α country to reduce emissions – a common assumption to focus the analysis on the interesting case of dominant freeriding incentives.

The number of α countries that are members of an agreement is denoted by $x \geq 0$, and those α countries that are not members by $y \geq 0$. The total number of β countries is $z \geq 0$, so that $x + y + z = n$. In our notation, aggregate emission of all countries belonging to a group g are accordingly denoted by e_g, while $e_{-g} = e - e_g$ denotes the aggregate emissions of all countries not belonging to that group.

In order to focus our analysis, we further impose for all α countries i the assumption:

$$\forall x, y \quad \exists e < n: \quad (x + y) \cdot D_i'(e) = b, \tag{50}$$

i.e. if all α countries would optimize their joint payoff, it is profitable to abate at least a little. Together with Eq. (48), this ensures that cooperation can yield gains, but unilateral action from single α country is never individually rational. Without these assumptions, we would also investigate uninteresting cases.

3.2.2 Game Structure

We determine the subgame perfect equilibrium of a three stage game. In the first stage (A) an agreement can be formed. Each α country anticipates the outcomes of the subsequent stages and choose individually whether it joins the agreement or not. The common solution concept we employ at this stage is internal and external stability (D'Aspremont et al., 1983). We assume that β countries do not become members. While the main reason

---

efficiency of adaptation (which in our model is implicit in the damage function) is connected to the emission level in a way that leads to nonconvexity in the damage function, in particular higher efficiency of adaptation at lower global emissions levels. Second the damage function can be concave if the damages are already close to a maximum level (cf. Baumol, 1972), for example in small island states or more generally countries with a lot of valuable infrastructure close to the coastline.

for this is tractability, our numerical experiments with allowing β countries to become members have shown that game equilibria do not substantially change.[7] In the second stage (B) the agreement with x members chooses the emissions of its members in order to optimize its joint payoff. In the third stage (C) the non-members (y α and z β countries) choose their emissions simultaneously.

Thus, the agreement acts as a Stackelberg leader committing to its emissions first, then the non-members play a Nash subgame. We thus follow the common rationale of Barrett (1994), and not the equally common of Carraro and Siniscalco (1993), where all emission decisions are made simultaneously. With the latter rationale, strategic complementarity would not effect the game equilibrium (Eisenack and Kähler, 2015). The paper analyzes the stages in reverse order by backward induction.

3.3 Game Equilibria

3.3.1 Stage C: Emissions of Non-Members

First, determine the best response correspondence of each of the y non-member α countries. By individually maximizing their payoff $\pi_i = be_i - D_i(e_i + e_{-i})$ for given emissions of all other countries e_{-i}, Eq. (48) implies the corner solution $e_i = 1$. This is a dominant strategy. Thus, the α countries which are not members of the agreement emit $e_y = y$ in total, independent from the decisions of the members of the agreement and the non-member β countries.

Second, turn to the best response correspondence of each of the z non-member β countries. They also individually maximize their payoff $\pi_i(e_i, e_{-i}) = be_i - D_i(e_i + e_{-i})$. Note that for β countries, $\frac{d^2\pi_i}{de_i^2} = -D_i'' > 0$, so that the first-order-condition would not yield a

---

[7] We have found that over various model specifications and for a very large range of parameter values no nontrivial agreement with β country participation is stable without transfer payments.

payoff maximum. Accordingly, a non-member β country compares the corner solutions. Define $\Delta(e_{-i}) := \pi_i(1, e_{-i}) - \pi_i(0, e_{-i}) = b - D_i(1 + e_{-i}) + D_i(e_{-i})$. The sign of Δ then determines the reaction. Observe that $\Delta' = D_i'(1 + e_{-i}) + D_i'(e_{-i}) > 0$ due to Eq. (49). Thus, Δ has at most one zero, is negative to the left of $\tilde{e}$, and positive to the right of $\tilde{e}$. We assume here and in the following that there exists an $\tilde{e}$ so that

$$\Delta(\tilde{e}) = 0. \tag{51}$$

This yields the best response correspondence

$$e_i = \begin{cases} 0 & \text{if } e_{-i} < \tilde{e}, \\ 1 & \text{if } e_{-i} > \tilde{e}, \\ \{0, 1\} & \text{if } e_{-i} = \tilde{e}. \end{cases} \tag{52}$$

While the β country chooses a unique corner solution in the first two cases, it is indifferent between them in the third case. Note that this intermediate result can be characterized as a generalized notion of strategic complements (Bulow et al., 1985). While the original definition rests on a best response function with a positive derivative, we have a non-decreasing correspondence in our case.

Further note that the existence of $\tilde{e}$ is not implied by the other assumptions made so far. However, the cases where it does not exist are not very interesting for our further analysis: If Δ would be always always positive, β countries would dominantly play $e_i = 1$, so that they would not behave differently from non-member α countries. If Δ would be always negative, β countries would dominantly play $e_i = 0$, so that they can be ignored and the analysis would be reduced to the common case without β countries.

Finally, turn to the Nash equilibrium in stage (C). The objective is to determine the aggregate emissions $e_y + e_z$ when all non-member of the agreement simultaneously chose their emissions, given the emissions e_x of the agreement members, and the choices of all non-members. The situation is simple for the non-member α countries since they have dominant strategies.

The situation is more tricky for a β country i. If the total emission of the α countries $e_x + e_z$ are already larger than $\tilde{e}$, all β countries would chose $e_i = 1$. In contrast, if the total emission of the α countries $e_x + e_z$ are so small that even $e_x + e_z + e_y < \tilde{e}$, all β countries would chose $e_i = 0$. But what happens in the case where $e_x + e_z < \tilde{e}$, but the choice of the other β countries would make a difference whether $e_{-i} \lesseqgtr \tilde{e}$? What if $e_{-i} = \tilde{e}$? This is clarified by the following proposition.

Proposition 7. *Assume that Eq. (48) and Eq. (49) hold,* $z \geq 1$*, and that* $\tilde{e}$ *exists according to Eq. (51). Let* e_x *be the given emissions of the agreement members. Then, the only Nash equilibria of stage (C) are:*

$$\textit{if } e_x \leq \tilde{e} - y \textit{ then } \forall \beta \textit{ countries } i : e_i = 0, \textit{ and } e_z = 0, e = e_x + y, \tag{53}$$

$$\textit{if } e_x \geq \tilde{e} - y - z + 1 \textit{ then } \forall \beta \textit{ countries } i : e_i = 1, \textit{ and } e_z = z, e = e_x + y + z. \tag{54}$$

Note that the Nash equilibrium is not always unique. If $\tilde{e} - y - z + 1 \leq e_x \leq \tilde{e} - y$, which is equivalent to $\tilde{e} - z + 1 \leq e_x + y \leq \tilde{e}$, the β countries either symmetrically chose $e_i = 0$ or $e_i = 1$. If the emissions of all α countries have a medium size, both a low emissions and a high emissions outcome are possible in equilibrium. Once one of those strategy profiles is given, no β country has an incentive to deviate from that. Consequently, the proof strategy is to show that both strategy profiles are consistent with the best response correspondence of each country. Finally, the proof in appendix 3.A shows that there are no further consistent strategy profiles.

For the remainder of the paper, we ease analysis by resolving the ambiguity of equilibria in the proposition. The proposition's result can be understood as a "response correspondence" of the aggregate of non-member α countries and β countries (that play non-cooperatively). For intermediate levels of e_x, this correspondence has a two-valued image. We chose a non-decreasing selection from this correspondence as follows. Let $\hat{e} \in [\tilde{e} - y - z + 1, \tilde{e} - y]$. We then assume that the stage (C) equilibria

$$e_z = \begin{cases} 0 & \text{if } e_x + y \leq \hat{e}, \\ 1 & \text{if } e_x + y > \hat{e}, \end{cases} \tag{55}$$

realize. It is further reasonable to consider only those cases in the paper where

$$0 < \hat{e} < x + y. \tag{56}$$

If $\hat{e}$ lies outside of these bounds, the results would be trivial because the non-convexity property of the β countries would not have any impact on the game.

3.3.2 Benchmark Solution

The results so far allow to determine the non-cooperative Nash solution, as it corresponds to $x = e_x = 0$. This will help to discuss the results of the three stage game equilibrium. There is a unique Nash equilibrium. The α countries emit $e_i = 1$ each, together $e_y = y$, as always due to dominant strategies. Due to Eq. (56), $e_x + y = y > \hat{e}$. Thus, the β countries emit $e_z = 1$ each, too. Therefore, global emissions are $e = n$. There is no abatement in the non-cooperative Nash solution.

3.3.3 Stage B: Emissions of Agreement Members

The agreement of α countries maximizes the aggregated payoff of all members. To do so, they coordinate and choose emissions of each member. We simplify this and let the agreement directly choose their aggregated emissions e_x[8]. Together with Eq. (55), the

---

[8] Due to linear benefits of emissions it is not relevant here how emissions are distributed among agreement members, as long as the benefits are distributed evenly.

agreement's optimization problem thus reads

$$\max_{e_x} \quad \Pi_x = e_x \cdot b - x \cdot D_i(e_x + y + e_z) \tag{57}$$

$$s.t. \quad e_x \in [0, x], \tag{58}$$

$$e_z = \begin{cases} 0 & \text{if} \quad e_x + y \leq \hat{e} \\ z & \text{if} \quad e_x + y > \hat{e} \end{cases} \tag{59}$$

Recall that the damage function D_i is identical for all members of the agreement, and that it has the properties Eq. (48) and Eq. (50).

The first-order condition for an interior solution would evaluate to

$$\frac{b}{x} = D_i'(e_x + y + e_z). \tag{60}$$

In our particular situation however, the function $D_i(e_x + e_y + e_z)$ has a discontinuity at $e_x = \hat{e} - y$ because there e_z changes from 0 to z. This results in a more complicated solution of the agreement's optimization problem. The agreement chooses the emissions of its members according to Prop. 8, proof is in appendix 3.B.

Proposition 8. *Assume that $x > 0$, $y, z \geq 0$ and $\hat{e} < x + y$. Let i be an α country. Define $f(e, x) = be - xD_i(e)$. Let F be the solution of*

$$b \equiv xD_i'(F(x)). \tag{61}$$

Then, F is strictly decreasing in x

$$F'(x) < 0. \tag{62}$$

The unique stage B equilibrium is given by Tab. 2. If $bz = f(F, x) - f(\hat{e}, x)$, then both the cases 5 and 6 in Tab. 2 are game equilibria in stage B.

case no.	condition $\hat{e}$	condition $F(x)$	outcome e_x^*	e_z^*	e^*
1	$\hat{e} < y$	$x + y + z < F$	x	z	$x + y + z$
2		$y + z < F \leq x + y + z$	$F - y - z$	z	F
3		$F \leq y + z$	0	z	$y + z$
4	$y \leq \hat{e} < x + y$	$x + y + z < F$	x	z	$x + y + z$
5		$\hat{e} < F \leq x + y + z$			
		and $bz < f(F, x) - f(\hat{e}, x)$	$F - y - z$	z	F
6		$\hat{e} < F \leq x + y + z$			
		and $bz > f(F, x) - f(\hat{e}, x)$	$\hat{e} - y$	0	$\hat{e}$
7		$y < F \leq \hat{e}$	$F - y$	0	F
8		$F \leq y$	0	0	y

Table 2: Stage B game equilibria.

In cases 1 through 3 of Tab. 2, the agreement can not reduce emissions sufficiently so that the β countries abate as well. Of the other cases, 4 and 8 are corner solutions of no and full abatement, respectively. Case 5 is a 'normal' internal solution (as common in the IEA literature without strategic complements). In case 6, the agreement reduces emissions so that global emissions fall below $F(x)$ (which would be optimal in the absence of strategic complements). The reduction is just enough to induce a choice of $e_i = 0$ by the β countries. In case 7 the 'normal' internal solution (similar to case 3) for the agreement is low enough that the β countries choose $e_i = 0$.

For the remainder of the paper, we focus our considerations on the cases No. 4 through 7 of (Tab. 2). The other cases are either very similar (case 1 is similar to case 4 and case 2 to case 5) or corner solutions that are not particular to our analysis of non-convexities (cases 3 and 8).

3.3.4 Stage A: Agreement Size

In this section we analyse the endogenous choice of agreement size x.

Every agreement size x yields certain global emissions $e^*(x)$ as equilibrium of stages B and C. Knowing these emissions, every α country can compare its payoff within the agreement for the actual agreement size $\frac{b}{x} - D_i(e^*(x))$ with the payoff it would get if it left the agreement $b - D_i(e^*(x-1))$. The difference between these is the value of the outside option $\Omega(x)$.

$$\Omega(x) = b - D_i(e^*(x-1)) - \frac{b}{x} + D_i(e^*(x)) \tag{63}$$

We assume that every country has a positive value of the outside option $\Omega(x)$ as long as global emissions are $e = F(x)$. A positive outside option means that a member of the agreement increases its payoff if it leaves (i.e. becomes a non-member).

$$\begin{aligned} &\text{If} && e < n, \\ &\text{then} && \Omega(x) = b - D_i(F(x-1)) - \frac{b}{x} + D_i(F(x)) > 0. \end{aligned} \tag{64}$$

This assumption is a stronger version of Eq. (48) and gives the game the form of a prisoner's dilemma; it makes cooperation (i.e. $e_i = 0$) a dominated strategy for all α countries as long as any others cooperate and emissions are $e = F(x)$ (i.e. like in the absence of strategic complements).

There is always at least one stable agreement in the stage (C) equilibrium. For $z = 0$ it is unique, for $z > 0$ there can be a second stable agreement. The size of stable agreements is given by propositions 9 and 10, proof is in appendices 3.C and 3.D.

Proposition 9. *Suppose that assumptions Eq. (48), (50), (64) hold and* $x + y > 0$. *Then there exists a stable agreement with size* $x > 1$. *The smallest abating agreement* $\bar{x}$ *that* *chooses emissions $e_x < x$ in stage B is internally stable. If $z = 0$ then this smallest abating agreement $\bar{x}$ is also externally stable and its size is unique. In this case, global emissions are $e = F(x)$.*

This proposition describes a small agreement, which is a standard result in the case without strategic complements (as Diamantoudi and Sartzetakis (2006) have shown for a wide variation of variables). The smallest abating agreement $\bar{x}$ forms, but no more than the minimum number of countries required for this enters the agreement. The agreement's choice of emissions is an internal solution, and global emissions are lower than in the business as usual case without a agreement. The agreement size is unique if there are no β countries.

We now come the paper's main result: If β countries take part in the game, then there can be a second stable agreement size (see Prop. 10). If this larger stable agreement is one country larger than the smallest abating agreement then the larger one is stable and the smaller one is not.

Proposition 10. *Assume that Eq. (48), (49), (50), (64) hold, $x + y > 0$, $0 < \hat{e} < x + y$ and $z > 0$. If $D_i(F(\underline{x} - 1)) - D_i(\hat{e}) > b \cdot \left(1 - \frac{\hat{e}-y}{\underline{x}}\right)$ holds for any agreement size $\underline{x}$, then an agreement of this size $\underline{x}$ is stable. No other agreements except those of size $\underline{x}$ and size $\bar{x}$ (see Prop. 9) are stable. An agreement of size $\underline{x}$ leads to global emissions of $e = \hat{e} < F(x)$.*

As both the (larger) agreement size $\underline{x}$ and the (smaller) stable agreement size $\bar{x}$ are stable, the latter is what we call an additional 'island of stability'. (Exception: if $\underline{x} = \bar{x} + 1$ holds, i.e. if the agreement size $\underline{x}$ from Prop. 10 is exactly one country larger than the agreement size $\bar{x}$ from Prop. 9, then $\underline{x}$ is stable and $\bar{x}$ is not.) Without further assumptions it is impossible to tell which of the two possible agreement sizes will be realized. This larger agreement $\underline{x}$ given in Prop. 10 is the smallest one that supports global emissions of

$e = \hat{e} < F(\underline{x})$. Global emissions in this case are lower than for the other agreement that is shown in Prop. 9.

Welfare is also improved for every country compared to the standard case $\bar{x}$, so the larger agreement size $\underline{x}$ is a Pareto-improvement. This is obvious for non-member α countries because they simply gain from lower damages and enjoy the same benefits. β non-members also gain from lower emissions: they lower their emissions (from $e_i = 1$ to $e_i = 0$) because it gives them an additional benefit over the already beneficial situation of the mitigation effort from the agreement (which lowers their damage in absolute terms even though it increases their marginal damage). The agreement members have gains from cooperation. In particular, global emissions are so low in this situation (due to mitigation by the β countries in addition to the agreement's mitigation) that their payoff is large enough to give them a negative outside option (which is why the agreement is stable).

Whether or not $\underline{x}$ is indeed internally stable, depends on $D_i(\hat{e})$ for α countries. If the damage is small enough (i.e. if the β countries reduce their emissions sufficiently between sections 5 and 6 of Tab. 2 to drive down global emissions significantly), then $\Omega(\underline{x})$ is negative and agreements of size $\underline{x}$ are internally stable.

3.4 Discussion

We have shown that the existence of countries with non-convex damage functions (β countries) can allow for an island of stability with non-conventional, larger agreement size $\underline{x} > \bar{x}$ than in the case without strategic complements. In such a game equilibrium, the agreement is just large enough to induce the β countries to chose emissions $e_i = 0$, even though they are not agreement members. In the equilibrium, global emissions are lower and the payoff is larger for every country.

