Essays on Climate Change

by

Anna Malova, MRes

Submitted in fulfillment of the requirements for the degree of
Doctor of Philosophy

Adam Smith Business School, College of Social Sciences
University of Glasgow

September 2023

Abstract

This thesis attempts to provide practical policy recommendations that may make a difference in the climate crisis. It is, therefore, grounded in non-cooperative game theory, and the author regards its simplicity as a virtue rather than a hindrance. By no means does it offer a comprehensive solution, but merely recommendations that are easy to implement.

Chapter 1 focuses on approaches to reform implementation in the face of political constraints, such as a net-zero transition. In a simple three-period model where a welfare-enhancing reform may never be implemented, we show how a winning coalition of voters can be identified and constructed before reform is proposed to enable its implementation. When domestic mobility frictions are small, the reform can be implemented immediately. Alternatively, in certain settings, a winning coalition of voters serves as a commitment device underpinning the credibility of a proposal to implement reform with delay. If nations are part of a network, the transition may be engineered in a small number of pivotal members, which will lower the cost of transition for other members and ensure the transition across all countries. We propose many different ways to identify and leverage a pivotal group of voters or countries.

In Chapter 2, we turn to the role of unilateralism and how it can be used to lower global emissions more effectively. In a model with n countries, we prove that the core is empty and thus, no multilateral agreement with full participation is immune to deviations. Extending the model to include technological spillovers in the form of discontinuity in the investment cost if enough nations adopt the technology early on leads to multiple subgame perfect equilibria, among which is a stable grand coalition. We survey the literature for ways to introduce technological trade, discuss the role of an international environmental treaty under these conditions and provide some policy implications.

Chapter 3 looks at a non-cooperative game of three interacting nations who pollute, consume, invest and bear delayed cost of their choices. The novelty of the model is in separating environmental damage into two terms, local and global, so that the two stocks have different effects on nations' payoffs. Instead of *assuming* that some countries would adopt new technology earlier than others, this approach provides a rationale for such choice. It is evident that already today, the consequences of climate change can be felt, but some countries have it worse than others, so the nations with greater present environmental damage will be the first to invest. Comparative statics suggests that merely an option of technological trade is enough to convince some nations to invest in green technology, but the country must be large in terms of its share of global emissions and suffer from local pollution already

at present. Under these circumstances, China is a reasonable candidate for the role of global transition leader.

Contents

List of Tables

List of Figures

Acknowledgements

When I started studying towards a PhD in Economics in 2018, I was convinced I would be done by 2021, having produced an absolute masterpiece of a thesis and saved the world from an environmental calamity. By 2020 it became clear to me that the plan was too ambitious and I should aim for 2022 instead. In 2021, life got in the way of that plan too: I got married and a month later found out that I was pregnant. In 2022, my daughter was born, and I had to request a leave of absence from the University to care for her, because, as it turns out, newborns do not sleep all that well. 2022 in general proved to be a very difficult year for my family, and the dissertation submission was again delayed. So, finally, five years after starting and two years later than expected, I am submitting my doctoral thesis.

Throughout all these years, if there was one person who believed in me more than I ever believed in myself, who saw something in me that I was never able to see, who never doubted my abilities and always inspired me to ask the most important and inherently unanswerable questions, it was my supervisor Sayantan Ghosal. His understanding of the world, his drive to make it a better place, to truly make a difference, but also his acceptance of the things that cannot be changed right now (or maybe ever), his curiosity and the extent of his economic knowledge are what shaped my research and me as a researcher. This thesis, I owe to him.

This thesis would also not be possible without the hard work of my husband, Pooria Panahi. For seven years he has been my consciousness, relentlessly reminding me of the right things to do, telling me the hard truths, supporting me wholeheartedly and accepting me for what I ever was, is, and will be. He has been my shoulder to cry on, my role model, my inspiration. But most importantly, he simply was there with me, always; and for that, I am eternally grateful.

I am also forever indebted to my mother, Alla Kachalova, without whom I would not be where I am today. Since the moment I was born, she did everything in her power to ensure that I would have the best education possible. She trusted me to make the right choice and helped deal with the consequences if it turned out to be the wrong one. She sacrificed her time, work and money, to see me through my studies. This thesis is her achievement as much as mine.

Finally, I want to dedicate this work to my daughter, Lydia Panahi, who, although mostly being a hindrance rather than help, managed to teach me more over a year than I have learnt in my whole life. The ability to prioritise, to function with a bare minimum of sleep, to understand the meaning of life and see what truly is important, all are thanks to her. As much as I don't want her to grow, I cannot wait to see what great things she will achieve in life. I hope this thesis will be proof enough for her that she can do anything she puts her mind to.

There are many more people my gratitude goes to, but to keep it brief, I will only say that studying towards a doctoral degree was by far the hardest thing I ever had to do.

Declaration of Authorship

I declare that, except where explicit reference is made to the contribution of others, that this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Printed Name: Anna Malova

Signature:

"*Necessity is the mother of invention*"

Latin proverb

Introduction

In an era defined by its interconnectedness and global challenges, few issues loom as ominously as climate change. With each passing day, the evidence of our planet's delicate ecological balance being disrupted becomes more undeniable, presenting a dire threat that transcends borders, economies, nations. It may seem surprising that even in the face of overwhelming scientific consensus, political action remains scant, if not inadequate. It is however expected in the world as complex as the one we live in, and humanity needs to pivot away from the policies that have yielded no result and agreements that have led nowhere.

The Paris Agreement, being a 180-degree turn from the bottom-up to top-down approach and succeeding dysfunctional Kyoto Protocol, has so far struggled to yield substantive results, laying bare the limitations of unilateral commitments in the face of a crisis that respects no national boundaries.

This thesis delves into the heart of this complex issue, weaving together the realms of game theory and political economy, and turning the gaze toward green technology as an instrument for enhancing international cooperation.

As political inertia continues to hinder collective action, which is evident from insufficient nationally determined contributions to the Paris Accord, the question emerges: how can nations be incentivized to transition toward sustainability, especially when self-interest and short-term priorities often overshadow long-term global well-being?

At its core, this thesis acknowledges the merit of unilateral actions as an essential building block for addressing the climate crisis. The failure of previous international agreements underscores the challenges of compelling sovereign states to commit to binding obligations that might conflict with their immediate interests. However, this approach has led to a paradoxical stasis, wherein ambitious pledges to achieve carbon neutrality by distant dates like 2050 can inadvertently provide political cover for delaying concrete action and transformative investments.

Yet, this thesis contends that while unilateral actions are a necessary starting point, they are insufficient to drive the profound societal shifts required for meaningful change. The linchpin lies not merely in economic models but in the realm of psychology and leadership. A pivotal step toward enacting effective policies rests in shaping the electorate's mindset, thereby cultivating a political landscape that empowers visionary leaders capable of steering nations toward sustainable future.

Research suggests that six out of nine vital planetary boundaries have been passed, two are very close to being broken, and only one is judged to be well outside the danger zone. Ironically, the one which is safe is the ozone layer, which was promptly fixed by the Montreal Protocol (Richardson et al. 2023). Why ironically? Because in fixing climate change, so much importance is placed on international cooperation, and the Montreal Protocol serves as a perfect example of what successful international cooperation looks like. The irony here is in the fact that the same approach would never work for solving global warming, and while it magnificently eliminated one big threat to human existence, so much so that it is the only boundary that has not been passed, it will same magnificently fail to eliminate the second big threat, which is evident in passing the other eight boundaries.

Paris Accords, unlike the Kyoto Protocol, did not make the mistake of being drafted in the likeness of the Montreal Protocol, but it, too, does not seem very promising. The focus on unilateral actions seems like a logical step to take after a top-down approach fell short of its promises, and it might be the only hope we have in surviving the twenty-first century. Still, the individual pledges are far too small to add up to the bare minimum required to cap the warming under 2°C, and even these insufficient targets are unlikely to be met. It is too little too late.

Does it mean we should stop wasting our resources on trying and live by the motto "après moi, le déluge"? To the author of this thesis, the answer is no. The future generations deserve a chance to live, and however little our attempts can achieve, the effort must be made. Having lost hope in an ambitious global international agreement that will once and for all cut greenhouse gas emissions, in this thesis, we look at unilateral localised actions that have the potential to make a difference.

In chapter 1, we will consider a simple political economy model with a benevolent policymaker who is interested in staying in power. To maximise her rent from being in office, she will want to engineer a transition within society, which will raise social welfare over time but will cost some citizens in the short run. To resolve the political bottleneck, we propose two solutions that are scalable to a network of countries. These solutions are not quick fixes but a steady road towards an environmental transition not only within the borders of one country but, if it is a pivotal country, within the borders of many others who will follow suit.

Since it is becoming evident that with conventional methods, the worst effects of climate change will not be averted, the discussion is shifting towards climate management in the form of geoengineering. This approach must, at worst, be taken with extreme caution, and at best, not at all (Biermann et al. 2022, Baum et al. 2022, McDonald 2023). Geoengineering is an example of an undesirable use of technology, but technology is not bad in and of itself; it is what we use it for that gives it meaning. And it so happens that it is necessary for the transition to a low-carbon economy. Fossil fuels must be substituted with something before we can eliminate them from our lives. In chapters 2 and 3, we will therefore direct the attention towards the role of technology in global efforts to curtail global warming. In chapter 2, we take a closer look at different ways technology has been introduced to the models of climate cooperation and propose our own to fill in the gaps in the literature. Relying on the argument of the core, we show that a treaty that sustains full participation is not possible if technology is not explicitly accounted for. On the other hand, it is but one of many subgame perfect equilibria when technology is introduced in the form of critical mass of nations whose investment at $t = 1$ lowers the cost of technology for everybody else in the future. The idea relies on strong technological spillovers, but can be interpreted as learning or shared R&D cost. To further incorporate technology transfer in the form of trade, we discuss a paper that does so and conclude with plans for future development.

In chapter 3, we continue to build on the results of chapter 2 and extend them to a richer modeling environment. We acknowledge that the window of opportunity for mitigation efforts is rapidly closing, and some damage to the environment is already done. This damage is felt unevenly between the nations, therefore the incentives to adapt also differ. We postulate that the nations who suffer more at present will be the hot spots of technological advancements from where it could spread to the countries who are more fortunate. We therefore propose a three-period model with immediate benefits, delayed cost and three asymmetric nations who engage in polluting activities, suffer from local as well as (delayed) global damages, and invest in technology. The cornerstone of the model is technological trade – nations who invest earlier to adapt to climate change in their region can sell the technology to the rest of the world thereby ensuring that global emissions are reduced later than ideally needed, but earlier than never. The versatility of the model in chapter 3 allows for different types of technology, such as mitigation, adaptation, or full-on geoengineering. The game is solved using the methods of non-cooperative game theory because the transfers between the nations are not assumed, to bring it more in line with the reality of the Paris Accord.

The analysis presented here is far from complete, and in the future, we hope to significantly expand it and add layers of complexity in the form of uncertainty about the cost of investment and the magnitude of environmental damages. Bayesian updating is also a possible way of extending the model in chapter 3. The thesis would also greatly benefit from an empirical section on technological spillovers to establish whether the assumption that underpins the model is valid.

Chapter 1

Mitigating domestic barriers to Pro-Climate Reform: The role of winning coalitions

1.1 Introduction

In 2015, Kyoto Protocol was officially abandoned in favour of the Paris Agreement. Its failure to achieve substantial multilateral greenhouse gas emissions cuts is the driving force behind the pivot in international negotiations towards a bottom-up approach to a global climate agreement. Indeed, it could be argued that the success of the Montreal Protocol was precisely that it leveraged unilateral commitments effectively (Barrett 2005a). The Paris Accord thus explicitly builds on unilateral emissions cuts and serves to coordinate and monitor such commitments.

Given this shift, the key issue is how countries can implement environmental reforms by mitigating domestic barriers to a shift to a low-carbon economy from an unsustainable inefficient status quo.

In the present paper, we adopt a viewpoint of practical mechanism design. Treating environmentally unsustainable society as a result of inefficient institutional persistence, we want to study how inefficient institutions can be replaced when national governments have no political interest in doing so, i.e., leaders have a stake in perpetuating the status quo. The discussion thus becomes part of a larger debate about persistence of inefficient institutions, especially in the context of entering and exiting international agreements, such as trade protectionism (the unwillingness to dismantle existing tariff barriers or the incentives to impose new tariff barriers) or climate change (the inability of national governments to commit to early adoption of low carbon technologies). Institutions are central for national growth and development, which makes persistent inefficient institutions a central issue in the political economy of development (Acemoglu and Robinson 2013). But institutions are also a consequence of historical conditions a nation finds itself in: “Men make their own history, but they do not make it as they please; they do not make it under self-selected circumstances, but under circumstances existing already, given and transmitted from the past" (Marx 1852). From this perspective, institutional change is far too gradual, subject to the vagaries of history, and cannot be easily remedied. Our approach of practical mechanism design suggests otherwise.

We analyse a setting where, following Fernandez and Rodrik (1991), an underlying domestic mobility friction creates individual-specific uncertainty about the ex-post identities of winners and losers when pro-climate reform is adopted. Such uncertainty could be due to bottlenecks in domestic capacity to adopt, simultaneously, both upstream and downstream low-carbon technologies to effectively abate emissions. An environmentally sound technology can be turned into a polluting one if the upstream technologies do not keep up with the transition. For instance, if the rate of adoption of electric vehicles exceeds the capacity of the electricity generation system to provide clean energy, environmental benefits may be forfeited.[1] Since winners from the reform cannot credibly commit to compensate the the losers, we assume limited transferability of payoffs between agents within countries, between time periods, across national boundaries. The resulting mobility frictions imply that either reform is implemented immediately, or never implemented at all. In the latter scenario, before the reform is proposed as a policy option, a winning coalition of voters can be identified and leveraged to mitigate mobility frictions that prevent the reform from taking place.

Under certain conditions, such a winning coalition of voters leads to an immediate reform adoption. For instance, Marechal and Lazaric (2010) note that the obstacles to wider implementation of efficient emission-reducing investments require targeting "lead users" who could serve as early adopters. Provided that domestic mobility frictions are not too large, this would be a specific instance of the mechanism discussed here.

In general, when it comes to promoting technological transitions, especially in the field of renewable technologies, governments may want to start from creating niche markets and managing them strategically to achieve a technological regime shift. This implies a government trying on a role of a catalyst and facilitator rather than a regulator or benefactor (Kemp et al. 1998).

When reform cannot be immediately implemented, a promise to do so in the future, with some delay, is made by the incumbent politician – the winning coalition of voters serves as a commitment device as they will hold the incumbent politician's

[1] Modifying electricity generation is likely to be incremental due to to path-dependence and carbon lock-in (Arthur 1989, Unruh 2000a, Markard et al. 2012).

"feet to the fire" if the promise is reneged on. Hence, the sequence of political measures is key. Before a programme can be introduced, building societal and strategic business support is necessary (Geels et al. 2017). Green policy constituencies in this case serve as a commitment device that ensures that the policy lives on. And, if revenues from a decarbonisation programme are reallocated to the public, public scrutiny will be strongly encouraged and the government will have less incentive to reverse or not comply with the policy (Brunner et al. 2012).

For instance, a politician may choose to subsidise complementary abatement technologies that are further upstream to affect the distribution of future costs and benefits. For example, as the first step in a transition to sustainable electricity system in Germany, increased R&D expenditures created advocacy coalitions that would later grow to be powerful enough to influence policy choices (Jacobsson and Lauber 2006). They included firms invested in wind and solar energy as well as associations and other organisations such as the German Solar Energy Industry Association and Eurosolar. The system of feed-in tariffs also created support amongst farmers and homeowners (Strunz 2014). The winning coalition thus identified and constructed, though not explicitly a goal of the German government, turned out to be key in the energy system transition.

As an extension of proposed mechanisms, considering the world as a global network of interconnected nations, we can also imagine an international policymaker who internalizes relevant global externalities and seeks to devise a policy with the aim of exploiting national governments' domestic constraints and catalyse a change in an accelerated time frame. In such a network, one nation's payoff from being in a certain political state may depend on the number of other countries in the same state. For instance, nations who wish to explore the potential of wind energy can share research and development costs.[2] An outside agent can manipulate the payoffs of pivotal countries to indirectly alter the preferences of a median voter in other countries to cause a cascade of political transitions. A quick example of this approach would be subsidising German car manufacturers to produce electric vehicles, so that the price and quality of them reaches an optimal point where people in other countries would find them a good substitute for a conventional vehicle.

The rest of the paper is organised as follows. The next section discusses relevant literature and our contribution to it. Then we describe the basic model and the benchmark result. The mechanisms for implementing an efficiency-enhancing reform by identifying and constructing winning coalitions are analysed in the following section. The last section provides a discussion and concludes.

[2] As a prerequisite, however, we must allow for technological spillovers, which might not be an easy task (Bayer and Urpelainen 2013).

1.2 Related Literature

Due to its interdisciplinary nature, this paper touches many strands of literature. Since it talks about political inertia and persistent suboptimal social states that are the result of a long sequence of past events, it is part of the literature devoted to institutions. Institutions, as defined by North (1991), can be a constraint for economic and political problems faced by society; institutions can be inefficient; institutions tend to persist, thereby defining and limiting the options of future trajectories for society's development.

Douglass North (1991, p.97) defined institutions as "the humanly devised constraints that structure political, economic and social interaction". Therefore, institutions determine the set of tools available to a society to effectively resolve its issues. Some institutions are correlated with better economic performance, such as better enforcement of property rights of a large fraction of the citizens and comparative equality before the law (North and Weingast 1989; Acemoglu et al. 2001). Extractive institutions are generally associated with poor economic performance and noticeably lower levels of economic development, although they still can emerge in equilibrium (Binger and Hoffman 1989; Acemoglu et al. 2002; Acemoglu and Robinson 2013). Less developed countries with institutions of poor quality may not benefit from trade, provided that trade and institutions interact bi-directionally (Segura-Cayuela 2006). This raises the question of institutional efficiency. How does one measure it, which institutions are more efficient, what prevents institutional evolution?

Discussion about institutional efficiency and how it relates to national economic performance dates back to Adam Smith (1963). Since then, a bewildering array of ideas has been introduced to public scrutiny about what affects economic performance. Among the candidates for the determinants of growth are geography and factor endowments (Diamond 1997), trade (Frankel and Romer 1999; Dollar and Kraay 2003), human capital (Glaeser et al. 2004), national leaders (Jones and Olken 2005), and, certainly, institutions (Acemoglu et al. 2001; Góes 2016). The preponderance of evidence suggests that the latter may be the largest contributor to economic success (Rosenberg et al. 1986; Hall and Jones 1999; Rodrik et al. 2004; Acemoglu et al. 2014), but institutions have to be efficient.

Traditionally, adequate enforcement of property rights for a large fraction of the society and equality before the law are mentioned when discussing economically efficient institutions (North and Weingast 1989; Knack and Keefer 1995; Hall and Jones 1999; Acemoglu et al. 2001). In the present research, we abstract from these ideas and slightly abuse the term 'institutions' by limiting it to underlying determinants of the payoff structure. Therefore, the notion of efficiency in this context is confined to purely social welfare account. As emphasized in Acemoglu et al. (2002), equilibrium institutions may be extractive and not possess conditions for economic growth. In our paper, we do not discuss how effective our newly established institutions are for long-run growth, but we are certain that they maximize social welfare at present.

Due to the vagueness of our use of the term "institutions", one can locally think of the process of institutional change as being synonymous to a political/economic reform, and hereafter, we will use terms reform, institution and policy interchangeably. The question we set out to tackle is best illustrated by Fernandez and Rodrik (1991). In their work, ex post beneficial reform does not carry the day ex ante due to uncertainty at the individual level about future gains and losses. This is the culprit that prevents the shift in our model as well, but rather than focusing on modelling such a political conundrum, we concentrate on the practical ways to promote an efficiency-enhancing reform in a democratic society with free and fair, simple majority elections. We thus introduce an international benevolent agent who, unlike in other recent similar models (Galiani et al. 2019), does not serve as a provider of funds to compensate the members of the coalition that bears all the costs, but eliminates the individual-specific uncertainty about the ex-post identities of winners and losers. She does so by revealing ex-ante who will be drawn from the losing to the winning group, so that these voters would be willing to abandon the status quo and create the majority in favour of the alternative.

Inefficient policies are infamous for their ability to persist for prolonged periods of time, even when a more efficient alternative is just around the corner. This may be due to already mentioned individual-specific uncertainty (Fernandez and Rodrik 1991), political failure to recognize cost of adjustment to the new policy as sunk which causes even more extensive support in the future (Coate and Morris 1999), or a holdup problem (Espín-Sánchez 2017, Battaglini and Harstad 2016). Inefficient water allocation in the cities of Lorca and Mula that are today parts of Spain was stealing in welfare from the people for more than 700 years (Espín-Sánchez 2017). Land policy in the New World may have prolonged higher extent of inequality (Sokoloff and Engerman 2000). Dealing with feeble institutions can be a dubious enterprise, especially if politicians are dynamically and/or time-inconsistent (Harstad 2020). Conventional wisdom, in this case, holds that inefficient institutions and political instruments are used strategically to the benefit of a ruling politician with little or no account for the inconsistency (Alesina and Tabellini 1990; Battaglini and Harstad 2020). In certain conditions, thus, autocratic regimes may do better, but not in others as they may, for instance, default on debt altogether where a democratic leader would pay it back for the possibility of returning to the office in the future (Amador 2003). Here, we do not assume dynamic inconsistency in voters or in the incumbent politician, but it nonetheless arises as a structural attribute of the model, suggesting that aggregate political inconsistency does not stem exlusively from agents' inconsistency. This is in line with the overwhelming majority of the literature that features the inability to commit to decisions of yesterday and uses it to explain inefficient institutional persistence. In section 1.4.2, we propose a solution for this deadlock, which we dub a political equivalent of behavioural economics' commitment device (Thaler and Benartzi 2004).

The problem of the commitment of the government to its promises is central to our analysis, along with exploiting network externalities when countries are interconnected. For the mechanism designed in section 1.4.3, necessary assumptions are voters' far-sightedness (Dutta et al. 2005) and spillovers to ensure that technology can spread between countries with no or minimal barriers[3]. There is no myopic adjustment as there are no shocks to the model, which is crucial to ensure the common knowledge in all periods and, consequently, to establish the equilibrium path to efficient institutions across the network of countries.

The paper is motivated by a seeming inability of national governments to promote emissions reduction in such a scale as to limit the average temperature increase to 2°C. The novelty of the approach used in the Paris Agreement, while supported by some (Harstad 2023), was heavily criticized for the lack of adequacy and ambition (Gollier and Tirole 2015). Proponents of carbon pricing believe it is the first-best solution to the game of climate change, however, for the complexity of the task, it does not seem attainable, and if so, we need some simple mechanisms that satisfy political constraints and at the same time deliver noticeable positive results in an accelerated timeframe. How do we model policy in such a way as to avoid a runaway climate change in the near future, when both politicians and voters implicitly prefer this option, but the world is locked up in the inefficient status quo? Our paper suggests three ways that are discussed in details in Sections 1.4.1, 1.4.2 and 1.4.3.

Assuming the main reason for the inefficient policy persistence is as suggested in Fernandez and Rodrik (1991), we show that, identifying and constructing, from within the majority in favour of the status quo, a coalition of voters who stand to be the "winners" of the reform, serves to mitigate political constraints that prevent

[3] Bayer and Urpelainen (2013) show that technology transfer is not always possible and certain conditions on the technology and on the host country must be met for the transfer to happen successfully.

it from being implemented. Moreover, even in the absence of dynamically inconsistent preferences, delaying reform may be essential for its implementation precisely because the winning coalition has incurred sunk cost and therefore is willing to vote out the incumbent politician from office if the promise of reform is reneged on. As noted by Thaler and Benartzi (2004), a commitment in advance entails performing an action in the present which causes the transition to be irrevocable. Strategic investment in adaptation and mitigation technologies can serve as such a commitment device (Heuson et al. 2015). In our paper, voters' coalitions play the role of commitment device. The identification and construction of a winning coalition of voters in a model of non-cooperative interaction is a political equivalent of such a commitment device.

1.3 The Model and the Benchmark Result

1.3.1 Voter payoffs

There are three time periods, $t = 1, 2, 3$. A mass one of voters is divided into two groups, G_1 (initial mass μ_0) and G_2 (mass $1 - \mu_0$), with $\mu_0 > 0.5$. At every t, a political state is $p_t \in \{a, b\}$, with a being the status quo, and b – the alternative (post-reform) political state. The game starts in state a, preferred by the voters in G_1 who constitute the majority.

Voters can "move" between the groups, i.e., they can change their behaviour if they find it utility-enhancing. Formally, per-period payoffs are defined as follows.

1. If $i \in G_1$, she obtains α if a, the status quo, is the current political state, 0 if b;

2. If $i \in G_2$, her payoff is $(\alpha - \delta)$ if a is the current political state, where $\delta > 0$, and β if b.

3. It costs c for a single person to "move" across groups. "Moving" between the groups is a flawed but convenient term to express the idea that voters can make changes to their lifestyles and become more or less environmentally friendly. We will therefore use it in quotes to avoid unnecessary confusion.[4]

In our context, a represents business-as-usual, b – climate reform; voters in G_1 thus lead a high-carbon lifestyle, and voters in G_2 have already made costly

[4]This is not to say that political shift directly triggers dynamics between the groups. It changes the incentives that voters have to belong to a group which forces them to reconsider. Therefore, if incentives (credibly) change for some external reasons prior to the shift or in the absence of such, voters can "move" from one group to the other, but only ε of them will be able to do so.

(reflected by δ) adjustments to reduce their carbon footprint.[5] Therefore, δ is a disutility of being environmentally friendly when high-carbon lifestyle is the norm (state a), and c is the cost of adjustment when a climate reform is passed (state b): the cost of buying an electric car, the inconvenience of commuting through a car-free zone, the opportunity cost of time spent learning the basic principles of recycling and plant-based diet, etc.