What happens if the number of α countries and β countries changes in the comparative statics sense? A larger number of β countries z means that emissions drop more sharply if the agreement forms at the island of stability (i.e. is large enough to achieve total emissions $e = \hat{e}$: case 6, Tab. 2). However if $\hat{e}$ remains constant, a larger z does not mean that global emissions fall to a lower level, but that they start falling from a higher level. This in turn means that less mitigation effort on behalf of the agreement is necessary to achieve the same (positive) result, and a smaller agreement may be able to do so.

This does not mean that the outside option for members of this agreement is necessarily smaller, even though it seems probable. It is possible that the border between cases 5 and 6 of Tab. 2 is reached for a smaller agreement size x, so the mitigation effort is distributed among fewer countries. If the agreement can increase emissions a little bit and still reach $\hat{e}$ the outside option for agreement size $\underline{x}$ could even grow (because $\underline{x}$ could sink). This means that even if the emissions reductions by β countries between cases 5 and 6 from Tab. 2 increase, agreement size $\underline{x}$ (which relies on case 6) may loose its stability.

Furthermore, consider that it were possible for non-member α countries to become β countries. This could stabilize the agreement because non-members would reduce more emissions, since y decreases, and $\hat{e}$ is not likely to increase. However, results depend on the dynamics between the β countries which result in the relationship between $\tilde{e}$ and $\hat{e}$. For an optimistic approach ($\tilde{e} = \hat{e}$) an α country which becomes a β country would indeed not change $\hat{e}$, thus increasing desirability of the $\underline{x}$ case. For a pessimistic approach ($\tilde{e} + z - 1 = \hat{e}$), $\hat{e}$ is reduced by 1 for every country that changes its type from α to β. This increases both costs and benefits for agreements of size $\underline{x}$; the effect on stability is ambiguous.

3.5 Conclusions

The paper investigates the equilibrium of an international emissions game for the case of two strategically distinct types of countries, some of which join a binding environmental agreement. While one country type has a conventional convex damage function (α countries), the other countries have concave damages (β countries). We assume that members of an agreement jointly act as Stackelberg leader, while the non-signatories of both types play a Nash game in the final stage of the game.

Due to their non-convex damage functions, β countries outside the agreement do not act as freeriders on mitigation efforts of the agreement. Instead they reduce their emissions if there is sufficient mitigation effort by the other countries. This is not a strategic choice (which could be non-credible), but individually rational. By anticipating this reaction, the agreement as Stackelberg leader has a novel incentive structure. If the emissions of its members are sufficiently low, then they can profit from the additional of cooperation by the β non-members.

We find that this leads to the possibility of a larger stable agreement which sufficiently reduces emissions to induce emissions reductions by the β countries. Then, global emissions are significantly lower than in the case without β countries, and it is also Pareto-superior. However, the smallest abating agreement size remains stable even if the larger one becomes stable, so it remains an open question which of the two potentially stable agreements would be realized.

Our model admittedly relies on the quite restrictive assumption of constant marginal gains from emissions. While this is not a very uncommon assumption (cf. Asheim et al., 2006; Barrett, 1999, 2001), it is still a strong one. However the main argument of our analysis does not rest critically upon this linearity. Instead we have chosen it because it helps keeping the model tractable. The main point of the analysis lies in the strategic complementarity of emissions abatement for some countries. Countries in and agreement can use their Stackelberg leadership position to exploit this strategic complementarity to the benefit of all countries. Therefore the results should carry over to models with diminishing marginal gains from emissions.

Further research could look into effects of β countries inside the agreement as well as multiple agreements. Based on current results and numerical experiments, we can begin to speculate about the potential outcome of such an analysis. Stable agreements contain only a small number of β countries, if any β country at all (supposed there are no transfer payments within the agreement). Eisenack and Kähler (2015) show that β countries voluntarily select the follower position to improve their payoff in the two countries case. To take this a little bit further, in a setting where countries with concave damage functions exist, a grand coalition is not required in order to come closer to the social optimum.

We thus think it is worth further exploring the effects of heterogeneous countries in international environmental agreements, in particular if countries exhibit qualitatively different strategic properties.

Appendix to 'Strategic Complements in International Environmental Agreements: a New Island of Stability'

3.A Proof of Prop. 7

Proof. First, suppose that $e_x \leq \tilde{e} - y$ and $e_i = 0$ for all β countries. Consider a specific β country i. Then,

$$e_{-i} = e - e_i = e_x + y \leq \tilde{e}. \tag{65}$$

Thus, $e_i = 0$ is a best response according to Eq. (52): country i cannot benefit from unilaterally changing its strategy.

Second, suppose that $e_x \geq \tilde{e} - y - z + 1$ and $e_i = 1$ for all β countries. Consider a specific β country i. Then,

$$e_{-i} = e - e_i = e_x + y + z - 1 \geq \tilde{e}, \tag{66}$$

so that $e_i = 1$ is a best response according to Eq. (52).

Third, exclude further equilibria. (i) Consider $e_x < \tilde{e} - y - z + 1 < \tilde{e} - y$. If there would be at least one β country i with $e_i > 0$, then, $e_{-i} = e - e_i < e_x + y \leq \tilde{e}$, so that $e_i = 0$ would be the best response, a contradiction. (ii) Consider $e_x > \tilde{e} - y > \tilde{e} - y - z + 1$. If there would be at least one β country i with $e_i < 1$, then, $e_{-i} = e - e_i > e_x + y + z - 1 \geq \tilde{e}$, so that $e_i = 1$ would be the best response, a contradiction. (iii) Consider $\tilde{e} - y > e_x > \tilde{e} - y - z + 1$, and assume that there is at least one β country i with $e_i = 0$, and at least one β country j with $e_j = 1$. The choice of i would only be a best response if $\tilde{e} \geq e_{-i} = e_x + y + e_z$. The choice of j would only be a best response if $\tilde{e} \leq e_{-i} = e_x + y + e_z - 1$. Both conditions cannot hold at the same time. □

3.B Proof of Prop. 8

Proof. [1] First, collect properties of F. If D_i fulfills the Inada-conditions, F always exists, and is positive. By taking the total differential, $F'(x) = -\frac{b}{x^2} D''_\alpha(F(x)) < 0$, so that F is strictly decreasing.

[2] Now observe that the conditions in Tab. 2 cover all possibilities in terms of x, y, z. Obviously, the cases are disjoint. They are also complete (the only missing case in the table, $bz = f(F, x) - f(\hat{e}, x)$, corresponds to non-unique equilibria).

[3] Now proceed to the main part of the proof. We go through all cases, and show that

the game equilibria are as given in Tab. 2. Generally, note that

$$d_{e_x}\Pi_x = b - xD'_\alpha(e_x + y + e_z)$$

$$\text{with } e_z = \begin{cases} 0 & \text{if } e_x \leq \hat{e} - y, \\ z & \text{if } e_x > \hat{e} - y, \end{cases} \tag{67}$$

$$d_{e_x e_x}\Pi_x = -xD''_\alpha(e_x + y + e_z) < 0. \tag{68}$$

[3.1] Here, $\hat{e} < y$ and $e_x \geq 0$ imply $e_z^* = z$. Thus, due to $F > x + y + z$ and the monotonicity of D'_α,

$$\forall e_x \in [0, x] : d_{e_x}\Pi_x = b - xD'_\alpha(e_x + y + z) > b - xD'_\alpha(F) = 0.$$

Thus, it is optimal so chose the corner solution $e_x^* = x$ in stage B.

[3.2] Again, the β-countries' reaction is $e_z^* = z$, so that $d_{e_x}\Pi_x = b - xD'_\alpha(e_x + y + z)$. Since $y + z < F \leq x + y + z$, the monotonicity of D'_α implies

$$D'_\alpha(y + z) < D'_\alpha(F) = b/x \leq D'_\alpha(x + y + z).$$

There exists thus, due to continuity of D'_α, a unique $e_x \in (0, x]$ so that $d_{e_x}\Pi_x = 0$. It follows from the definition of F that this solution is characterized by $F = e_x + y + z$. Sufficiency is then guaranteed by the concavity of Π_x, so that $e_x^* = F - y - z$.

[3.3] Again, the β-countries' reaction is $E_z^* = z$. The monotonicity of D'_α implies

$$\forall e_x \in (0, x] : d_{e_x}\Pi_x = b - xD'_\alpha(e_x + y + z) < b - xD'_\alpha(y + z) \leq 0.$$

The last inequality follows from $F \leq y + z$ and the monotonicity of D'_α. Thus, it is optimal so chose the corner solution $e_x^* = 0$ in stage B. Then, total emissions amount to $e^* = y + z$ in the stage B equilibrium.

[3.4] Generally,

$$\forall e_x \in [0, x] : d_{e_x}\Pi_x = b - xD'_\alpha(e_x + y + e_z) \geq b - xD'_\alpha(x + y + z).$$

The last expressions is strictly positive in case 4, so that it is optimal to choose the corner solution $e_x^* = x$ in stage B. Since $x + y > \hat{e}$, we obtain $e_z^* = z$.

[3.5 / 3.6] Now consider case 5 and case 6. Observe that

$$\forall e_x \in [0, \hat{e} - y] : d_{e_x}\Pi_x = b - xD'_\alpha(\hat{e}) > b - xD'_\alpha(F) = 0.$$

Thus, the corner solution $e_x = \hat{e} - y$ is the local maximum on the interval $[0, \hat{e} - y]$. It is yet also possible to select $e_x \in (\hat{e} - y, x]$, where $d_{e_x}\Pi_x = b - xD'_\alpha(e_x + y + z)$, so that Π_x is locally maximized at $e_x + y + z = F$ (recall the concavity of Π_x). It yet needs to be determined whether the corner solution or the interior solution is the global maximum. It holds that

$$\Pi_x(\hat{e} - y) = b(\hat{e} - y) - xD_i(\hat{e}) \tag{69}$$

$$> b(F - y - z) - xD_i(F) = \Pi_x(F - y - z) \tag{70}$$

$$\Leftrightarrow bz > f(F, x) - f(\hat{e}, x). \tag{71}$$

Thus, the last inequality implies $e_x = \hat{e} - y$, which is case 6. Otherwise, case 5 applies. If the left-hand-side and the right-hand side are equal, the payoff in the corner and the internal solution is equal, so that both decisions are game equilibria.

[3.7] Due to $y < F \leq \hat{e}$ and the definition of F, it holds that $xD'_\alpha(y) < b \leq D'_\alpha(\hat{e})$. Thus, monotonicity and continuity of D'_α guarantees $\exists^1 e_x \in (0, \hat{e} - y] : xD'_\alpha(y + e_x) = b$. This just states, by the definition of F, that $e_x = F - y$ and $e_z = 0$ together fulfill the first-order condition. This also fulfils the second-order condition since Π_x is concave. This choice is consistent with Eq. 67 and $e_x \in [0, x]$ since $0 < e_x = F - y \leq \hat{e} - y$ in case 7. Thus, the stage B equilibrium is $e_x^* = F - y$ with $e^* = F$.

[3.8] It generally holds that

$$\forall e_x \in (0, x] : d_{e_x}\Pi_x = b - xD'_\alpha(e_x + y + e_z) < b - xD'_\alpha(y).$$

The last expression cannot be positive since $F \leq y$. Thus, it is optimal so chose the corner solution $e_x^* = 0$ in stage B. Thus $e_x \leq \hat{e} - y$ since $y \leq \hat{e}$ in case 8, so that $e_z^* = 0$. Then, total emissions amount to $e^* = y$.

□

3.C Proof of Prop. 9

Proof. We first show that the smallest abating agreement $\bar{x}$ is stable and that for a scenario without β countries ($z = 0$) it is unique. Then we prove stability for $z > 0$.

Stability for $z = 0$: All agreements smaller than $\bar{x}$ do not abate (i.e. case 1 or 4 in Tab. 2) because $\bar{x}$ is the smallest abating agreement. A non-abating agreement always exists because at least for an agreement size of $x = 1$ there is no incentive to abate (this follows directly from Eq. (48) because a single α country plays dominantly $e_i = 1$). Agreements that do not abate give no advantage for members, so there is no incentive to join. Therefore no agreement smaller than $\bar{x}$ is internally stable.

If x is larger, F is smaller due to Eq. (62). Because of Eq. (50), positive abatement (i.e. $F(x) < x$) is chosen at some agreement size. Since we know that abatement is chosen because it is profitable (F is optimal by definition), the smallest abating agreement is stable. Formally, this is the point where the outside option is negative, i.e.

$$\text{for} \quad F(\bar{x}) < n : \quad \Omega(\bar{x}) = b - D_i(n) - \frac{b}{\bar{x}} + D_i(F(\bar{x})) < 0, \tag{72}$$

because in the case of $(x = \bar{x} - 1)$ there is no abatement at all.

We know that for $z = 0$ all abating agreements larger than $\bar{x}$ are not internally stable because α countries always have a positive outside option according to (64).

To summarize for $z = 0$: Agreements smaller than $\bar{x}$ are not stable, agreements larger than $\bar{x}$ are not internally stable and a agreement of size $\bar{x}$ is beneficial for its members. In other words $\bar{x}$ is the only agreement size that gives a negative outside option. Therefore it is stable and unique.

Stability for $z > 0$: The only (possible) difference to the case of $z = 0$ is that the smallest abating agreement $\bar{x}$ could fall into case 6 of Tab. 2, so $e \neq F(x)$. If this is true, then the outside option would change to

$$\Omega(\bar{x}) = b - D_i(n) - \frac{b}{\bar{x}} + D_i(\hat{e}). \tag{73}$$

This is still negative for the same reason that applies if emissions $F(x) - y - z$ are chosen by the agreement: Any positive abatement chosen by the agreement maximizes the payoff of the members, therefore it is preferable to $e = n$. The agreement is benefitial; in other words the outside option is negative and internal stability is given, just as in the case of $z = 0$.

□

3.D Proof of Prop. 10

Proof. We show here that for $z > 0$ exactly one other agreement can be stable. This is true because for the smallest agreement size in case 6 of Tab. 2 the outside option is larger than in the case of $z = 0$.

The outside option Ω for the smallest agreement size $\underline{x}$ that supports case 6 from Tab. 2

is

$$\Omega(\underline{x}) = b \cdot \left(1 - \frac{\hat{e} - y}{\underline{x}}\right) - D_i(F(\underline{x} - 1)) + D_i(\hat{e}). \tag{74}$$

Now we compare this with the outside option in the standard case (i.e. within section 5). If the outside option at agreement size $\underline{x}$ is smaller for emissions $e = \hat{e}$ than for $F(\underline{x})$ (i.e. is smaller for a voluntary choice by the agreement of section 6 over section 5 in Tab. 2), then the following must hold:

$$\begin{aligned} & b \cdot \left(1 - \frac{\hat{e} - y}{\underline{x}}\right) - D_i(F(\underline{x} - 1)) + D_i(\hat{e}) \\ & \qquad\qquad < \\ & \qquad\qquad b \cdot \left(1 - \frac{F(\underline{x}) - y - z}{\underline{x}}\right) - D_i(F(\underline{x} - 1)) + D_i(F(\underline{x})) \end{aligned} \tag{75}$$

$$\Leftrightarrow \quad -\frac{\hat{e} - y}{\underline{x}} \cdot b + D_i(\hat{e}) < -\frac{F(\underline{x}) - y - z}{\underline{x}} \cdot b + D_i(F(\underline{x})) \tag{76}$$

$$\Leftrightarrow \quad \frac{b}{\underline{x}} \cdot (F(\underline{x}) - z - \hat{e}) < D_i(F(\underline{x})) - D_i(\hat{e}) \tag{77}$$

This corresponds exactly to the definition of section 6 (in comparison to section 5). Therefore it holds if $\underline{x}$ lies in section 6 and $(\underline{x} - 1)$ in section 5. It follows that the outside option for agreement size $\underline{x}$ is smaller than for $e = F(x)$.

Whether or not $\underline{x}$ is indeed internally stable depends on $D_i(\hat{e})$. If it is small enough (i.e. if the β countries reduce their emissions sufficiently between sections 5 and 6 of Tab. 2 to drive down global emissions significantly), then $\Omega(\underline{x})$ is negative and agreement size $\underline{x}$ is internally stable. Formally, when $\Omega(\underline{x})$ is negative, the agreement is internally stable:

$$\Omega(\underline{x}) = b \cdot \left(1 - \frac{\hat{e} - y}{\underline{x}}\right) - D_i(F(\underline{x} - 1)) + D_i(\hat{e}) < 0 \tag{78}$$

$$\Leftrightarrow \quad D_i(F(\underline{x} - 1)) - D_i(\hat{e}) > b \cdot \left(1 - \frac{\hat{e} - y}{\underline{x}}\right) \tag{79}$$

External stability is not an issue for agreement size $\underline{x}$ because other agreements (except for $\bar{x}$ as described in Prop. 9 are not stable. The proof works just like the corresponding one for Prop. 9:

If an α country enters the agreement within section 6 global emissions $e = \hat{e}$ do not change. In effect, damage does not change by entry here but the entering country will have to bear part of the abatement costs. Therefore entry is never attractive within section 6 of Tab. 2.

The smallest agreement of section 7 is not internally stable, as well. Compared to the outside option for $e = F(x)$, a country has a damage if it leaves the agreement $D_i(\hat{e})$ that is at least as high as $D_i(F(x-1))$ while the rest of the terms are equal.

Within sections 5 and 7 of Tab. 2 no agreement except $\bar{x}$ is stable due to (64).

Therefore no other agreement than $\bar{x}$ and $\underline{x}$ can be stable. $\square$

Interlude II

The global public good nature of greenhouse gas mitigation has led to a strong focus on an IEA as the optimal solution. Local emissions influence the global climate dynamic. Benefits of greenhouse gas emissions are local but damages occur everywhere - albeit not equally dire. If an agreement on a global level is both feasible and ambitious, it obviously offers the best chance to tackle climate change. Therefore the optimisation of negotiation structures for IEAs certainly is a valuable approach in fighting climate change.

Strategic complements for example can lead to a much better outcome both in terms of emissions and welfare, as we have shown. Exploiting adaptation for the creation of strategic complements can offer new possibilities for unilateral action that does not lead to freeriding. These findings can help to design negotiations. For example: If there are indeed multiple equilibria for stable coalition sizes (as our 'Island of Stability' article suggests), then it is important to be careful in order to reach the more favourable one.

However the IEA approach might be reaching the limits of its potential. The Paris agreement shows broad nominal support for emissions reductions and a lot of good intentions but the lack of commitment to binding mission reduction targets remains a large issue. It may be necessary to look beyond the usual framework of one coalition of nation states. Political science has brought up several interesting ideas. Empirical observation finds the UNFCCC framework dominant but other structures are growing. Subnational governments, NGOs and cities show initiative of their own. Multiple international agreements coexist.

Economic methods which have so far been used to analyse IEAs can help evaluate alternative structures as well. Game theory in particular can give valuable insights on incentives to cooperate, conditions for coalition stability, and expected welfare shifts under different agreement regimes. We need to understand the potential and limits of such TEAs. We should also analyse who are the winners and losers of different kinds of agreements; who pays, who freerides and who is affected indirectly.

The following article puts together ideas from political science literature. We also develop exemplary ideas for the application of economic models to to such new approaches. Two of these approaches are put into concrete form that can serve as primers for modelling certain kinds of TEAs. Finally we give an outlook on promising avenues for further research in the field.

4 Transnational Environmental Agreements with Heterogeneous Actors

Abstract

This paper explores transnational environmental agreements on climate change as complements to the United Nations Framework Convention on Climate Change (UNFCCC) process. We discuss scope and limits of the current economic literature on international environmental agreements. We argue that future game theoretical research would benefit from extending the analysis to consider (i) actors that are not nation state governments, and (ii) multiple environmental agreements. We underpin this claim by suggesting two proposals for economic models that analyse climate clubs and city alliances. The results show that transnational environmental agreements can be individually rational and can improve effectiveness of climate policies.

4.1 Introduction

There is unequivocal scientific agreement on the dangerous interference of anthropogenic greenhouse gas emissions with the climate. But efforts to find cooperative solutions on an international level have been mostly unsatisfactory so far. The recent UN climate negotiations in Paris have led to some agreement about global targets, but not about the individual nations' contributions to the global public good. This state of affairs motivates the search for complementary approaches for global emissions reductions. Some suggestions are in the air. For example, some authors think about minilateralism (Eckersley, 2012), climate clubs (Widerberg and Stenson, 2013; Falkner, 2015) or a building blocks approach (Stewart et al., 2013). Lobby groups and NGOs influence climate and energy policy. City alliances grow in parallel to nation state based coalitions. This chapter aims at exploring some of such transnational initiatives or patterns of cooperation. Although there has been some research on those patterns in the global governance literature (related to political science), we aim at making this topic conducive for economic analysis, in particular game theory. How can such patterns of cooperation be explained? Can we expect cooperation to be effective?

In this chapter, we call a contract that stipulates rules for contributions to a global environmental good 'transnational environmental agreement' (TEA) if it has heterogeneous contracting parties, i.e. of different type. Parties can be national, subnational, international, or of different quality. Such contracts can be explicit or implicit. They might directly aim at emissions reductions, or only indirectly (e.g. by stipulating monitoring procedures). We chose the term 'transnational' to generalize from the established 'international' environmental agreement (IEA) framing. Transnational agreements are not undertaken within single jurisdictions (which would not be international either), but the main actors involved do not necessarily need to be national governments (cf. Andonova et al., 2009; Hale and Roger, 2014).

TEAs are not an invention from the theory. For example, the C40 Cities Climate Leadership Group (C40, 2015) with more than 80 megacities (from the South and the North) took leadership in signing the Greenhouse Gas Protocol for Cities in 2014. As of December 2015, the number of signatories increased to 428 cities (GHG Protocol, 2015). Weischer et al. (2012) map 17 climate clubs, being non-universal and partially overlapping agreements of nation states that cooperate on climate change. In total, 122 countries are members of at least one of those clubs. Some of these clubs include nonnation state partners. A first study roughly estimates that non-state initiatives might reduce greenhouse gas emissions by three gigatons in 2020 (UNEP, 2015). Although the empirical fact that many TEAs already exist might seem impressive at the first glance, some sceptical questions warrant attention. It is well known, after all, that global public goods suffer from freerider incentives. So what does motivate actors then to be frontrunners and sign a nonuniversal TEA? And if they do so for some reason, why shouldn't they not just pretend to reduce greenhouse gas emissions? These questions will be further explored in this chapter.

There are only few papers in economic journals that address TEAs, some of which are discussed in more detail below. The theme of city alliances seems to be broadly neglected (but see Sippel, 2010; Millard-Ball, 2012, for some data analysis). Subnational emission reductions are not analysed, to our knowledge, from the perspective of cooperation between actors from different countries. The exception is the game theoretic literature on environmental agreements that explains non-universal cooperation (more on that below). Studies that admit for multiple climate clubs are sparse (e.g. Asheim et al., 2006; Finus, 2008; Hagen and Eisenack, 2015). National lobby groups are addressed by Marchiori et al. (2016), Habla and Winkler (2013) and Hagen et al. (2016), but not from a transnational perspective (for a literature review on the political economy of the formation of international environmental agreements, see Wangler et al., 2013). The chapter is not intended to fill all these gaps, but contributes by arguing for the relevance of this research field. It provides structure in transferring insights from global governance research, where much more has been published on transnational climate governance than in economics, to game theory. First, we report on the global governance literature and empirical examples of emerging transnational climate agreements. Then we give an overview of the existing economic literature on the scope and limits of international environmental agreements. Building on these two pillars, we follow up with two proposals for game theoretic models. They analyse strategic effects of climate clubs and city alliances as examples for TEAs. We then take a look at the larger picture again and contextualize these approaches in an outlook on promising future research.

4.2 Current transnational approaches in the global governance literature

This section puts together some selected and documented empirical observations of transnational environmental agreements, and summarizes relevant publications from the global governance literature. Climate clubs can be understood as ’Club-like arrangements between states that share common climate-related concerns, and sometimes in partnership with non-state actors such as companies and Non-Governmental Organizations’ (Widerberg and Stenson, 2013, 1). Climate clubs are also coined as ’minilateralism’ (Eckersley, 2012). They are currently analysed in the discourse on fragmented global governance (e.g. Biermann et al., 2009; Keohane and Victor, 2011; Isailovic et al., 2013). This literature acknowledges that there is no monolithic and rational global governance architecture, but a carpet of loosely coupled international institutional arrangements and regimes, not all being universal but many overlapping. Although they may address multiple issues, their scope can be synergistic, cooperative or conflctive. One set of overarching questions addresses the conditions under which fragmentation is conducive or detrimental to regime effectiveness (e.g. Gehring and Oberthür, 2008; Biermann et al., 2009).

Weischer et al. (2012) analyse existing climate clubs and explore their contribution to climate action as well as the incentives for becoming club members and taking action. Similarly, Widerberg and Stenson (2013) find different types of clubs, from political and technical dialogue forums to country strategy and project implementation groups. Examples are the Asia-Pacific Partnership on Clean Development (2006 – 2011, including the US and China) and the International Energy and Climate Initiative – Energy+ (since 2010, International Energy and Climate Initiative – Energy+ 2015). The latter, led by Norway, has 16 national government members (from Africa, Asia and Europe), and multiple non-governmental partners, e.g. the World Bank and the World Business Council for Sustainable Development (WBCSD). It aims at promoting energy efficiency and renewables by incentivizing commercial investments. While some papers focus on the legitimacy of climate clubs (e.g. Karlsson-Vinkhuyzen and McGee, 2013), others focus on their effectiveness (see Moncel and Asselt, 2012, for an overview).

Different arguments are put forward to underpin the potential of climate clubs. It might be easier to reach agreement in smaller clubs of countries that are more willing to push forward climate protection (based on the argument of Olson, 1971). Falkner (2015) distinguishes three dominant rationales of climate clubs. First, club benefits are created for the members. Second, a re-legitimation of the climate regime by giving great powers a privileged position in the negotiations while acknowledging their greater responsibility at the same time. Third, the potential of climate clubs to enhance the bargaining efficiency of the international negotiations by facilitating agreement amongst smaller groups of players. Further pros and cons of climate clubs will be discussed below.

Another case for TEAs is contracts between cities from different countries. City networks on sustainability issues have some tradition. The International Council for Local Environmental Initiatives (since 1990) has more than 1,000 cities, towns and metropolises from all continents as members (ICLEI, 2015). Over 1,700 cities and municipalities are members of the Climate Alliance (since 1990, Climate Alliance, 2015), and have voluntarily committed to reduce greenhouse gas emissions by 10 per cent every five years. The C40 Cities Climate Leadership Group (since 2005) pushed the Compact of Mayors (2015), which is currently signed by cities with more than 5 per cent of the global population. The Compact of Mayors has adopted a common monitoring, reporting and verification standard, the Greenhouse Gas Protocol for Cities (GHG Protocol, 2015). The standard is built on experience with a private sector initiative, the Carbon Disclosure Project (CDP, 2015), and has established a joint carbon registry.

As with climate clubs, there is also some research on city alliances. A special issue in Local Environment reviewed the early studies (Betsill and Bulkeley, 2007). Interesting questions are the motivations for joining city alliances, and their environmental effectiveness. The early literature is mostly descriptive in nature and undertakes single or comparative case studies. For example, Betsill and Bulkeley (2004) show for six case studies of municipalities in the UK that membership in Cities for Climate Protection (CCP) is mostly motivated by the availability of additional financial and political resources, and not so much by transfer of technical and best practice knowledge. International recognition of the local engagement and the re-framing of existing measures in terms of climate change helps increase legitimacy and place those activities higher on the local agenda. Gustavsson et al. (2009) explore the potential of city networks for Swedish cities. Kern and Bulkeley (2009) analyse modes of cooperation in three transnational municipal networks (Climate Alliance, CCP and Energie-CitÃl's). Members are active to quite different degrees in terms of information and communication, funding, recognition, benchmarking and certification.

Bulkeley and Broto (2013) collected an impressive database with more than 600 'urban climate change experiments' from 100 systematically selected global cities. All these experiments are explicitly targeted at reducing greenhouse gas emissions or at adapting to climate change. Most experiments are found in Europe, Latin America and Asia. Less of them relate to adaptation, but many to urban infrastructure, the built environment and energy. Half of the experiments involve partnerships, for example between local governments and the private sector. More recently, Hakelberg (2014) collected a sample of 274 European cities of which 41 per cent became members of city networks until 2009. The econometric analysis shows that membership in a city network increases the likelihood of adopting a local climate strategy. In contrast, there is no such effect on geographically neighbouring cities. Top-down governmental policies have a stronger effect on local climate strategies than city network membership.

Some studies explore the reasons why city alliances exist and might (not) be effective.

Bulkeley (2010) generally stresses the changing role of cities and states in political systems, and highlights political economy reasons. Furthermore, urban areas are expected to be particularly vulnerable to climate change, though some more so than others (e.g. IPCC, 2014b; Corfee-Morlot et al., 2009; Gill et al., 2007; Campbell-Lendrum and Corvalan, 2007). This might contribute to urgency in climate change adaptation and mitigation in some cities. Generally the local approach offers potentially easier stakeholder engagement, concrete action, resource mobilization and investment, mostly because actors are directly involved (e.g. Corfee-Morlot et al., 2009; Sippel and Jenssen, 2009). On the other hand, urban action cannot be understood as being disconnected from national law. While the latter sets the context for the former, the former can help enforcing national action by contracts, building trust and through the political process. As a further reason, there might be local co-benefits due to investments, local pollution, or first-mover advantages if a city specializes in technological solutions (although e.g. Urpelainen, 2009, shows that local co-benefits are not sufficient to motivate local frontrunners). Further pros and cons of city alliances will be discussed below.

Approaches to study city alliances, climate clubs, and other modes of transnational environmental agreements resonate with different literature streams. Some scholars study subnational climate policies from the multi-level perspective (e.g. Betsill and Bulkeley, 2006; Monni and Raes, 2008). Hooghe and Marks (2003) disentangle different modes that might be helpful to characterize different transnational governance patterns. Type I governance refers to hierarchically nested arrangements (like in a classic federal system), while Type II governance refers to arrangements that cross hierarchies or overlap between jurisdictions. The literature on fiscal federalism (Oates, 1972, 2005) and environmental federalism (Shobe and Burtraw, 2012) uses more economic concepts to study the allocation of policies between subsidiarity and centralization. This approach might be helpful to study TEAs.

The debate on transnational climate governance got further impetus from Elinor Ostrom after her Nobel laureate speech (Ostrom, 2010, 2012). She rooted the considerations on addressing climate change both down from the top and up from the bottom in the concept of polycentric governance. In such governance modes many centres of decision making, which are formally independent from each other, make mutual adjustments for ordering their relationships (Ostrom et al., 1961). This line of inquiry was taken up further by Cole (2011) and recently by Jordan et al. (2015).

4.3 Scope and limits of international environmental agreements

International environmental agreements with a focus on climate agreements have been analysed in the economic literature since the 1990s. This has led to the development of various models that serve as a starting point for the analysis of TEAs. This section gives an overview of this strand of research and its main assumptions and results.

The literature on IEAs started with the seminal work of Carraro and Siniscalco (1993) and Barrett (1994). The basic idea is to transfer concepts from the theory of economic cartels (D'Aspremont et al., 1983; Chander and Tulkens, 1995) to the study of stable coalitions that contribute to a public good. A large set of publications that refined the first contributions followed suit, with further analytical and simulation studies up to date. Most of this research is based, inter alia, on the following propositions:

1. Global environmental problems are about provision of public goods.
2. Players are aspiring and achieving individually rational decisions in a game theoretic framework.
3. International environmental agreements need to be self-enforcing.
4. Players are nation states; their payoffs are determined by national welfare.

5. Full global cooperation (the grand coalition) would yield the first-best outcome.
6. The social optimum is ideally achieved, in principle, by a single global policy instrument (e.g. a uniform carbon tax or an emission trading scheme).

Based on these propositions, some standard insights have been consolidated over a broad range of settings. Some of them can be stated in a stylized way as follows: [i] The social optimum cannot be achieved due to freerider incentives. [ii] If some countries or coalitions undertake unilateral emission reductions, their effect is dissipated due to carbon leakage. [iii] Cooperation is either broad but shallow, or deep but small. Thus, if we assume that reducing carbon emissions is associated with high mitigation costs and small damage reductions, a stable coalition will not have many signatories.

Although scientifically robust, these results are politically mostly frustrating. They do a good job in explaining the long-lasting stalemate and questionable effectiveness of the climate negotiation process under the UNFCCC.

Taking on that, two questions remain. First, if these results are valid for the climate case, is there any chance of averting the greatest market failure ever (Stern, 2007) or, more pathetically, loss of life and quality of life for billions of people? Is there no alternative to accepting the inevitable? Second, are these results indeed valid for the climate case?

Some sceptical remarks may deserve attention. For example, some studies have determined social costs of carbon of just a few dollars per ton (in particular for higher discount rates, IPCC, 2014b). Several other studies have shown that the costs of mitigating emissions to limit greenhouse gas concentrations below 430 – 480 ppm by 2100 lead to a reduction of consumption growth by 0.04 to 0.14 percentage points over the twenty-first century (IPCC, 2014b), i.e. these costs might be relatively low. If at least one of these kinds of conclusions is valid, it seems that the gains from cooperation are shallow. The theory would thus imply broad cooperation. This implication is falsified by over 20 years of slow progress in climate negotiations.

Furthermore, the empirical examples of TEAs outlined above cannot be explained by the standard insights. Why should multiple climate clubs on overlapping issued be formed? Why do some climate clubs engage, although probably on a low level, in unilateral action? Why do cities from different countries start cooperation on emissions reductions, although most of their national governments do not, although there is no (single) global policy instrument in place, and although there are still many cities that do not participate in city networks? Instead, theory would predict cities to be freeriders.

Solving such puzzles seems to be important both for climate protection and for scientific inquiry. One starting point for analysis could be to reconsider some of the six propositions outlined above. In the following, we want to explore how proposition (3), (4) or (5) might be relaxed, while keeping the remaining propositions.

4.4 Proposals for theoretical analysis

In this section we give two selected proposals for economic models that concentrate on transnational environmental agreements: climate clubs and city alliances. Both can be observed empirically. However, both have got little attention in the economic literature so far even though they offer interesting concepts. We give general outlines for these two approaches that can serve as seeds for further model development. In addition, we give a detailed outlook on promising lines of further research in these and related areas.