We assume that $2\beta - c > \beta > \alpha > \beta - c > 0$, so that voters in G_1 have no individual incentives to vote for b and "move" to G_2 immediately, unless b is expected to persist for at least another period after that. We also assume that at the beginning of the game, no voter belonging to a specific group prefers to move across to the other group, i.e., $\alpha - c < \alpha - \delta$, or, equivalently, $c > \delta$.[6]

For simplicity of exposition, we assume that voters do not discount future payoffs.[7]

1.3.2 The evolution of group membership

Mobility across groups is triggered whenever there is a shift in the prevailing political state but not otherwise. This is because the payoffs are altered as a result, and being in G_2 becomes more attractive. Once the prevailing political state has shifted, we assume that all voters who belong to the disadvantaged group simultaneously decide whether or not to "move" across the groups. Of all those voters who choose to "move", a fraction ε is chosen at random. Formally, group membership evolves as follows:

1. if $p_t = p_{t-1}$, $\mu_{t+1} = \mu_t$ (set $p_0 = a$);

2. for some externally defined and fixed $\varepsilon \geq 0$, if $p_t \neq p_{t-1}$,

   (a) $\mu_{t+1} = \min\{\mu_t + \varepsilon, 1\}$ if $p_t = a$, where with probability one, a voter in G_1 remains in G_1, and with probability $\max\left\{1 - \dfrac{\varepsilon}{1-\mu_t}, 0\right\}$, a voter in G_2 remains in G_2, and

[5] We assume that voters prefer to belong to G_2 voluntarily, knowing that, in the foreseeable future, there is no guarantee that b would be implemented. This could be due to higher intrinsic value of living a low-carbon lifestyle, e.g., for moral or ideological reasons.

[6] A different approach would entail that the longer a persists, the more voters from a disadvantaged group would want to "move" to G_1.

[7] We have solved the model with exponential discounting and the results differ only quantitatively. We could have assumed quasi-hyperbolic discounting which is inherently time-inconsistent, and the conclusions would inevitably be time-inconsistent too. But then, the result would be dictated by the assumption, carrying less value than having the same result with the assumption of time consistency in voters. We therefore settle on the simplest version of the model which still features time inconsistency in its outcomes.

(b) $\mu_{t+1} = \max\{\mu_t - \varepsilon, 0\}$ if $p_t = b$, where with probability one, a voter in G_2 remains in G_2, and with probability $\max\left\{1 - \dfrac{\varepsilon}{\mu_t}, 0\right\}$, a voter in G_1 remains in G_1.

Hence, $\dfrac{\varepsilon}{\mu_0}$ is a measure of the underlying mobility friction, so that $\varepsilon = 0$ is a situation with maximum friction and $\varepsilon = \mu_0$ is a situation with no mobility frictions. Exogeneity and rigidity of ε reflect the inertia of the economy, which for climate reform purposes can be its dependence on fossil fuels for providing employment opportunities or budget revenues. As Muttitt and Kartha (2020) compare, the speed of transition away from coal in China and in Germany differs drastically and is defined by the share of workforce the industry employs. They also suggest that countries which rely heavily on oil revenues are constrained in their transitional pace by how quickly these can be replaced: in Saudi Arabia, for instance, oil share of budget revenue has decreased by 10 percentage points over the last 40 years, and accounts for 85%. This can provide a rough estimate of an individually sustainable transition pace.

Unfortunately, even in the countries with negligible oil revenues, the pace of transition is still limited. Consider, for instance, the market for electric vehicles (EV). For starters, the market would never be able to produce enough cars if suddenly everyone decided to go green, but even assuming no capital and production constraints as a thought experiment, the environmental benefits of driving electric cars would be reduced to naught. There is a whole field of research devoted to measuring efficiency and environmental worth of EVs.[8] They have uncovered several drivers: electricity mix that charges the car, charging time, climate, and production intensity. We will focus on electricity mix as it is a parameter that, unlike climate, can be altered, but, unlike charging time, does not have a quick fix. To compare GHG emissions of an EV to emissions of a car with an internal combustion engine (ICE), one will either calculate marginal emissions (Graff Zivin et al. 2014) or life cycle emissions (Archsmith et al. 2015). Both approaches have its challenges and strengths, and often yield contradicting results. Life cycle emissions are more suitable for long-term planning and are sensitive to improvements in overall generation mix, so for the near-term policy considerations, marginal emissions are arguable more useful.

Marginal emissions estimation does not require one to know the fuel mix, and also typically excludes renewables from the calculations. The results will only be

[8]Most of empirical research in this area builds on the US data which may affect the generalisability of the results. However, the analyses of life cycle emissions that are performed using the European or Chinese data generally are not in contradiction with those based on the US data (Hawkins et al. 2013, Ji et al. 2012).

precise if it is possible to identify the plant 'on the margin' and the sequence of dispatching electricity generation sources is clear. As it stands, this condition holds in the US, where nuclear and coal baseload plants are dispatched first, and if electricity demand is not met, the "cleaner" plants with higher cost become engaged (McCarthy and Yang 2010).[9] This implies that charging EVs during the *recommended* off-peak hours leads to higher marginal emissions than charging them during peak hours (Graff Zivin et al. 2014). However, marginal emissions also depend on the region the power plant is located in. For instance, only in two regions (Western interconnection and Texas) are emissions from charging an EV at night lower than those from driving a hybrid vehicle (Graff Zivin et al. 2014), and only there, there are environmental benefits from driving electric cars (Holland et al. 2016, Clinton and Steinberg 2019)[10]. This necessitates careful geographical and timely EVs market penetration: not all of the states are technologically prepared to preserve environmental benefits associated with driving electric cars.

However, even those states and countries that can currently accommodate a rapidly growing number of EVs still run the risk of losing this ability if the EVs fleet is expanded too quickly. For instance, in Los Angeles it is predicted that by 2030, EVs can achieve a significantly deeper market penetration and thus put more energy load on the grid, while simultaneously lowering marginal emissions contingent on coal phase-out in the region by 2025 (Kim and Rahimi 2014). Due to grid overload, charging during the night may result in inadequate cooling of the system's transformers and their subsequent degradation, necessitating frequent maintenance (Blumsack et al. 2008). Charging EVs during peak hours may lead to electricity shortages and even blackouts. This may entail energy imports with higher carbon intensity which will again raise marginal emissions negating the benefits of coal phase-out.

Thus, the adoption rate of EVs should be on par with a transition in the energy system, and since modifying the latter is a more effortful and complicated process,

[9]Dispatching model of generation sources in the UK differs from the one adopted by the US, and coal is almost phased out. Moreover, the share of renewable energy derived from sun and wind is higher, but due to the intermittent nature of these sources they can be dispatched at different times of day. Therefore, emission intensity during the day may not have such a significant variation and is expected to be lower than the US average. There are electricity providers in the UK that guarantee 100% clean energy from renewable sources delivered to one's house and charging point. There are also trials of a vehicle-to-grid programme which uses plugged in EVs as batteries to store renewable energy when it is generated until it is needed elsewhere. This reduces generation emissions further.

[10]Archsmith et al. (2015) find that when accounting only for life cycle emissions, EVs will undoubtedly emit less GHGs than a comparable petrol car, which is in contradiction with the results of Graff Zivin et al. (2014). This is due to including non-fossil fuel sources in their calculations in the former study. However, they also establish that temperature effects are significant, and in regions with colder climate and higher coal usage in power generation sector EVs can produce similarly sized emissions to ICE vehicles.

the substitution of ICE vehicles with EVs must happen at an exogenously defined rate.

Therefore, even with a slow adoption, an environmentally sound technology can be turned into a polluting one if the upstream technologies do not keep up with the transition. This suggests that promoting a downstream technology such as EVs should be done at a rate not exceeding the capacity of the electricity generation system to provide the cars with clean energy. Modifying a system as complex as electricity generation will inevitably run into many challenges. Power generation is an example of what Gregory Unruh called a techno-institutional complex (2000a) and Jochen Markard and co-authors – a socio-technical system (2012). Transition in such systems can be described as socio-technical transition that calls for infrastructural, social and institutional shift and is likely to be incremental due to path-dependence and carbon lock-in (Arthur 1989, Unruh 2000a), that together define the extent to which politicians can exercise their power to make a change (Unruh 2002). Thus, the rate of the transition is predefined by how quickly all of the elements of the system can adapt to the change.

Structural changes of the scale required to mitigate climate change will necessarily take time, and the speed of the transition for every country is mostly predetermined by the degree of its fossil fuel dependency as well as social and institutional inertia. The transition pace can be only marginally affected, and thus should be treated as given and unchangeable, at least in the short term.

1.3.3 Voting, timeline and optimality

A policy p^t at time period t is simply a proposal to implement either of the two political states within that time period.

There is an incumbent politician at the beginning of $t = 1$ who proposes a policy p^1. After observing the policy chosen by the incumbent politician, an entrant (drawn from an exogenous set of possible candidates) can choose to challenge the incumbent by paying a cost $h > 0$. If there is more than one challenger, one is chosen at random with equal probability. If there is no challenger, the incumbent politician gets re-elected with probability one. If there is a challenger, she proposes an alternative policy q^1. Each voter votes for their preferred policy. The politician with the majority of votes wins; if there is a tie, either candidate has a probability 0.5 of being elected.[11]

[11] We could have assumed that the incumbent politician has an advantage over any newcomer which is translated into her having a higher probability of being elected when there is a tie, but this assumption is irrelevant to the results of the model and we thus prefer to ignore it.

The elected politician implements her policy and if it differs from the previous period policy, voters "move" across the groups defining the distribution of the voters in the subsequent period.

For simplicity of exposition, we assume that the incumbent politician in office then obtains a per-period office rent consisting of a share $e > 0$ of the social payoff at the implemented political state.[12] We assume that $h > \hat{h} \equiv 3\beta(1 - \mu_0 + \varepsilon) - c\varepsilon$, so that in no subgame an entrant will choose to challenge the incumbent when both make identical policy proposals. Therefore, in equilibrium, as long as the incumbent makes a proposal that commands a majority, she will not be challenged.[13] The elected politician from the preceding period becomes the incumbent in period $t = 2$, at which point the political process described above is iterated.

The timeline within a period is depicted in Figure 1.1 below.

1. The share of voters in each group, μ_t, is defined.

2. An incumbent proposes a programme p^t.

3. A challenger (if any) makes an alternative policy proposal q^t and pays the cost h (if several, then, first, they propose their policies, then, one of them is randomly picked, finally, she pays the cost and gets to compete the incumbent).

4. Voters choose a new incumbent by the simple majority rule.

5. The winning politician implements proposed policy and if there is a change in the status quo, voters from the disadvantaged group queue to "move" to the advantaged group, ε of them gets picked. This defines μ_{t+1} for the next period.

6. Voters receive payoffs in accordance with the (new) state, politician receives her share of the social payoff. This completes the process and in the next period the game repeats from the top.

[12]This makes for a benevolent ruler whose payoff grows proportionally to social welfare. While in some countries this may well be the case, corruption among government leaders is widespread and states interested mostly in political rent do not seek to maximize (and consequently never maximize) social welfare (Shleifer and Vishny 1993, Mauro 1995). More on this in section 1.6.

[13]We understand that this deprives the model from an element of political competition that would normally be expected in the political economy models. However, not assuming an absolute advantage for the incumbent at the very least would not introduce any interesting implications to the model because it is unimportant for the solution. We could alter the model to have more interaction between competing politicians, elements of lobbying and strategic policy proposals, but this, intriguing as it is, is beyond the authors' concern in this paper.

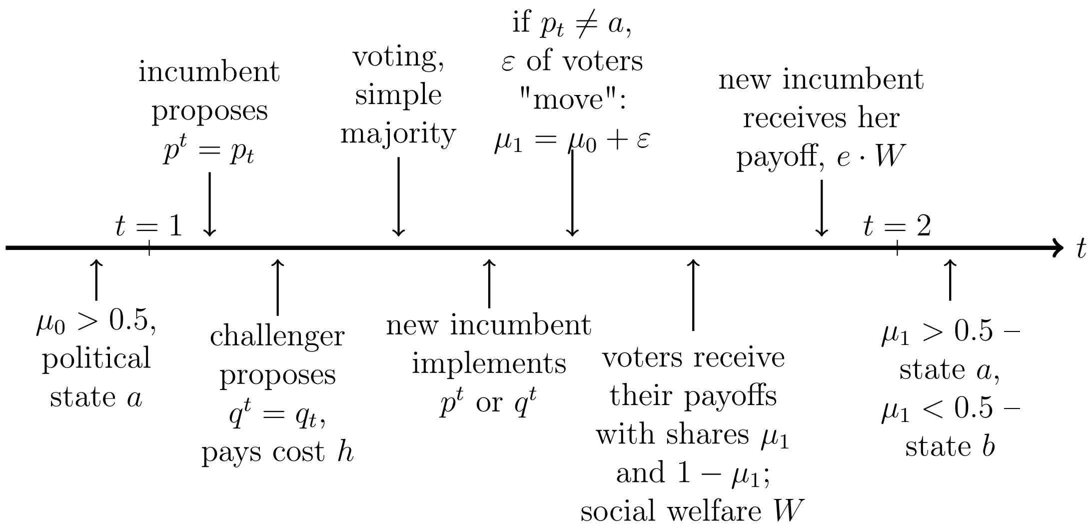

FIGURE 1.1: The order of play within one ($t = 1$) period.

Before we proceed to the solution, we introduce one final assumption: we assume that b is a socially efficient state in a sense that it is characterized by higher social welfare. This is achieved by assuming that

$$3\beta - \mu_0 c > 3(\alpha - (1 - \mu_0)\delta) \tag{1.1}$$

and

$$3\beta - \mu_0 c > 3\alpha - (1 - \mu_0)c \tag{1.2}$$

The former inequality guarantees that if the group mobility was absolute and everyone from G_1 could "move" to G_2 at $t = 1$, then social welfare across the whole game would be higher than if the voters stayed where they are. The latter compares two extremes and states that social welfare is higher when everyone from G_1 "moves" to G_2 at $t = 1$ rather than when everyone "moves" in the opposite direction. With such definition of a socially optimal political state, we can never reach the first-best solution, unless $\varepsilon \geq \mu_0$, which is a trivially uninteresting case. Therefore, we aim at achieving a constraint-optimal solution, which we define as a political strategy that features the highest social welfare under existing constraints.

Unfortunately, social optimality of b does not imply its constraint optimality: the change in social welfare is non-linear, and depends on the relative values of the parameters and the starting point, i.e., lower ε and higher μ_0 would imply a significant loss when shifting to b and non-commensurate gains that would not compensate for the losses even across the three periods. We therefore introduce a notion of constraint-optimality. If the transition to b is constraint-optimal, then it will lead to social welfare increase with given parameters. This assumption is reflected by the following inequality:

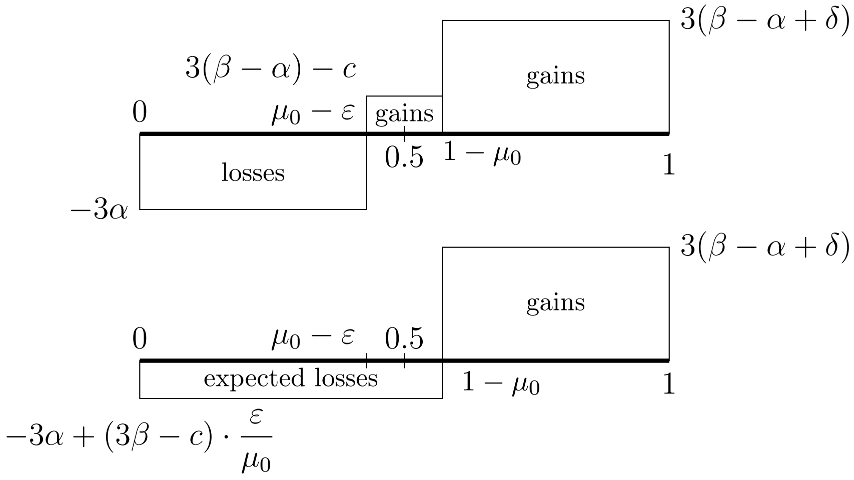

FIGURE 1.2: Relative gains and losses ex post (top figure) and ex ante (bottom figure) implementation of b. Since expected losses are negative, b will not be voted for.

$$3(1 - \mu_0)\,(\beta - \alpha + \delta) + \varepsilon\,(3(\beta - \alpha) - c) > 3\alpha(\mu_0 - \varepsilon). \tag{1.3}$$

The inequality ensures that the gains relative to the status quo are larger than losses, justifying the transition to b at the earliest stage. The expression does not calculate social welfare in absolute terms and rather compares the relative values of gains and losses as a result of the political transition.[14]

Equation 1.3 also defines the social planner solution. Without this assumption social planner solution would not exist, and any state could be constraint-optimal depending on the relative parameter values. The assumption thus establishes the earliest implementation of b as the welfare optimising policy.

Interestingly, Equation 1.3 alone does not mean that b will be adopted at $t = 1$. If voters in G_1 expect losses, they will not vote for b. In Figure 1.2, we show schematically gains and losses of the two groups of the society when b is a welfare-superior political state but does not carry through to being implemented. Voters in G_1 are shown to have expected losses rather than gains, which is not necessarily the case and depends on the mobility across the groups: as will be shown in the solution below, if $\varepsilon < 3\mu_0 \dfrac{\alpha}{3\beta - c}$, then expected payoff would indeed be negative and efficient institutions would not be adopted. Alternatively, if $\varepsilon \geq 3\mu_0 \dfrac{\alpha}{3\beta - c}$, then voters in G_1 expect to gain relative to the status quo and will vote for b. In the figure, we illustrate the situation when mobility is too low.

[14] A similar condition in Fernandez and Rodrik (1991) is used to establish that reform is welfare enhancing.

However, mobility being high enough, i.e., $\varepsilon \geq 3\mu_0 \frac{\alpha}{3\beta - c}$, is a sufficient (but not a necessary) condition for the transition to be constraint-optimal. Assumption 1.3 is thus obsolete if the transition does take place at $t = 1$, but necessary if it does not but should, because it would be welfare improving.[15]

Therefore, in what follows, we maintain that assumptions 1.1, 1.2 and 1.3 hold, together with $2\beta - c > \beta > \alpha > \beta - c$ and $c > \delta$. It is worth noting that these many restrictions hold for a non-empty open set of parameters. As the first example, assume $\mu_0 = 0.6$ and the mobility is $\varepsilon = 0.2$. Whether the transition happens depends on the relative values of other parameters. For instance, if $\alpha = 0.2$, $\delta = 0.1$, and $c = 0.8$, then b will be adopted at $t = 1$ if $\beta = 0.9$ (expected gains of 0.03, the boundary on ε equals 0.19), but it will never be adopted if $\beta = 0.85$ (expected losses of -0.017, the boundary mobility is 0.206). In both cases, however, net change of social welfare is positive: 0.98 in the former example, and 0.89 in the latter.

In the above example, the range of β is limited to $\beta \in (0.8, 1)$, as calculated from $\beta > \alpha > \beta - c$ with fixed values of α and c. Ceteris paribus, any value from this range will result in a positive social welfare change and will be constraint-optimal, but the reform will only be adopted without an intervention for $\beta > 0.87$. Below this threshold, expected payoff to the majority is negative and uncertainty elimination is needed to see the reform through.

Larger range for values of β will be available if we the overall magnitude of the parameters is increased. For instance, consider $\alpha = 0.5$, $\beta = 1.3$, $c = 0.85$, and $\varepsilon = 0.3$, while keeping the rest unchanged. It is clear that for $\beta \in (0.85, 1.35)$, any specific parameter value will render the reform an improvement over the status quo, but will only be adopted for $\beta > 1.28$, i.e., for the values of β for which the majority's expected payoff is positive.

For a more rigid economy with higher mobility frictions, it is easy to find parameter values that would fit the description of the model where a reform would be an improvement, but it almost never contains the set of parameters where the reform would be adopted on its own. Consider $\alpha = 0.85$, $\beta = 1$, $c = 0.9 > \delta = 0.8$, and $\mu_0 = 0.55$ with $\varepsilon = 0.1$. Tweaking these variables while keeping them within the welfare-improving range will never result in the threshold value of the mobility frictions below $\varepsilon^* = 0.67$, implying that the reform will never pass unless engineered,

[15]To contrast this with Fernandez and Rodrik (1991), in our model, the transition is never welfare decreasing if it is voted for. This results from the way the payoff structure is set up: expected payoff to the majority is always less than the total societal gains. Therefore, if expected payoff is negative, ex post social net welfare gains may or may not be positive. But, if expected payoff is positive, so is the net social welfare change. We cannot ever observe a reversal to the status quo in our model.

in which case it will be adopted within one period since 10% of the voters are more than enough to tip the majority from 55% to 45%.

1.3.4 Benchmark result

Here, we establish the following benchmark result characterizing the subgame perfect equilibria of the three-period sequential game of reform.

Proposition 1.1. *There is a cutoff value for domestic mobility frictions, call it $\hat{\varepsilon}$, such that the transition takes places for all $\varepsilon \geq \hat{\varepsilon}$.*

(a) Along the equilibrium path of play, if $\varepsilon < \max\{\mu_0 - 0.5, 3\mu_0 \frac{\alpha}{3\beta - c}\} \equiv \hat{\varepsilon}$, then $\{p_t = a : t = 1, 2, 3\}$; at each $t \geq 1$, the incumbent proposes $p^t = a$, and gets re-elected. There is Inefficient Institutional Persistence.

(b) Along the equilibrium path of play, if $\varepsilon \geq \max\{\mu_0 - 0.5, 3\mu_0 \frac{\alpha}{3\beta - c}\} \equiv \hat{\varepsilon}$, $\{p_t = b : t = 1, 2, 3\}$; at each $t \geq 1$, the incumbent proposes $p^t = b$, and gets re-elected. Efficiency-enhancing reform is adopted at an early stage.

Proof. Note first that $\mu_t \leq \mu_0$ for each $t = 1, 2, 3$. We solve the model using backward induction.

At $t = 3$, for any $p_2 \in \{a, b\}$, we examine four subgames.

1. If $\mu_2 > 0.5$, we need to consider two scenarios:
   - If $\mu_2 - \varepsilon > 0.5$, the incumbent politician will make a policy proposal $p^3 = a$, no challenger will arise and the incumbent will be re-elected with probability one and implement a;
   - If $\mu_2 - \varepsilon < 0.5$, then those in G_1 will compare α from voting for a and $\frac{\varepsilon}{\mu_2}(\beta - c) + \left(1 - \frac{\varepsilon}{\mu_2}\right) \times 0 = \frac{\varepsilon}{\mu_2}(\beta - c)$ from voting for b. Since $\alpha > \beta - c$ and $\frac{\varepsilon}{\mu_2} < 1$, we have $\alpha > \frac{\varepsilon}{\mu_2}(\beta - c)$, and thus voting for b is not associated with any gain for those in G_1, and they will therefore vote for a. The incumbent politician will propose and implement a. Notice that the politician should only be concerned with the payoff to the winning group (i.e., the majority at t), because she cannot maximize social payoff directly by imposing a Pareto superior alternative on the society.
2. If $\mu_2 < 0.5$, again, there are two scenarios:

- If $\mu_2 + \varepsilon < 0.5$, the incumbent politician will make a policy proposal $p^3 = b$; no challenger will arise and the incumbent will be re-elected with probability one and implement b.
- If $\mu_2 - \varepsilon < 0.5$, then those in G_2 can expect β from voting for b, and $\frac{\varepsilon}{1-\mu_2}(\alpha - c) + \frac{\varepsilon}{1-\mu_2}\beta < \frac{\varepsilon}{1-\mu_2}\beta + \frac{\varepsilon}{1-\mu_2}\beta = \beta$, which means that voters in G_2, once b is implemented, will not want to reverse this decision at $t = 2$. Hence, the politician will propose b and get re-elected.

Hence, a will be proposed and implemented when $\mu_2 > 0.5$, b will be proposed and implemented when $\mu_2 < 0.5$, regardless of the size of ε.

At $t = 2$, for any $p_1 \in \{a, b\}$, anticipating the outcome of the subgame at $t = 3$, there are several possible cases to look at.

1. If $\mu_1 > 0.5$, and
   - $\mu_1 - \varepsilon > 0.5$, then the incumbent politician will propose $p^2 = a$; no challenger will arise and the incumbent will be re-elected with probability one.
   - $\mu_1 - \varepsilon < 0.5$, then voters in G_1 anticipate that voting for a proposal $p^2 = b$ will result in the expected payoff

   $$\frac{\varepsilon}{\mu_1}(2\beta - c) + \left(1 - \frac{\varepsilon}{\mu_1}\right) \cdot 0 = \frac{\varepsilon}{\mu_1}(2\beta - c),$$

   while voting for proposal $p^2 = a$ will result in the expected payoff of 2α. Therefore, the incumbent will make the policy proposal $p^2 = b$ if and only if

   $$\frac{\varepsilon}{\mu_1}(2\beta - c) \geq 2\alpha \iff \varepsilon \geq 2\mu_1 \frac{\alpha}{2\beta - c};$$

   otherwise, the incumbent will make the policy proposal $p^2 = a$; no challenger will arise and the incumbent will be re-elected with probability one.

2. If $\mu_1 < 0.5$, and
   - $\mu_1 + \varepsilon < 0.5$, then the incumbent politician will propose the policy $p^2 = b$; no challenger will arise and the incumbent will be re-elected with probability one.

- $\mu_1 + \varepsilon > 0.5$, then voters in G_2 anticipate that voting for a proposal $p^2 = a$ will result in the expected payoff

$$\frac{\varepsilon}{1-\mu_1}(2\alpha - c) + \left(1 - \frac{\varepsilon}{1-\mu_1}\right)(2(\alpha - \delta)) = 2(\alpha - \delta) + \frac{\varepsilon}{1-\mu_1}(2\delta - c),$$

while voting for a proposal $p^2 = b$ will result in the expected payoff of 2β. Therefore, the incumbent will make the policy proposal $p^2 = a$ if and only if

$$2(\alpha - \delta) + \frac{\varepsilon}{1-\mu_1}(2\delta - c) \geq 2\beta;$$

otherwise, she will make the policy proposal $p^2 = b$. However, note that $2\beta > 2\alpha > 2\alpha - 2\delta + \frac{\varepsilon}{1-\mu_1}(2\delta - c)$, because $-2\delta + \frac{\varepsilon}{1-\mu_1}(2\delta - c) < 0$, and thus, voters in G_2 will never vote for a once b is implemented. Therefore, the incumbent politician will propose b, no challenger will arise and the incumbent will be re-elected with probability one.

At $t = 1$, given that the status quo is a and $\mu_0 > 0.5$, anticipating the outcome of the subgame at $t = 2$, there are only two possible scenarios:

- If $\mu_0 - \varepsilon > 0.5$, then the incumbent politician will propose $p^1 = a$, no challenger will arise and the incumbent will be re-elected with probability one. Note that we obtain this conclusion even though each incumbent politician, conditional on remaining in power, prefers the alternative b to a, but any attempt to alter the status quo implies that the incumbent politician loses power with probability one.

- If $\mu_0 - \varepsilon < 0.5$, then voters in G_1 anticipate that expected payoff from voting for a proposal $p^1 = a$ will be

$$\alpha + \frac{\varepsilon}{\mu_0}(2\beta - c) \text{ if } \varepsilon \geq 2\mu_0 \frac{\alpha}{2\beta - c},$$
$$3\alpha \text{ if } \varepsilon < 2\mu_0 \frac{\alpha}{2\beta - c},$$

while from voting for the proposal $p^1 = b$ it will be $\frac{\varepsilon}{\mu_0}(3\beta - c) + \left(1 - \frac{\varepsilon}{\mu_0}\right) \times (3 \cdot 0) = \frac{\varepsilon}{\mu_0}(3\beta - c)$. Therefore, voting will take place as follows:

$$\text{if } \varepsilon \geq 2\mu_0 \frac{\alpha}{2\beta - c}, \text{ vote } b \text{ iff } \varepsilon \geq \mu_0 \frac{\alpha}{\beta}, \text{ otherwise vote } a,$$
$$\text{if } \varepsilon < 2\mu_0 \frac{\alpha}{2\beta - c}, \text{ vote } b \text{ iff } \varepsilon \geq 3\mu_0 \frac{\alpha}{3\beta - c}, \text{ otherwise vote } a.$$

Note that $\mu_0 \frac{\alpha}{\beta} < 3\mu_0 \frac{\alpha}{3\beta - c} < 2\mu_0 \frac{\alpha}{2\beta - c}$. Thus, the incumbent will make the policy proposal $p^1 = b$ if and only if

$$\varepsilon \geq 3\mu_0 \frac{\alpha}{3\beta - c},$$

otherwise, the incumbent will make the policy proposal $p^1 = a$, no challenger will arise and the incumbent will be re-elected with probability one.

The cut-off value for ε is therefore $\hat{\varepsilon} = \max\{\mu_0 - 0.5, 3\mu_0 \frac{\alpha}{3\beta - c}\}$. This completes the proof. ■

Proposition 1 calculates the bounds on the mobility friction above which there is inefficient institutional persistence and below there is early adoption of efficient institutions. The closer μ_0 is to 0.5, the lower is the bound; the lower is α (the instantaneous loss in payoff to a voter in G_1 when the political state shifts from a to b), the lower is the bound; the higher is $3\beta - c$ (the future gain in payoff to a member of G_1 who successfully becomes a member of G_2 following a shift in the political status quo from a to b), the lower is the bound.

1.4 Winning coalitions and reform

The benchmark result identifies the parameter values for which reform is never implemented even though it is constraint-optimal to do so: when $\varepsilon < \max\{\mu_0 - 0.5, 3\mu_0 \frac{\alpha}{3\beta - c}\}$. In this section, we study mechanisms where, before reform is proposed as a policy option, a group of voters from within the majority that prefer the status quo is identified to enable implementation of the reform.

In what follows, it is assumed that the incumbent cannot directly manipulate the values of the underlying payoff parameters and takes mobility frictions as given: $\varepsilon < \max\{\mu_0 - 0.5, 3\mu_0 \frac{\alpha}{3\beta - c}\}$ so that, by Proposition 1, the reform will never take place. Parameter combinations corresponding to different mechanisms are depicted in Figure 1.3.

1.4.1 Engineering majority in favour of the alternative

Consider an incumbent whose goal is to stay in power while trying to implement the reform.

From the proof of Proposition 1, note that when $\mu_0 - 0.5 < \varepsilon < 3\mu_0 \frac{\alpha}{3\beta - c}$, if there is a way to engineer a switch to b at $t = 1$, in equilibrium, at $t = 2$, b

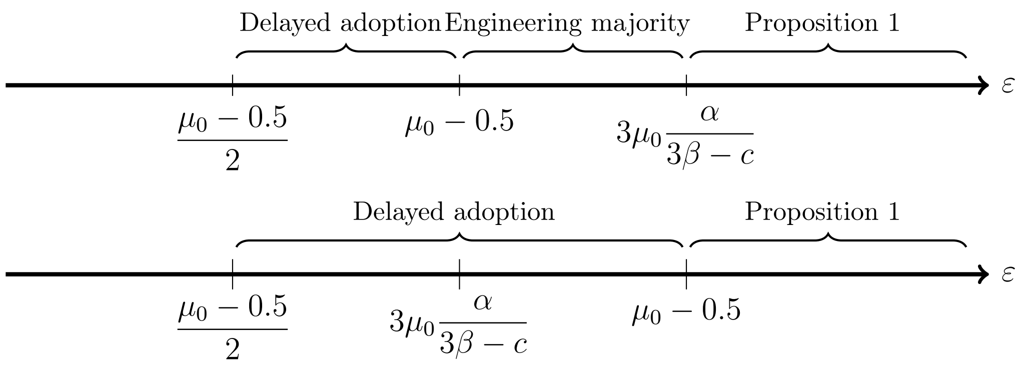

FIGURE 1.3: The ranges of ε that permit a political transition as a result of applying a particular mechanism.

continues to persist. The reason for the persistence of the inefficient status quo in the above scenario is that although voters in G_1 realise that the switch to b at $t = 1$ will ensure that b prevails thereafter, the probability of any individual voter in G_1 of being able to actually "move" to G_2 is small and expected gain is insufficient for b to payoff-dominate a.

So how can the majority in favour of the reform be engineered?

The intervention that might work in this setting is to eliminate individual-specific uncertainty by ensuring that all voters belonging to a specific subset of G_1, say G_1', of mass ε, are guaranteed to become reform winners if they vote for b (at $t = 1$). Straightforwardly, once such a subset is identified, computations symmetric to those underlying Proposition 1 will ensure that if offered a choice between a and b, all voters in $G_1' \cup G_2$ will vote for b at $t = 1$. Specifically, those in G_1', predicting the stability of b in the future, expect with certainty to receive $3\beta - c$ from voting for b, and assuming that $3\beta - c > 3\alpha$, they will vote for b. Together with voters in G_2, they form the mass sufficient to pass the reform, since $1 - \mu_0 + \varepsilon > 0.5$.

Such transition could also take place at $t = 2$ (but not at $t = 3$, due to insufficient time left to compensate those who "moved" and had to pay the cost, as $\beta - c < \alpha$), however, transition at $t = 1$ is welfare-superior and therefore preferred.

Proposition 1.2. *Provided that* $\mu_0 - 0.5 < \varepsilon < 3\mu_0 \frac{\alpha}{3\beta - c}$ *and* $3\beta - c \geq 3\alpha$, *there exists* $G_1' \subset G_1$ *of early "movers", such that* $G_1' \cup G_2$ *has a mass greater than 0.5, which ensures that the reform is adopted at* $t = 1$.[16]

Proof.

At $t = 3$, the solution is the same as in Section 1.3.4: as long as $\mu_2 > 0.5$, for any $p_2 \in \{a, b\}$, the incumbent will suggest $p^3 = a$ and win. This is due to insufficient

[16] Since all the sets in our model are subsets of the continuum of voters, we cannot apply the concept of cardinality to compare them. Therefore, we resort to term "mass", which is synonymous to voters' share.

time for the "movers" to make up for the foregone payoff at $t = 3$: $(\beta - c)$ is under no circumstances preferred to α, and whether ε of the voters in G_1 are aware of their "move" to G_2 in advance has no bearing on their decision.

It is worth pointing out that since the earliest adoption of reform is welfare dominant, we focus on $t = 1$. This implies that $\mu_1 = \mu_0 - \varepsilon < 0.5$, and at $t = 2$, we only need to consider two subgames. To simplify further, recall that when $\mu_1 < 0.5$, both subgames result in the same outcome, namely, b persists once adopted. Therefore, at $t = 2$, conditional on the engineering having taken place at $t = 1$, b will continue to prevail. Note, that here, $\varepsilon < 3\mu_0 \dfrac{\alpha}{3\beta - c}$, and therefore, it is also true that $\varepsilon < 2\mu_0 \dfrac{\alpha}{2\beta - c}$. Thus, there is no way b is voted for at $t = 2$ without prior engineering, and the subgames where $\mu_1 > 0.5$ will always see a implemented.

At $t = 1$, given the status quo is a, in contrast with the benchmark, we will only look at one case, because, by the limitation of this mechanism, it is necessary to have $\mu_0 - \varepsilon < 0.5$. Thus, $\varepsilon \subset G_1$ who form G_1', predicting the stability of b in the future, expect with certainty to receive $3\beta - c$ from voting for b. Provided that $3\beta - c > 3\alpha$, those in G_1' will vote for b, so will those in G_2, and since $1 - \mu_0 + \varepsilon > 0.5$, the incumbent politician will propose $p^1 = b$, which will subsequently be implemented.

■

Contrast Proposition 1.2 with an example of a trade reform found in Fernandez and Rodrik (1991). They describe a small economy which is considering a decrease in a tariff. This will change the relative prices of the two goods and make working in one sector more attractive. Relocating between the sectors is costly for individuals, but the cost is not known in advance; it is revealed only after the relocation has taken place. An example is constructed to show that there exist circumstances under which a welfare-enhancing trade reform would not be adopted when workers do not know their relocation cost in advance, but would if this uncertainty was somehow eliminated. Based on the heterogeneity of individual costs, they demonstrate that there exists a value of the cost for which everyone with lower cost changes the sector and votes for the reform, and everyone with higher cost does not, but that the reform is passed. The group of workers for whom it is "cheaper" to relocate are the ones ensuring that the majority is in favour of the reform. Unlike in our model, this group is not of a fixed size, and the reform adoption is ensured by establishing that there exists such value of the tariff. Proposition 1.2, on the other hand, establishes existence of such group, taking the threshold value as given. In Fernandez and Rodrik (1991), there is no limit on how many workers can relocate between the sectors, but equally there is no lower bound in a sense that a new tariff value has to be carefully chosen to ensure that enough workers decide to relocate. In contrast, in our model, if there is a chance of engineering the reform, there are always more than enough voters to choose from to form the ex post majority. Thus, unlike in Fernandez and Rodrik (1991), the reform itself does not matter in our paper (i.e., the tariff value does not have to be chosen strategically), the state of the economy does, which allows for a wider range of social issues to be resolved within our framework. And most importantly, Fernandez and Rodrik (1991) do not discuss how the uncertainty can be eliminated. Below, we propose several approaches to this crucial question.

The marker identifying the subset G_1' could, for instance, be political orientation. Research suggests that liberals are more likely than conservatives to be concerned about sustainable consumption and act accordingly (Watkins et al. 2016). This is a consequence of their moral values such as concern for fairness and social justice. The authors conclude that since liberals are more responsive, a policy should target them specifically with the messages that appeal to the moral foundations of their choices.

Liberals are also more likely to progress up the consumption hierarchy to harder sustainable consumption practices that require more commitment, discipline and personal sacrifice. According to Wooliscroft et al. (2014), consumers pass through certain stages in their ethical (sustainable) choices. People progress from easier ethical consumption actions to harder ones (Watkins et al. 2016) and can be in different stages for different behaviours. Particular ethical consumption actions tend to be synchronised. Policymakers can thus promote behaviour that is usually paired with another ethical choice in an area if they know that the latter is already being performed there. This is in line with suggestions drawn from evolutionary economics by Marechal and Lazaric (2010), who propose to incentivize consumers using non-market mechanisms that are tailored to specific characteristics of targeted groups.

Alternatively, if agents have positive but small heterogeneous queuing costs (for consistency with our existing computations, we will need to assume that the queuing costs are small enough so that it is risk-dominant for all voters in the disadvantaged group to queue up whenever there is a shift in the political status quo), G_1' could be constructed on the grounds of economic efficiency, i.e. the subset of agents in G_1 whose moving costs across the two groups is lowest. Another possibility is to create a first mover advantage in moving out of G_1 by compensating the first ε voters in the queue.

Acting on the supply side, politicians could also identify a specific industry (or firms within an industry) and target it with subsidies to ensure a switch to a low carbon technology. In this case, governments may want to start from creating niche markets and managing them strategically to achieve a technological regime shift. This implies a government trying on a role of a catalyst and facilitator rather than a regulator or benefactor (Kemp et al. 1998). Marechal and Lazaric (2010) also note that the obstacles to wider implementation of efficient emission-reducing investments are likely of non-economic nature and advise to target "lead users", i.e., early adopters, which is another way to identify a pivotal group of consumers for the mechanism discussed here. Similarly, when it is easier to tax the "losers" than to pick "winners", the regulator can tax a technology without a learning potential rather than trying to identify and subsidise the most dynamically efficient one (Kalkuhl et al. 2012). This can also happen if the benefits of a policy are spread among many, while the costs fall on the few. It might be wise in such case to compensate all or selected (pivotal) "losers" to weaken anti-regulatory coalitions (Oye and Maxwell 1994).

In effect, the argument above demonstrates that by discriminating between agents in G_1 and favouring voters belonging to a carefully chosen subset G_1' over other agents in G_1, a majority in favour of b is constructed and reform is implemented immediately.

1.4.2 Delayed Adoption as a Commitment Device

In this subsection, we consider a situation where the critical mass required to implement the reform can only be reached if the mobility constraint within a period is violated. We continue to assume that the values of the underlying payoff parameters cannot be directly manipulated.

By assumption, voters have no incentives to make costly adjustments to their lifestyle preceding the changes to climate legislation. However, as seen from the benchmark solution, b can be implemented at $t = 2$, provided that ε of farsighted voters in G_1 "move" to G_2 at $t = 1$, before the reform is proposed. They will therefore need a guarantee that the reform will be carried out in the next period. And although commitment issues are ubiquitous in policy making, here, the group of early adopters of low-carbon lifestyle act themselves as a commitment device: they will hold politician's feet to the fire if the policy is not enacted. They are the ones who forfeit some of their benefits in the current period in anticipation of recovering these losses in the future when the status quo changes. Existence of such subset of early adopters is necessary, because the transition *cannot and does not* take place in the same period.

This idea can be re-framed in terms of Coate and Morris (1999) who hypothesize that voters take actions to benefit from a newly introduced economic policy, which increases their willingness to pay for the policy in the future, which in turn translates

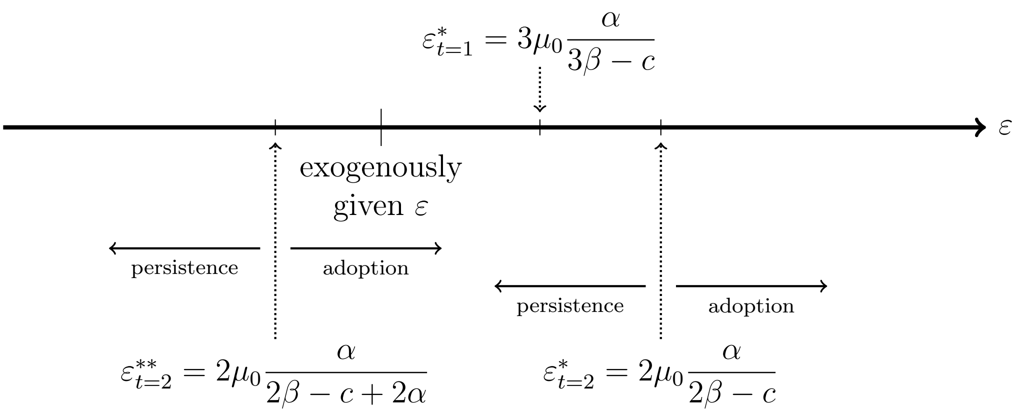

FIGURE 1.4: Comparison of the boundary values for ε with and without ex ante displacement of ε of voters to G_2. Delayed adoption mechanism permits a political transition to take place with lower exogenously given mobility. Note that $\varepsilon^*_{t=1}$ must be greater than ε that is given for Proposition 1 (inefficient institutional persistence) to hold. At the same time, $\varepsilon^{**}_{t=2}$ must be below ε and therefore below $\varepsilon^*_{t=1}$ too. The latter would be true if $\alpha > \frac{c}{6}$.

into pressure to retain the policy. They use this argument to explain why inefficient policies persist, but it can equally be used to forcefully make desirable policies stick. They model a situation where a policy that provides temporary efficient benefits would be forgone because voters predict its persistence once implemented, which gives rise to political failure. However, by the same logic, if a policy was *somehow* passed which forces some voters to pay the cost to benefit from it ("move" between the groups in our model, or relocate between the sectors in the terminology of Coate and Morris (1999)), then it would persist without external help, in line with Proposition 1.1.

Delayed reform implementation can be described as follows. At $t = 1$, the incumbent can propose that an ε fraction of agents be given an option to "move" between the groups. In other words, a carefully identified share of voters among the current majority G_1 is targeted to alter their lifestyle while the status quo is maintained. Then, as we know from the benchmark solution, at $t = 2$, voters will choose b if $\varepsilon \geq 2\mu_1 \frac{\alpha}{2\beta - c}$. Provided that some voters "moved" to G_2 at $t = 1$, $\mu_1 = \mu_0 - \varepsilon$, and, to implement b, we now require $\varepsilon \geq 2\mu_0 \frac{\alpha}{2\beta - c + 2\alpha}$. As long as this condition holds, b will be implemented at $t = 2$ (see Figure 1.4).

The only thing left to check are the incentives of the early adopters – their expected future gains have to exceed immediate losses. If b is implemented at $t = 2$, they compare $\alpha - \delta - c + 2\beta$ with 3α, and so long as $\beta \geq \alpha + \frac{\delta + c}{2}$, they will agree to "move" to G_2 at $t = 1$ at their own (temporary) expense.

Proposition 1.3. *Along the equilibrium path of play, if* $\mu_0 - \varepsilon > 0.5 > \mu_0 - 2\varepsilon$, $\beta \geq \alpha + \frac{\delta + c}{2}$ *and* $\varepsilon > 2\mu_0 \frac{\alpha}{2\beta - c + 2\alpha}$, *then* $p_1 = a$, $p_2 = p_3 = b$; *at* $t = 1$, *the incumbent politician proposes* $p^t = a$ *and* ε *fraction of voters in* G_1 *become early adopters at* $t = 1$, *while at* $t \geq 2$, *she proposes* $p^t = b$ *and gets re-elected. Socially efficient policy is adopted with a delay of one period.*

Proof. At $t = 3$, the solution is the same as in Section 1.3.4.

At $t = 2$, we only consider the subgame where $\mu_1 > 0.5$ and $\mu_1 - \varepsilon < 0.5$. As we know from the benchmark solution, voters will choose b if $\varepsilon \geq 2\mu_1 \frac{\alpha}{2\beta - c}$. Thus, when some voters "move" to G_2 at $t = 0$, $\mu_1 = \mu_0 - \varepsilon$, and, to promote b, we now require $\varepsilon \geq 2\mu_0 \frac{\alpha}{2\beta - c + 2\alpha}$. As long as this condition holds, b will be implemented at $t = 2$. At $t = 1$, assuming $\mu_0 - \varepsilon > 0.5 > \mu_0 - 2\varepsilon$, we only need to examine the incentives of the early adopters – they have to find future payoff gains sufficient to offset immediate losses. Since they expect the transition to take place at $t = 2$, they compare $\alpha - \delta - c + 2\beta$ with 3α, and so long as $\beta \geq \alpha + \frac{\delta + c}{2}$, expected payoff from moving to G_2 at $t = 1$ is higher than from choosing not to do so. They will prefer to embrace personal changes ahead of the political transition. ■

At first glance, after ε of voters relocate at $t = 1$, the society ends up in the same conundrum as in Proposition 1.1 with $\varepsilon < 3\mu_0 \frac{\alpha}{3\beta - c}$, where the reform will still not pass without further intervention. However, the Proposition 1.3 works not by just relocating ε of voters, but by changing the lower bound required for the reform to come to pass on its own. Once the new boundary value on mobility frictions is below the one given exogenously, the transition will automatically take place because voters' expected payoff from supporting the reform is now positive. Therefore, the share of voters who have been inconvenienced prior to the reform, perform the role of a commitment device in a sense that they lower the boundary value which allows the transition to take place, and should an incumbent politician offer to stay in the old political state at $t = 2$, she will be challenged and lose the office, because the society now is ready for a change.

It is helpful to illustrate the argument with an example. Consider parameter values in a similar range as before: $\alpha = 0.35$, $\beta = 1.5$, $c = 1.2$, $\delta = 0.8$, $\varepsilon = 0.2$, and $\mu_0 = 0.71$. Evidently, it takes two voters' relocation to see the pro-reform share, $1 - \mu_t$, exceed half, therefore the reform can only be adopted with a delay, if at all. Without an intervention, the boundary on ε is equal to $\varepsilon^* = 0.226 > \varepsilon = 0.2$. However, once ε share of voters pay the cost and move at $t = 0$, the new boundary value become $\varepsilon^{**} = 0.199 < \varepsilon$, and thus, at $t = 2$, the transition will take place as per (slightly modified) Proposition 1.1, without any further meddling.

One can also imagine an alternative narrative to Proposition 1.3 in a form of endogenous mobility frictions. Instead of "moving" ε of voters from G_1 to G_2 ahead of the transition, and if manipulating the rigidity of the economy was possible, an incumbent could invest at $t = 1$ in lowering future mobility frictions, so that, with exogenously given boundary value $\varepsilon^*_{t=2}$, ε increases. Note that since the game only has three periods, the boundary values on mobility frictions calculated with the reform passing at $t = 1$ and at $t = 2$ are different, and the later the transition takes place, the higher the required value is. This way, instead of adjusting the boundary value, it is taken as given, and mobility frictions are changed.

Using numerical example from above, the boundary value for the transition to take place at $t = 1$ equals 0.226, but, ceteris paribus, the same value for the transition to take place at $t = 2$ equals 0.276. Therefore, to pass the reform in the second period, the investment at $t = 1$ would have to increase ε from o.2 to above 0.276.

If such manipulation is possible, then it can serve as a signal of commitment of the incumbent to the new green policy. This is similar to the results of the analysis conducted by Harstad (2020) who suggests that, when politicians are time-inconsistent, the best course of action for them is to invest strategically in technologies that are complements to future investments or are further upstream in supply chain. Referring to previously discussed example with electric vehicles, this approach would imply investing in renewable electricity to ensure that more electric cars can hit the road in the coming years without losing environmental benefits that come with it.

As it stands, however, Proposition 1.3 relies on a well-established idea of building constituency in favour of the transition; those with vested interest in moving away from the status quo. Green policies, for instance, create constituencies that are invested in decarbonisation. As time passes, such groups will form powerful coalitions that are opposed to policy regression, thus providing credibility to government's intentions. For lengthy transitions such as global decarbonisation that are likely to take decades, credible policy commitment is critical.

Credible policy commitment is critical for climate policy. Green innovations will not take off being solely driven by the market (Meadowcroft 2011). Such transition is a long-term process, which would benefit from targeted industry-specific innovation policies (Meckling et al. 2015). Any policy creates those with a vested interest in perpetuating it. Green policies create constituencies that are invested in decarbonisation. As time passes, they will create coalitions that are opposed to policy regression and are in favour of tougher regulations. Positive feedback will ensure that such coalitions get stronger over time allowing policies to be passed smoother.

To do so, initially a policy should be specific and targeted at a certain industry or group of people. It can be a market policy that creates or restructures incentives or groups empowerment through social movement (Bernstein and Hoffmann 2018).

To create groups with vested interest, a policy initially should be specific and targeted at a certain industry or group of people. It can be a market policy that creates or restructures incentives, empowers groups through social movement (Bernstein and Hoffmann 2018). Later, when the policy is established and well rooted, it can become broad and less targeted. The sequence of political measures is of utmost importance here, and before a programme can be introduced, building societal and strategic business support is necessary (Geels et al. 2017). And, if revenues from a decarbonisation programme are reallocated to public, public scrutiny will be strongly encouraged and the government will have less incentive to reverse or not comply with the policy (Brunner et al. 2012).

This parallels the conclusions made by Harstad (2020) about commitment-like role of technology. A time-inconsistent policymaker can commit to future decisions by strategically investing in capital to alter the costs or benefits of future actions, such as a subsidy on green technology. Moreover, in line with our recommendation to create niche market and target early adopters, he proposes to invest more in upstream technologies because they influence all the subsequent steps.

1.4.3 Exploiting network externalities

When a techno-institutional transition is engineered in one country, how does this reflect on domestic policies and economies of other nations? In other words, can a transition be engineered in more than one country when an outside agent only targets one?

Almost all of the countries today are part of one or more networks for trade, geographical, ethnic or other reasons. There is a multitude of international agreements that are signed and ratified by those countries, legally binding them to act in a certain way and defining constraints on the actions that they do have discretion in. Crucially, actions of a significantly large and important members of the network commonly have effect on other members' incentives and possibly even choice sets; i.e., introduction of trade barriers will have consequences for the importing country as well as for the importers themselves. This directed link can be explored and exploited.