4.4.1 Climate clubs

One way to open up the classical approach of one single international environmental agreement is to allow heterogeneous countries to form climate clubs. As described in the global governance literature, climate clubs may have different effects and may improve over one monolithic agreement through different rationales (cf. Falkner, 2015).

The aspect of club benefits for the members of a climate coalition is analysed by Nordhaus (2015). He finds that a climate club that imposes trade sanctions on non-participants can induce a larger stable coalition with more abatement than a coalition without sanctions. Asheim et al. (2006) model the case of symmetric countries and two coexisting agreements. The countries are partitioned in two regions and can choose whether they sign an agreement for that region or not. They conclude that a larger number of cooperating signatories can be sustained, compared to the standard case of a single IEA. The case of two coexisting TEAs is further analysed in a numerical study by Osmani and Tol (2010) who additionally consider two asymmetric country types in a three-stage sequence of play between the coalitions and the non-signatories. Their results show that the possibility of two coalitions could increase as well as decrease emission abatement in comparison to the standard case with one coalition.

Going beyond numerical examples, Hagen and Eisenack (2015) study the effect of multiple coexisting climate clubs in an analytical game theoretic setting. The paper allows for asymmetric countries and investigates if global cooperation for emissions abatement can be improved if countries can form coexisting TEAs. This very general analytical approach to climate clubs helps to get insights in the effects of negotiating coexisting climate clubs without being bound by specific assumptions on the concrete costs and benefits of countries emissions abatement. The rationale of this analysis will be introduced for the simplest version of this game theoretic climate clubs model. Its main results are derived and discussed.

The model is set up in the widely used two-stage game structure with countries first choosing to join a coalition or not (e.g. Carraro and Siniscalco, 1993). In the second stage the members of a coalition decide cooperatively on the amount of emissions abatement that is undertaken by the coalition. The game is solved by backward induction. In contrast to the bulk of the existing literature, coexisting agreements are possible. Each stage of the model is set up as a simultaneous Nash game.

The simplest version of the model already allows for important insights to the idea of climate clubs. It considers two types of asymmetric countries and two possible TEAs. The number of abating countries of type $i(i = 1, 2)$ is denoted by z_i. We assume linear benefits of global emissions abatement and a binary choice for countries between abatement, which is associated with abatement costs c, and pollution. An abating country of type i gets the payoff $\pi_i^\alpha = -c + \alpha_i(z_1 + z_2)$. Asymmetric benefits of the countries are expressed by the parameter α_i where α_2 is normalized to $\alpha_2 = 1$ and $\alpha_1 \epsilon [0, 1]$. A type 1 country therefore benefits less or at most as much as a type 2 country from abatement. The net benefit of own abatement of each country is negative since $c > 1$. Thus, playing pollute is the dominant strategy if there is no TEA and all countries play pollute in the non-cooperative Nash equilibrium. In the first stage of the game countries decide about their TEA-participation. The case of one agreement is compared to that of two coexisting agreements. In the first case, countries of both types can choose to join or not to join the agreement. In the other case, each agreement consists of similar countries, representing e.g. regional agreements (cf. Asheim et al., 2006). Solving the second stage of the game first, the agreements cooperate internally in their decisions about their emissions abatement. In the two agreements case the agreements take their decisions independently and simultaneously. Maximization of the respective joint payoffs yields the second stage equilibrium with agreement i playing

$$z_i^* = k_i \text{ (abate)} \qquad \text{if} \quad \alpha_1 k_1 + k_2 > c \tag{80}$$

$$z_i^* = 0 \text{ (pollute)} \qquad \text{if} \quad \alpha_1 k_1 + k_2 < c \tag{81}$$

with k_i denoting the number of type i signatories. This result already shows that the decision of each agreement depends on the number of its members, but not on the abatement decisions of the other countries. The application of the criteria of internal and external stability solves the first stage of the game. As playing pollute is a dominant strategy for non-signatories, internal stability is only given if the members of an agreement choose to abate and would change from abate to pollute if one country left the agreement so that $c > \alpha_1(k_1^* - 1) + k_2^* > \alpha_1 k_1^* + (k_2^* - 1)$. The stability conditions together with this linchpin condition indicate that a stable abating agreement may consist of countries of both types with the number of signatories satisfying

$$c + \alpha_1 > \alpha_1 k_1^* + k_2^* > c. \tag{82}$$

Setting either the number of type 1 or of type 2 members in the agreement to zero, we get the results for the size of the single agreement if it consists only of type 1 (83) or of type 2 (84) countries:

$$c + \alpha_1 > \alpha_1 k_1^* > c \tag{83}$$

$$c + 1 > k_2^* > c. \tag{84}$$

As the abatement decisions in the case of two agreements are mutually independent, the total number of abating countries in this case can be found by adding (83) and (84), and thus has to satisfy

$$2c + \alpha_1 + 1 > \alpha_1 k_1^{**} + k_2^{**} > 2c. \tag{85}$$

By comparing the equilibrium abating stable coalitions in the case of one single and two coexisting agreements, we find that two agreements lead to a greater number of agreement members as well as to a greater amount of global emissions abatement and welfare. This effect would be replicated for any larger number of admitted climate clubs. It is caused by the coalitions' and the outsiders' dominant abatement strategies that stem from the linear payoff-structure of the model.

As shown by Hagen and Eisenack (2015), linear benefits of abatement always lead to dominant abatement strategies, while other cost and benefit structures from emissions abatement may lead to non-dominant reaction functions. In the extreme case of linear costs and concave benefits from abatement, only one agreement would undertake emissions abatement while all other countries do not abate any emissions regardless of their potential membership in other agreements. The findings of Eisenack and Kähler (2015), who show that individual countries with convex benefits from abatement may have increasing reaction functions so that emissions abatement becomes a strategic complement, give rise to the question about the strategic behaviour of clubs that consist of such countries. In light of the previous analysis and the already existing economic literature, we may conclude that climate clubs improve the outcomes of climate negotiations in some cases. Even in the least desirable cases we find that the outcome of negotiations with climate clubs leads to the same amount of global emissions abatement as would be achieved with one single IEA.

4.4.2 City alliances

Cities are important actors regarding global climate change, both on the emitting and on the damage side. It might generally make sense that they organize an alliance among themselves in order to tackle these problems. In our proposed model we focus on the economic arguments of vulnerability, local co-benefits and enforceability.

The problem of enforcing an environmental agreement can be greatly diminished as cities are not 'above the law' like nation states in the international system. They can be bound to abide to contracts by national laws. This makes trust, compliance and enforcement less challenging problems. Generally, there are political, social and cultural links between rural and urban areas of one country. Additionally, a city alliance can introduce a voluntary and legal link between urban areas of multiple countries. The combination of these links might yield more cooperation than the usual economic approach of considering only a voluntary and self-enforcing agreement between countries.

Cities are potentially more vulnerable to climate change than other regions (Hallegatte and Corfee-Morlot, 2011). Therefore they have stronger incentives to reduce climate change impacts. There can also be local co-benefits in mitigation, e.g. the removal of air pollution (Bollen et al., 2009; Harlan and Ruddell, 2011) or a specialization on business opportunities from technological solutions like green energy (Jochem and Madlener, 2003). Particularly early movers may have an advantage here.

For technical reasons, we characterize the actors in this section by their benefits and damages from emissions (in contrast the model specifications in section 4.4.1). In our model each country i consists of one $city$ and one $rural$ region. The payoff of each city $\pi^i_{city}(e^i_{city}, e) = B^i_{city}(e^i_{city}) - D^i_{city}(e)$ and each rural region $\pi^i_{rural}(e^i_{rural}, e) = B^i_{rural}(e^i_{rural}) - D^i_{rural}(e)$ depends on the benefits B^i_{city}/B^i_{rural} from its own emissions e_i, and, as usual, on the damage D^i_{city}/D^i_{rural} from global emissions $e = e^i_{city} + e^i_{rural} + e^{-i}$. The local emissions are an essential (but partly substitutable) factor of industrial production; they are linked to local benefits. Global emissions change the climate, which in turn creates local damages. In line with standard IEA literature (e.g. Hoel, 1991), we assume for all regions positive but decreasing marginal benefits from local emissions $B^{i\prime}(e^i) > 0, B^{i\prime\prime}(e^i) < 0$ and positive and increasing marginal damages from global emissions $D^{i\prime}(e) > 0, D^{i\prime\prime}(e) > 0$.

We further assume the following properties of the benefit and damage functions:

$$D^{i\prime}_{city}(e) > D^{i\prime}_{rural}(e), \tag{86}$$

$$B^{i\prime}_{city}(e^i_{city}) > B^{i\prime}_{rural}(e^i_{rural}), \tag{87}$$

$$B^{i\prime\prime}_{city}(e^i_{city}) > B^{i\prime\prime}_{rural}(e^i_{rural}). \tag{88}$$

The first property corresponds to the comparatively higher vulnerability of cities. The second and third inequalities result from assuming local co-benefits from emissions reductions in cities (e.g. lower air pollution or a head start in green technology development). These co-benefits compensate for the loss of benefits from emissions reduction, and therefore, lead to a lower net loss of benefits from local greenhouse gas production.

The model comprises two stages: First, each city decides whether it wants to participate in the TEA by entering an alliance with all other willing cities. Second, each country decides on the emission level of its city and rural region. The entry decision $(c^i \epsilon \{A, \neg A\})$ in the first stage is based only on the payoff of the city: Is π^i_{city} higher as an alliance member? The payoff of the rural regions or the other cities does not enter consideration here.

In the second stage of the game, countries choose the emissions that maximize their respective payoffs Π^i. If the city region of a country has entered an alliance, we assume that the country considers the damages to foreign cities of the alliance $D^{A\backslash i}_{city}(e)$ to some degree. This works similar as in stable agreements between nation states that fully internalize all damages from the emissions to all other agreement members. The degree of internalization of foreign cities in an alliance where domestic cities are members is represented by a weight $x \epsilon]0, 1[$ because cities may not be able to force their national governments to fully integrate a city alliance into their emissions planning.

The optimization problem of each country i in the second stage is:

$$\max_{e^i_{city}, e^i_{rural}} \Pi^i(e^i_{city}, e^i_{rural}, e^{-i}) \tag{89}$$

$$= \begin{cases} \text{if } c^i = \neg A & : \quad \pi^i_{city}(e^i_{city}, e) + \pi^i_{rural}(e^i_{rural}, e) \\ \text{if } c^i = A & : \quad \pi^i_{city}(e^i_{city}, e) + \pi^i_{rural}(e^i_{rural}, e) - x \cdot D^{A\backslash i}_{city}(e) \end{cases} \tag{90}$$

We assume that all countries simultaneously play a Nash game at this stage. In the first stage all cities determine membership simultaneously in a Nash game

$$\underset{c^i \epsilon A, \neg A}{max} \pi^i_{city}(e^i_{city}, e) = B^i_{city}(e^i_{city}) - D^i_{city}(e) \tag{91}$$

It is obvious that an alliance between cities is easier to reach than an agreement between countries. Due to their high vulnerability, cities value emissions reductions more; at the same time, they are more likely to accept emission reductions because they have lower marginal benefits from emissions.

The largest part of the emissions reductions (in comparison to a status without any agreement) is borne by the cities in the alliance, because their benefits are reduced least if they lower emissions. They also have the largest reduction in damages. The rural areas (of the countries in which the cities are in the alliance) have to make some emission reduction effort as well, but their main contribution is not allowing any leakage. In a negotiation that only allows for nation states to form an agreement, even rural areas might want an agreement, but ree rider incentives are much higher for them than for cities. Therefore they would prefer others to form an agreement and stay singletons themselves.

The national government is important in our model insofar as it ensures that no leakage of 'dirty' industry from cities to rural areas occurs. Of course, the willingness of governments to engage in local climate policy is important as well. However, in this model they don't have to enforce large emission reductions in (unwilling) rural areas, they only have to prevent them from increasing their emissions. Maintaining a status quo is more feasible in many political cases than enforcing unwanted change.

We conclude that if cities can form a mitigating alliance which national governments consider to some degree in their policy decision making, more cooperation and larger emission reductions can result. Cities have an incentive to enter a city alliance because they expect higher damages from climate change, and have lower costs of emission reduction than other regions (particularly taking into account co-benefits from greenhouse gas mitigation).

4.4.3 Outlook

There are many further approaches to transnational environmental agreements in addition to the analysis of those proposed above. We think that they offer promising extensions of the state of the art in research on international environmental agreements. We sketch some of them in the following.

Concerning climate clubs, one could think of overlapping clubs as an alternative to the proposed setting of coexisting disjoints clubs. If countries would be signatories of more than one climate club, this would change the strategic interaction of the clubs and possibly also the reaction functions in the game. Another way to include climate clubs as disjoint coalitions in the climate negotiations is to allow countries to form sub-coalitions in a first stage, followed by multilateral negotiations between the coalitions and remaining nonsignatories. Possible effects of climate clubs in a broader sense include the generation of club-benefits as proposed, for instance, by Nordhaus (2015). By the creation of such benefits that only favour signatories of a climate agreement, the incentives to join are strengthened. This could be implemented through issue-linkage. Existing international agreements on other topics as, for example trade, would then be linked to climate agreements. Existing research on IEAs and trade (e.g. Eichner and Pethig, 2015) could serve as a starting point here. Such multi issue clubs as well as climate clubs that do not negotiate on emission reductions but other issues like monitoring or technology sharing are a challenging but interesting modelling task. With regard to transaction costs we can say that, on the one hand, a shift towards smaller clubs of negotiating countries could possibly lower the transaction costs of forming a climate agreement, while possible interactions between clubs could impose additional transaction costs.

There are several economic arguments for an alliance between cities for emissions reductions. In our modelling approach we use vulnerability, local co-benefits and enforceability. In addition to these assumptions, we suggest three more possible reasons in favour of city alliances. First, transaction costs are potentially lower. The implementation of policy measures might be easier on a subnational than on a national level. Second, there is presumably less reason to behave opportunistically in moral hazard situations. The problem of individually rational but collectively harmful behaviour can be reduced if people directly observe each other. It might even be argued that urban areas are more likely to have a clientele that shares common norms, such as a collective commitment to behave responsibly and to abstain from opportunistic behaviour. Within such a group, information asymmetries are less problematic in a moral hazard configuration. Third, there can be learning effects. Transfer of policies between cities or even from a subnational to a national level could be modelled.

Our modelling proposal for city alliances can be combined with research on climate clubs. Cities within countries with low ambition could join climate clubs and exert their influence on the respective countries to join such agreements and take climate action. We actually observe that there are multiple city alliances in place, so these are, in our terminology, coexisting climate clubs of cities. What is the rationale and environmental effectiveness of cities forming coexisting TEAs, and how might cities strategically interact with national governments in heterogeneous TEAs where both cities and countries are members?

Apart from city alliances and climate clubs, there are many other actors that could participate in TEAs. Non-state actors play an important role for adaptation to climate change as well as for mitigation of emissions. Industry lobbies and transnational NGOs influence governments and groups of countries in different ways while subnational governments and internal politics also play an important role for the decisions national governments take.

Involving these actors in transnational agreements might open up new possibilities for negotiations and climate action but also raise threats to effective agreements. Whether they are within an agreement between nation states or within a coalition only consisting of non-nation state actors, their interests differ substantially so that the effects of heterogeneity on their outcomes are not clear. These effects should not be neglected and deserve more attention in further research.

4.5 Conclusions

This chapter provides an exploration of some transnational initiatives for climate cooperation. The global governance literature finds ample empirical evidence for emerging TEAs. These can be only partially explained by the conventional economic literature that emphasizes the role of nation states with freerider incentives. We thus propose that more research is needed to understand and evaluate the role of TEAs in order to contribute to deal with climate change. We argue that this particularly requires to consider the strategic interaction of heterogeneous actors, not only nation state governments, and to consider coexisting and possibly overlapping contracts that stipulate emission reductions or other institutions that are conducive to this aim.

To illustrate and underpin this claim, we extend already existing game theoretic approaches to IEAs in order to analyse the strategic effects of TEAs. Our two examples show that both climate clubs and city alliances may be able to lead to an increase in emissions abatement and in global welfare. Climate clubs offer an opportunity to cooperate in more than one agreement at the same time. Cities can form alliances in which they agree to mitigate greenhouse gases; the effectiveness of such TEAs will depend on the political influence cities have on national governments.

We find that cooperation can be individually rational, even in the presence of freerider incentives. Depending on the characteristics of the actors, negotiation structures can facilitate cooperation. Multiple agreements, for example, can stimulate more countries to cooperate than a single IEA. National political and legal institutions can be used to avoid the problem of non-binding agreements if actors other than nation states cooperate. Cities, rural regions and other subnational actors can be compelled by law to enact an agreement. Both examples of TEAs have shown that such agreements may indeed be effective and improve over the standard single IEA consisting only of nation states. Depending on the structure of costs from mitigation efforts and damages from climate change, the example of climate clubs shows that it is not in any case clear if TEAs take climate action beyond lip service.

Beyond these two examples there are various other settings of heterogeneous actors that might be conducive to tackle climate change. Other forms, mechanisms and players in TEAs, like NGOs, issue linkage, policy learning, moral hazard and political economy warrant further attention. Also cooperative game theory may be used to model TEAs.