This section is a natural extension of the mechanisms introduced above. If a transition is possible to engineer in one country, by capitalizing on the connections nations have with one another, the transition may be possible to engineer in a network of countries, using only limited resources. Below we discuss the conditions for this to be achievable and examples that illustrate our argument.

Assume now that there are n countries, $i = 1, ..., n$, that make up a connected network[17]. Within each country, the game is largely the same: there are two groups of voters, G_1^i (initial mass μ_0^i) and G_2^i (mass $1 - \mu_0^i$), who choose between two political states, $p_t^i \in \{a, b\}$, starting at a. Political state of the whole network at t is a vector $\mathbf{p}_t = (p_t^i : i = 1, ..., n)$. There are no alterations to the original model other than an obvious change of notations.

The key modelling change is that the payoffs within each country will now depend on the network structure. For tractability, only one payoff value is state-dependent.[18]

The per-period payoffs to voters in country i are:

1. if $k \in G_1^i$, $\alpha^i(\mathbf{p}_t)$ if a is the current political state, 0 if b is the current political state;

2. if $k \in G_2^i$, $\alpha^i(\mathbf{p}_t) - \delta^i$ if a is the current political state, β^i otherwise. The payoff is therefore increasing in the number of countries that have undergone the transition.

A helpful illustration of the current setup is an introduction of tougher environmental production standards or a complete ban of certain (common) chemicals from the production cycle by a country that relies on import to fully satisfy its domestic demand. This unilateral action forces importers to readjust causing their incentives structure to change. Their payoff from enacting similar environmental standards is now higher.

Let $\mathbf{p}_t^a = (p_t^i = a : i = 1, ..., n)$.

The costs of "movement" and mobility frictions are country-specific and there is no change in the way we specify how group membership evolves. Political state b is still assumed socially and constraint-optimal for the entire network of countries at $t = 0$, although within each country any member of the advantaged group is strictly worse off under the status quo if the network as a whole shifts to the alternative.

[17]The concept of connectivity is borrowed from the graph theory but the terms are used here in a more general sense.

[18]Pure simplification that has no bearing on the result. The choice of the parameter that is network-structure-dependent is also strategic to simplify the search for the borderline condition. Any other payoff structure could be used and interpretation of the results should not be built on a particular specification used.

Suppose that an outside agent has a limited quantity of resources to play with. Also, assume that there is a pivotal country in this network which sets in motion the process of transition for the whole network[19].

To illustrate how such a mechanism might work, we need some notations. For any non-empty subset of countries C, let

$$\mathbf{p}_t^C = \begin{cases} p_t^i = b \text{ if } i \in C \\ p_t^i = a \text{ if } i \notin C. \end{cases}$$

For each country i, assume that there exists a non-empty subset of countries $C_i = \{i^{'} < i\}$ such that $\alpha_1^i \left(\mathbf{p}_t^{C_i}\right) < 0$. C_i may thus be interpreted as a set of countries that for i, if all switch to b, make up necessary and sufficient critical mass to push the majority payoff to turn negative, which consequently makes b more attractive. Note that under this assumption, country 1 is a pivotal member of the network and political shift in it triggers a shift in all other countries. Interestingly, mobility frictions in all but a pivotal country are unimportant. Their defining role is essentially assumed away since the increase in the social payoff is of the magnitude that matches actual mobility and its boundary value for the transition to take place.

Depending on mobility frictions of a pivotal country, different policies are available to the outside agent. If ε^1 is not too small, $\varepsilon^1 > \mu_0^1 - 0.5$ and $\varepsilon > 3\mu_0 \frac{\alpha^1(\mathbf{p}_t^a)}{3\beta^1 - c^1}$, b will be implemented trivially. If either of the conditions does not hold, the transition must be engineered. This can be achieved using any appropriate policy, including the two mechanisms described above.

The easiest to illustrate is a subsidy, s. An outside agent would have to subsidize voters in G_1^1 or G_2 so that

1. either $\varepsilon^1 > \dfrac{3\mu_0^1\alpha^1(\mathbf{p}_t^a)}{3\beta^1 + s - c^1}$, if $\varepsilon^1 > \mu_0^1 - 0.5$,
2. or $\alpha_1^1\left(\mathbf{p}_t^a\right) < s$ otherwise.

Although one option may be cheaper, the outside agent will have to choose based on the circumstances, but both approaches ensure that, assuming all other countries vote for a, it is a dominant strategy for voters in country 1 to vote for b at $t = 1$. If the transition in country 1 takes place, voters in other countries, correctly predicting each other's behaviour and payoff, will vote for b at $t = 2$. Assuming that transition to b at that period is still constraint-optimal, an outside agent thus secures a switch to the socially efficient political state globally.

[19]Might be a set of such pivotal countries. This assumption is irrelevant from the conceptual point of view.

Therefore, in our model, a (benevolent) international agent always improves on the political situation in a country, and, by extension, in a group of countries. Contrast this with the findings of Galiani et al. (2019) who, as a result of their analysis, posit that foreign funds make reforms as well as reversals more likely, especially if an international agent is myopic and not fully aware of potential reversals. The welfare in that case can be lower than with no intervention at all. The result is due to the possibility of transfers from the winners to the losers of the reform, that are announced before the election and are meant to provide incentive for future losers to still vote for the reform. As in Fernandez and Rodrik (1991), uncertainty about the ex-post welfare distribution is present, but unlike there, their identities are known; what is not known is who among the losers will get compensated, since not everyone will. The last part bears more resemblance to our model, if we ignore different lexical choices between the two papers. The idea that a share of losers would have to be compensated is the same as saying that a share of voters would be drawn from one production sector to another, or would have to "move" groups, because those are precisely the voters who will benefit from the new policy as if they were compensated. The major difference is that we do not assume the possibility of transfers, since the ruling party cannot credibly commit to these after the reform has passed. This assumption is however relaxed in the case of external agency and we allow for a subsidy as a viable option of inducing a transition. It seems more likely and more readily accepted by the public in a pivotal country that an international organisation provides funds rather than a way to handpick future reform beneficiaries.

For instance, a successful policy implemented by one country can encourage other countries to copy the approach, providing necessary feedback to spur a rapid development of a relevant market on an international scale. Such was the tale of the solar energy sector: the dynamics between Germany and China, their different stages of development and complementarity provided the stage for technological innovation system growth and global expansion in the field of solar PV panels, encouraging other countries to adopt similar policies, further contributing to the global market development (Quitzow 2015). While this particular example illustrates more of a bidirectional influence between two economies, it generally falls under the umbrella of the international technology transfer, knowledge spillovers and innovation diffusion. There is a bewildering amount of work done in these areas, but taking a cursory look, some research suggests that before being adopted worldwide, a technological innovation is accepted by one country, from which it may or may not spread to others (Beise 2004). This country will then become a lead market. There are ways to assess a lead market potential for an international agent to then act based upon the assessment results (Beise and Cleff 2004). The authors establish several key attributes of a country that increase the chance of successful innovation adoption. Although the article is written from an organisational perspective, it can be of use for an international policymaker looking to start a rapid transition away from fossil fuel intensive technologies to yet more expensive and less developed environmentally friendly technologies. Transnational linkages will then transfer the impact from the lead market to the followers (Gosens and Lu 2013).

Engineering transition in a market with a global lead potential is a different question, and sometimes subsidising it is inevitable. Even though subsidies can be suboptimal to taxes, due to political and institutional constraints, they may sometimes be the best available option. Welfare efficiency of different subsidies is not easily estimated empirically, especially when distortionary taxes are present (Parry 1998), and having an external funder therefore bypasses this efficiency ambiguity. With funds coming from an international organization such as Green Climate Fund, Children's Investment Fund Foundation, World Bank, etc., welfare improvement in the recipient country is near certain, and the main focus shifts to monitoring how the funds are being utilised. Provided that conditional aid and monitoring are possible, subsidising a pivotal country in a network may well be justified.[20]

Generally speaking, being part of an international network means that the benefits from implementing more sustainable practices are higher than in a stand-alone situation and the burden is shared by using a second-mover advantage and relying on experience of other members. Our model thus suggests that instead of focusing on creating an agreement which most countries would ratify, an international institution such as the United Nations should focus on convincing pivotal countries to make a transition first, so that the rest of the world would follow without much more nudging. The US and China as the biggest emitters have to be on board with the rest of the world's intention to mitigate climate change; without them, any agreement to decrease global GHG emissions is doomed. The change will have to start from somewhere, as it did in the case of the Montreal Protocol, if the world is to have any hope to avert the worst.

[20] Funding of such international organizations is a separate issue which we do not consider here, but it is worth noting that the decision about being a donor at a domestic level is a complex interplay between many factors, including public opinion and economy's capacity (Halimanjaya and Papyrakis 2015, Pickering and Mitchell 2017), and it is not a given that such fund would possess enough resources to initiate a transition or to monitor how the funds are used, or indeed perform as expected (Bracking 2015).

1.5 The T period model

In this section, we study a $T > 3$ period extension of the three-period model studied in the main body of the paper. We assume that the reform is constraint-optimal in a game with $T < \infty$ periods and that the sequence of events within each period remains unchanged. In addition, we assume that each voter has a discount factor $\sigma \in [0, 1]$. Under the assumption that the incumbent politician has impact on the evolution of group membership, Proposition 1.1 can now be extended to a game with arbitrary number of periods.

Proposition 1.4. *For each value of the discount factor $\sigma \in [0, 1)$, there is a cutoff value for domestic mobility frictions, ε^*, such that, for $\varepsilon \geq \varepsilon^*$, efficiency-enhancing reform is implemented at $t = 1$, and for all $\varepsilon < \varepsilon^*$, it is never implemented. Moreover,* $\lim_{\sigma \to 1} \varepsilon^* = \dfrac{T\mu_0\alpha}{T\beta - c}$.

Proof. See appendix. ■

The only scenario in which the reform is voted for occurs when $\varepsilon \geq \varepsilon^*$. As in the three-period game, the reform is implemented at $t = 1$ or never adopted at all.

Voters who benefit from the status quo trade-off the transition costs of pro-climate reform with the present value of future benefits from such reform. If the magnitude of the domestic mobility constraint is below a threshold value, the shift is immediate; if, to the contrary, its is above the threshold value, the shift never happens.

Note that when $\sigma = 1$ and $T = 3$, the value of ε^* coincides with the one derived for the three-period case. For larger values of σ, the boundary on ε is smaller. Moreover when $\sigma = 1$ and T tends to infinity, ε^* tends to zero so that there is immediate agreement to adopt pro-climate reform. In other words, the preceding proposition shows that voters with a longer time horizon and more patience will tend to adopt a pro-climate reform sooner rather than later.

The generalised game of reform adoption results in a qualitatively similar conclusion: without active interference in the group transition process by the incumbent politician, the reform is either adopted at the very beginning of the game, or never. The attractiveness of the transition depends on the size of the discount factor as well as on the number of periods the reform lasts for. Longer game duration and less future discounting implies higher likelihood that the reform will be voted for, given some mobility frictions ε.

Next, we generalize the one period delayed adoption result in the three period model to allow for the possibility of a delay of $k < T$ periods. As before, we assume that incumbent politician cannot influence the magnitude of ε but within such a constraint, can impact the evolution of group membership over time.

Proposition 1.5. *Along the equilibrium path of play, if* $\mu_0 - k\varepsilon > 0.5 > \mu_0 - (k+1)\varepsilon$, *there exists* $\varepsilon(\bar{k})$ *such that if* $\varepsilon \geq \varepsilon(\bar{k})$, *and* $-\frac{c(1-\sigma)}{1-\sigma^T} + \frac{\beta\sigma^k\left(1-\sigma^{T-k}\right)}{1-\sigma^T} \geq \alpha$, *then* $p^i = a$ *for* $i = \overline{1,k}$ *and* $k < T$, *while* $p^j = b$ *for* $j = \overline{k+1,T}$; *the reform is implemented with a delay of* k *periods.*

Proof. See Appendix. ■

The above proposition demonstrates that in the T-period model, when pro-climate reform is not adopted immediately, we show that identifying and acting on a domestic coalition of voters with a stake in pro-climate reform can lead to delayed adoption after $k < T$ periods.

The delay in adopting pro-climate reform is driven by the limited time horizon of politicians and voters. The only way to compensate early movers is by allowing them to reap the benefit of the economy running in a low-carbon mode for a maximum number of periods. Moreover, the present value of delayed future gains must dominate current losses for early movers.

Hence, the period of delay involved in implementing pro-climate reform is limited by the value of the discount factor, the time horizon T and the individual cost of transition. Evidently, when $\sigma \longrightarrow 0$, the present value of future gains is dominated by current losses for early movers even if the transition period is short.

For a fixed value of T, when $\sigma = 1$, the conditions in the above proposition simplify to $T > c/\alpha$, $k \leq T(1-\alpha) - 1$, and $k < 2c/\delta$. When $\sigma \longrightarrow 1$, and the period of delay involved in implementing pro-climate reform is not too long, the present value of delayed future gains dominates losses for early movers. The length of delay is determined by the individual cost of transition δ with a smaller value of δ being associated with a longer transition period k.

1.6 Discussion

Being a reduced form representation of a complex idea, our model has some limitations.

We assume that all the voters are aware of the preferences of all other voters, the politician is also aware of their preferences, and they are all aware of each other's awareness. This assumption, while certainly convenient for computational purposes, may not be adequately reflecting the reality. A model without a common prior or incomplete information would be a different computational exercise and would have different results. The focus of this paper is, however, not a mathematical difficulty, but political applicability achieved with minimum complications.

Another potentially problematic assumption is that of an international policymaker interfering in internal affairs of a given country. Having good intentions is not sufficient to justify meddling with a sovereign nation's domestic policy. It may be prudent of an international agency to sponsor a local government or to support it in other non-fiscal ways to engineer the transition. An international policymaker can conduct discussions and conversations with embassies, trade officials and foreign ministries to promote its position and explain the importance of unilateral actions like the US did back in 1986, prior to the Montreal Protocol (Haas 1992). An international agency is limited in what actions it can take with regards to domestic policies.

It is also worth acknowledging that often funding for such bodies comes from donations of developed nations, and thus, at least in part, serves their interest. There is a large amount of literature devoted to the allocation of international aid, and there is some evidence that it is not necessarily free from a donor interest bias (Neumayer 2005). Turning back to the Montreal Protocol example, at the beginning, the UK and France were among the staunchest opponents to the US's position because due to the history of supersonic transportation development they believed that the US had concealed interest in concocting an international agreement on ozone protection (Morrisette 1989). The US's true intentions did not matter; the beliefs of its counterparts did. Therefore, as helpful as it may be to have a pivotal negotiation party with an upper hand, it can also be counterproductive when other powerful actors believe that there is impurity of intentions and biases that are not in their favour. Caution and common sense must be exercised when applying our framework to real life situations.

We assume a simplified view on the superiority of political states. Namely, in our model one state is clearly and unambiguously inferior to the other, which is a level of simplification rarely observed in reality. However, it is our belief that where global public goods such as environment are involved, there are universally right and wrong choices. For instance, business-as-usual scenario is the worst choice that governments today can make, and a transition would definitely be a superior option. We do not apply our findings to morally grey areas and situations that do not constitute a clear, scientifically established welfare improvement.

Last but not least is the assumption of a benevolent politician who derives her payoff entirely from social welfare. Often, incumbents are interested in prolonging their rule rather than engaging in a hefty transition that may cost them their position. This pattern has been observed in many nations over the course of human history (Acemoglu and Robinson 2000a), as well as in the field of international policymaking (Battaglini and Harstad 2020). Our modelling choice is rather normative and reflects a 'should' rather than 'is'. Although considering a re-election-oriented incumbent would not affect the results of the present study (our mechanisms aim at reforming public preferences, disregarding the incumbent), it would be interesting to alter our framework to model a case of a selfish incumbent whose political rent does not directly depend on social welfare. This is a task for future research.

The analysis done in this paper suggests that politicians should focus on non-market policies of upstream technologies. Currently, however, most policies are market-based and target a downstream technology such as EVs. The policies include income tax rebates and exemption, direct purchase rebates, financial incentives for charging infrastructure, sales tax waivers and income tax credits. Some have been shown to be more effective than the others, e.g., Gallagher and Muehlegger (2011) estimate that sales tax waivers are better than income tax credits, but all are marginally effective in increasing adoption of electric vehicles (Clinton and Steinberg 2019). The more important question is about their environmental efficiency and welfare implications. The former is defined by multiple characteristics of the vehicle operational environment, such as, for instance, colder climate, the electricity source and charging time (Blumsack et al. 2008, Archsmith et al. 2015). The latter is more of a cost benefit analysis type exercise which involves social cost of carbon, potential damage assessment, spillover effects and long-term benefits. At present, although the policies predominantly succeed in encouraging the uptake of EVs, they are neither environmentally beneficial, with the exception of a handful of the US states (calculated as the difference between the foregone upstream emissions and the downstream electricity generation emissions)[21] (Graff Zivin et al. 2014, Holland et al. 2016, Clinton and Steinberg 2019) nor welfare enhancing (Beresteanu and Li 2011, Michalek et al. 2011, Holtsmark and Skonhoft 2014, Holland et al. 2016, Clinton and Steinberg 2019)[22]. Holland et al. (2016) go as far as to demonstrate that subsidies are not even the second-best solution. In their model, the first-best is a (differentiated) Pigouvian tax, while the second-best is a differentiated subsidy, which is, on average, negative across the country, i.e., it is also a tax.

The environmental efficacy of government incentives will continue to increase as energy generation across the world becomes less fossil fuel intensive. As for the cost efficiency, none of these papers explicitly accounts for spillovers and other long-term benefits of stimulating the sales of electric vehicles under current circumstances, but

[21] The analysis is limited to the emissions of carbon dioxide and mostly ignores other potent green house gases, except for Holland et al. (2016).

[22] The metrics differ from one model to the other. In Holland et al. (2016), emissions from different types of vehicles are mapped into damages or environmental benefits using the social cost of carbon estimated by the EPA and then compared on the geographical basis. In Clinton and Steinberg (2019) and many others, the cost of avoided CO_2 per ton is calculated and compared across incentive programmes whose cost effectiveness is estimated elsewhere.

Holland et al. (2016) and Clinton and Steinberg (2019) do acknowledge this as a limitation and hypothesize that their welfare analyses would likely yield different results if network externalities and increased returns to scale were accounted for. Spillover effects and feedbacks loops can in fact be significant but also uneven on the two sides of the market. For instance, Li et al. (2017) show that subsidising charging stations exhibits greater returns. Before implementing a financial incentive programme, such externalities should be carefully considered.

Non-market policies promoting EVs have also so far not been as effective as is desired. In Bento et al. (2014), they study the welfare effects of a non-market policy implemented in Los Angeles to encourage the uptake of hybrid vehicles that gave access to high occupancy lanes (HOV) to low-emission car owners. The analysis demonstrates that the policy increased implicit congestion costs for carpoolers since congestion externalities were not appropriated by hybrid owners, which resulted in negative welfare effects. Moreover, the policy was regressive as it made carpoolers worse off and all the rents of the programme were likely collected be the purchasers of hybrid cars. The policy which was initially thought of as "free" turned out to be welfare-expensive due to unaccounted, unpriced and unappropriated policy externalities. However, the policy was environmentally beneficial.

In Norway, a policy which was a combination of market and non-market incentive led to a dramatic EV sales growth, and while its environmental benefits are difficult to calculate because of the dependence on the electricity source, in terms of cost, it was more expensive than purchasing a corresponding amount of permits from the EU ETS (Holtsmark and Skonhoft 2014). However, and even more importantly, the indirect effects of the policy included EVs being acquired as an additional car by wealthy households and the change in driving patterns, crowding out public transport, walking and cycling. This evidence supports the case in point – focusing on upstream technologies and on social transition that takes place in people's minds is the first step in implementing a successful long-term green reform.

1.7 Concluding remarks

Since 1896 it has been known that doubling atmospheric CO_2 would raise the Earth's temperature by approximately 5-6 degrees Celsius (Arrhenius and Holden 1897). By 2020, the world has seen two international agreements signed, ratified and entered into force, and no real improvement where carbon dioxide and other greenhouse gases are concerned. The danger is imminent and by now quite well understood, and yet, the transition away from business-as-usual is still a dream rather than a reality. Such transition has a socio-technical nature and involves changes in technology, institutions, behaviour. Hindering the transition are institutional inertia, technological lock-in, habits and social norms, inequality and unfairness of the situation.

In this paper, we take a tentative, initial step towards addressing these issues. In the presence of limited transferability of payoffs and mobility frictions, we explore three mechanisms to engineer techno-institutional transition. Our suggestions are based upon minimal financial resources involvement and political feasibility. We therefore propose that a governing politician eager to catalyse a societal transformation eliminates uncertainty about the winners' identity as a result of the reform by strategically creating coalitions within a society. One way is to identify a group of early adopters or those who are more likely to embrace the transition at an early stage, and target these groups with non-market policies, creating a public entity with vested interested in the new social order. These groups can benefit from the transition immediately, if it happens to take place soon after the policy has been enacted, or with a delay. In the latter case, they would have to first go through a period of inconvenience, until the transition occurs and the benefits are realised. What would make them take immediate loss in exchange for a distant prospect of a potential benefit? Since the incumbent is unable to commit to any future policy, she can use such coalitions as a commitment device to constrain her future choice set. Since the coalitions of newly created policy constituencies have enough power to resist a transitional rollback, their immediate loss is justified and a political shift is achieved.

With only limited resources, such shift can also be engineered in a network of countries where there are knowledge and technology spillovers. By identifying a pivotal member of the network, – i.e., a market with the greatest potential, – an outside agent can initiate a transition that would over time spread worldwide. Similarly, a pivotal member can use a threat of unilateral actions to practically force other nations into compliance. But this is conditional on a pivotal member being benign and having world's best interest in mind.

Motivated by the current emphasis on unilateral commitment by nations in global climate change negotiations, in this paper, we propose two mechanisms to implement a pro-climate reform, and our research is of direct relevance to policymakers who have become signatories to the Paris Accord and are now bound by it to deliver on their net-zero promises. Nationally Determined Contributions are the main mechanism for global emissions reduction (see, for example, the published net zero plans by different countries at the United Nations website). Prominent evaluations of Nationally Determined Contributions focus on how ambitious (the quantum of, and delay in, mitigation efforts) such plans are (e.g. Emissions Gap

Report 2022 by the UN environment program). However, beyond ambition, there remains the issue of whether the governments can actually deliver on their goals. The NDCs that make up the Accords are already far below what is needed to achieve the emissions level defined by the agreement, but even these goals will not be reached without proper policies in place.

Our analysis was conducted with a net-zero transition in mind. A general logic of such a transition is to first replace fossil fuel energy with its renewable counterpart, and then ensure mass electrification of existing energy system. A politician in a modern democracy cannot commit to a policy that spans over 20 years, thus creating powerful winning coalitions with a stake in perpetuation of renewable energy is key for net zero to become reality.

One such example can be found in Denmark. According to some economists (Buen 2006, Singh 2012), one of the reasons behind its wind energy industry success is the policy's ability to create constituencies in favour of wind energy early on that grew to become sufficiently influential to ensure that the course of the policy stayed stable long enough to allow the industry to reach a self-sustainable state. This was done through subsidising individual wind turbine owners and co-operatives that play a major role in Danish energy generation.

Another instance comes from a country with a nascent renewable energy industry, yet well illustrates the idea of deliberate creation of vested interest for future green transition. Boute and Zhikharev (2019) argue that in Russia, government's preference for solar energy over any other renewables, coupled with the incentives granted to the industry, led to the creation of strong entities invested in solar power who now have influence over future energy policies and are thus capable of promoting the agenda of renewable electricity along such vital sectors as oil and gas. Policy's skewness towards solar energy allowed the benefits to be concentrated on the few companies/individuals which led to them becoming more powerful than they could have been having the policy spread over multiple industries.

Extending our model to settings with incomplete information and infinite horizon are topics for future research.

1.A Appendix

1.A.1 Proof of Proposition 1.4

We first demonstrate that the reform will never be reversed once implemented.

Recall that, due to the game structure, the share of voters belonging to group G_1 at any time period, μ_t, can only take two values, $\mu_0 > 0.5$ and $\mu_0 - \varepsilon$, $\forall t$. Denote

$\mu = \mu_0$ and $\mu' = \mu_0 - \varepsilon$. Assume that the reform has been implemented at some period t. This implies that the share of voters belonging to G_2 at the beginning of $t+1$ has changed to $\mu' > 0.5$, and pro-environmental groups now constitute a majority. The reversal to a will take place if expected payoff to the majority from voting for a is larger than that from voting for b. Note that the reversal can potentially take place with a delay.

Assume the reform lasts for k periods from now (including the current period which is not discounted). Consider immediate reversal, no reversal, and a reversal delayed by $\tau < k$ periods, meaning the reform lasts for another τ periods including the current one and then the reversal lasts for $k - \tau$ periods. The discount factor is $\sigma \leq 1$.

$$\mathbf{E}(\text{no reversal}) = \sum_{t=1}^{k} \sigma^{t-1}\beta$$

$$\mathbf{E}(\text{immediate reversal}) = \frac{\varepsilon}{1-\mu'}\left(\sum_{t=1}^{k}\sigma^{t-1}\alpha - c\right) + \left(1 - \frac{\varepsilon}{1-\mu'}\right)\sum_{t=1}^{k}\sigma^{t-1}(\alpha-\delta)$$

$$\mathbf{E}(\text{delayed reversal}) = \sum_{t=1}^{\tau}\sigma^{t-1}\beta + \frac{\varepsilon}{1-\mu'}\left(\sum_{t=\tau+1}^{k}\sigma^{t-1}\alpha - \sigma^{\tau}c\right) + \\ + \left(1 - \frac{\varepsilon}{1-\mu'}\right)\sum_{t=\tau+1}^{k}\sigma^{t-1}(\alpha-\delta)$$

Compare first $\mathbf{E}$(immediate reversal) and $\mathbf{E}$(delayed reversal).