Although we have shown that game theoretic analysis might well be helpful to better understand the formation and effects of TEAs, it is clear that it also has its limitations. Some aspects like the re-legitimation of the climate regime (cf. Falkner, 2015) or potentially irrational behaviour of agents are difficult to analyse in a game theoretic setting and might be better researched by other means. One can also question the legitimacy of TEAs with non-state actors in contrast to multilateral IEAs negotiated by national governments. Nevertheless, we argue that especially with regard to the slow progress of the international climate negotiations, and in light of the empirical development already going on, it is important to include non-state actors complementary to an IEA.

Non-cooperative game theory offers a conservative view on agreements, i.e. it tends to underrate cooperation incentives (Carbone et al., 2009). Therefore our positive findings carry a particularly heavy meaning; we expect a real potential for TEAs. Institutions and negotiation structures for climate governance can improve if they allow for transnational actors. A combination of different scientific approaches sharpens the view. The global governance literature widens the horizon for economic analysis and challenges the conventional theory of IEAs, as it offers observations that cannot easily be explained by existing models. This is both a provocation and great opportunity for further theory building. Economics offer rigorous methods for the analysis of incentives for cooperation, and model results can give new ideas for TEA structures and negotiation processes.

Understanding TEAs is of the highest importance, particularly in the light of the Paris agreement of 2015 which does not provide binding emission reduction targets for nation states. This challenges both the negotiating actors and research. Our study sketches several promising policy options and avenues for further research.

Interlude III

IEAs are probably not sufficient to limit climate change to an acceptable level, at least in the current form. In Hagen et al. (2017) we have shown concepts of transnational solutions which potentially can improve on the performance of IEAs. Some of them are merely avenues for further research while others are already put into practice. Theoretical research on the latter is still important because we know little about their impact. Understanding the effect that they will have on other actors, global emissions and welfare is important for their development.

City alliances seem particularly promising. While they can probably not replace agreements on the nation state level, they can probably complement them very well. From a practical perspective one important point about city alliances is that they already exist. C40, the Covenant of Mayors and others are alliances with many large cities.

From a theoretical perspective there are two main points which favour city alliances. Cities are important actors in global climate policy because they are drivers of a large part of global greenhouse gas emissions. Emission heavy industry and electricity production are often located in cities or close to them. Transport of many people is often emission heavy as well, particularly in areas with a lot of traffic congestions (which also tend to be in and around cities).

Cities are often particularly vulnerable to climate change damages. Many of them are located directly at the coast of the ocean or major rivers so flooding is a real risk. The dense population leads to valuable infrastructure that is both expensive and vital to many people.

So it is possible to accept city alliances as an institution which is already starting to take action in emission reductions. Doubts about the existence or stability of city alliances can be put aside reasonably. The question of agreement size and behaviour towards non-members is not trivial, though. Membership is increasing but it is not self-evident whether all major municipalities will be part of one city alliance or another soon. The status of smaller towns is even more at question. If cities get a voice in transnational climate policy by virtue of collaboration, what becomes of municipalities that are not part of such an agreement - whether by their own choice or by the institutional design of the city alliances? Representation of citizens could be skewed on the international level.

The legal status of cities is not as easy as it may seem at first glance, too. On the one hand they are subnational entities. In this regard they are bound by national law and often have legal power of their own to some extent, mostly in the form of government on the community level. This power however is restricted by the constitution on the national level, i.e. a higher hierarchy. On the other hand cities take action in an international context in a way that is similar to nation states. This is not unprecedented historically but is not easy to accommodate in the current system of international law (cf. Aust, 2015). On an international level only nation states have formal sovereignty. Cities have been allowed to form contracts of their own (particularly more recently in the environmental context) but the formal binding power of these treaties is not necessarily equal to treaties between nation states.

In the following I will deal with some specific questions about what the effects of a city alliance could be. I focus on emissions but take welfare considerations into account as well. My central question is: What happens if many cities come together and reduce emissions by a significant amount? How will national governments react? Will they claim the emission reductions of cities as part of their own pledges to the international community? Or will they take them as additional and still reduce emissions in other parts of their respective countries? How will other subnational regions react? Will these other (rural) regions act as freeriders and welcome the carbon-heavy industry that is now virtually banned from the cities? Or will they follow the example of the cities and reduce their own emissions? There are good arguments either way.

5 The Effects of a City Alliance on Rural Emissions

Abstract

The problem of greenhouse gas-induced global warming is best addressed in cooperation of all relevant authorities on a global level. However, international cooperation is obviously difficult to agree upon. While national governments are negotiating reduction targets, subnational actors already sign agreements like city alliances. Other subnational actors that are not part of an agreement are still influenced by such actions in climate policy. On the one hand they enjoy the lower global emissions. On the other hand their industry is affected by a decline in dirty technology use in city regions that they are politically and economically connected with. This article analyses the effect of emission reductions by a city alliance on the emissions of such non-member regions that are connected to member regions. I find that such regions act as freeriders if the city alliance engages in modest mitigation but start following the example of the city alliance when the agreement members engage in more ambitious emission reductions.

5.1 Introduction

Anthropogenic climate change is a problem that needs to be addressed by coordinated action. Single actors are unable to make a difference. Therefore the international community needs to find a way to mitigate greenhouse gas emissions. This is what conventional wisdom tells us and what seems sensible. And indeed there are several agreements in which many countries have pledged to lower their emissions. The Paris Agreement from 2015 seems to have finally brought all important parties together.

International environmental agreements (IEAs) like this one are already widely studied in various scientific fields. Economic literature mostly uses analytic modelling (e.g.

Barrett, 1994; Carraro and Siniscalco, 1993) and calibrated simulation (e.g. Nordhaus and Yang, 1996) to describe them and predict results. Links with other questions have been made; topics like trade (e.g. Eichner and Pethig, 2013; Helm et al., 2012) and adaptation to climate change (e.g. Zehaie, 2009; Eisenack and Kähler, 2015) have become more prominent in recent years.

However there are doubts that cooperation on the nation state level will be enough. One of the main concerns is that the emission targets of the Paris Agreement are not binding. If it turns out that the UNFCCC approach is indeed insufficient to reduce climate change to an acceptable level[9], other measures will have to be taken in order to avoid devastating consequences for the global ecosystems, humans, cultural sites and landscape.

Transnational approaches to the greenhouse gas problem are already being considered in the social sciences (e.g. Betsill and Bulkeley, 2006; Bansard et al., 2016), economic literature is lagging behind. It is important to realise that national governments are not the only important actors and that successful agreements can take on diverse forms (Hagen et al., 2017). Subnational governments (like cities), supranational entities (like the EU), lobby groups, NGOs and others influence the UNFCCC negotiations and formulate agreements of their own (Hagen et al., 2016; Dietz et al., 2012).

A prominent example for coordinated responses to climate change are city alliances. Will they be able to make a meaningful contribution? One important issue here is the interaction with other actors. Will cities that join a city alliance and pledge greenhouse gas mitigation crowd out other actors, like their own national and regional governments or initiatives in adjacent rural areas of the same country, who will then mitigate less? Or will they be frontrunners of climate protection and others will follow their lead? There are

---

[9] Such a level is not universally agreed upon. 2 degrees Celsius are commonly called acceptable, while 1.5 degrees are also in the debate. Both numbers seem quite arbitrary since we can only guess at the 'tipping points' in the global ecosystem.

good arguments either way.

The case for freeriding is mainly the well-known carbon leakage: if some region (here: urban) makes an effort (jurisdictive or otherwise) to reduce emissions unilaterally, carbon heavy industry is expected to move jurisdictions more favourable to them. These other jurisdictions have little incentive to limit the influx of carbon-heavy industry.

On the other hand geographically close and economically (and politically) connected regions profit from similar or at least complementary technology use. Therefore rural areas could benefit from following the example of the original urban mitigator in adopting 'green' technology as well. They would then lower their emissions even without joining an agreement.

In my model I find that both effects interact in a way that crowding out dominates for high emissions from the city regions and complementarity-induced quasi-cooperation dominates for low city emissions. In other words the rural regions emit more if the city alliance lowers emissions by a little but if the city emissions sink sufficiently, then the rural regions lower theirs as well. There is one unique tipping point between these two types of behaviour which depends on the closeness of the complementarity effect and other parameters of the payoff function of the rural regions. This means that once a city alliance lowers emissions sufficiently, the connected rural regions follow their lead and wil not turn back to freeriding.

In this article I consider the main economic drivers of emissions in rural regions. There are, of course, other factors (cultural, social etc.) which are not included here. Also I analyse only the consequences of the emergence of a city alliance, not the reasons thereof.

In the following section I present the model starting with the main premises. There is a detailed discussion of complementarity of greenhouse gas emissions in a city region and a geographically close and economically (and politically) connected rural region because it is essential to the model and novel in the field of environmental economics. After that I explain the functional forms and specifications of the model and solve it. Then there are two short sections on extensions on independent freeriding and welfare considerations. I conclude with an interpretation of the results and a summary.

5.2 Model

The central question of this article is: How do rural regions change their emissions if cities coordinate and lower their emissions?

We do not know yet how the interaction of a city alliance with other actors (like the UNFCCC level) in the 'game' of emissions work (e.g. Bulkeley, 2010). City alliances are quite diverse with respect to membership requirements, emission reduction commitments, monitoring and other factors (Bansard et al., 2016). However there are some with substantial mitigation ambitions, namely the Compact of Mayors and the Climate Alliance (Bansard et al., 2016). The question whether these commitments are additional to national emission reduction targets (e.g. under the UNFCCC framework) or whether they are claimed by their respective national governments is crucial. If a city alliance can claim additional emission reductions, the overall effect could be quite powerful. If on the other hand unilateral emission reduction by the city regions lead to less ambitious climate policy in other regions of the countries involved, a city alliance has probably very little impact overall.

5.2.1 Basics

The world in the model consists of an equal number of city and rural regions. They are symmetric among each type. Each region has an industry which produces emissions as a byproduct. Higher local emissions (c.p.) lead to increasing benefits, but with decreasing marginal benefits. Each region also suffers damages from global emissions. These damages rise in a convex relation to the global emissions. These model assumptions are very similar to those common in economic literature on climate change agreements (e.g. Hoel, 1991; Barrett, 2001; Altamirano-Cabrera and Finus, 2006; Asheim et al., 2006).

What makes my model special is the additional assumption that each rural region is connected to one city region; the productivity of emissions in the rural region (positively) depends on the emissions level of the city region. This is discussed in detail in section (5.2.2).

I take an alliance of the city regions as given and only look into its effect on rural non-members. There are two main reasons for this. First it allows me to have a certain amount of complexity in the functional forms and interactions that would make the solution of an additional game stage with agreement formation very difficult. So the simplification allows me to focus on the stage and question that I'm interested in. Second in the real world there already are city alliances with regard to climate change action (e.g. C40, the Compact of Mayors and the Covenant of Mayors for Climate & Energy). Therefore the question whether such an alliance would form or be stable is already answered by observation of the actual political situation. The effect of their actions is not so clear yet because emission reductions are just starting.

Starting with an exogenous city alliance I look into the effect of changes in their emission level on the emissions of the (rural) non-members. I assume that the non-member regions act individually rational, taking the emissions of the cities and all other rural regions as given. Therefore I model their actions as a Nash game (basically a subgame after the city regions have formed an alliance and decided on their emissions), again in line with the literature on IEAs (e.g. Barrett, 1994; Carraro and Siniscalco, 1993).

5.2.2 Complementarity

A central assumption of this article is that emission reductions in urban regions create co-benefits to mitigation in the respective connected rural regions. Here I want to explain and justify this premise.

Why should production functions be complementary between urban and (connected) rural regions? Generally, similar production methods mean similar machinery and intermediate products, which in turn leads to cheaper purchasing and more efficient production of these. Also there is less demand for energy intensive goods as environmental consciousness rises in the cities.

Kessides (2006) finds that the ”availability of urban-based activities” can be economically beneficial for the rural hinterland by ”providing knowledge and resources”. Development of ’green’ technology in cities spreads via knowledge spillovers to adjacent rural regions. This means that green technology becomes more efficient there, too, making dirty technology more expensive in comparison. There is evidence that knowledge spillovers are localized (e.g. Jaffe et al., 1993), primarily intranational in scope (Branstetter, 2001), and aided by urban features like high job density (Carlino et al., 2007). However these claims are not uncontested (e.g. Thompson and Fox-Kean, 2005).

While the historical origin claim ’cities first’ by Jacobs (1969) is archaeologically disproved (Smith et al., 2014), her idea of knowledge spreading out from (urban) centres toward a (rural) hinterland is generally accepted (e.g. Cooke, 2008). Still it is not entirely clear whether specialization or diversification of industries within a regional cluster leads to more local knowledge spillovers (e.g. Feldman and Audretsch, 1999; van der Panne, 2004; van der Panne and van Beers, 2006). The line of literature going back to Jacobs (1969) sees diversification as dominant; knowledge spillovers take place where different industries have complementarities. The line of literature going back to Marshall (1890) argues in favour of specialization; positive externalities work via economies of scale and scope.

Both lines of theory can be used to argue in favour of complementarity within local ’green’ industries. Marshallian specialization obviously favours similar technology directly. Jacobian diversification - though less direct - also needs complementarity of technologies which is arguably higher within ’green’ or ’dirty’ industries rather than between the two.

There are some more arguments in favour of a production function that links productivity of emissions in a region to the emissions used in production in a connected city. Geographically close regions benefit from common infrastructure, which can be green/dirty specific like rail vs. street transport systems, electricity storage (for the integration of ’green’ energy) and a good electricity transport grid (for compensation of fluctuations in renewable energy production). Education in the production and application of certain technologies can be made available for a wider populace with relatively little additional expenditure. The labour market for skilled workers can be integrated better if similar technology is used.

Based on these arguments I assume for the model that higher emissions in a city region increase the productivity of emissions in a connected rural region. The exact origin of this complementarity is not important for the model. I explicitly do not include joint action in emission control but allow for the rural regions to decide their emissions independently, taking into account the complementarity described here but not bound by any agreement.

5.2.3 Specifications

The model uses concrete functional forms for the welfare payoff of rural regions. The variables and functions are specified as follows.

The world consists of $k > 0$ countries with one rural ri and one city region pi each, so $2 \cdot k$ regions total. A city region pi which is partnered to rural region ri has emissions of e_{ci}, which are exogenous in this model. Each rural region ri has emissions e_{ri}, the emissions of each other rural region are e_{r-i}. Global emissions E are the sum of all regional emissions, so $E = k \cdot (e_{ri} + e_{ci})$.

The welfare payoff function of a single rural region has the following form:

$$\pi_{ri}(e_{ri}, e_{ci}, e_{r-i}) = a \cdot e_{ri} \cdot e_{ci}^{b} - \frac{1}{2} \cdot e_{ri}^{2} - d \cdot E^{2} \tag{92}$$

The first term ($a \cdot e_{ri} \cdot e_{ci}^{b} - \frac{1}{2} \cdot e_{ri}^{2}$) is the benefit part, the second term ($d \cdot E^{2}$) signifies damage, both as explained in section (5.2.1). The productivity of local emissions a, the marginal damages d and the connection intensity b fulfil the following conditions:

$$a > 0, \tag{93}$$

$$d > 0, \tag{94}$$

$$0 < b < 1. \tag{95}$$

Note that a higher b signifies a stronger connection between the emission level of the city region and the emissions efficiency of the rural region. I also assume that emissions of each region are positive:

$$e_{ri} > 0. \tag{96}$$

5.2.4 Solution

In this chapter I determine the emissions of the rural regions as a reaction on the emissions of the cities. Results are concrete forms for the reaction function, the slope of the reaction function, a tipping point in the reaction function and an observation about absolute change in emissions by the rural regions.

All cities form the city alliance. They then collectively choose their emissions. Since they are symmetric, an efficient spread of these emissions yields equal emissions e_{pi} for each city. Rural regions then choose their emission levels individually, taking the emissions level of the cities and the emission decisions of the other rural regions as given. Region ri optimizes its payoff as given in Eq. (92):

$$\underset{e_{ri}}{max} \qquad \pi_{ri}(e_{ri}, e_{ci}, e_{r-i}) = a \cdot e_{ri} \cdot e_{ci}^{b} - \frac{1}{2} \cdot e_{ri}^{2} - d \cdot E^{2} \tag{97}$$

Individually optimal emissions e_{ri}^{*} for rural region ri can be found via the first derivative of this payoff function for the local emissions e_{ri}:

$$0 = a \cdot e_{ci}^{b} - e_{ri} - 2 \cdot d \cdot (e_{ri} + (k-1) \cdot e_{r-i} + k \cdot e_{ci}) \tag{98}$$

$$e_{ri}^{*} = \frac{a \cdot e_{ci}^{b} - 2 \cdot d \cdot ((k-1) \cdot e_{r-i} + k \cdot e_{ci})}{2 \cdot d + 1} \tag{99}$$

This emission level e_{ri}^{*} as given in Eq. (99) depends on the (endogenous) emissions of the cities e_{ci} and of the other rural regions e_{r-i} and on (exogenous) parameters like the number of regions. Since the regions are assumed to be symmetric, the Nash game at between the rural regions can be resolved in a manner that yields emissions e_{ri}^{*} as a reaction function $R_{ri}(e_{ci})$ of each rural region ri depending only on the emission level e_{ci} of each city region pi:

$$e_{ri}^{*} = R_{ri}(e_{ci}) = \frac{(a \cdot e_{ci}^{b} - 2 \cdot d \cdot k \cdot e_{ci})}{1 + 2 \cdot d \cdot k}. \tag{100}$$

The denominator of this is positive. Therefore the whole term has the algebraic sign of the numerator. If negative emissions and corner solutions shall be excluded, the following condition has to hold:

$$e_{ci}^{1-b} < \frac{a}{2 \cdot d \cdot k}. \tag{101}$$

To find the slope of the reaction function, I differentiate R_{ri} for e_{ci}.

$$R'_{ri}(e_{ci}) = \frac{\mathrm{d}e_{ri}}{\mathrm{d}e_{ci}} = \frac{(a \cdot b \cdot e_{ci}^{b-1} - 2 \cdot d \cdot k)}{1 + 2 \cdot d \cdot k} \tag{102}$$

This slope of the rural reaction function is critical for understanding the relationship between a city in an alliance and it's rural hinterland. It determines whether the rural region increases or decreases emissions if the city alliance decreases the member emissions. The former behaviour is known as freeriding. I call the latter behaviour 'quasi-cooperation' because on the one hand the rural region in question reduces its emissions if the cities in the alliance reduce theirs. On the other hand the rural region does not sign a treaty or coordinate its emission reductions with any other region in an explicit way. Instead the emission reduction in this special situation follows from self interest.