Eliminating similar terms from the first two expectations and using the fact that $\sum_{t=1}^{k}\sigma^{t-1}\beta - \sum_{t=1}^{k}\sigma^{t-1}\alpha > 0$, we are left to compare

$$\frac{\varepsilon}{1-\mu'}(-\sigma^{\tau}c) \gtrless \frac{\varepsilon}{1-\mu'}\left(\sum_{t=1}^{\tau}\sigma^{t-1}\alpha - c\right) - \\ - \sum_{t=1}^{\tau}\sigma^{t-1}\delta - \frac{\varepsilon}{1-\mu'}\sum_{t=1}^{\tau}\sigma^{t-1}\alpha + \frac{\varepsilon}{1-\mu'}\sum_{t=1}^{\tau}\sigma^{t-1}\delta.$$

Rearranging reduces the comparison to

$$0 \gtrless \frac{\varepsilon}{1-\mu'}(-c + \sigma^{\tau}c) - \left(1 - \frac{\varepsilon}{1-\mu'}\right)\sum_{t=1}^{\tau}\sigma^{t-1}\delta.$$

Since $\mu' < 0.5$, we have $\dfrac{\varepsilon}{1-\mu'} < 1$, and both terms it the expression on the right-hand side are negative, resulting in delayed reversal being preferred for any τ.

Similarly, it can be demonstrated that $\mathbf{E}(\text{no reversal}) > \mathbf{E}(\text{delayed reversal})$, and thus the reform will persist, once implemented.

Now, voters will choose to vote for the reform if expected payoff from the transition is larger than from the absence of such. Since the reform is never reversed, the expected payoff will depend on how many periods are left until the last period T. Assume, as before, that the reform is not reversed for $k \leq T$ periods, including the current one.

Also, note that it may be preferable to implement the reform with a delay of τ periods ($1 \leq \tau < k$).

$$\mathbf{E}(\text{immediate reform}) = \frac{\varepsilon}{\mu}\left(\sum_{t=1}^{k}\sigma^{t-1}\beta - c\right) = \frac{\varepsilon}{\mu}\left(\beta\frac{1-\sigma^k}{1-\sigma} - c\right)$$

$$\mathbf{E}(\text{no reform}) = \sum_{t=1}^{k}\sigma^{t-1}\alpha = \alpha\frac{1-\sigma^k}{1-\sigma}$$

$$\mathbf{E}(\text{delayed reform}) = \sum_{t=1}^{\tau}\sigma^{t-1}\alpha + \frac{\varepsilon}{\mu}\left(\sum_{t=\tau+1}^{k}\sigma^{t-1}\beta - \sigma^{\tau}c\right) = $$

$$= \alpha\frac{1-\sigma^{\tau}}{1-\sigma} + \varepsilon\frac{\sigma^{\tau}}{\mu}\left(\beta\frac{1-\sigma^{k-\tau}}{1-\sigma} - c\right)$$

Immediate reform will be preferred to no reform if

$$\varepsilon \geq \frac{\alpha\mu}{\beta - c\dfrac{1-\sigma}{1-\sigma^k}} \equiv \varepsilon_1.$$

On the other hand, delayed reform is preferred to immediate if

$$\varepsilon \geq \frac{\alpha\mu}{\beta - c(1-\sigma)} \equiv \varepsilon_2.$$

It is easy to show that $\varepsilon_1 \geq \varepsilon_2$ with equality reached for $\sigma = 0$. In other words, if the future is fully discounted, then it has no bearing on the choices made today and having the reform delayed is no different that not having it implemented at all. Also note that when k is large enough and $\sigma \in (0,1)$, the two boundary values coincide: $\lim_{k\to\infty}\varepsilon_2 = \varepsilon_1$.

This results in the following conclusion.

- For $\varepsilon < \varepsilon_2$, the reform is never adopted.

- For $\varepsilon > \varepsilon_1$, the reform is always adopted as early as possible.
- For $\varepsilon_2 < \varepsilon < \varepsilon_1$, immediate implementation is preferred to delayed implementation, but no reform is preferred to both. Hence, the reform is not adopted either.

Hence, the proof is complete by setting $\varepsilon^* \equiv \max \left\{ \mu_0 - 0.5, \dfrac{\mu_0 \alpha}{\beta - c\dfrac{1-\sigma}{1-\sigma^T}} \right\}$.

1.A.2 Proof of Proposition 1.5

Consider a delay of $k < T$ periods, i.e., the reform is engineered to take place in period $t = k+1$ with $k\varepsilon$ voters having made prior adjustments. Note that this requires $\mu_0 - k\varepsilon > 0.5 > \mu_0 - (k+1)\varepsilon$. For delayed reform implementation to take place, it has to satisfy incentive constraints of the two types of voters:

1. those who have made costly adjustments prior to the reform, and
2. those who will be voting for the reform after other voters have made costly adjustments.

Starting from the latter group, upon reaching the period $t = k+1$, the share of voters left in G_1 is equal to $\mu - k\varepsilon > 0.5$. They will vote for b if the following holds.

$$\frac{\varepsilon}{\mu_0 - k\varepsilon}\left(\sum_{t=k+1}^{T} \beta\sigma^{t-k-1} - c\right) \geq \sum_{t=k+1}^{T} \alpha\sigma^{t-k-1}$$

$$\iff \frac{\varepsilon}{\mu_0 - k\varepsilon}\left(\frac{\beta(1-\sigma^{T-k})}{1-\sigma} - c\right) \geq \frac{\alpha\left(1-\sigma^{T-k}\right)}{1-\sigma}$$

This requires that per-period transitional capacity ε is larger than a certain boundary value.

$$\varepsilon \geq \frac{\mu_0 \alpha}{\beta - \dfrac{c(1-\sigma)}{1-\sigma^{T-k}} + k\alpha}$$

Note that the boundary value coincides with the one established by the Proposition 1.4, if k was equal to 0.

On the other hand, there are k groups of voters who make ex-ante adjustments, and there are therefore k incentive constraints to be satisfied.

$$\begin{cases} -c+\sum_{t=k+1}^{T}\beta\sigma^{t-1} \geq \sum_{t=1}^{T}\alpha\sigma^{t-1} \\ -c+\sum_{t=k+1}^{T}\beta\sigma^{t-2} \geq \sum_{t=2}^{T}\alpha\sigma^{t-2} \\ \dots \\ -c+\sum_{t=k+1}^{T}\beta\sigma^{t-k} \geq \sum_{t=k}^{T}\alpha\sigma^{t-k} \end{cases}$$

$$\iff \begin{cases} -c+\dfrac{\beta\sigma^{k}\left(1-\sigma^{T-k}\right)}{1-\sigma} \geq \dfrac{\alpha\left(1-\sigma^{T}\right)}{1-\sigma} \\ -c+\dfrac{\beta\sigma^{k-1}\left(1-\sigma^{T-k}\right)}{1-\sigma} \geq \dfrac{\alpha\left(1-\sigma^{T-1}\right)}{1-\sigma} \\ \dots \\ -c+\dfrac{\beta\sigma\left(1-\sigma^{T-k}\right)}{1-\sigma} \geq \dfrac{\alpha\left(1-\sigma^{T-k}\right)}{1-\sigma} \end{cases}$$

Note however that as t changes from 1 to k, the left-hand side in the expressions above increases, while the right-hand side decreases, which implies that it is enough to demand that only the first inequality holds.

Thus, for the reform delayed by k periods to be implemented, we require that the following holds:

$$\begin{cases} \varepsilon \geq \dfrac{\mu_0\alpha}{\beta-\dfrac{c(1-\sigma)}{1-\sigma^{T-k}}+k\alpha} \equiv \varepsilon(\bar{k}) \\ -\dfrac{c(1-\sigma)}{1-\sigma^{T}}+\dfrac{\beta\sigma^{k}\left(1-\sigma^{T-k}\right)}{1-\sigma^{T}} \geq \alpha \end{cases} \tag{1.4}$$

The first inequality sets a new boundary on ε. Recall the boundary calculated in Proposition 1.4, ε^*, which marks reform implementation without an intervention, and notice that the new boundary must be lower, since for delayed transition we have $\varepsilon < \varepsilon^*$.

$$\frac{\mu_0\alpha}{\beta-\dfrac{c(1-\sigma)}{1-\sigma^{T}}} \geq \frac{\mu_0\alpha}{\beta-\dfrac{c(1-\sigma)}{1-\sigma^{T-k}}+k\alpha}$$
$$\iff \frac{c(1-\sigma)}{1-\sigma^{T}}+k\alpha \geq \frac{c(1-\sigma)}{1-\sigma^{T-k}}$$

Studying the expression for corner values of σ and $k \leq T-1$, it is easy to see that the inequality is always satisfied for $\sigma \longrightarrow 0$. For $\sigma \longrightarrow 1$, we are limited to $T > c/\alpha$, which only matters if $c > \alpha$.[23]

[23] If, as before, $\alpha = 0.2$ and $c = 0.8$, the game should have more than 4 periods for the condition to be satisfied.

In other words, for large enough T and σ close to 1, the new boundary on mobility frictions is lower, which means that at $t = k+1$, voters in G_1 will decide to implement b, having $k\varepsilon$ people transitioned in advance.

The second inequality in equation 1.4 simply requires σ to be large enough, depending on the delay, game length and other parameters. Thus, the condition is never satisfied for $\sigma \longrightarrow 0$, but for $\sigma \longrightarrow 1$, only a reform with a delay $k \leq T(1-\alpha) - 1$ can be engineered.

Therefore, for both incentive constraints to hold, provided that σ is close to 1, we require that the game is long enough, $T > c/\alpha$, and the delay is not too much compared to the total number of periods, $k \leq T(1-\alpha)-1$. In other words, provided that the game lasts for T periods, the delay cannot be too long, since it must leave a certain number of periods to compensate those who have transitioned early on.

What is left to check is the optimality of such delay. Compare two social welfare levels, one without a reform, SW_1, and one with a delay of k periods, SW_2.

$$SW_1 = \sum_{t=1}^{T} \sigma^{t-1}(\alpha - \delta + \delta\mu_0) = (\alpha - \delta + \delta\mu_0)\frac{1-\sigma^T}{1-\sigma}$$

$$\begin{aligned} SW_2 &= \sum_{t=1}^{k} \sigma^{t-1}\left[\alpha - \delta(1-\mu_0 + t\varepsilon)\right] - \sum_{t=1}^{k+1} \sigma^{t-1}\varepsilon c + \sum_{t=k+1}^{T} \sigma^{t-1}\beta\left[(k+1)\varepsilon + (1-\mu_0)\right] \\ &= (\alpha - \delta(1-\mu_0))\frac{1-\sigma^k}{1-\sigma} - \delta\varepsilon\sum_{t=1}^{k}\sigma^{t-1}t - c\varepsilon\frac{1-\sigma^{k+1}}{1-\sigma} + \\ &+ \sigma^k\beta\left[(k+1)\varepsilon + (1-\mu_0)\right]\frac{1-\sigma^{T-k}}{1-\sigma} \end{aligned}$$

The difference between the two, $SW_2 - SW_1$, should be non-negative.

$$\delta\varepsilon\sum_{t=1}^{k}\sigma^{t-1}t - c\varepsilon\frac{1-\sigma^{k+1}}{1-\sigma} + \sigma^k\frac{1-\sigma^{T-k}}{1-\sigma}(\beta\left[(k+1)\varepsilon + (1-\mu_0)\right] - (\alpha - \delta(1-\mu_0)) \geq 0$$

When $\sigma \longrightarrow 0$, the present value of gains is always dominated by current losses so that $\lim_{\sigma\longrightarrow 1}(SW_2 - SW_1) = \varepsilon(\delta - c)$. When $\sigma \longrightarrow 1$, and $k\delta < 2c$, we have that:

$$\begin{aligned} &\lim_{\sigma\longrightarrow 1}(SW_2 - SW_1) = \\ &\delta\varepsilon\frac{k(1+k)}{2} - c\varepsilon(k+1) + (T-k)\cdot(\beta\left[(k+1)\varepsilon + (1-\mu_0)\right] - (\alpha - \delta(1-\mu_0)) \geq 0 \end{aligned}$$

Chapter 2

Technology in international environmental agreements

2.1 Introduction

Climate change is recognised as *the* threat to the stability of the economic system and world order by the economists (Nordhaus 2019). Since climate is a public good, human and industrial activities affecting it typically extend beyond market dynamics and are not reflected in market prices. Climate change is also a global externality which transcends the control of national governments and markets, affecting individuals worldwide in a way that is not proportional to their contributions. However, the global community has had limited success in reversing the changes wrought by global warming. Recent times have seen the establishment of different international agreements to address global issues, but no significant progress has been observed on the front of global greenhouse gas emissions reduction.

The provision of good governance at both individual and collective national levels plays a pivotal role in addressing global externalities (Zhou et al. 2020). However, Panhans et al. (2017) and Calel and Dechezleprêtre (2016) point out that existing international environmental laws do not provide an adequate legal framework to compel nations disinterested in pollution reduction to participate in addressing global externalities. While Zhou et al. (2020) emphasize the importance of effective governance in environmental management, Panhans et al. (2017) and Calel and Dechezleprêtre (2016) argue that current international legislation falls short in ensuring the enforcement of climate change policies among individual nations. They advocate for international institutions to focus on formulating and implementing mandatory and environmentally friendly legislation and policies applicable to all countries. However, in the current context, persuasion remains the primary option to foster cooperation among nations worldwide and control greenhouse gas emissions and other activities contributing to global pollution.

Under the premise of weak global governance over pollution, the agreement between the nations should be self-sustainable (Barrett 1994). It is a game between self-interested independent players whose decisions are influenced by their individual economic and political constraints. In this context, climate treaties play a secondary role in mitigating global emissions, as the studies have long established that achieving substantial provisions in such agreements is challenging because all treaties are susceptible to deviations, unless some modifications are made to the models, such as incorporating technological improvements that can spill over from one nation to another through infrastructure purchase or technology investment. In a sustainable equilibrium scenario, a few nations may innovate and develop new emission-mitigating technology. Others will emulate these innovative nations, seek access to the technology, and ultimately contribute to global emission reduction. The adoption of innovative technology by other countries relies on technology transfer within and across nations and a careful evaluation of the consequences of ignoring mitigation measures (Roth et al. 2022). Nations must collaborate to investigate and develop innovative strategies for reducing carbon emissions and implement policies that support the sharing of information about such technology to create a pollution-free global environment.

According to Li and Wu (2023), the development of new technologies for lower carbon emissions has become a priority to limit global temperature increases and prevent adverse consequences of carbon emissions on the climate system. However, experts and researchers have struggled to provide substantial guidance to policymakers in the energy and climate fields due to limited knowledge of sources of diffusion and innovation (Roth et al. 2022). Although existing academic literature on technological change suggests that technical progress cannot be predicted independently because it is not exogenous, recent years have seen a growing interest among energy researchers in introducing endogenous technological changes into energy models (Lee et al. 2022; Aoyama and Silva 2022). Endogenous technological changes, influenced by public policy and prevailing energy market conditions, can lead to improved efficiency and cost reductions over time (Lindman and Söderholm 2012; Li and Wu 2023). National governments should intensify investment in endogenous technological development in energy models to combat climate change effectively. Bottom-up models have introduced technology change through technology learning rates, which quantify the relationship between technology costs and cumulative experience with the technology (Jiao et al. 2018). Technology learning rates play a crucial role in reducing the costs of new low-carbon technologies, making them more attractive and efficient than existing options. High learning rates are particularly important for new low-carbon technologies, as they incentivize upfront investments in technology to realize economic returns from learning (Wang et al. 2021). National governments should collaborate in allocating a specific percentage of their budgets to research and development funds to seek new technological solutions for carbon emission reduction within the global environment.

This chapter delves into the extensive body of work relevant to abatement technology spillovers and aims to uncover gaps in the current literature pertaining to the role of technology in climate change. By pinpointing areas that previous articles have not addressed properly, the chapter is meant to contribute to bridging these gaps in the current body of work. Its conclusion rests on a typical for the literature premise that a number of nations both benefit from their pollution-related activities and bear the future costs proportional to overall environmental pollution. The externalities caused by individual nation's pollution are of course not internalized by the global market. These countries possess a crucial opportunity to invest in green technology to achieve economic goals while minimizing environmental pollution. Assuming that the emission strategies adopted by one group of nations significantly influence the choices of subsequent groups, the scenario can produce multiple equilibria. For instance, it may lead to a situation where no country minimizes its greenhouse gas emissions, or every nation does so. A central premise of this model is that a significant group of nations must reduce emissions early to motivate others to follow suit. Achieving this relies heavily on the adoption of advanced technology for environmental pollution mitigation. The model, therefore, hinges on technology spillover to drive global emissions reduction.

While seemingly simplistic, the proposed model is not far from reality. A pivotal factor behind the banning of ozone-depleting chemicals was the availability of substitutes to mitigate environmental pollution. There were various technological options that could promote emissions reduction on a national and global scale. In light of the absence of a similar easy fix for global warming and nascency of technological substitutes, the Kyoto Protocol fell short of global expectations regarding emissions reduction. The Paris Agreement is the latest institution addressing climate change, but there remain questions and predictions to test regarding its effectiveness.

The proposed model encourages exploration of issues related to technology spillover. The crucial starting point is formulating relevant research questions. The model posits that the adoption of specific green technologies by a few countries can make it easier for others to follow suit. To quantify this relationship, it's vital to consider how much technology country j needs to adopt at period t to facilitate country i's adoption in period $t+1$. Research suggests positive technology spillovers and rates of learning (Langniß and Neij 2004), but this should be tested in the specific settings of the present study.

2.2 Green technology in the literature

It has been long understood that a decrease of emissions by one country can lead to an increase of such by another country. This, however, is true if low carbon technology is regarded as exogenous. Making it instead endogenous and allowing for spillovers change this conclusion in a not so obvious way and depends on the model used by a particular study. It is possible to imagine a scenario where a country with a low environmental harm undertakes an investment in a low-carbon technology merely because other countries have done so, lowering everybody's abatement cost. In fact, this is exactly the question we are interested in here.

There are multiple ways to approach the problem in focus, but in their core, they boil down to only a few genuinely different specifications.[1] The first is a one-period stage game of coalition formation, written in terms of abatement. There is no accumulation of greenhouse gases and it is a relatively static approach where technology is often implied rather than introduced explicitly. The second is to use emissions instead of abatement and have a dynamic game of treaty adoption with several periods and possibly infinite time horizon. Introducing technology to the second set of models is mostly done by adding a stock variable that represents investment. Neither of the approaches is satisfactory to answer the research question of the present study. The former approach cannot accommodate the delayed nature of the benefits of a low-carbon transition and the immediate cost of such, while the second cannot adequately discuss technological trade.[2] Some mixture of the two is therefore needed to move past the limitations of the previous studies.

The former thread of literature takes roots in the works of Hoel (1992), Carraro and Siniscalco (1993), and Barrett (1994). Individual levels of abatement add up to form a global level of abatement which then enters each country's benefit function, so that the individual payoff is the difference between the benefit of global abatement and the cost of individual abatement. The stage game of coalition formation is then solved where countries decide whether to join the abating coalition before they take action. Typically the game is assumed to be a prisoner's dilemma type game with a unique inefficient Nash equilibrium. Signatories are assumed to comply with the treaty conditional on having made the decision to abate. The game is solved backwards, maximizing either individual (for non-signatories) or aggregate payoff (for signatories), resulting in equilibria with side payments.

[1] A significant amount of models with spillovers come from the theory of economic growth. There are many such models discussed in Klenow and Rodriguez-Clare (2005). For the approach used in the present study, these are not particularly relevant.

[2] Not diminishing the merit of all the previous work done on technological trade and spillovers, the author of the study has a very specific problem in mind which necessitates a merge of the two approaches.

This approach of sharing the aggregate payoff maybe be problematic when it comes to making policy recommendations, so the latter strand of research may be of more relevance to the real world.

It originates from Dockner et al. (1996) and is further developed by Dutta and Radner (2009) and Chander (2017).

Dockner et al. (1996) analyse 2 countries whose emissions are a byproduct of domestic production activity, $e_t^i = y_t^i$, in discrete infinite time. Pollution is treated as a stock, $p_{t+1} = y_t^1 + y_t^2 + (1-\beta)p_t$, and it is assumed that the ecosystem will collapse if a certain pollution threshold $\bar{p}$ is reached. Each country maximizes an infinite sum of per-period utility functions that are a difference between production and the cost of pollution. As a benchmark, they find two different Markov perfect equilibria, one with convex and monotone strategies, the other with a chaotic dynamics. Neither can be shown to be superior, but the second one can be used to construct a trigger strategy equilibrium that would support an efficient cooperative solution.

In Dutta and Radner (2009), they extend the analysis to I countries and replace the linearity of the per-period utility with concave benefits and linear costs of pollution. Their benchmarks are a global Pareto optimum which is obtained by maximizing the weighted sum of country payoffs, and a symmetric stationary Markov perfect equilibrium which is equivalent to business-as-usual, with constant emission levels over time. They then proceed to find a trigger strategy equilibrium that maximizes a weighted sum of country utilities when the emission policies are incentive-compatible under the threat of reversion to business-as-usual. This equilibrium is strictly welfare-improving over the benchmark business-as-usual solution. They also characterize a set of subgame perfect equilibria and consider some asymmetry between the countries on a reduced model of two countries.

Chander (2017) extends the approach to introduce nonlinearity and asymmetry in damages and nonlinearity in benefits. He proves the existence of a unique subgame perfect equilibrium which represents a business-as-usual scenario and shows that asymmetry can lead to a trigger strategy equilibrium not be an improvement over business-as-usual. He then finds a cooperative solution that can improve upon the benchmark but only if there are transfers between the countries.

The approach is further developed by Harstad (2012), Harstad (2016), Battaglini and Harstad (2016). They explicitly add technological investment to the model. Since the analysis is based on n countries, this leads to having them having $n+1$ stock variables (one for pollution and n for the investment stocks). In the solution they however usually reduce it to two by establishing the payoff-irrelevance of individual investment levels and only relying on the total investment. For them, it is also a stage game within a period where countries decide how much to invest, how much to pollute and whether to sign an agreement, but the horizon is infinite. Battaglini and Harstad (2016) generalises their other papers by assuming that technology matures in some arbitrary period of time (not necessarily one period), and the time between the repeated stages of the game is also arbitrary. The solution is always a Markov perfect equilibrium for its simplicity and tractability.

A note of caution before we proceed with the model. Spillovers are often understood differently than what I mean. For me, it is the option of *not* investing in green technology and benefiting from someone else's investment, mostly through trade, but others have used weak intellectual property rights as a spillover channel. This way, emissions can be reduced globally even if just one nation invests.

Spillovers can also be thought of as learning, which comes in various forms (see, for instance, Lindman and Söderholm (2012) for a review of learning models in wind power). It can take place at various stages of technological development, such as R&D (Sagar and Van der Zwaan 2006; Ulph and Ulph 2007) and the diffusion stage (Guo and Fan 2017). It can also refer to the unknown cost of developing the technology (Tarui and Polasky 2005) or uncertainty of environmental damages (Kolstad and Ulph 2011).

A more relevant way to introduce spillovers is used in Harstad (2016). For him, it is the effect someone else's investment has on $i's$ utility function (the time subscript has been suppressed because there is no difference in functions between the periods, bar stock variables such as investment, R_i, whose law of motion would include a previous period value, which is then denoted using minus, $R_{i,-}$).

$$u_i = B_i\left(y_i\right) - C(G) - kr_i + e \sum_{j \in N \setminus i} r_j, \text{ where}$$

- $B_i\left(y_i\right) = -\frac{b}{2}\left(\bar{y}_i - y_i\right)^2$ is the benefit of consumption of energy, $y_i = g_i + R_i$, concave and increasing up to $\bar{y}_i$, where g_i comes from polluting sources such as fossil fuels, and R_i is clean, renewable energy;
- $C(G) = \frac{c}{2}G^2$ is environmental damage associated with the stock of greenhouse gas stocks $G = q_G G + \sum_{i \in N} g_i + \theta$, with θ_t being a time-varying shock, Independent and identically distributed with mean 0 and variance σ^2;
- k is the private cost of investment, r_i is country i's investment in the current period; note that the technology stock R_i evolves as a result of each period's investment and natural depreciation of capital at rate q_R, so that $R_i = q_R R_{i,-} + r_i$;
- e captures possible externalities, or *spillovers*, as a result of other nations' investment;

- total utility is measured over time with continuation value $U_{i,t} = \sum_{\tau=t}^{\infty} \delta^{\tau-t} u_{i,\tau}$

The way externalities are introduced in the model allows to interpret it as "traditional technological spillovers, diffusion, imitation, licensing, or trade" (Harstad 2016, p.725). He thus discusses how to extend the model to incorporate imitation, licensing, and tariffs. Without changing the model, he describes what e would be in each of the cases. Subsequently solving it with the newly defined parameters, he concludes that intellectual property rights (IPRs) and a treaty on emissions/investments are strategic substitutes. If a satisfactory treaty cannot be signed, then IPRs should be strengthened, technological trade between the nations – facilitated, and the technology itself – subsidised.

In a similar fashion, technological externalities are introduced in Harstad et al. (2019). It is a repeated extensive-form game with infinite horizon, with investment and pollution in each stage within a time period.

$$u_i = b_i\left(g_i, r_i\right) - h_i c\left(r_i\right) \sum_{j \in N} g_j - k_i r_i, \text{ where}$$

- $g_i \in \underline{g}, \bar{g}$ denotes emissions;
- benefit (from consuming energy, from whichever source) is $b_i\left(g_i, r_i\right)$, increasing and concave in technology r_i;
- country-specific harm $h_i c\left(r_i\right) \Sigma g_j$, with $c(r_i)$ decreasing and convex in r_i;
- $k_i \equiv \delta^e \hat{k}_i$ - present value of marginal investment cost

To introduce spillovers, they substitute r_i in the per-period utility with a variable $z_i \equiv (1-e)r_i + \dfrac{e}{n-1}\sum_{j \neq i} r_j$, where $e \in (0,1)$ is the benefit nation i gets from other nations' investment at t. The utility then becomes

$$u_i = b_i\left(g_i, z_i\right) - h_i c\left(z_i\right) \sum_{j \in N} g_j - k_i r_i$$

This specification can be best thought of as the strength of IPRs and not as an open, possibly paid, exchange between the countries. The results suggest that with homogeneous countries, large spillovers discourage investments, so strengthening IPRs should be prioritised, while with heterogeneous countries, larger spillovers are preferred, to encourage more reluctant countries to decrease their emissions.

Being a stock variable in the studies described above, technological trade is difficult to model. For this, building a model around abatement technology can be more sensible. The papers in this strand of research are commonly a static stage game, where countries decide on their abatement technology and on whether to be part of a coalition that would go on to sign an international treaty.

The paper that is closest to my research question is by Barrett (2006). It is written in a standard fashion of self-enforcing international environmental agreements where the decision variable is abatement, and the stage game defines how many nations would join an agreement.

Denoting, as is standard, global abatement as a sum of individual abatement decisions, $Q = \sum_{i=1}^{n} q_i$, and individual payoff as $\pi_i = bQ - \frac{cq_i^2}{2}$, the basic Nash equilibrium of the game predicts $q_i = \frac{b}{c}$, while a full cooperation outcome as $q_i = \frac{bN^2}{c}$ (if full commitment was possible). The result suggests that in equilibrium, only three countries would sign a treaty, which will not make much of a difference to the state of the climate. To remedy that, Barrett proposes two alternative treaties: one that focuses on cooperative R& D of a breakthrough technology, the other – on its adoption.

To model the first treaty, he introduces a technology Y and lets i's payoff be dependent on the new technology in the following manner:

$$\pi_i = b_y \left(y_i + \sum_{j \neq i}^{N} y_j \right) - c_y y_i + b \left((1 - y_i) q_i + \sum_{j \neq i}^{N} (1 - y_i) q_j \right) - \frac{c_0 (1 - y_i) q_i^2}{2}$$

The technology is modelled as binary, $y_i \in \{0, 1\}$, with b_y being the benefit from adopting Y, c_0 and c_y – the costs of old and new technology respectively. Note that, as formulated, the new technology is a substitute for existing technology, which implies that it is also a substitute to pollution abatement, which is criticised and remedied in Urpelainen (2014).

Assuming the game is a typical prisoner's dilemma, so that all countries are better off collectively adopting Y, but each country is individually better off not, the question is, whether a treaty could promote the adoption of the new technology. Note that by "adoption" here he means i's contribution to global R& D efforts, treating Y as a global public good.

Applying the concepts of collective and individual rationality, we can write i's expected payoff from contributing m_i to collecting R&D effort, if k_y^* countries decide to adopt Y:

$$\pi_i = y \left[\frac{k_y^*}{N} (N b_y - c_y) - \frac{b^2}{2c_0} (2N - 1) \right] - m_i$$

The result suggests that Y is a threshold public good, meaning only if enough countries invest, will it be globally adopted,

$$\begin{cases} Y = 1 \text{ if } \sum_{i=1}^{N} m_i \geqslant \bar{M}_y \\ Y = 0 \text{ otherwise} \end{cases}$$

If global contribution exceeds $\bar{M}_y$, then the treaty is needed to *coordinate* contributions, not to enforce them.

Unfortunately, the number of signatories to a technological agreement can only be large if many countries move first and adopt Y, driving further adoption. But it is precisely under these circumstances that the signatories will not be substantially better off, and the treaty will not make much of a difference.

To bypass these limitations, he then imagines technology X with increasing returns to adoption, which is again binary, $x_i, x_j \in \{0, 1\}$.. Expected payoff of nation i is then

$$\pi_i = b_x \left(x_i + \sum_{j \neq i}^{N} x_j \right) - \frac{c_x}{N} \left(N - \sum_{j \neq i}^{N} x_j \right) x_i + b \left((1 - x_i)\, q_i + \sum_{j \neq i}^{N} (1 - x_j)\, q_j \right) - \\ - \frac{c_0 (1 - x_i)\, q_i^2}{2}$$

Assume again that X will not be adopted spontaneously, but that all countries would adopt if enough others do. Denote this critical mass as z – the agreement will only enter into force if z countries ratify it. Such a treaty may be more effective than the one above, but it is crucial that the new technology exhibits increasing returns to scale, which might be true in case of hydrogen fuel vehicles, but not so in electricity generation. The results of the study are thus not promising in providing a ready solution to climate change.

Capitalizing on Barrett's framework, with a few tweaks, Hoel and De Zeeuw (2010) arrive at a different conclusion. Instead of setting a fixed adoption cost and allowing for adoption only if a certain investment level is reached, they make the cost endogenous by making it a function of the R&D investment. The more research is done prior to adoption, the cheaper it is to implement the new technology. As a result, there emerges a non-cooperative equilibrium with full adoption and sufficiently high R&D levels. A treaty that coordinates research contributions can improve upon the outcome by lowering the adoption cost through higher R&D investment.

Another recent paper that modifies the standard framework above and also arrives at a more promising conclusion is by Zavaleta (2016). Instead of allowing for one coalition of adopting countries, the author permits two different consortiums – the countries that invest and own the rights for the technology, and the countries that decide whether to consume the technology developed by the first group. When the consuming coalition is large, the gains from developing the technology are the most significant, so unlike in Barrett (2006), large gains from cooperation are an advantage rather than a hindrance. Creating an international market is key for almost full adoption of the new technology. Who owns the bargaining power is important, but whether it rests solely with the producing consortium or is shared between the two, the outcome is still welfare-improving. This conclusion however relies on assuming transferability of payoffs between the members of the producing coalition, which is a reasonable theoretic assumption that may not stand up to scrutiny in the real world, where nations are very reluctant to share, as is evident from the failure of the Kyoto protocol. From this point of view, the results in Hoel and De Zeeuw (2010) bear more importance for international negotiations as they rely purely on non-cooperative behaviour. This is closer to the reality of the Paris agreement that leans only on individually determined contributions without reliance on international cooperation.

Important thing to note about the abatement-based models above is that technology is treated as a public good, with possibly private gains, but public nonetheless in a sense that developing it requires international cooperation. It is thus assumed that no country has an incentive to invest in technology unilaterally; only collective investment is possible. In emissions-based models, private investment is sensible but the technology is a stock, so technological trade is difficult to impossible to model. But what if some nations do find investment attractive and the model is formulated in such a way that allows for spillovers in ways other than weak intellectual property, licensing and tariffs? What role an international agreement would take then, if any?

2.2.1 Issues and alternative narratives

As is evident from the literature reviewed earlier, the approach based on abatement relies on cooperative game theory and studies the maximum size of a stable coalition of abating countries. Transferability of payoffs is therefore essential for a coalition of any size to form. The results of the basic models are however disheartening – the stable coalition is small, especially when the gains of cooperation are high. The approach based on emissions yields a more promising conclusion of almost any desired equilibrium being sustainable under the threat of reversion to business-as-usual. The dynamic game is usually Markovian with symmetric countries, so that the basic equilibria are symmetric and stationary and only rely on the total, not individual, stock of technology. Adding technology to the first group of models changes the predictions depending on how the technology is modelled, but mostly the outcome is improved with a stable large coalition and almost full adoption. The technology is however introduced only as a public good, and the transferability of payoffs is still implied, save for Hoel and De Zeeuw (2010) where a purely non-cooperative outcome is also shown to be satisfactory.

I see several problems with both approaches. A dynamic game with infinite horizon is not consistent with the idea of a runaway climate change. Researchers recently identified nine planetary boundaries that delimit safe operating space for humanity (Steffen et al. 2015), and by now six of them are broken (Richardson et al. 2023). A related concept of tipping points also suggests that we are approaching, or perhaps have already crossed, some of the critical thresholds that render our environment habitable and life-sustaining as we know it (Lenton 2011). Infinite-horizon dynamic games assume the cost of global pollution simply keeps mounting as time goes on, without any discontinuity. Dockner et al. (1996) are the only ones whose model includes a tipping point, but there is no technology in it. The theoretical model should thus be similar to the experimental studies by Barrett and Dannenberg (2012), Barrett and Dannenberg (2014), Dannenberg et al. (2015).

Due to the uncertainty around the impacts of climate change, it is not clear whether the runaway climate change is a possibility, or the affects will be steadily and continuously growing over the years. I believe the tipping points exist and crossing them would mean some serious irreversible damages. This in turn implies that the damages are not evenly distributed throughout time, and the bulk of it is delayed until a further point in the future. There is thus a clear difference between environmental damages that are closer in time and those that are far, which cannot be captured by using the same function for both with the only difference resulting from the global pollution stock. Therefore, instead of modelling it like Dockner et al. (1996) did, one could assume a finite number of periods with full damage realised only at the end of the game. To be sure, the game must end at some point, to be descriptive of the idea of a tipping point. Making one period represent a decade would decrease the number of periods further, and since most people are unlikely to plan that far ahead in the future, three period seems sufficient to illustrate the basic idea. More periods than that would not add value to the model, while considering a large T would make it similar to infinite horizon models.

Another questionable assumption is that of a trigger equilibrium which sustains a welfare improving treaty (as in Dutta and Radner (2009)). A reversion to business-as-usual is a reasonable idea in theory, and if it is in the set of subgame perfect equilibria, then it is a credible threat, but in practice, reversion to high emissions may not be feasible after the transition to green technology has taken place. Capital cannot mature in a day or be thrown away, so if nation i has invested in lowering emissions of its energy sector by developing solar and wind energy and decommissioning old coal reactors, it would not be able to punish nation j for deviation by instantly destroying the wind turbines and solar panels and building the coal plants in a day. What would likely happen if one nation were to deviate *after* another nation has already committed to a new course of action is nothing. Sustaining a treaty by a threat of punishment which is not based on trade and similar international relationships does not look like a practical idea that can be implemented in real life.

It could be possible, in theory, to modify the approach in Harstad (2016), by adding the tipping point as a constraint, but then the model would likely not be tractable, as it is already extremely complicated. Moreover, as discussed above, technological trade is hard to incorporate into similar models. For this, technology should be a binary variable akin to Barrett's breakthrough technology 2006. Then again, abatement-based models cannot accurately represent the delay in damages. The other problem with these models (I am only referring to Barrett (2006), Hoel and De Zeeuw (2010), Zavaleta (2016)) is that the breakthrough technology is treated is a public good, albeit with private benefits in some cases. In the world of sophisticated sovereign states it is hard to imagine that a technology would necessarily need cooperation from several countries to be developed. Large technologically advanced nations such the US and China are likely capable of developing whatever technology they may need, unless the technology is ahead of its time and cannot be developed regardless of how many countries chip in.

Yet another issue is reliance on transferability of payoffs. Years of international climate negotiations have demonstrated that countries are unlikely to share, be it a burden or a benefit, even if it is "fair" to do so because they bear a historical responsibility for emissions or it is unethical for them not to. The Paris agreement seems to have grasped that idea and instead of trying to impose external limits on national CO_2 production it allows countries to determine how much they want to pollute. The solution to climate crisis thus lies in the realm of unilateral choices and non-cooperative game theory. The most important question is thus not how to incentivize the countries to join an abatement coalition, but under what conditions a non-cooperative outcome can be improved upon.

In other words, without technology incorporated in the model, it could reasonably be hypothesized that countries put their national interests above the global public good, and unless their domestically chosen emission levels are socially efficient, there is no international agreement that is global welfare-optimizing (although any agreement could be marginally welfare improving). Indeed, in the simple model suggested below, using the argument of the core of a system with negative externalities (Shapley and Shubik 1969), such an international agreement does not exist. From the emptiness of the core, I conclude that there is no agreement which is immune to individual and group deviations, and the grand coalition is not stable. This result is in line with the predictions in the literature on self-enforcing international environmental agreements.

On the other hand, once technology is introduced, even in the most primitive way, there are multiple equilibria possible, including the grand coalition. Technology in the model can either be developed or acquired for a set price from another country. Under certain assumptions, some countries would invest in green capital, and some would be willing to buy access to it for a low enough price, so that cumulative emissions reduction can eventually be observed. This however raises a question of equilibrium selection, which should be the main role of an international agreement: not to alter the choice set (as this is an impossible task), but to provide the means to choose the best equilibrium once it becomes available.

2.3 The model

The main concern of the present paper is the role technology transfer in its purest form may play in national abatement decisions and its potential for global emissions reduction. To answer this question, I propose a simple but novel model with technology incorporated differently than in most papers. The added benefit is its simplicity – it only requires three periods to demonstrate the point. But first, to have a basic reference point, I will consider a variation of the model where technology is assumed rather than explicitly introduced.

There are three time periods, $t = 1, 2, 3$ and $N = \{1, ..., n\}$ nations. At $t = 1, 2$, $g_i^t \in \{0, 1\}$ denotes the emissions of greenhouse gases by nation i, so that at any t, $g_i^t = 0$ corresponds to emissions cut, while $g_i^t = 1$ refers to a situation where nation i persists with high carbon activities. Each country has a mass $m_i \geq 0$ and the total mass across countries is $M = \sum_i m_i$. The mass is introduced to differentiate the nations by size since emissions are binary.

Let $G^t = \sum_i m_i g_i^t$ be the total emissions at time t, and $G = G^1 + G^2$ – total emissions across the two time periods.

Define $M_{-i} = \sum_{j \neq i} m_j$.

Assume that at each $t = 1, 2$, there is a benefit $\alpha_i g_i^t$ from high emission activities. One can think of $alpha_i$ as a nation-specific parameter reflecting the value of carbon-intensive economy. To borrow the idea from Harstad (2016), it can denote the benefits of producing energy from non-renewable sources, for example. Higher α_i would then mean that a country is deeply invested in fossil fuels and would find it hard divesting from them.

The cost of emissions is borne at $t = 3$ only and each nation i bears a cost that is proportional to the total emissions across the two time periods $c_i G$.

A typical assumption to make as well is that of low emissions being socially efficient but not individually preferred choice. This is achieved by imposing the following restrictions on the parameters: $\sum_i \alpha_i - M \sum_i c_i < 0$, so that it is socially efficient to set $g_i^t = 0$ for any i, t, and therefore, necessarily, at a core allocation, $g_i^t = 0$ for any i, t; and $\alpha_i - m_i c_i > 0, i = 1, ..., n$ so that if emission levels were chosen simultaneously and non-cooperatively in each t, it is a dominant action for each i to choose $g_i^t = 1$ for all i, t.

No discounting is assumed for simplicity and because the results would remain unchanged if exponential discounting was introduced.

Each nation realises its payoff once per game at $t = 3$. The payoff to i is therefore equal to

$$\pi_i = \alpha_i g_i^1 + \alpha_i g_i^2 - c_i G = \alpha_i (g_i^1 + g_i^2) - c_i (G^1 + G^2) = \alpha_i (g_i^1 + g_i^2) - \\ - c_i \left(\sum_j m_j g_j^1 + \sum_j m_j g_j^2 \right).$$

A nation that chooses not to pollute receives a purely negative payoff at the end of the game equal to

$$-c_i \left(\sum_{j \neq i} m_j g_j^1 + \sum_{j \neq i} m_j g_j^2 \right),$$

while a nation that pollutes throughout the whole game receives

$$2(\alpha_i - c_i m_i) - c_i \left(\sum_{j \neq i} m_j g_j^1 + \sum_{j \neq i} m_j g_j^2 \right).$$

Note that at $t = 1, 2$, a polluting nations enjoys a payoff of α_i, while a non-polluting one gets nothing.

As is standard, the burden of global emissions is shared by all, so the concerns over free riding are valid, because in an efficient scenario where $g_i^t = 0$ for all i, t, a unilateral deviation by nation i can capture the benefits from choosing $g_i^t = 1$ at any t without bearing the social cost of doing so. This might mean an agreement with no explicit punishment, but not imperfect monitoring. This is a game with perfect information, where nations are capable of recognizing a deviation.

This setup is typical of the literature on global pollution abatement, and yet it is not similar to any one paper. Without technology explicitly added to the model, emissions reduction simply takes place, or it does not. Can a multilateral agreement to cut emissions be stable – immune to individual and coalitional deviations? To answer this question, we resort to the argument of the core and show that it is empty.[3]

Any allocation from the core has to satisfy individually and group rationality. The only allocation which satisfies the latter by way of all-player set is if everyone plays $g_i^t = 0$. Thus, if the core is non-empty, it contains only this set of strategies, and everyone receives the payoff of 0. To show that the core is empty, it thus suffices to show that each nation has a unilateral incentive to deviate from an efficient action profile at some t.

When externalities are present, it is necessary to define what nation j does following a unilateral deviation of nation i. There are two scenarios to consider.

- Scenario 1. All other nations do nothing, i.e., the do not change their own action in response to a unilateral deviation by nation i;

- Scenario 2. All other nations choose a best response to any action chosen by nation i, including a unilateral deviation.

Under the first scenario, the core is empty. Suppose nation i contemplates a unilateral deviation at $t = 1$ to $g_i^1 = 1$. As $g_j^t = 0$, $j \neq i$, and the nations do not change their behaviour in response to i's deviation, for each t and $\alpha_i - m_i c_i > 0$, nation i will always deviate.

Under scenario 2, however, the core is also empty. To see this note that whether or not a nation i contemplates a unilateral deviation, at each t, $j \neq i$ will best-respond by choosing $g_j^t = 1, t = 1, 2$, assuming the deviation is detected immediately and the best response takes place in the same period. Therefore, by choosing $g_i^t = 0$ for each t, nation i gets a payoff of

$$-c_i \left(\sum_{j \neq i} m_j g_j^1 + \sum_{j \neq i} m_j g_j^2 \right) = -c_i \cdot 2 \sum_{j \neq i} m_j,$$

while a deviation to $g_i^t = 1$ in either period yields a payoff of

[3]Note that in the game with only two nations, the coalition consisting of both will always pick a Pareto efficient action profile as any other profile can be jointly improved upon.

$$2\alpha_i - c_i \left(\sum_j m_j g_j^1 + \sum_j m_j g_j^2 \right) = 2\alpha_i - c_i \cdot 2 \sum_j m_j = 2(\alpha_i - c_i m_i) - c_i \cdot 2 \sum_{j \neq i} m_j$$

and as $\alpha_i - m_i c_i > 0$, the latter is a better option.[4] The result is consistent with the prediction by Shapley and Shubik (1969) for an economic system with negative externalities.

We summarize the above discussion as the following result.

Proposition 2.1. *Assume that* $\sum_{i=1}^n \alpha_i - N \sum_{i=1}^n m_i c_i < 0$ *while* $\alpha_i - m_i c_i > 0, i = 1, ..., n$. *Then, the core is empty and no multilateral agreement with full participation is possible.*

Proof. Follows from the discussion above. ■

2.3.1 Endogenous technology

Consider now a variation of the preceding model where the choice of g_i^1 impacts the incentives for other nations to cut emissions at $t = 2$. This is a similar to the way technology is regarded in Barrett (2006) if it had increasing returns to scale. Therefore, the conclusion should match as well – a treaty should take on a coordinating function to ensure enough countries adopt the technology early on. Whereas Barrett interprets the approach as shared R&D investment with increasing returns, one can also think of it as a learning process where watching enough countries implement the technology leaves no gaps in the understanding of how it functions and how it can be applied elsewhere. It can be exemplified by some contradictory technology such as spraying sulphur dioxide in the atmosphere to dim the sun, the consequences of which are difficult to predict on paper, and natural experiments are impossible to conduct. Before resorting to this option, a nation must be convinced that the technology is safe to use, and only by observing many other nations doing so will the evidence be enough.

A simple way of modelling it is as follows.

Assume that now α_i is time dependent for each i, so that $\alpha_i^1 = \alpha_i$, and

$$\alpha_i^2 = \begin{cases} \alpha_i \text{ if } \tilde{M} < \bar{M} \\ \bar{\alpha}_i \text{ if } \tilde{M} \geq \bar{M} \end{cases}$$

[4] Note that as $\alpha_i - m_i c_i > 0$, allowing for transfers between the nations makes no difference: the core will be empty in this case as well.

where $\bar{\alpha}_i - m_i c_i < 0$ for all i and $\tilde{M} = \sum_{i \in \left\{i \in N : g_i^1 = 0\right\}} m_i$, the critical number of adopting countries which is between 0 and M.

The interpretation is that if the critical mass of nations cutting emissions at $t = 1$ exceeds a certain threshold $\bar{M}$, then it becomes a dominant action for each nation to cut emissions at $t = 2$. This assumption models the idea that an initially limited, unilateral commitment to cut emissions by a small group of nations will stimulate innovative activity in technologies that lower the relative cost of low carbon activities for all nations. The Paris accord, by relying on unilateral commitments, can hope for such future transitional cost reduction. For instance, developing wind and sun energy harvesting technologies lowers their cost over time, which at some point will make their price compatible with the price of traditional energy sources and even cheaper, for every country. This feeds back to the idea proposed in chapter 1 about engineering the transition in a network of countries. If there is indeed learning to such an extent that the cost of a new technology will fall quickly and substantially when adopted by many (pivotal) nations (or, as in Barrett (2006), the technology has increasing returns to adoption), then the following logic applies. Since the end price charged to the consumer is usually higher in the case of a new technology (one can think of renewable electricity price as compared to conventional sources of energy), and assuming that a representative consumer has no internal motivation to choose a green technology over brown because she makes her choice exclusively based on price, the green technology will never take off without government support. As discussed in chapter 1, a nation whose citizens are endowed with such mentality will have to first change the mentality, which takes decades, so lowering the price of a new technology is likely to be a faster way to ensure it is adopted. Alternatively, a government may not be interested in a transition at all. Then, an international institution which possesses the funds, can subsidise the end consumer in some nations to stimulate the adoption of a new technology there, and once the critical mass is reached, the cost of adopting it elsewhere will be low enough to make it a dominant action. Reinterpreting the model in terms of learning rather than shared R&D costs completely changes the nature of results. Instead of requiring z nations to ratify the treaty before it enters into force, as Barrett suggests 2006, it can specify which countries get subsidized first. Such formulation is likely to make countries more cooperative and willing to sign the agreement.

The subgame perfect equilibrium scenarios are described by the following proposition.

Proposition 2.2. *Let* $\hat{N} = \left\{i \in N : 2\alpha_i - \sum_{j \in N} m_j c_j < 0\right\}$. *Then:*

(i) If $\sum_{i \in \hat{N}} m_i < \bar{M}$, *at any SPE,* $g_i^t = 1$, *for all* i, t;

(ii) If $\sum_{i\in\hat{N}} m_i \geq \bar{M}$, and $\bar{M} \leq \inf_{i\in\hat{N}} m_i$, at any SPE, there exists $i \in \hat{N}$, such that $g_i^1 = 0$ with $g_j^1 = 1$ for all $j \neq i$ and $g_i^2 = 0$ for all i;

(iii) If $\sum_{i\in\hat{N}} m_i \geq \bar{M}$, and $\bar{M} > \inf_{i\in\hat{N}} m_i$, there are two types of SPE:

a) $g_i^t = 1$, for all i, t;

b) there exists $N^* \subseteq \hat{N}$, such that $g_i^1 = 0$ for $i \in N^*$ with $g_j^1 = 1$ for all $j \notin N^*$ and $g_i^2 = 0$ for all i.

Proof. Consider a subgame perfect equilibrium where $\hat{N}$ is non-empty and $\sum_{i\in\hat{N}} m_i \geq \bar{M}$. Then, there exists $N^* \subseteq \hat{N}$ such that $M^* = \sum_{i\in N^*\subseteq\hat{N}} m_i$, where $M^* \geq \bar{M}$ and $M^* - m_i < \bar{M}$ for all $i \in N^*$.

Let $g_i^1 = 0$ if and only if $i \in N^*$.

If $i \in N^*$, consider a unilateral deviation by i so that $g_i^1 = 1$. In this case, following a deviation by i, the dominant strategy for each j at $t = 2$ is to switch from choosing $g_j^2 = 0$ to $g_j^2 = 1$. By computation, it is checked that i will not find it unilaterally profitable to switch if and only if

$$-c_i \sum_{j\in N\setminus N^*} m_j > 2\alpha_i - c_i m_i - c_i \sum_{j\neq i\in N\setminus N^*} m_j - c_i \cdot M$$

In other words, a deviation is not profitable if and only if $i \in N^* \subseteq \hat{N}$.

Next, if $\hat{N}$ is non-empty with $\sum_{i\in\hat{N}} m_i < \bar{M}$, then $\sum_{i\in\hat{N}} m_i - m_k < \bar{M}$ and $g_k^1 = 0$ for each $k \in \hat{N}$, so that following a unilateral deviation by some $k \in \hat{N}$ to $g_k^1 = 1$, it is a dominant strategy for each i to choose $g_i^2 = 1$, while k gains $\alpha_k - m_k c_k > 0$, with k's continuation payoff unchanged. Therefore, at any SPE where a non-empty subset of individuals choose $g_i^1 = 1$, $\hat{N}$ must be non-empty with $\sum_{i\in\hat{N}} m_i \geq \bar{M}$ and $i \in N^* \subseteq \hat{N}$, where N^* is such that $M^* = \sum_{i\in N^*\subseteq\hat{N}} m_i$, $M^* \geq \bar{M}$ and $M^* - m_i < \bar{M}$ for all $i \in N^*$.