Proposition 11. *If there are k pairs of one city and one rural region each, and if all city regions are ex ante symmetric, and if all rural regions are ex ante symmetric, and if Eq. (92) gives the payoff for rural region ri, and if conditions (93), (94) and (95) hold, and if all rural regions determine their emissions in a Nash equilibrium, then the slope $R'_{ri}(e_{ci})$ of the reaction function of rural region ri to the emissions of city region ci is*

$$\textit{positive if} \qquad e_{ci}^{1-b} < \frac{a \cdot b}{2 \cdot d \cdot k} \tag{103}$$

$$\textit{negative if} \qquad e_{ci}^{1-b} > \frac{a \cdot b}{2 \cdot d \cdot k} \tag{104}$$

$$\textit{zero if} \qquad e_{ci}^{1-b} = \frac{a \cdot b}{2 \cdot d \cdot k}. \tag{105}$$

If the damage part of the reaction function dominates, the slope is negative and the rural region acts as a freerider: emission reduction by the city region and it's allies will lead to higher emissions in the rural region. If the benefit part dominates, the slope is positive and the rural region quasi-cooperates: the rural region will decrease emissions if the cities do so.

Another finding from proposition (103) is important: It signifies a unique tipping point. If the emissions of the partnered city region are above a certain threshold $\bar{e}_{ci}$ with

$$\bar{e}_{ci} = \left(\frac{a \cdot b}{2 \cdot d \cdot k} \right)^{\frac{1}{1-b}} , \tag{106}$$

then the slope of the rural region's reaction function is negative; if the emissions of the city region are below $\bar{e}_{ci}$, then the slope is positive.

Technically the tipping point is quite similar to condition (101) for positive emissions. The slope of $R_{ri}(e_{ci})$ is always positive if optimal rural emissions are non-negative and b were equal to 1. (If b equals 1, then then there is no tipping point. Trying to calculate one would result in division by 0.)

What is the meaning of this tipping point? If the city regions lower their emissions sufficiently, the rural regions will abandon their (probable initial) strategy of freeriding and switch to semi-cooperative mitigation. The tipping point describend in Eq. (106) signifies the threshold in e_{pi} between these two kinds of behaviour. This threshold between freeriding and quasi-cooperation is higher (i.e. easier to reach for the city regions) if the rural regions have low marginal damages, are few in number, have a high marginal productivity of emissions and if their marginal productivity of emissions is strongly connected to the city emission level.

How do emissions of the rural regions change in absolute terms? So far I have only looked at marginal changes in emissions. It is of course also important to know how they change in absolute terms if the city alliance changes emissions by more than a marginal amount.

Let e_{ci}^{or} and e_{ri}^{or} be the (original) starting point of the city and rural emissions respectively. The terms e_{ci} and e_{ri} are used for the actual emissions (after any potential adjustment by the cities and the reactions of the rural regions).

Lemma 1. *If there are k pairs of one city and one rural region each, and if all city regions are ex ante symmetric, and if all rural regions are ex ante symmetric, and if Eq. (92) is the payoff for rural region ri, and if conditions (93), (94), (95) and (106) hold, and if $R_{ri}(e_{ci}^{or}) > 0$, and if all rural regions determine their emissions in a Nash equilibrium, then there exists a unique $\underline{e}_{ci}$ with $0 < \underline{e}_{ci} < \bar{e}_{ci}$ for which the following holds:*

$$\text{If} \quad e_{ci} < \underline{e}_{ci} \quad \text{then} \quad R_{ri}(e_{ci}) < R_{ri}(e_{ci}^{or}). \tag{107}$$

This means that even if the city emissions start at an amount high enough that $R'_{ri}(e_{ci})$ is negative, the absolute change in rural emissions will be negative if the cities change their emissions by a sufficiently large amount.

5.2.5 Independent Freeriding

How does the situation change if there is an independent freerider? For this section I assume that there exists player which does not reduce emissions together with the cities and is not connected economically to them (like the rural regions) either. Instead this player freerides on emission reductions of the others.

Since I don't want to distort results by making the world larger, I reduce the number of rural regions by 1 at the same time as adding the new player. Effectively I now assume that the special connection from city emissions to productivity of rural emissions does not hold for one special rural region. Global emissions E in this case of course include emissions from the freerider region, i.e. $E = k \cdot e_{ci} + (k-1) \cdot e_{ri} + e_f$.

The independent freeriding region f decides on its emissions at the same stage as the other regions that are not part of the city alliance. It is therefore part of the Nash game that the non-members play. In the equilibrium of this Nash game the independent freerider has a reaction function $R_f(e_{ci})$ which is negatively sloped. [10]

$$-1 < R'_f(e_{ci}) < 0 \tag{108}$$

In this game which includes an independent freerider the representative rural region ri then has the following payoff function:

$$\pi^f_{ri}(e_{ri}, e_{ci}, e_{r-i}, e_f) = a \cdot e_{ri} \cdot e^b_{ci} - \frac{1}{2} \cdot e^2_{ri} - d \cdot E^2 \tag{109}$$

Note that this looks almost the same as Eq. (92) but global emissions E include e_f for this case as stated before. Each rural region ri now has the following reaction function $e^{f*}_{ri} = R^f_{ri}(e_{ci})$ in the Nash subgame equilibrium:

$$e^{f*}_{ri} = R^f_{ri}(e_{ci}) = \frac{(a \cdot e^b_{ci} - 2 \cdot d \cdot (k \cdot e_{ci} + R_f(e_{ci}))) \cdot (2 \cdot d \cdot (2 - k) + 2)}{(2 \cdot d + 1)^2 - (2 \cdot d \cdot (k - 2))^2}. \tag{110}$$

Note that in comparison to the basic case in Eq. (100) there is an additional term $(R_f(e_{ci}))$ and that there are other changes to account for the missing rural region that was displaced by the new freerider.

The slope of the reaction function of the representative rural region now is:

$$R^{f\prime}_{ri}(e_{ci}) = \frac{\mathrm{d}e_{ri}}{\mathrm{d}e_{ci}} = \frac{(a \cdot b \cdot e^{b-1}_{ci} - 2 \cdot d \cdot (k + R'_f(e_{ci}))) \cdot (2 \cdot d \cdot (2 - k) + 2)}{(2 \cdot d + 1)^2 - (2 \cdot d \cdot (k - 2))^2}. \tag{111}$$

Like in the basic scenario the algebraic sign of this slope is important.

Proposition 12. *If there are k city regions, $k-1$ rural region and one freerider region, and if all city regions are ex ante symmetric, and if all rural regions are ex ante symmetric, and if Eq. (109) is the payoff for rural region ri, and if conditions (93), (94) and (95) hold,*

---

[10] Following standard literature (e.g. Hoel, 1991) I assume that the freerider does not overcompensate, i.e. does not react so strongly that the overall result (in global emissions) is a reversal of the original change.

and if all rural regions and the freerider region determine their emissions in a Nash equilibrium, then the slope $R^{f\prime}_{ri}(e_{ci})$ of the reaction function of rural region ri to the emissions of city region pi is

$$\textit{positive if} \qquad e_{ci}^{1-b} < \frac{a \cdot b}{2 \cdot d \cdot (k + R'_f(e_{ci}))} \tag{112}$$

$$\textit{negative if} \qquad e_{ci}^{1-b} > \frac{a \cdot b}{2 \cdot d \cdot (k + R'_f(e_{ci}))} \tag{113}$$

$$\textit{zero if} \qquad e_{ci}^{1-b} = \frac{a \cdot b}{2 \cdot d \cdot (k + R'_f(e_{ci}))}. \tag{114}$$

Since $R'_f(e_{ci}) < 0$ by definition, (112) is a weaker condition for a positive slope of the reaction function of $R^{f\prime}_{ri}$ than (103), i.e. the threshold is at a higher e_{ci}.

The reason behind this is that the freerider increases its emissions as the cities reduce theirs. Therefore the marginal damage of of the each other rural region ri (which is $2 \cdot d \cdot E$) sinks by less (if the cities reduce emissions) than in the scenario without the freerider. The resulting relatively higher marginal damage in turn makes decreasing own emissions more attractive for those rural regions that are partnered to city regions because optimal emissions e^*_{ri} depend negatively on the marginal damage.

5.2.6 Welfare

What effect do emission reductions of the cities have on welfare in the rural regions? I assume that the cities in the alliance choose an emission level for its members that is optimal for their welfare. However if the actual emission level of the cities is lowered (e.g. in comparison with a status quo before the cities form the alliance), how does this then reflect in the payoff of the connected rural regions?

There are competing influences. The lower global emissions lead to lower damage, which affects the welfare positively. On the other hand the lower emissions of the partnered city region lead to a reduction in benefits from domestic emissions in each rural region. Then there is the reaction of the rural regions which potentially changes both benefits and damages.

The payoff of a (representative) rural region takes the following form if the reactions of the rural regions are endogenized:

$$\pi_{ri}(e_{ci}) = a \cdot R_{ri}(e_{ci}) \cdot e_{ci}^{b} - \frac{1}{2} \cdot R_{ri}(e_{ci})^2 - d \cdot k \cdot (R_{ri}(e_{ci}) + e_{ci})^2 \tag{115}$$

To study the effect of an emission change by the cities, the derivative with respect to city emissions is of interest.

$$\begin{aligned}\frac{\mathrm{d}\pi_{ri}(e_{ci})}{\mathrm{d}e_{ci}} =& a \cdot b \cdot e_{ci}^{b-1} \cdot R_{ri}(e_{ci}) + a \cdot e_{ci}^{b} \cdot R'_{ri}(e_{ci}) - R_{ri}(e_{ci}) \cdot R'_{ri}(e_{ci}) \\ & - 2 \cdot d \cdot k \cdot (R_{ri}(e_{ci}) + e_{ci}) \cdot (R'_{ri}(e_{ci}) + 1)\end{aligned} \tag{116}$$

Unfortunately the algebraic sign of this term is ambivalent. It is not even necessarily monotonic. Therefore the influence of an emission reduction by the cities on the welfare of the connected rural regions remains generally ambiguous.

One interesting detail is clear however: At the point $\bar{e}_{pi}$ (i.e. where $R'_{ri} = 0$ and where the rural regions change their behaviour between freeriding and quasi-cooperation) the derivative is negative ($\frac{\mathrm{d}\pi_{ri}(e_{ci})}{\mathrm{d}e_{ci}} < 0$). This means that at this point (and presumably in the close vicinity thereof) reductions in emissions by the city regions increase the welfare of the rural regions.

5.3 Interpretation

The effect of a unilateral emission reduction by regions of a city alliance on connected rural regions is ambiguous. For high city emissions the 'partnered' rural region has a negatively sloped reaction function. This means that it increases emissions if the city region unilaterally decreases its own emissions by only a little. For low city emissions the 'partnered' rural region has a positively sloped reaction function, i.e. it imitates a unilateral emission reduction by the city region.

The tipping point between the two behavioural states of the rural region is unique, which means that once the emissions of the partnered city region are low enough, further reductions will always result in an emission reduction by the rural region. This tipping point is high (i.e. easy to reach by relatively small emission reductions) if it is strongly connected to the city. If the productivity of emissions in a rural region strongly depends on the emissions level of the connected city region, then the rural regions can be expected to quasi-cooperate even at moderate mitigation by the cities.

Some policy conclusions can be drawn here. If a rural region freerides on small emission reductions this does not mean that it will continue to do so for larger emission reductions. If the reductions become larger (and the connection is strong enough) the incentive to freeride will decrease significantly and will eventually even turn into an incentive for mitigation (as compared to the business as usual before the unilateral action). Therefore it can be advisable not to abandon unilateral emission reductions even if some (local) carbon leakage is detected. If the emission reductions by the city alliance are strong enough the rural regions will not only stop freeriding but eventually their emissions can even be expected to sink below the level they had before the first emission reductions.

How does the tipping point depend on the parameters?[11] A rural region switches between freeriding and imitation of mitigation if the emissions of the partnered city region fall below a specific point. This point is high (i.e. easy to reach by relatively small emission reductions) if it is strongly connected to the city and if the rural region faces low marginal

---

[11] The answer to this question is highly sensitive to the model specification like functional forms (in particular the multiplicative form of complementarity and a constant exponent) and symmetry between regions. Note that therefore generality can not be claimed for this part of the results.

damages from global emissions and if the rural regions has relatively high productivity of emissions (this is non-intuitive, the reason is a multiplication of local productivity with the influencing factor in the model, could be an artefact).

This means that we can expect more cooperative behaviour (i.e. switching from freeriding to quasi-cooperation in mitigation at a still-higher level of city emissions) if the rural economy in question is connected strongly to the city which reduces emissions, if its marginal utility of emissions (which it looses if the city reduces emissions) is high and if the marginal damages are low (because then the damage reduction induced by the original emission reduction is low, which leads to a low incentive to increase the region's own emissions).

5.4 Summary

Climate action on a subnational level in general and city alliances in particular are an obvious and well-liked alternative to more conventional international climate protection treaties. There is hope that urban actors can succeed where national governments might not. City alliances are already forming and growing. Some of them are quite large and have ambitious emission targets. In the absence of binding emission targets in more traditional international environmental agreements the possibility of a solution on a subnational level gains importance.

In my model I have looked into the effect of such unilateral action in climate protection by a city alliance on other actors - (rural) regions that are not part of the agreement but which are politically and economically connected to individual member cities. I find that for modest emission reductions we can expect the rural regions to act as freeriders, taking the damage reductions from the city alliance's action as a windfall and even inceasing their own emissions. However if the emission reduction of the city alliance is strong enough, the economic connection of the rural regions to the (close) cities that are members of the mitigating alliance gives them a stronger incentive to follow suit and lower their emissions, too.

There is a unique tipping point between the two types of rural reaction. This means that once the rural (non-member) regions start to react quasi-cooperatively to unilateral action by the cities, further emission reductions by the city alliance will lead unambiguously to lower emissions from the rural regions as well. Connected rural non-members will never go back to increasing emissions if the city alliance lowers emissions even more.

This is true not only with regard to the effect of marginal changes in city emissions; the rural regions lower their emissions also in absolute terms if the reductions by the cities are sufficient. Therefore once the emissions of the city alliance are low enough, the rural (non-member) regions will unambiguously decrease their emissions as well.

If there is a third kind of region which does not belong to the city alliance and is not economically connected to a member, then this region will always act as a freerider. This region enters the Nash game of the non-members and changes the incentives of those rural non-members that are connected to the cities. If the city alliance lowers emissions, the damage and marginal damage of the connected rural regions does not sink as much as in the basic scenario (without the additional freerider) while the incentive to follow the mitigating policy of the 'partnered' city regions stays the same. This in turn moves the tipping point to a higher level of city emissions, which means that rural regions (which are connected to city alliance members) start to lower their emissions already for less ambitious emission reductions by the city alliance.

The effect of city alliance emission reductions on the welfare of rural regions is ambivalent. On the one hand damage decreases, on the other hand benefits from local emissions sink. Only at the tipping point I find that a marginal reduction in city emissions leads to higher welfare in the rural regions unambiguously.

Overall the coordinated emission reductions by a city alliance can lead to emission reductions by (non-member) rural regions, too, if the original reductions are large enough and if the economic connection between city and rural regions is strong enough. If we observe that rural regions act as freeriders that needs not necessarily discourage further reduction ambitions. Instead we can expect quasi-cooperation if the impulse of emissions reduction by the city alliance gets large enough.

Appendix to 'The Effects of a City Alliance on Rural Emissions'

5.A Proof of Prop. 11

Proof. First, maximize $\pi_{ri}(e_{ri}, e_{ci}, e_{r-i})$ as given in Eq. (92) with respect to e_{ri}. The first order condition is satisfied only for e_{ri}^* is given in Eq. (99), namely

$$e_{ri}^* = \frac{a \cdot e_{ci}^b - 2 \cdot d \cdot ((k-1) \cdot e_{r-i} + k \cdot e_{ci})}{2 \cdot d + 1}. \tag{117}$$

The second order condition is fulfilled as well because $\pi_{ri}(e_{ri}, e_{ci}, e_{r-i})$ is strictly concave in e_{ci}:

$$\frac{\mathrm{d}\pi_{ri}(e_{ri}^*, e_{ci}, e_{r-i})^2}{\mathrm{d}^2 e_{ri}} = -1 - 2 \cdot d < 0. \tag{118}$$

Therefore e_{ri}^* represents the unique local maximum of $\pi_{ri}(e_{ri}, e_{ci}, e_{r-i})$ for each given combination of e_{ci} and e_{r-i}. Since there are no limits on e_{ri} at this point, e_{ri}^* is also the global maximum.