Now, consider $i \notin N^*$ and suppose $g_i^1 = 0$. In this case, following a unilateral deviation by i to $g_i^1 = 1$, it still remains a dominant strategy for each $j \in N*$ to choose $g_j^2 = 0$ at $t = 2$. Therefore, i gains $\alpha_i - m_i c_i > 0$, while $i's$ continuation payoff remains the same. Clearly, if $\sum_{i\in\hat{N}} m_i \geq \bar{M}$, and $\bar{M} \leq \inf_{i\in\hat{N}} m_i$, at any SPE, there exists some $i \in \hat{N}$, such that $g_i^1 = 0$, with $g_j^1 = 1$ for all $j \neq i$ and $g_i^2 = 0$ for all i. However, if $\sum_{i\in\hat{N}} m_i \geq \bar{M}$, and $\bar{M} > \inf_{i\in\hat{N}} m_i$, if $g_i^t = 0$ for all i, t, for any $k \in \hat{N}$, following a unilateral deviation by k to $g_k^1 = 0$, it still remains a dominant strategy for each i to choose $g_i^2 = 1$. Therefore, k suffers a loss of $\alpha_k - m_k c_k > 0$ without changing her continuation payoffs. Thus, if $\sum_{i\in\hat{N}} m_i \geq \bar{M}$, and $\bar{M} > \inf_{i\in\hat{N}} m_i$, there is always a subgame perfect equilibrium with $g_i^t = 0$ for all i, t. ■

Following the Proposition 2.2, the question ensues, which countries would be the first to transition to low carbon economy. Intuitively, in equilibrium, only those for whom it is cheaper to switch to low emissions, will do so, and they might be enough to make a difference in the cost of the technology in the future period, so that the rest of the countries will free ride in the first period but follow suit in the second.

An example of a major transition that has started in one country and led to a successful international treaty is the Montreal Protocol. Signed in 1987, it was a product of several years of negotiations with the US having a markedly different view on the cuts required to mitigate the ozone depletion than the rest of the world. The protocol was an outcome of many forces confronting each other, but from a bird's-eye view, it was the unbending position of the US that provided incentives for other nations to reconsider their stance (Morrisette 1989). At the time, the US was the largest producer as well as consumer of chlorofluorocarbons (CFCs), and the US-based chemical corporation, DuPont, was the company with the largest market share, domestically and worldwide (Haas 1992). Although the country was not large enough to single-handedly mitigate the issue, due to the size of the market it did have the upper hand in the negotiations: the threat of a unilateral ban on imported products containing CFCs represented a worse option for the European manufacturers than a treaty with caps and cuts (Haas 1992). Stringent regulations were instituted for the good of everybody, when originally only one country was interested in this outcome.

The US was the key player in the CFCs market whom other nations could not afford to ignore. The European companies already had manufacturing surplus and could not have risked losing access to the US market, while the US, having its largest producer ready to invest in alternative chemicals and move away from CFCs once the alternatives were found, would close its market unilaterally if satisfactory pollution reductions were not agreed on (Haas 1992). The fact that DuPont was in favour of strict regulations was a boon to the US's bargaining power and not only added credibility, but actually transformed the market conditions thereby incentivising smaller firms to seek alternative chemicals to move away from CFCs.

It appears that the US lowered the cost of technological substitutes for CFCs through raising the cost of business-as-usual, and it was just enough to ensure that the efficient equilibrium was chosen and sustained.

The development of the international ozone protection agreement can be viewed as a two-stage process, where in the first stage, stratospheric ozone depletion was a domestic issue for the US and some other countries (Morrisette 1989). There, it was already high on a political agenda and the public was in favour of pollution regulations. This, and the fact that the US (being a pivotal member of the international community) was the first to initiate the transition away from CFCs, were effectively driven by the circumstances without directed traceable intervention. As part of the supersonic transportation programme, ozone layer damage had been looked into and established as an environmental threat (Morrisette 1989). Moreover, those in power were either trained in science or sympathetic to scientists, which facilitated the spread of knowledge and ensured that potential consequences of inaction were well understood (Haas 1992). The influence of the scientific community cannot be disregarded. Also not to be overlooked is the galvanizing effect of the discovery of the Antarctic ozone hole which was not predicted by any model. Together with bilateral conversations the US had been conducting prior to 1987, these and possibly some other unaccounted circumstances resulted in a treaty that is still up and running rather successfully.

The issues underpinning the adoption of an international emissions reduction agreement are more complicated. There are more parties to the negotiations with considerable bargaining power; the phaseout of fossil fuels and GHGs pollution abatement is not merely about finding substitutes like it was with CFCs; the consequences of inaction are uncertain and hard to quantify with adequate precision. Climate change is simply a much more complicated problem than the ozone depletion ever was. However, the story of network externalities and political spillovers still has a part to play, especially in emerging markets, nascent industries and innovative technologies.

Following, there are several policy implications that arise from the model.

To identify the nations who are likely to be in $\hat{N}$, suppose that the linear payoffs used in the example are actually linear approximations of intertemporal expected utility, so that α_i is proportional to the current marginal utility from high emission activities today evaluated at the current level of the per capita consumption, whereas c_i is proportional to the expected future utility loss in future per capita consumption. This likely to be true if the nation enjoys high per capita consumption at $t = 1$, so that α_i is small and c_i is high. Unsurprisingly, this implies that $\hat{N}$ likely consists of wealthy nations who can afford reducing their emissions.

The assumption that emission cuts by a certain number of countries is needed to lower the costs of reducing emissions for all other countries is, in part, an assumption about the strength of positive technological spillover effects (Grafström and Lindman 2017). It rests on the idea that any unilateral commitment to cut emissions will stimulate policy/institutional/technological innovation that has positive spillover effects across nations. However, this is also an assumption that rests on the design of appropriate policy mechanisms – efficient information sharing and design of targeted subsidies to facilitate technology transfer are some examples. An international environmental agreement should, if not be based entirely on technology development and sharing, then at least incorporate articles that facilitate investment and learning (Lindman and Söderholm 2012). Wind energy generation, for instance, has achieved a maturity level at which learning has transformed from domestic to international, so it is advised that public policies are coordinated internationally and target research, development and deployment (Langniß and Neij 2004). The treaty therefore should be a coordination device that chooses a subgame perfect equilibrium that Pareto dominates the rest of them, since, as the proposition above suggests, this is a game with multiple equilibria. It will likely be a narrow treaty that only a few nations will ratify, and initially the scope of emission cuts will not be very different from existing unilateral initiatives, but in the future, it is expected to lead to considerable global emission reduction. In practice, building on and operationalizing such a spillovers-based agreement may require a global funding mechanism to subsidize both technology transfer and efficient information sharing.

Another channel through which a transition in one country can inspire a similar shift in others is by reducing uncertainty about the cost of transition. Nations can learn by simply acquiring a technology developed elsewhere and reverse-engineering it, or by forming networks with a purpose of, for instance, addressing such pressing matters as climate change. Such networks have become rather widespread in recent years, and its actors range from communities and cities to whole nations. Among the most prominent are the C40 Cities Climate Leadership Group, Asian Cities Climate Change Resilience Network, Climate Mayors. Members of these networks cooperate to exchange knowledge, provide technical assistance, share best practices and facilitate their implementation. This insures cost sharing between countries/cities and also creates more benefits from otherwise costly actions, such as alterations in transportation system, upgrading existing buildings to be more energy efficient, etc. C40, for instance, has a third-party implementation partner that in turn partners with private companies to supply members with resources at lower price provided that explicit planning for reduced emissions takes place (Román 2010). Not only the costs are actually lower for certain policies for cities that are part of the group, but by sharing experience and knowledge, the members can more accurately predict the cost of a particular decision. The structure of the C40 is such that within the network, there are smaller groups comprised of a handful of cities that focus on an isolated environmental issue. Often, the cities form these groups on the basis of facing the issue as their individual most pressing problem. London, for instance, as the most polluted European city, leads the network on tackling air pollution by electrifying its public transport and introducing low emission zones. This sets a precedence for other network members and allows them to learn not only the cost but also the benefits of the policy.

Finally, an important conclusion of the model is that the costs associated with high emissions cannot be completely avoided. Therefore, contrary to some prior research (Harstad et al. 2019), the model suggests that in addition to mitigation, sufficient investment is required to ensure that nations are in a position to adapt to climate change.

2.4 Discussion and further developments

So far, the process of technology transfer has not been assumed explicitly, apart from the fact that it exists. Technological trade is not commonly found in models of international environmental agreements. Below I discuss a paper by Chatterji and Ghosal (2009) that attempts to do that and is the only paper that I was able to find close to my vision.

In a similar model as the one introduced earlier, minus the difference weights countries have in the global emissions, the authors study how technology affects the outcomes of the game. They assume that at $t = 1$, each country makes a choice of whether to invest in a technology, $x_i \in \{0, 1\}$, that lowers the relative cost of cutting emissions at $t = 2$. The payoffs are linear and additive, so the way it lowers the cost is by adding exactly x_i to the country's payoff (no discounting is assumed). In order to invest, a nation has to pay an up-front cost of $k_i(x_i)$ with $k_i(0) = 0$ and $k_i(1) = k_i > 0$. Alternatively, nation i can obtain access to the technology developed by nation j, in which case both nations will have the same benefit x_i, but only one will have paid the investment cost.

The analysis is performed with only 2 nations.

They consider the case when $\alpha_1 - c_1 < 0$ but $\alpha_2 - c_2 > 0$, so that without an agreement, Nation 1 will always cut emissions, while Nation 2 will never do so. The question they want to answer is whether there is such a price p_{21} that nation 2 would find it in its own self-interest to acquire the technology from nation 1. Since its payoff from doing so is equal to $1 - p_{21}$, it will access the technology if and only if $p_{21} < 1 - (\alpha_2 - c_2)$. To ensure incentive-compatibility for nation 1, $p_{21} \geq -c_1 + k_1$ must hold. Note that k_2 must be large, $k_2 > 1$, to deter nation 2 from investing unilaterally.

Deterring a nation from investing may sound counter-intuitive, but it is simply a matter of efficiency – investing twice in the same technology, while one nation could do it and pass the technology to the other for free, is not rational. This is what

Barrett calls an overinvestment and hopes to avoid with the help of an international treaty (Barrett 2006).

Therefore, under certain conditions on parameters, nation 1 would invest and nation 2 would acquire the technology from nation 1 at a price p_{21}, which can be positive or negative, in which case it is a subsidy. The agreement would specify the price, so that with (possible subsidized) technology transfer, there is a path to cumulative emissions reduction.

The most interesting case of both nations benefiting from continuing with high carbon activities is not discussed.

The paper is incomplete in many ways and it leaves the reader with several intriguing questions. One obvious question is about the outcome of the game when neither of the nations wants to cut emissions unilaterally. Would the presence of technological trade encourage one of them to do so? How would an investing nation be chosen? Would they need an agreement to determine the price? Another question is, if one nation decided to invest unilaterally, what is its motivation? Could other nations be motivated to invest in a similar fashion?

It also lacks an empirical confirmation of the strength of technological spillovers pertaining specifically to the context of the chapter. The key empirical question which one would want to ask is by how much does adoption of a "green" technology by country j at period t increase the likelihood of adoption of the same technology by country i at period $t+1$. The main problem in trying to answer such question would be data. What green technology should the focus be on? What policy interventions have "speeded up" the adoption of that technology?

The most obvious choice would be to look at a green technology which has been around for many years and which has become widely adopted over time. The two possibilities are wind and solar energy. Higher investment in either of these technologies leads to lower emissions in the energy sector, but initially, these green technologies were more expensive per unit than fossil fuel technologies.

World Bank, the International Energy Agency and the OECD would be the first point of call to look for data on cumulative installed capacity for different countries over time. One would identify individual countries as leaders (e.g., Denmark or Germany for wind energy), and then look at whether the relationship between investment in wind in either of these countries has resulted in higher rates of adoption in other countries. That of course means trying to control for other factors driving green technology adoption, such as the price of fossil fuels, foreign direct investment inflow, net electricity imports, GDP per capita, energy import, energy intensity, greenhouse emissions, etc.

One could also look at policy instrument such as carbon or energy taxes – have these aided the rate of adoption?

These questions I will try to answer in the next chapter and in the future research.

2.5 Conclusion

In this chapter, I seek to find a way to introduce technology to the model to answer several important questions. Observing the climate crisis unfolding despite international best effort begs the question: what are the treaties of Kyoto and Paris doing wrong and how do we fix it? It is evident from the literature and my model that basic models that do not explicitly account for technological change in an endogenous way, do not paint a pretty picture – the size of a stable coalition is small, especially when gains from cooperation are large. Thus, technology must be introduced, and it has been in many models, but there are only two main approaches that I have identified. One is abatement-based, the other is emissions-based and technology is a stock. Both approaches are problematic for answering the questions above. The first cannot model accumulation of damages and delay, the second does not suit for modelling technology transfer by means of trade.

I then introduce my own simple model with and without technology and verify its predictions against the literature. There is no stable multilateral agreement which sustains full cooperation in a variation without technology, while there is multiplicity of equilibria in a variation with endogenous technology and even the grand coalition is stable. Even this model however does not allow for technological trade, so I then discuss an introductory version of a model that does by Chatterji and Ghosal (2009) and conclude with open questions that I intend to answer in the next chapter.

Chapter 3

Technology, Unilateral Actions and Global Emissions Reduction: Dynamic Game of Climate Change

3.1 Introduction and related literature

The game of climate change has been studied extensively by economists over the past decades, and many approaches have been used to find the optimal mitigation solution (for a comprehensive review refer to Carattini et al. (2019)). The classical approach relies on dynamic games and considers international environmental agreements as the outcome of a stage game where countries first decide on whether to join a coalition and then on how much to emit given the decisions of others (Barrett 1994). The models of this type have been modified over time to study the effects of uncertainty about the benefits and costs of mitigation efforts and the role of learning (Finus and Pintassilgo 2013; Ulph 2004; Kolstad 2007). They have also been turned into repeated games where countries invest in technology (clean or otherwise) and pollute on an infinite time horizon, with the classic stage game representing one period. The studies utilising this approach either focus on identifying an often unique Markov perfect equilibrium (Harstad 2012[1]; Harstad 2016; Battaglini and Harstad 2016), or, aiming to solve for self-enforcing treaties, look for subgame perfect equilibria (Harstad et al. 2019).

The models described above are the main tools in studying global cooperation in climate change mitigation. Other approaches make use of capital accumulation games (Chander (2017), building on Dockner et al. (1996)), mechanism design (Martimort and Sand-Zantman 2016), macroeconomic models of growth where climate is included as a parameter of the production function (Nordhaus 1993a; Nordhaus

[1]In this model, the MPE coincides with SPE if the time was finite but approached infinity.

1993b; Nordhaus and Yang 1996), and Nash bargaining solution (Carraro and Siniscalco 1993). The concept of the core can also be applied to the game of climate change. Although, it has been shown that the core is empty in economies with negative externalities (Shapley and Shubik 1969), under certain assumptions and with some modifications to the concept of the core, it can be non-empty when public "bad" such as environmental degradation is present, and the solution will then be analogous to Lindahl equilibrium (Chander and Tulkens 2006b; Chander and Tulkens 2006a).

Empirically, public goods games have been employed to test the theoretical predictions (Barrett and Dannenberg 2012; Barrett and Dannenberg 2014; Barrett and Dannenberg 2017).

The primary narrative of global climate policy has thus been collective action theory. According to this view, climate change is a public "bad", whose mitigation is therefore a public good, free-riding is the culprit to global cooperation and all efforts of the global commons should therefore be directed at deterring it. Climate change is a prisoner's dilemma type game and requires commitment mechanisms to force players to cooperate on a Pareto efficient outcome (for a more detailed discussion see Barrett (2005b)). Any treaty that aims at deep global emission cuts should focus on re-engineering the incentives and changing the underlying game structure, relying on external penalty such as import tariffs and taxes (Barrett and Stavins 2003; Nordhaus 2015). A pledge review process that is incorporated in the Paris agreement may be marginally better than no review, but so long as it does not restructure the payoffs, the treaty is of limited help to the environment (Barrett and Dannenberg 2016).

There are other reasons why prisoner's dilemma is the basis of modern international climate treaties analysis. For instance, since every country's contribution to global environmental damage is small relative to the overall temperature rise and it can only claim a small proportion of global mitigation benefits, every nation has incentives to free-ride on the efforts of others. Also, increasing marginal abatement costs and a hold-up problem due to worsened future bargaining position of the nation that heavily invests in green capital at present make countries more reluctant to cooperate and participate in global environmental treaties (Beccherle and Tirole 2011; Harstad 2012; Harstad 2016).

Settling on this perspective has implications. First, framing the problem in terms of burden sharing makes national decisions interdependent to a point when the choices of one country are conditional on the choices of the others. Surely, climate change is the story of negative externalities with emissions of one country inevitably entering the payoff functions of all other countries, but this is not the same as assuming that their policies are *conditioned* on each others. There are reasons to believe that public support for domestic policies and mitigation actions is not affected by the mitigation efforts (or lack of thereof) of other nations (Bernauer and Gampfer 2015; Bernauer et al. 2016; Beiser-McGrath and Bernauer 2019). Moreover, actions of national leaders tend to be unconditional as well (Aklin and Mildenberger 2020)[2]. Governments can also act on climate change due to domestic pressure or to address pressing environmental domestic issues (Keohane and Oppenheimer 2016).

Second, assuming that the underlying game is a prisoner's dilemma implies that every country has a dominant strategy of not reducing its emissions, under any circumstances. While this approach could explain the observed inaction in global climate policy, it does not explain ample actions taken by some countries or group of countries to reduce their carbon footprint, such as the climate policy of the European community or aggressive uptake of renewable energy in China. On the other hand, the lack of action could be due to carbon lock-in (Unruh 2000b), desire of leaders to enjoy political rent by staying in power at the expense of social welfare (Acemoglu and Robinson 2000b), uncertainty about the post-transitional wealth distribution (Fernandez and Rodrik 1991), agents' short-termism and many more. Free-riding concerns, while may be of some importance to the parties involved, do not have to be invoked to explain global climate mitigation procrastination (Aklin and Mildenberger 2020).

Additionally, a different view is offered by Carraro and Siniscalco (1993), who hypothesize that the underlying game is rather the game of chicken, which is a coordination game with multiple equilibria and no dominant strategy. The observation they make is that "all countries have an incentive to let the others cooperate" (Carraro and Siniscalco 1993, p. 322), which puts an emphasis on coordinating one's policy choice so as to maximize individual gains, which may happen with emissions unchanged as well as being a part of a stable environment-preserving coalition. They contend that, unlike in most of the models where nations are assumed symmetric and identical, there is asymmetry in practice and countries differ in their preferences, technology, abatement costs, etc., which leads to different preferred climate strategies. For instance, countries with lower abatement costs may choose to form a coalition, while those with higher abatement cost may want to allow them to do so using transfers. The equilibrium selection process is not in question here,

[2]To support their argument, the authors invoke two episodes of climate policymaking in the US that are mostly interpreted as support for conditional cooperation, but propose that in those instances, the parties involved were in fact "unconditional non-cooperators" who were not interested in a climate treaty in the first place and used collective action account and free-riding concerns to justify their position.

but the multiplicity is obvious, which is not a feature of a prisoner's dilemma type game.

Coordination game can also arise if the model incorporates the idea of an environmental tipping point - there is a possibility of experiencing a catastrophic damage if the threshold global abatement level is not achieved (Barrett 2013). However, uncertainty about the abatement needed to avoid the runaway climate change is likely to turn the coordination game into, once again, a prisoner's dilemma.

Another reason why prisoner's dilemma may not accurately represent the reality of climate negotiation is reversal to business-as-usual as a punishment for non-compliance with the terms of the treaty. When environmental agreements are modelled as repeated games and players have a dominant strategy of not reducing emissions in any one stage of the game, a penalty is needed to prevent them from defecting from a Pareto optimal equilibrium. External motivation in a form of a concomitant trade agreement which permits import bans and tariffs would be the best non-compliance deterrent (Nordhaus 2015). The Montreal protocol thrived on a unilateral threat of trade bans by the pivotal player in the market of chlorofluorocarbons, which was credible and could potentially wreak havoc on non-complying national economies (Haas 1992). However, climate change is not caused by a single polluting substance, nor does the solution lie in simply finding alternative chemicals without altering the fundamentals of the economic and societal systems, so no major climate change mitigation agreement is tied to trade sanctions. This leaves internal compliance mechanisms as the only option. Alas, those are neither renegotiation-proof, nor credible in the form they are usually assumed in the literature. Let me elaborate on this point. It is a common practice in the models of international environmental agreements to assume that, upon observing a defection, other members of the treaty immediately revert to business-as-usual to penalise the defector (Dutta and Radner 2009). Since this is never observed in equilibrium, the particular form of punishment is not that important as long as it is credible. But this is in theory. In practice, however, reversal to business-as-usual seems problematic. Implicitly embedded in a nation's ability to increase emissions at the first sight of non-compliance is the assumption that capital can be built overnight and thrown away once acquired. Realistically speaking, none of these actions are possible or rational. More importantly, the specificity of climate change mitigation lies in immediate costs in exchange for delayed benefits. So, once the costs have been incurred, would a nation stop pursuing a low-emission strategy merely because another nation somewhere did not fulfil its promise? Unlikely.

Collective action theory as the basis of international environmental agreements also abstracts too much from national political agenda. Political scientists have been arguing that such approach, although being a reasonable simplification, should not be so widely adopted, and domestic and international politics are intertwined and should be considered together.

International negotiations should thus be analysed as a two-level game where participants have to reconcile international and domestic needs. It would be wrong to consider each level in isolation (Putnam 1988). Some economists internalized this idea in search of the explanation for the "paradox of weak agreements" introducing re-election concerns into the model (Battaglini and Harstad 2020). The agreement there is signed strategically so as to affect the preferences of the median voter and maximize the probability of staying in power. Adding domestic political concerns thus leads to different results and may prove to be indispensable if we are to address global warming effectively.

Lastly, collective action approach and especially free-riding concerns lack empirical support in global climate politics. To be precise, the experiments designed to analyse the relationship society has with climate change mitigation do so within a collective action framework (Barrett and Dannenberg 2012; Barrett and Dannenberg 2014; Barrett and Dannenberg 2016), but this does not provide evidence supporting free-riding as the main driver behind the weakness of global climate action. Domestic distributional conflict can explain the observed inaction equally well (Aklin and Mildenberger 2020). The authors do not discard free-riding as a non-existent issue. Rather, they believe it not to be of utter importance. Climate change policies will create winners and losers whose identities cannot be known in advance. This inevitable wealth redistribution holds governments' hands tied while they try to keep current constituencies happy. This, and not free-riding concerns, prevents them from implementing stringent climate policies that could deliver a socio-economic transition in the shortest period of time.

Taking importance of domestic distributional conflict and re-election concerns at par value, it can be deduced that nations mostly act *unilaterally* - national leaders strive to stay in power even if it does not maximize social welfare (by signing weak suboptimal treaties, for example). This view would be the polar opposite to collective action account, but still extreme and unlikely to reflect the reality well. Merging these two approaches may yield the most accurate predictions and insights about global climate cooperation.

In this paper, I propose a model that does not draw heavily on collective action approach, neither does it presume the governments are being driven primarily by the re-election concerns. National leaders are still assumed to act unilaterally. However, the term is not understood here as it is in most of the literature. The difference is subtle but important – each nation does not act out of concern for global environment, but this does not imply that it does not act at all. As Nordhaus (2015) briefly pointed in his work on climate clubs, the nations who would start the club may be the ones with the lowest abatement cost or highest environmental damage. Therefore, in my model, asymmetric environmental damages are key. Under these circumstances, some countries will inevitably start to experience the damage earlier than others: local pollution may become so high that it affects public health already at present, the sea level rise may submerge entire areas of a country and threaten to submerge more, etc. It is these countries who have incentives to act unilaterally even if no other nation does. Their leaders are not driven by desire to stay in power. They are simply addressing domestic pressing environmental matters, which they have direct interest in mitigating. The question I am thus interested in is what effect, if any, these acting countries have on nations with no such natural incentives.

Mine is not the first paper to study unilateral emission reductions and the reasons behind it. It has been suggested elsewhere that when there is private information about asymmetric environmental damages, a country may want to reveal its type through signalling high damage cost with early emission cuts (Brandt and Nannerup 2013). This action may, however, lead to carbon leakage, i.e., increased emissions from other nations, but only if technological advancement and cross-country spillovers are considered exogenous or non-existent. On the other hand, a treaty that relies on adoption of a green technology with increasing returns can be successful in sustaining high level of international cooperation (Barrett 2006). Similarly, if countries differ in their preferences towards the environment, and technology spillovers are allowed, it is possible to observe emissions reduction in all nations when the preferences change in only one of them (Golombek and Hoel 2004).

Another driver of early unilateral actions may be heterogeneous (but known) values put on environmental protection. This way, the burden of contributing to the global public good lies mostly on the nations with the highest valuations, and if technology sharing is facilitated, then in equilibrium, their green investment can be even larger (Elsayyad and Morath 2016).

I develop this idea further by introducing uncertainty allowing for explicit technological exchange. The issue of interest is whether cumulative global emissions reduction is possible without an international agreement or with one of a largely voluntary nature such as the Paris agreement: if the cost of developing a green technology is not known in advance, but once developed, it can be acquired at a set price. By assuming the presence of technological spillovers (by acquiring access to it and not through weak intellectual property rights), I analyse global cumulative carbon emissions dynamics, and whether significant reductions are possible when the nature of actions is rather voluntary.

In this chapter, we propose a novel approach to a standard question of global emissions reduction path. It is standard in the literature to assume that nations suffer from global pollution, disproportionally to their contribution, but according to their individual damage ratio. We instead assume that each nation presently suffers only from its own accumulated pollution, while the damage from global emissions is delayed until the third (last) period. This allows us to model the incentives some nations might have to invest in emission-cutting technology. We assume that nations only act non-cooperatively, in their best interest, and will not reduce emissions for the good of the world if this action is dominated. There is no utility or cash transfers between the nations, which is an extreme assumption on their selfish nature, which probably only partially holds true. But the opposite assumption is even more dangerous because then, the solution would rely on the existence of the transfers, and, as Chander (2017) put it, they would be a necessity and not a choice. If the reality proves that transfers are impossible, the model that assumes otherwise would be invalidated; but if the transfers were possible, the model that did not rely on their existence for its predictions would only be enhanced.