Now, examine the Nash equilibrium. Due to symmetry the Nash equilibrium of all rural regions in emissions is determined by $e_{ri}^* = R_{ri}(e_{ci})$ as given in Eq. (100), namely

$$e_{ri}^* = R_{ri}(e_{ci}) = \frac{(a \cdot e_{ci}^b - 2 \cdot d \cdot k \cdot e_{ci}) \cdot (2 \cdot d \cdot (2-k) + 1)}{(2 \cdot d + 1)^2 - (2 \cdot d \cdot (k-1))^2}. \tag{119}$$

The derivative of $e_{ri}^* = R_{ri}(e_{ci})$ for e_{ci} is given in Eq. (102):

$$R_{ri}'(e_{ci}) = \frac{\mathrm{d}e_{ri}}{\mathrm{d}e_{ci}} = \frac{(a \cdot b \cdot e_{ci}^{b-1} - 2 \cdot d \cdot k) \cdot (2 \cdot d \cdot (2-k) + 1)}{(2 \cdot d + 1)^2 - (2 \cdot d \cdot (k-1))^2} \tag{120}$$

The algebraic sign of $R'_{ri}(e_{ci})$ is determined by $(a \cdot b \cdot e_{ci}^{b-1} - 2 \cdot d \cdot k)$ and follows

$$R'_{ri}(e_{ci}) \begin{cases} > 0 & \text{if} \quad e_{ci}^{1-b} < \frac{a}{2 \cdot d \cdot k} \\ < 0 & \text{if} \quad e_{ci}^{1-b} > \frac{a}{2 \cdot d \cdot k} \\ = 0 & \text{if} \quad e_{ci}^{1-b} = \frac{a}{2 \cdot d \cdot k}. \end{cases} \tag{121}$$

□

5.B Proof of Lemma 1

Proof. Payoff $\pi_{ri}(e_{ri}, e_{ci}, e_{r-i})$ is strictly concave in e_{ci} for an optimal choice of e_{ri}. Therefore it has only one maximum point, which is defined by $\bar{e}_{ci}$ as given in Eq. (106):

$$\bar{e}_{ci} = \left(\frac{a \cdot b}{2 \cdot d \cdot k} \right)^{\frac{1}{1-b}}. \tag{122}$$

Also $R_{ri}(0) = 0$ holds. It follows that $\pi_{ri}(e_{ri}, e_{ci}, e_{r-i})$ is monotonous in e_{ci} for $0 \leq e_{ci} < \bar{e}_{ci}$.

Therefore if the original rural emissions are positive according to $R_{ri}(e_{ci}^{or}) > 0$, then there exists $\underline{e}_{ci}$ with $0 < \underline{e}_{ci} < \bar{e}_{ci}$ and with $R_{ri}(\underline{e}_{ci}) < R_{ri}(e_{ci}^{or})$.

□

5.C Proof for Prop. 12

Proof. First, take $\pi_{ri}^{f}(e_{ri}, e_{ci}, e_{r-i}, e_f)$ as given in Eq. (109), namely:

$$\pi_{ri}^{f}(e_{ri}, e_{ci}, e_{r-i}, e_f) = a \cdot e_{ri} \cdot e_{ci}^{b} - \frac{1}{2} \cdot e_{ri}^{2} - d \cdot E^2 \tag{123}$$

The first order condition

$$0 = a \cdot e_{ci}^{b} - e_{ri} - 2 \cdot d \cdot E \tag{124}$$

leads to unique optimal emissions e_{ri}^{f*}. The second order condition confirms that e_{ri}^{f*} is a local maximum because $\pi_{ri}^{f}(e_{ri}, e_{ci}, e_{r-i}, e_f)$ is strictly concave in e_{ci}.

Now consider the Nash equilibrium of all rural regions and the freerider region. Optimal emissions for a rural region in this Nash equilibrium are given by Eq. (110):

$$e_{ri}^{f*} = R_{ri}^{f}(e_{ci}) = \frac{(a \cdot e_{ci}^{b} - 2 \cdot d \cdot (k \cdot e_{ci} + R_f(e_{ci}))) \cdot (2 \cdot d \cdot (2-k) + 2)}{(2 \cdot d + 1)^2 - (2 \cdot d \cdot (k-2))^2}. \tag{125}$$

The derivative of $e_{ri}^{f*} = R_{ri}^{f}(e_{ci})$ with respect to e_{ci} is given by Eq. (111):

$$R_{ri}^{f\prime}(e_{ci}) = \frac{\mathrm{d}e_{ri}}{\mathrm{d}e_{ci}} = \frac{(a \cdot b \cdot e_{ci}^{b-1} - 2 \cdot d \cdot (k + R_f'(e_{ci}))) \cdot (2 \cdot d \cdot (2-k) + 2)}{(2 \cdot d + 1)^2 - (2 \cdot d \cdot (k-2))^2}. \tag{126}$$

The algebraic sign of $R_{ri}^{f\prime}(e_{ci})$ is determined by $(a \cdot b \cdot e_{ci}^{b-1} - 2 \cdot d \cdot (k + R_f'(e_{ci})))$ and follows

$$R_{ri}^{f\prime}(e_{ci}) \begin{cases} > 0 & \text{if} \quad e_{ci}^{1-b} < \frac{a \cdot b}{2 \cdot d \cdot (k + R_f'(e_{ci}))} \\ < 0 & \text{if} \quad e_{ci}^{1-b} > \frac{a \cdot b}{2 \cdot d \cdot (k + R_f'(e_{ci}))} \\ = 0 & \text{if} \quad e_{ci}^{1-b} = \frac{a \cdot b}{2 \cdot d \cdot (k + R_f'(e_{ci}))}. \end{cases} \tag{127}$$

□

6 Overall Conclusion

I have taken two new approaches to climate policy. The focus is on economical incentives and strategic choices. My method is game theory, I have used analytical models of global emissions.

The first approach follows from the idea that adaptation to climate change can give new strategic opportunities which can be beneficial. The option of adaptation can lead to non-convexities in the effective damage function of some countries. This in turn makes emissions strategic complements from the perspective of these countries. Therefore their reaction functions are positively sloped. If other countries reduce emissions unilaterally, countries with this special property reduce their own emissions in a quasi-cooperative way, instead of freeriding.

We have found that the role distribution (of Stackelberg leader and follower positions) necessary to exploit this is chosen endogenously in a two player game. In a multi-player game of coalition formation there are two equilibria - one with a small coalition which corresponds to conventional wisdom and one 'island of stability' with a larger coalition, lower global emissions and higher global welfare.

These findings have a somewhat optimistic implication. If strategic complements are recognised as such and if negotiation structure are designed to reflect this, then global emissions can be reduced and welfare increased. This would be a Pareto-improvement over the standard outcome which does not allow for special circumstances like strategic complements. However it is also possible that worse scenarios play out and global emissions are much higher than necessary if there is no such fitting agreement design. Therefore awareness for strategic complements and optimal negotiation design is particularly important.

The second approach broadens the perspective from a single international agreement between nation states to a broader view of cooperation forms. City alliances, multiple agreements and supranational NGOs offer new ways to tackle emission reductions. Such Transnational Environmental Agreements (TEAs) offer several advantages over more traditional International Environmental Agreements (IEAs) like greater flexibility, more direct consumer participation and new strategic options.

City alliances in particular are already forming. In my analysis of the effect of emissions reductions by a city alliance on (rural) non-member regions I have found that there are conflicting influences - one crowding out mitigation by non-members and another one exemplary and inspiring quasi-cooperation by means of technology leadership. The positive effect dominates the negative one for larger emissions reductions by the city alliance, which means that even though at firs there may be freeriding, further and more ambitious climate action by a city alliance can be expected to lead to imitating behaviour by non-members.

There are still many open questions regarding the efficiency and stability of TEAs as well as their distributional welfare effects, overall effectiveness in greenhouse gas mitigation. However my work here already leads to some conclusions. TEAs are a political reality and will probably grow in importance. They will probably not replace IEAs outright but have the potential to complement them very well. The negotiation design and some member properties are crucial for successful effects regarding emissions reductions. If the right actors come together and if agreements are structured well, then ambitious targets can be achieved. Also it is important not to let seemingly small effects in early mitigation phases discourage the endeavour of further climate action. Increasing mitigation effort can turn the tide and lead to lower freeriding among non-members.

There are several promising avenues for further research. The concept of strategic complements can be examined in different negotiation settings and other contexts. Empirical work on the effects of adaptation on agreements about mitigation could also be very interesting, although the relationships are probably difficult to isolate. The field of transnational environmental agreements offers many more approaches for climate policy research. City alliances in particular could (and should) be subject to closer economic research in terms of stability, size, ambition and welfare effects.

Global warming is a large problem and the climate is already changing. The task of coordinating greenhouse gas mitigation across many jurisdictions, all nations and across a long time horizon is very challenging. Negotiations show quite a lot of goodwill but only small successes that go beyond rhetoric so far. Nonetheless there are possibilities to overcome these challenges. Both unilateral and coordinated action can work out if organised well. Understanding the incentives of diverse actors and collaborating in new ways can help in designing successful agreements to tackle climate change.

7 Thanks

Like so many projects, my dissertation has taken place in a complex environment, which in my case was very supportive. I want to thank my supervisors and co-workers, first of all Klaus Eisenack and Achim Hagen. We have shared many ideas and worked together very productively. Other colleagues from Oldenburg have offered valuable input as well: It was a pleasure to work with Anna Pechan, Nils Marscheider and Micha Steinhäuser. Carsten Helm, Heinz Welsch and Michael Trautwein have given interesting and encouraging feedback, too. My everyday working life was very much enriched by the more recent colleagues Linda Neubauer, Paul Neetzow and Jasper Meya.

Now that I have physically moved to Berlin I have found a warm welcome here. Mattheo Rogero, Katherine Farell, Dimitrios Zikos and Roman Mendelevitch have proven to be both competent scientists and very nice co-workers. Nor should the administative staff go unmentioned: Catherina Schramm, Ines Jeworski, Renate Judis and Sigrid Heilmann have been greatly supportive in the background.

I have been granted a scholarship by Evangelisches Studienwerk Villigst which helped me both financially and personally. This way I have found the opportunity to meet many interesting and nice people, to engage in professional and meta-scientific dialogue. I am also very glad for the opportunities to create and shape exciting workshops on diverse topics.

Finally there are many more friends, people from university, the student congregation in Oldenburg, from my family and many others who have supported me in several small and not-so-small ways. Thank you all for your time, assistance and patience. Thank you for your backing, and sometimes distracting me. Thank you for questioning me and making me question, for encouraging me, and for helping me direct my curiosity.

References

Altamirano-Cabrera, J.-C. and M. Finus (2006). Permit trading and stability of international climate agreements. *Journal of Applied Economics 9*(1), 19.

Andonova, L. B., M. M. Betsill, and H. Bulkeley (2009). Transnational climate governance. *Global environmental politics 9*, 52–73.

Asheim, G. B., C. B. Froyn, J. Hovi, and F. C. Menz (2006). Regional versus global cooperation for climate control. *Journal of Environmental Economics and Management 51*(1), 93–109.

Aust, H. P. (2015). Shining cities on the hill? the global city, climate change, and international law. *European Journal of International Law 26*(1), 255–278.

Bandyopadhyay, S. (1997, February). Demand elasticities, asymmetry and strategic trade policy. *Journal of International Economics 42*(1-2), 167–177.

Bansard, J. S., P. H. Pattberg, and O. Widerberg (2016). Cities to the rescue? assessing the performance of transnational municipal networks in global climate governance. *International Environmental Agreements: Politics, Law and Economics*, 1–18.

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

Barrett, S. (1999). A theory of full international cooperation. *Journal of Theoretical Politics 11*(4), 519–541.

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

Barrett, S. (2008). Dikes v. windmills: Climate treaties and adaptation. *John Hopkins University Discussion Paper*.

Baumol, W. J. (1972). On taxation and the control of externalities. *American Economic Review 62*(3), 307–322.

Baumol, W. J. and D. F. Bradford (1972). Detrimental externalities and non-convexity of the production set. *Economica 39*, 160–176.

Benchekroun, H., W. Marrouch, and A. R. Chaudhuri (2011, November). Adaptation effectiveness and free-riding incentives in international environmental agreements. Technical Report CentER Discussion Paper Series No. 2011-120, Tilburg University.

Betsill, M. and H. Bulkeley (2004). Transnational networks and global environmental governance: The cities for climate protection program. *International Studies Quarterly 48*, 471âĂŞ493.

Betsill, M. M. and H. Bulkeley (2006). Cities and the multilevel governance of global climate change. *Global Governance: A Review of Multilateralism and International Organizations 12*(2), 141–159.

Betsill, M. M. and H. Bulkeley (2007). Looking back and thinking ahead: A decade of cities and climate change research. *Local Environment 12*(5), 447–456.

Biermann, F., P. Pattberg, H. Van Asselt, and F. Zelli (2009). The fragmentation of global governance architectures: A framework for analysis. *Global Environmental Politics 9*(4), 14–40.

Bollen, J., B. Guay, S. Jamet, and J. Corfee-Morlot (2009). Co-benefits of climate change mitigation policies. *OECD Economics Department Working Papers 693*.

Bosetti, V., C. Carraro, M. Galeotti, E. Massetti, and M. Tavoni (2006). Witch: A world induced technical change hybrid model. *The Energy Journal 27 (Special Issue 2)*, 13–38.

Bosetti, V. and E. de Chian (2013). A good opening: The key to make the most of unilateral climate action. *Environmental and Resouce Economics 56*(2), 255–276.

Branstetter, L. G. (2001). Are knowledge spillovers international or intranational in scope?: Microeconometric evidence from the us and japan. *Journal of International Economics 53*(1), 53–79.

Bulkeley, H. (2010). Cities and the governing of climate change. *Annual Review of Environment and Resources*, 229–253.

Bulkeley, H. and V. Broto (2013). Government by experiment? global cities and the governing of climate change. *Transactions of the Institute of British Geographers 38*(3), 361–375.

Bulow, J. I., J. D. Geanakoplos, and P. D. Klemperer (1985, June). Multimarket oligopoly: Strategic substitutes and complements. *Journal of Political Economy 93*(3), 488–511.

Buob, S. and S. Siegenthaler (2011). Does adaptation hinder self-enforcing international environmental agreements? In *Proceedings of EAERE 2011 Conference*.

Butler, R. V. and M. D. Maher (1986). The control of externalities: Abatement vs. damage prevention. *Southern Economic Journal 52*, 1088–1102.

C40 (2015). C40 cities climate leadership group. Retrieved from C40 web site: `www.c40.org/`.

Campbell-Lendrum, D. and C. Corvalan (2007). Climate change and developing-country cities: Implications for environmental health and equity. *Journal of Urban Health: Bulletin of the New York Academy of Medicine 84*(1), 109–117.

Carbone, J. C., C. Helm, and T. F. Rutherford (2009). The case for international emission trade in the absence of cooperative climate policy. *Journal of Environmental Economics and Management 58*(3), 266–280.

Carlino, G. A., S. Chatterjee, and R. M. Hunt (2007). Urban density and the rate of invention. *Journal of Urban Economics 61*(3), 389–419.

Carraro, C., J. Eyckmans, and M. Finus (2006). Optimal transfers and participation decisions in international environmental agreements. *The Review of International Organizations 1*(4), 379–396.

Carraro, C. and D. Siniscalco (1993). Strategies for the international protection of the environment. *Journal of Public Economics 52*(3), 309–328.

CDP (2015). Carbon disclosure project. Retrieved from CDP web site: `www.cdp.net`.

Chander, P. and H. Tulkens (1995). A core-theoretic solution for the design of cooperative agreements on transfrontier pollution. *International tax and public finance 2*(2), 279–293.

Climate Alliance (2015). Celebrating 25 years of the climate alliance! Retrieved from Climate Alliance Web site: `www.climatealliance.org`.

Cole, D. H. (2011). From global to polycentric climate governance. *Climate Law 2*(3), 395–413.

Cooke, P. (2008). Regional innovation systems, clean technology & jacobian cluster-platform policies. *Regional Science Policy & Practice 1*(1), 23–45.

Copeland, B. R. and M. Taylor (2005). Free trade and global warming: a trade theory view of the kyoto protocol. *Journal of Environmental Economics and Management 49*(2), 205–234.

Corfee-Morlot, J., L. Kamal-Chaoui, M. Donovan, I. Cochran, A. Robert, and P.-J. Teasdale (2009). Cities, climate change and multilevel governance. OECD Publishing.

D'Aspremont, C. A., J. Jacquemin, J. Gabszeweitz, and J. A. Weymark (1983). On the stability of collusive price leadership. *Canadian Journal of Economics 16*, 17–25.

de Bruin, K., H. Weikard, and R. Dellink (2011). The role of proactive adaptation in international climate change mitigation agreements. Working Paper 9, CERE, Centre for Environmental and Resource Economics.

Diamantoudi, E. and E. S. Sartzetakis (2006, 05). Stable international environmental agreements: An analytical approach. *Journal of Public Economic Theory 8*(2), 247–263.

Dietz, S., C. Marchiori, and A. Tavoni (2012). Domestic politics and the formation of international environmental agreements. *Grantham Research Institute on Climate Change and the Environment Working Paper* (87).

Ebert, U. and H. Welsch (2009). Optimal response functions in global pollution problems can be upward-sloping: Accounting for adaptation. Technical report, University of Oldenburg.

Ebert, U. and H. Welsch (2011). Optimal response functions in global pollution problems can be upward-sloping: accounting for adaptation. *Environmental Economics and Policy Studies 13*(2), 129–138.

Ebert, U. and H. Welsch (2012). Adaptation and mitigation in global pollution problems: Economic impacts of productivity, sensitivity, and adaptive capacity. *Environmental and Resource Economics 52*(1), 49–64.

Eckersley, R. (2012). Moving forward in the climate negotiations: Multilateralism or minilateralism? *Global Environmental Politics 12*(2), 24–42.

Eichner, T. and R. Pethig (2013). Self-enforcing environmental agreements and international trade. *Journal of Public Economics 102*, 37 – 50.

Eichner, T. and R. Pethig (2015). Self-enforcing international environmental agreements and trade: taxes versus caps. *Oxford Economic Papers 67*(4), 897–917.

Eisenack, K. (2012). Adaptation financing in a global agreement: is the adaptation levy appropriate? *Climate Policy 12*(4), 491–504.

Eisenack, K. (2014). The inefficiency of private adaptation to pollution in the presence of endogenous market structure. *Environmental and Resource Economics 57*(1), 81–99.

Eisenack, K. and L. Kähler (2015). Adaptation to climate change can support unilateral emission reductions. *Oxford Economic Papers DOI:10.1093/oep/gpv057*, in press.

Eisenack, K., S. Moser, E. Hoffmann, R. Klein, C. Oberlack, A. Pechan, M. Rotter, and C. Termeer (2014). Explaining and overcoming barriers to climate change adaptation. *Nature Climate Change 4*, 867–872.

Falkner, R. (2015). Towards minilateralism. *Nature of Climate Change 5*, 805–806.

Fankhauser, S. and N. Martin (2010). The economics of the cdm levy: Revenue potential,tax incidence and distortionary effects. *Energy Policy 38*, 357–363.

Feldman, M. P. and D. B. Audretsch (1999). Innovation in cities:: Science-based diversity, specialization and localized competition. *European economic review 43*(2), 409–429.

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*, 29–67.

Finus, M. and B. Rundshagen (1998). Renegotiation-proof equilibria in a global emission game when players are impatient. *Environmental and Resource Economics 12*, 275–306.

Finus, M., E. van Ierland, and R. Dellink (2006). Stability of climate coalitions in a cartel formation game. *Economics of Governance 7*, 271–291.

Fredriksson, P. R. and D. L. Millimet (2002). Is there a 'california effect' in us environmental policymaking? *Regional Science and Urban Economics 32*, 737–764.

Füssel, H.-M. and R. J. T. Klein (2006). Climate change vulnerability assessments: an evolution of conceptual thinking. *Climatic Change 75*(3), 301–329.

Gehring, T. and S. Oberthür (2008). Interplay: Exploring institutional interaction. *Institutions and environmental change: Principal findings, applications, and research frontiers*, 187–224.

GHG Protocol (2015). Greenhouse gas protocol for cities, global standard for measuring greenhouse gas emissions. Retrieved from GHG Web site: `www.ghgprotocol.org/`.

Gill, S., J. Handley, A. Ennos, and S. Pauleit (2007, March). Adapting cities for climate change: The role of the green infrastructure. *Built Environment 33*(1), 115–133.

Gustavsson, E., I. Elander, and M. Lundmark (2009). Multilevel governance, networking cities, and the geography of climate-change mitigation: two swedish examples. *Environment and Planning C: Government and Policy 27*, 59–74.

Habla, W. and R. Winkler (2013). Political influence on non-cooperative international climate policy. *Journal of Environmental Economics and Management 66*(2), 219–234.

Hagen, A., J.-C. Altamirano-Cabrera, and H.-P. Weikard (2016). The influence of political pressure groups on the stability of international environmental agreements. *Oldenburg Discussion Papers in Economics*.

Hagen, A. and K. Eisenack (2015). International environmental agreements with asymmetric countries: climate clubs vs. global cooperation. *FEEM Working Paper*.

Hagen, A., L. Kähler, and K. Eisenack (2017). *Economics of International Environmental Agreements: A Critical Approach*, Chapter Transnational Environmental Agreements with Heterogeneous Actors. Routledge.

Haites, E. (2011). Climate change finance. *Climate Policy 11*(3), 963–969.

Hakelberg, L. (2014). Governance by diffusion: Transnational municipal networks and the spread of local climate strategies in europe. *Global Environmental Politics 14*(1), 107–129.

Hale, T. and C. Roger (2014). Orchestration and transnational climate governance. *The Review of International Organizations 9*, 59–82.

Hallegatte, S. and J. Corfee-Morlot (2011). Understanding climate change impacts, vulnerability and adaptation at city scale: an introduction. *Climatic Change 104*(1), 1–12.

Hamilton, J. H. and S. M. Slutsky (1990). Endogenous timing in duopoly games: Stackelberg or cournot equilibria. *Games and Economic Behavior 2*(1), 29–46.

Harlan, S. L. and D. M. Ruddell (2011). Climate change and health in cities: impacts of heat and air pollution and potential co-benefits from mitigation and adaptation. *Current Opinion in Environmental Sustainability 3*(3), 126–134.

Helm, C. and R. C. Schmidt (2015). Climate cooperation with technology investments and border carbon adjustment. *European Economic Review 75*, 112–130.

Helm, D., C. Hepburn, and G. Ruta (2012). Trade, climate change, and the political game theory of border carbon adjustments. *Oxford review of economic policy 28*(2), 368–394.

Heugues, M. (2012a). Endogenous timing in pollution control: Stackelberg versus Cournot-Nash equilibria. *Strategic Behavior and the Environment 2*(2), 133–158.

Heugues, M. (2012b). International environmental cooperation: a new eye on the greenhouse gas emissions' control. *Annals of Operations Research 220*(1), 239–262.

Heuson, C., E. Gawel, O. Gebhardt, B. Hansjürgens, P. Lehmann, V. Meyer, and R. Schwarze (2012). Fundamental questions on the economics of climate adaptation. Technical Report UFZ Report 05/12, Helmholtz Centre for Environmental Research (UFZ).

Hoel, M. (1991). Global environmental problems: The effects of unilateral actions taken by one country. *Journal of Environmental Economics and Management 20*, 55–70.

Hof, A. F., M. G. den Elzen, and A. M. Beltran (2011, October). Predictability, equitability and adequacy of post-2012 international climate financing proposals. *Environmental Science and Policy 14*(6), 615–627.

Hooghe, L. and G. Marks (2003). Unraveling the central state, but how? types of multi-level governance. *American political science review 97*(02), 233–243.

ICLEI (2015). Iclei - the global cities network. Retrieved from ICLEI web site: `www.iclei.org/iclei-members.html`.

Ingham, A., J. Ma, and A. Ulph (2007). Climate change, mitigation and adaptation with uncertainty and learning. *Energy Policy 35*, 5354–5369.

IPCC (2013). Climate change 2014: The physical science basis. Intergovernmental Panel on Climate Change: Cambridge University Press, Cambridge, UK.

IPCC (2014a). Climate change 2014: Impacts, adaptation, and vulnerability. Intergovernmental Panel on Climate Change: Cambridge University Press, Cambridge, UK.

IPCC (2014b). Climate Change 2014: Mitigation of Climate Change. Intergovernmental Panel on Climate Change: Cambridge University Press, Cambridge, UK.

Isailovic, M., O. Widerberg, and P. Pattberg (2013). Fragmentation of global environmental governance architectures: a literature review. *Available at SSRN 2479930*.

Jacobs, J. (1969). *The Economy of Cities*. Vintage Books, New York.

Jaffe, A. B., M. Trajtenberg, and R. Henderson (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics*, 577–598.

Jochem, E. and R. Madlener (2003). The forgotten benefits of climate change mitigation: Innovation, technological leapfrogging, employment, and sustainable development. In *Workshop on the Benefits of Climate Policy: Improving Information for Policy Makers*. Citeseer.

Jordan, A. J., D. Huitema, M. Hildén, H. van Asselt, T. J. Rayner, J. J. Schoenefeld, J. Tosun, J. Forster, and E. L. Boasson (2015). Emergence of polycentric climate governance and its future prospects. *Nature Climate Change*.

Kähler, L. and K. Eisenack (2016). Strategic complements in international environmental agreements: a new island of stability. *University of Oldenburg, Department of Economics Working Papers 393*(16).

Karlsson-Vinkhuyzen, S. I. and J. McGee (2013). Legitimacy in an era of fragmentation: The case of global climate governance. *Global Environmental Politics 13*(3), 56–78.

Keohane, R. O. and D. G. Victor (2011). The regime complex for climate change. *Perspectives on politics 9*, 7–23.

Kern, K. and H. Bulkeley (2009). Cities, europeanization and multi-level governance: Governing climate change through transnational municipal networks. *Journal of Common Market Studies 47*(2), 309âĂŞ332.

Kessides, C. (2006). *The urban transition in Sub-Saharan Africa: Implications for economic growth and poverty reduction*. World Bank Africa Region Working Paper Series No. 97.

Lessmann, K., R. Marschinski, and O. Edenhofer (2009). The effects of tariffs on coalition formation in a dynamic global warming game. *Economic Modelling 26(3)*, 641–649.

Marchiori, C., S. Dietz, and A. Tavoni (2016). Domestic politics and the formation of international environmental agreements. *Journal of Environmental Economics and Management September 2016*.

Marrouch, W. and A. R. Chaudhuri (2011). International environmental agreements in the presence of adaptation. working paper 35, Fondazione Eni Enrico Mattei (FEEM).

Marshall, A. (1890). *Principles of economics*. Macmillan, London.

McGinty, M. (2007). International environmental agreements among asymmetric nations. *Oxford Economic Papers 59*(1), 45–62.

McKitrick, R. and R. A. Collinge (2002). The existence and uniqueness of optimal pollution policy in the presence of victim defense measures. *Journal of Environmental Economics and Management 44*, 106–122.

Millard-Ball, A. (2012). Do city climate plans reduce emissions? *Journal of Urban Economics 71*(3), 289–311.

Moncel, R. and H. Asselt (2012). All hands on deck! mobilizing climate change action beyond the unfccc. *Review of European Community & International Environmental Law 21*(3), 163–176.

Monni, S. and F. Raes (2008). Multilevel climate policy: the case of the european union, finland and helsinki. *Environmental Science & Policy 11*(8), 743–755.

Moser, S. C. and J. A. Ekstrom (2010). A framework to diagnose barriers to climate change adaptation. *Proceedings of the National Academy of Sciences 107*(51), 22026–22031.

Nagashima, M., R. Dellinka, E. van Ierlanda, and H.-P. Weikarda (2009). Stability of international climate coalitions - a comparison of transfer schemes. *Ecological Economics 68*, 1476–1487.

Neary, J. P. (1994). Cost asymmetries in international subsidy games: Should governments help winners or losers? *Journal of International Economics 37*, 197–218.

Nordhaus, W. (2015). Climate clubs: overcoming free-riding in international climate policy. *The American Economic Review 105*(4), 1339–1370.

Nordhaus, W. D. (1993). Rolling the 'dice': an optimal transition path for controlling greenhouse gases. *Resource and Energy Economics 15*, 27–50.

Nordhaus, W. D. and J. Boyer (2000). *Warming the World: Economic Models of Global Warming*. Cambridge: MIT Press.

Nordhaus, W. D. and Z. Yang (1996). A regional dynamic general-equilibrium model of alternative climate-change strategies. *American Economic Review 86*, 741–65.

Oates, W. E. (1972). *Fiscal federalism*. Edward Elgar Publishing.

Oates, W. E. (2005). Toward a second-generation theory of fiscal federalism. *International tax and public finance 12*(4), 349–373.

Olson, M. (1971). *The Logic of Collective Action : Public Goods and the Theory of Groups*. Cambridge (MA, USA): Harvard Univ. Press.

Ortiz, R., A. Golub, O. Lugovoy, A. Markandya, and J. Wang (2011). Dicer: A tool for analyzing climate policies. *Energy Economics 33*(Supplement 1), 41–49.

Osmani, D. and R. Tol (2010). The case of two self-enforcing international agreements for environmental protection with asymmetric countries. *Computational Economics 36*, 93–119.

Ostrom, E. (2010). Polycentric systems for coping with collective action and global environmental change. *Global Environmental Change 20*, 550–557.

Ostrom, E. (2012, February). Nested externalities and polycentric institutions: must we wait for global solutions to climate change before taking actions at other scales? *Economic Theory 49*(2), 353–369.

Ostrom, V., C. Tiebout, and R. Warren (1961). The organization of government in metropolitan areas: A theoretical inquiry. *American Political Science Review 55*(4), 831–842.

Pielke, R., G. Prins, S. Rayner, and D. Sarewitz (2007). Lifting the taboo on adaptation. *Nature 445*, 597–598.

Shobe, W. and D. Burtraw (2012, January). Rethinking environmental federalism in a warming world. Resources for the Future Discussion Paper RFF DP 12-04, Center for Economic and Policy Studies, Charlottesville, Virginia (USA).

Sippel, M. (2010, Juni). Cities in germany and their climate commitments: More hype than substance? Discussion Paper MPRA Paper No. 23011, Institute of Energy Economics and Rational Energy Use, University Stuttgart.

Sippel, M. and T. Jenssen (2009, February). What about local climate governance? a review of promise and problems. Discussion Paper MPRA Paper No. 20987, Institute of Energy Economics and Rational Energy Use, University Stuttgart.

Smith, M. E., J. Ur, and G. M. Feinman (2014). Jane jacobs' âĂŸcities firstâĂŹ model and archaeological reality. *International Journal of Urban and Regional Research 38*(4), 1525–1535.

Stern, N. (2007). *The Economics of Climate Change: The Stern Review*. Cambridge, UK: Cambridge University Press.

Stewart, R. B. (2008). States and cities as actors in global climate regulation: unitary vs. plural architectures. *Arizona Law Review 50*, 681.

Stewart, R. B., M. Oppenheimer, and B. Rudyk (2013). Building blocks for global climate protection. *Stanford Environmental Law Journal 32*(2), 341–392.

Thompson, P. and M. Fox-Kean (2005). Patent citations and the geography of knowledge spillovers: A reassessment. *American Economic Review*, 450–460.

Tol, R. S. J. (2005). The marginal damage costs of carbon dioxide emissions: an assessment of the uncertainties. *Energy Policy 33*, 2064–2074.

UNEP (2015). Climate commitments of subnational actors and business: A quantitative assessment of their emission reduction impact. Nairobi.

Urpelainen, J. (2009). Explaining the schwarzenegger phenomenon: Local frontrunners in climate policy. *Global Environmental Politics 9*(3), 82–105.

van der Panne, G. (2004). Agglomeration externalities: Marshall versus jacobs. *Journal of Evolutionary Economics 14*(5), 593–604.

van der Panne, G. and C. van Beers (2006). On the marshall–jacobs controversy: it takes two to tango. *Industrial and Corporate Change 15*(5), 877–890.

Wangler, L., J.-C. Altamirano-Cabrera, and H.-P. Weikard (2013). The political economy of international environmental agreements: a survey. *International Environmental Agreements 13*(3), 387–403.

Warren, R., C. Hope, M. Mastrandrea, R. Tol, N. Adger, and I. Lorenzoni (2006). Spotlighting impact functions in integrated assessment. Tyndall Working paper 91, Tyndall Centre for Climate Change Research. Research Report Prepared for the Stern Review on the Economics of Climate Change.

Watkiss, P. (2011). Aggregate economic measures of climate change damages: explaining the differences and implications. *Wiley Interdisciplinary Reviews - Climate Change 2*(3), 356–372.

Weischer, L., J. Morgan, and M. Patel (2012). Climate clubs: Can small groups of countries make a big difference in addressing climate change? *Review of European Community and International Environmental Law 21*(3), 177–192.

Widerberg, O. and D. E. Stenson (2013). *Climate clubs and the UNFCCC - Complement, bypass or conflict?* Fores.

Zehaie, F. (2009). The timing and strategic role of self-protection. *Environmental and Resource Economics 44*, 337–350.

Selbständigkeitserklärung

Die vorliegende Dissertation habe ich selbständig und ohne fremde Hilfe verfasst. Neben der von mir zitierten Literatur ist die vorliegende Arbeit auf der Grundlage von insgesamt vier Artikeln erarbeitet worden. Davon ist ein Artikel veröffentlicht, ein Artikel im Erscheinen und zwei Artikel sind eingereicht.

(Ort, Datum) Leonhard Kähler

Erklärung zur Übereinstimmung von gedruckter und digitaler Fassung

Hiermit erkläre ich an Eides statt, dass die gedruckte und die elektronische Fassung dieser Arbeit inhaltlich und formal übereinstimmen.

(Ort, Datum) Leonhard Kähler