The question we ask is under what circumstances we can observe significant global emission cuts, and what role does a treaty have to play when nations only act non-cooperatively.

The model gives rise to some interesting dynamics. Nations' decisions change depending on their proportional share of global emissions. As such, a small nation is less likely to invest in technology than a large one is. This is because a large nation stands to gain more from its own reduction in the future. Also, in some setting, investing earlier is an inferior option to investing later, despite the nations being time-consistent. But in line with standard folk theorems, investing at $t = 1$ would be a dominant action for a large enough discount factor. Technological trade makes nations more willing to invest because it allows them to recover some part of the investment, and if, in the absence of trade, no nation would invest, when the trade is allowed, full adoption over time is possible to observe. This outcome suggests that the intellectual property rights should be strengthened and protected to ensure that an investing nation can capture the benefits associated with sharing the technology down the line. The agreement thus only specifies the technology price and defines the property rights around it. There is no bigger role for it in our model. Unless some considerably sized nations suffer from early environmental damages, there is naught a treaty could do to encourage investment.

Recently, top-down Kyoto protocol has been substituted with a bottom-up Paris agreement, and neither seems to be satisfactory for many scholars. While Kyoto is obviously a failure, Paris is only predicted to be one, but is already being heavily criticized for being inefficient, unable to deter free-riding, and just "an inadequate response to climate change" (Gollier and Tirole 2015, p.12). Other authors are slightly more optimistic and believe that the improvement in global climate policy may occur because states, industries and public recognize the severity of the problem, and their interaction at the domestic and international levels will create incentives for them to act (Keohane and Oppenheimer 2016). Its success will depend on whether national interest groups who hold the power will be interested in supporting stringent environmental regulations. With this paper, I would like to provide even more reasons to be optimistic. All is not lost because grand coalition is not stable and free-riding is not successfully deterred. Nations are not playing a prisoner's dilemma game, so the solution will not be what we expect either. Does not mean it will not be enough to avoid a calamity.

3.2 Model with local damage

The model builds on the following preliminaries. All countries perform activities that emit greenhouse gas as a byproduct and derive some benefit from such activities. At the same time, accumulated pollution affects countries differently, and some may experience consequences earlier than others. Any country can choose to invest in green technology and in this way reduce its emissions. Other countries can then choose to acquire this technology for a price and reduce their emissions as well, even though there is no immediate and obvious consequences for them of continued pollution (yet). We want to study under what conditions this process will lead to significant global emissions reduction.

There are some key elements such model has to have to fill in the gaps in the literature.

There must be explicit asymmetry which generates different strategies and patterns. It is common to introduce country-specific coefficients or indices while keeping functional form of a per-period utility function the same (Dutta and Radner 2009, Harstad 2012, Harstad 2016, Battaglini and Harstad 2016), which results in a stationary Markov perfect equilibrium generally referred to as business-as-usual (BAU). From this benchmark, the papers proceed to analyse different levels of cooperation, length and depth of agreements, best contractual variables (emissions, investments, or both).

Some papers deal with asymmetry more explicitly. Barrett (2001) considers two types of countries depending on their parameters, i.e., one type benefits more from global abatement than does the other. Apart from this work, most efforts have been directed to study asymmetry in the context of uncertainty about environmental damages and how learning may affect international cooperation (Na and Shin 1998, Kolstad and Ulph 2011).

This second type of models rely, however, on the stage game approach with no stock variables or green capital, that both are the two indispensable elements of our model. These have been taken into account elsewhere. Technology, for instance, has been incorporated in the works of Harstad and Battaglini (Harstad 2012, Harstad 2016, Battaglini and Harstad 2016). There, it is treated as a stock and it is not easy to re-interpret the model to reflect the element of technological trade in a way that would considerably alter the behaviour of the purchaser of the technology (although Harstad (2016) does include a section devoted to trade policies and intellectual property rights). Harstad (2016) also introduces a stochastic element in the form of time-independent shock to global greenhouse gases stock and technological externalities as part of a per-period utility function. Since the shock affects all countries in the same way, this does not generate interesting new dynamics as compared to the case without stochasticity. Note also that in all of his papers on this topic, Harstad reformulates the model from being a game with stocks to a repeated game by rewriting the instantaneous utility function to not contain any stocks explicitly, which is only possible because the stocks are not payoff relevant in the Markov perfect equilibrium sense.

It is also important to notice that in the literature on environmental agreements and cooperation, there is an idea that emissions reduction undertaken by a group of countries can lead to a treaty with wider participation and considerable global cuts. Barrett (2006) sees such treaties as a result of increasing returns to scale of the new technology, Nordhaus (2015) – as a consequence of creating a climate club with exclusive benefits and punishment for non-participation. Another reason can be multiplicity of equilibria, so that a treaty would play a role of a coordination device and some countries, somehow, would be the first to join. However, there is no game-theoretic rationale as to how these countries are chosen and why. I would like to fill this gap by modelling asymmetry explicitly through experienced environmental damage function. Technology in this case is the only channel through which countries can influence each other's behaviour: a presently suffering country who has developed a costly technology can grant access to it (for a fee) to a nation with less interest in reducing its emissions. Such uninterested nation may wish to acquire the technology if the price is low enough.

The complex interaction between nations, sequentiality and heterogeneity imply that the number of agents will have to be limited if we want to characterize a closed solution rather than solving the model numerically, and each period will need to be described and considered separately from the others. Thus, while a stationary equilibrium can possibly occur at a benchmark solution, it is unlikely to represent optimal behaviour. In an optimum, a change in countries' actions is expected to take place, i.e., the nation would be expected to switch from positive emissions to none at $t = 2$ if another nation has invested in technology at $t = 1$.

For simplicity, the are three nations, i, j and k, and three periods. Every country derives benefit from individual consumption, but production comes with pollution as an externality. A country may choose to invest in a technology that would lower emissions while still allowing it to keep the consumption level unchanged. While the most attractive way to model this could be similar to Harstad (2012), as discussed earlier, this formulation does not permit an easily definable technological trade (but it does permit spillovers in other ways, such as licensing and intellectual property rights). Thus, clean technology needs to have a more defined threshold of implementation; a point after which technology is operational and available for sale. There are multiple ways of introducing technology to satisfy this. For one, it can be a switch style variable, which simply changes emissions from high to low without affecting the output. Another way would be to follow Harstad but add separate R&D cost to the utility function if a nation decides to go green. Once paid, this sunk cost allows emissions to be reduced proportionally to accumulated green capital as well as selling this technology to other countries.

As an interesting extension, the research and development costs can be not known in advance, only roughly estimated. They can turn out to be surprisingly low or prohibitively high, and this is the uncertainty that deters countries from investment in the first place. This would also be a clear advantage of acquiring technology over developing it – the price of purchase is predefined with no uncertainty. This can be incorporated into the model by assuming that the cost of technology can be high or low with probabilities p_i and $(1 - p_i)$ respectively. The probabilities can and should vary from country to country.

As in all of the literature on international environmental agreements, there is also global pollution cost, which is an increasing function of accumulated emission stock, but the global damage is delayed to $t = 3$. In addition to this, I introduce local pollution which is a function of only local emissions, and thus nations have significant control over it. As another interesting extension, it may be affected by a shock as in Harstad (2016) which is a function of green technology. In the simplest scenario, it can be completely eliminated as green technology is developed or acquired. Local damage can also be uncertain in its magnitude.

In the first period, the shock to local pollution is realised and every country observes the damage. The shock occurs once at the beginning of the game and local pollution is realised for the whole game. This step can be interpreted as Nature's move with each country learning its own type – high or low local damage. Once the type is determined, it persists for the rest of the game, which, coupled with the fact that local damage also depends on the stock of local emissions, implies that a country with a high local damage will only suffer more over time if mitigation/adaptation technology is not developed. So, at this stage, such technology can be invested in, but it will only become operational in the next period. It is assumed that once developed, it does not require any more investment. Therefore, we are not talking about green capital, but rather about breakthrough technologies as discussed in Barrett (2006) that require large research and development expenses and have unknown total development cost ex ante.

The game description is as follows.

At t=1, there are 3 countries, $i = 1, 2, 3$. Their production $y_{i,t}$ can be by means of clean or dirty technology. Harstad (2016), for instance, talks about energy that can be3 derived from renewable and non-renewable sources. The latter is by default; the former requires special technology to be developed internally or acquired externally from a nation who had developed it in the previous period (there can thus be no acquisition at $t = 1$). Whichever production method is chosen, it will cover the needs of the economy, but with different efficiency, so that $y_{i,t} = g_{i,t}$ when no clean technology is employed, and $y_{i,t} = e_{i,t}$ otherwise. The benefit from different types of technology is $b_i g_{i,t} > a_i e_{i,t}$. The variables are binary, so that at any time, the country can decide to either be green, $e_{i,t} = 1$, $g_{i,t} = 0$, or vice versa, but never both at the same time. As modelled, old dirty technology is simply more efficient.

As a result of production activity, there is cumulative *national* emissions, $G_{i,t} = q_G G_{i,t-1} + g_{i,t}$ with normalised initial level $G_{i,0} = 0$ for simplicity, and cumulative *global* emissions, $G_t = q_G G_0 + \sum_i m_i g_{i,t} = \sum_i m_i G_{i,t}$, where the mass, $\sum_i m_i = M$, is necessary since pollution is binary. In every period, $1 - q_G$ share of global pollution is dissipated in the atmosphere, while $q_G \in (0, 1]$ survives till next period.

At this stage, each country suffers from its own cumulative pollution, but to s different extent. Similar to Na and Shin (1998), there is some $\theta_i \in \{d_1, d_2, d_3\}$, with $d_1 < d_2 < d_3$, which defines country i's local damage from its own economic activity. For instance, China suffers significantly from air pollution caused by domestic industry, so it could be said to have $\theta_{China} = d_3$, while Norway has very low level of local pollution, so $\theta_{Norway} = d_1$. The damage is linear and equal to $\theta_i G_{i,t}$.

Any nation can choose to invest in green technology which costs k_i to develop and takes one period to mature (if invested at $t = 1$, country i will stop polluting and move to clean production at $t = 2$). An intriguing modification could introduce some uncertainty here and assume that technology may or may not mature in the next period, at which point a nation can either sink the cost and not continue with the investment or invest another k_i and wait for the next period, when it may again not mature, with some probability. This would reflect the fact that it is not possible to know at the R&D stage how soon a new technology will become operational and how much investment it will require. For the basic model however it is not needed.

At t=2, those nations who have not invested before can choose to acquire the technology from the ones who have, at a price p_{ij}, if i is purchasing the technology from j. As before, it takes time for green capital to be built, so the the technology becomes operational at $t = 3$. Note that if the uncertainty about maturity time is introduced as suggested above, there would be no uncertainty with the acquisition, and that would be a clear advantage that might discourage nations from investing, waiting and hoping that somebody else would take on a role of a market leader and resolve the uncertainty for themselves. Without these modifications, the main advantage technological acquisition has over its development is in the ownership itself. Some of the cost can be reimbursed by the buyer to the developer. This advantage would be lost if intellectual property rights are not enforced properly.

At t=3, global damage is realised (while local damage simply keeps accumulating throughout the game), linear in global pollution, $c_i G_t$. There is no decision to be made. It is the last period when nations get the payoffs based on the decisions they had already made.

It is important to define how the technology works. The model is versatile and can incorporate different types of green technology.

1. Mitigation technology such as clean energy, for instance, eliminates emissions from the moment it becomes operational, but does nothing about cumulative emissions from the previous periods. It does not reduce global damage either other than through reduced global emissions. Thus, such technology would set $g_{i,\tau+1} = 0$, $e_{i,\tau+1} = 1$, if τ is the period when investment/acquisition was made. Nation i's damage is then $\theta_i G_{i,\tau} > 0$, $\theta_i G_{i,\tau+1} = \theta_i G_{i,\tau}$, $c_i G_t > 0$. In essence, it freezes the severity of damages at a level that was experienced in period when investment was made. The environment does not deteriorate locally further, but it does not heal either, and global damage is experienced in full, minus own emissions following the technology implementation.

2. Technologies such as carbon capture and storage, on the other hand (which we would classify separately from mitigation and adaptation types), eliminate

future local emissions as well as previous ones, so that cumulative local pollution is negated, $g_{i,t} = 0$, $\forall t$, and so is local damage, $G_{i,t} = 0$. However, global damage as a term is still not eliminated as long as any nation pollutes and does not similarly clean up after itself, so that $c_i G_t = c_i G_{-i,t}$, where $G_{-t} = \sum_{j \neq i} m_j G_{j,3}$.

3. Lastly, an adaptation technology can fully eliminate local damage as well as global damage, not associated emissions, so that $\theta_i G_{i,\tau+1} = 0$, $c_i G_{\tau+1} = 0$, but only for those nations who have the technology. Unlike carbon capture and storage, it does not remove previously accumulated emissions from the atmosphere, and unlike mitigation technology, it does not reduce future missions. Instead, it eliminates both global and local damages. We can think of it as reducing c_i and θ_i all the way down to zero.

The game can be solved with several interesting modifications.

- Uncertainty about technology maturity, sunk cost and repetitive investment, as discussed above. After the first investment is made, technology may or may not mature in the next period, at which point a nation can ignore the sunk cost and not continue with the investment, or invest another k_i and wait for the next period, when it may again not mature, with some probability.

- Uncertainty in the benefits of new technology, a_i (similar to Zhou et al. (2020)). If spillovers are defined similarly to Zhou et al. (2020) as well – learning the value of the parameter through someone else's experience with it – then the model would be closer to the literature on learning (Ulph and Ulph 1996, Ulph 2004, Kolstad 2007, Kolstad and Ulph 2008). The benefit a_i can also be defined as a function of time to reflect the idea that the more time passes, the more efficient technology becomes.

- Uncertainty about the timing of global damage. The game does not have to finish at $t = 3$. It can go on forever, with countries being hit by global damage differently, at some unknown period in the future, and forever from then on.

- A domestic decision layer can be added in a fashion of Battaglini and Harstad (2020), by introducing a median voter who re-elects a new incumbent in every period. This implies that whoever rules at t is only in power for one period, so a commitment to long term policy is impossible. In a model where the maturity time is uncertain, this modification may prove a serious deterrent to green transition as compared to an authoritarian leader who is in power for the duration of the game.

3.3 Benchmark solution

For the benchmark solution there is no uncertainty other than in the magnitude of the local damages, and even that is resolved before any decision is made.

We start by summarizing the payoffs and relevant notations. Define:

$$
\begin{aligned}
&G_{-i,3} = \sum_{j \neq i} m_j G_{j,3} \\
&M = \sum_i m_i \\
&M_{-i} = \sum_{j \neq i} m_j \\
&G_{i,1} = g_{i,1} \\
&G_{i,2} = q_G g_{i,1} + g_{i,2} \\
&G_{i,3} = q_G^2 g_{i,1} + q_G g_{i,2} + g_{i,3} \\
&G_3 = m_i G_{i,3} + G_{-i,3} = m_i G_{i,3} + \sum_{j \neq i} m_j G_{j,3} \\
&G_{-i,3} = \sum_{j \neq i} m_j G_{j,3} = m_j G_{j,3} + m_k G_{k,3} = \\
&\quad = q_G^2 (m_j g_{j,1} + m_k g_{k,1}) + q_G (m_j g_{j,2} + m_k g_{k,2}) + (m_j g_{j,3} + m_k g_{k,3})
\end{aligned}
$$

The following sets of equations describe the payoffs at different periods, for different types of technology.

3.3.1 Mitigation technology

Define $\pi_{i,3}(1)$ as i's payoff from having invested at $t = 1$, measured at $t = 3$. Similarly, $\pi_{i,3}(2)$ is the payoff at $t = 3$ from having acquired or invested (in) technology at $t = 2$; $\pi_{i,3}(n)$ is the payoff at $t = 3$ from never reducing emissions. Payoffs at t=3 are in general as follows.

$$\pi_{i,3}(n) = B(g_{i,3}) - \theta_i C(G_{i,3}) - c_i G_3 = b_i g_{i,3} - \theta_i G_{i,3} - c_i(m_i G_{i,3} + G_{-i,3}) =$$
$$= b_i g_{i,3} - (\theta_i + c_i m_i) G_{i,3} - c_i G_{-i,3} = b_i g_{i,3} - (\theta_i + c_i m_i)(q_G^2 g_{i,1} + q_G g_{i,2} + g_{i,3}) -$$
$$- c_i G_{-i,3} = (b_i - \theta_i - c_i m_i) g_{i,3} - (\theta_i + c_i m_i)(q_G^2 g_{i,1} + q_G g_{i,2}) - c_i G_{-i,3}$$
$$\pi_{i,3}(2) = A(e_{i,3}) - \theta_i C(G_{i,3}) - c_i G_3 = a_i e_{i,3} - \theta_i G_{i,3} - c_i(m_i G_{i,3} + G_{-i,3}) =$$
$$= a_i e_{i,3} - (\theta_i + c_i m_i) G_{i,3} - c_i G_{-i,3} = a_i e_{i,3} - (\theta_i + c_i m_i)(q_G^2 g_{i,1} + q_G g_{i,2}) - c_i G_{-i,3}$$
$$\pi_{i,3}(1) = A(e_{i,3}) - \theta_i C(G_{i,3}) - c_i G_3 = a_i e_{i,3} - \theta_i G_{i,3} - c_i(m_i G_{i,3} + G_{-i,3}) =$$
$$= a_i e_{i,3} - (\theta_i + c_i m_i) G_{i,3} - c_i G_{-i,3} = a_i e_{i,3} - (\theta_i + c_i m_i) q_G^2 g_{i,1} - c_i G_{-i,3}$$

Eliminating the variables for emissions, $e_{i,t}$ and $g_{i,t}$, and doing so for all the payoffs below, payoffs at $t = 3$ are

$$\pi_{i,3}(1) = a_i - (\theta_i + c_i m_i) q_G^2 - c_i G_{-i,3}$$
$$\pi_{i,3}(2) = a_i - (\theta_i + c_i m_i)(q_G^2 + q_G) - c_i G_{-i,3}$$
$$\pi_{i,3}(n) = (b_i - \theta_i - c_i m_i) - (\theta_i + c_i m_i)(q_G^2 + q_G) - c_i G_{-i,3}$$

Define $\pi_{i,2}(1)$ and $\pi_{i,2}(n)$ in a similar way. Note that $\pi_{i,2}(2, p_{ij})$ and $\pi_{i,2}(2, k_i)$ are the payoffs from purchasing the technology at $t = 2$ and investing at $t = 2$, respectively. Also, we assume that future per-period payoffs are discounted at rate δ, common for everyone. We further eliminate q_G from the model by setting it equal to 1, because its existence does not have a value of its own, and in short time frames, only negligible amount of carbon dioxide is decayed naturally.

Payoffs at t=2:

$$\pi_{i,2}(1) = -[\theta_i + \delta(\theta_i + c_i m_i)] + [a_i] + [\delta a_i] - \delta c_i G_{-i,3} + p_{ji} + p_{ki}$$
$$\pi_{i,2}(2, p_{ij}) = -[\theta_i + \delta(\theta_i + c_i m_i)] + [(b_i - \theta_i) - \delta(\theta_i + c_i m_i)] + [\delta a_i] - \delta c_i G_{-i,3} - p_{ij}$$
$$\pi_{i,2}(2, k_i) = -[\theta_i + \delta(\theta_i + c_i m_i)] + [(b_i - \theta_i) - \delta(\theta_i + c_i m_i)] + [\delta a_i] - \delta c_i G_{-i,3} - k_i$$
$$\pi_{i,2}(n) =$$
$$= -[\theta_i + \delta(\theta_i + c_i m_i)] + [(b_i - \theta_i) - \delta(\theta_i + c_i m_i)] + \delta[b_i - \theta_i - c_i m_i] - \delta c_i G_{-i,3}$$

Colour-coded to stand out, in the first equation we have the possibility of selling the technology to both j and k if none of them has invested at $t = 1$. These terms however may not be part of the payoff if both or even one other nation has also invested in technology.

Lastly, payoffs at t=1 are

$$\pi_{i,1}(1) = [b_i - \theta_i - \delta\theta_i - \delta^2(\theta_i + c_i m_i)] + [\delta a_i] +$$
$$+ [\delta^2 a_i] - \delta^2 c_i G_{-i,3} - k_i + \delta(p_{ji} + p_{ki})$$
$$\pi_{i,1}(2) = [b_i - \theta_i - \delta\theta_i - \delta^2(\theta_i + c_i m_i)] + [\delta(b_i - \theta_i) - \delta^2(\theta_i + c_i m_i)] +$$
$$+ [\delta^2 a_i] - \delta^2 c_i G_{-i,3} - \delta p_{ij}$$
$$\pi_{i,1}(2) = [b_i - \theta_i - \delta\theta_i - \delta^2(\theta_i + c_i m_i)] + [\delta(b_i - \theta_i) - \delta^2(\theta_i + c_i m_i)] +$$
$$+ [\delta^2 a_i] - \delta^2 c_i G_{-i,3} - \delta k_i$$
$$\pi_{i,1}(n) = [b_i - \theta_i - \delta\theta_i - \delta^2(\theta_i - c_i m_i)] + [\delta(b_i - \theta_i) - \delta^2(\theta_i + c_i m_i)] +$$
$$+ [\delta^2(b_i - \theta_i - c_i)] - \delta^2 c_i G_{-i,3}$$

Solving the game by backward induction, we first look at the decision at $t = 2$. If nation i acquires or invests at $t = 2$, then it cannot pass the technology down to any other nation. This is due to the game only take place across three periods and may be eliminating an interesting scenario to model where nation j develops the technology at $t = 1$, nation i buys it at $t = 2$ and sells it to nation k at $t = 3$. This bypasses intellectual property rights law and allows country i to capture the benefits that it has no right for. Country j would therefore be partially discouraged to invest. This case is however beyond the scope of a three-period model.

Lemma 3.1. *If $p_{ij} \leq \delta(a_i - (b_i - \theta_i - c_i m_i))$, then i would prefer to acquire the technology from j at $t = 2$ to never having done so; similarly, if $k_i \leq \delta(a_i - (b_i - \theta_i - c_i m_i))$, then i would prefer to invest in the technology at $t = 2$. If, additionally, $p_{ij} < k_i$, then acquiring the technology from j at $t = 2$ is preferred to developing it, which is preferred to never owning it.*

Proof. Compare $\delta a_i + b_i - \theta_i - \delta(\theta_i + c_i m_i)(q_G + q_G^2) - \theta_i q_G - \delta c_i G_{-i,3} - p_{ij}$ and $\delta(b_i - \theta_i - c_i m_i) + b_i - \theta_i - \delta(\theta_i + c_i m_i)(q_G + q_G^2) - \theta_i q_G - \delta c_i G_{-i,3}$. Note that i's decision to purchase the tech does not affect other nations' decisions about technology, so we can assume their strategies are the same in these two scenarios of i's behaviour, thus $\delta c_i G_{-i,3}$ is the same in both expressions. The comparison further reduces to $\delta a_i - p_{ij}$ and $\delta(b_i - \theta_i - c_i m_i)$. ■

Lemma 3.1 suggests that j can always induce i to purchase the technology by charging only a small price. This would make j better off if she would choose to invest in the absence of trade. However, if j's desire to invest relied mostly on the possibility of selling it later, then she may choose not to invest if she cannot charge a satisfactory price.

Let us now look at the variation of the first period if no trade was allowed.

Lemma 3.2. *Assuming no spillovers, nation i would invest in technology at $t = 1$ if and only if $k_i \leq \frac{\delta(a_i - (b_i - \theta_i)) + \delta^2 q_G(\theta_i + c_i m_i)}{1-\delta}$. If the condition does not hold, then it would invest at $t = 2$ if and only if $k_i \leq \delta(a_i - (b_i - \theta_i) + c_i m_i)$. It would never invest otherwise.*

Proof. To establish the result, it suffices to compare the option of developing technology at $t = 1$ and at $t = 2$ for any player, ignoring the total emissions, since they are the same in both cases. The technology will be invested in at $t = 1$ if

$$[b_i - \theta_i - \delta\theta_i q_G - \delta^2 q_G^2(\theta_i + c_i m_i)] + [\delta a_i] + [\delta^2 a_i] - \delta^2 c_i G_{-i,3} - k_i \ \geq$$
$$\geq \ [b_i - \theta_i - \delta\theta_i q_G - \delta^2 q_G^2(\theta_i + c_i m_i)] + [\delta(b_i - \theta_i) - \delta^2 q_G(\theta_i + c_i m_i)] + [\delta^2 a_i] - \delta^2 c_i G_{-i,3} - \delta k_i$$

The first and the third period emissions are identical, so reduce the above to

$$[\delta a_i] - k_i \ \geq \ [\delta(b_i - \theta_i) - \delta^2 q_G(\theta_i + c_i m_i)] - \delta k_i.$$

Technology would be developed at $t = 1$ if and only if

$$k_i \leq \frac{\delta(a_i - (b_i - \theta_i)) + \delta^2 q_G(\theta_i + c_i m_i)}{1-\delta} = \frac{\delta(a_i - (b_i - \theta_i))}{1-\delta} + \frac{\delta^2 q_G(\theta_i + c_i m_i)}{1-\delta}$$

Note that it would be developed at $t = 2$ rather than never (provided it is not at $t = 1$) if $k_i \leq \delta(a_i - (b_i - \theta_i) + c_i m_i)$.

Can it be said with certainty that investing earlier is preferred?

$$\frac{\delta(a_i - (b_i - \theta_i))}{1-\delta} + \frac{\delta^2 q_G(\theta_i + c_i m_i)}{1-\delta} \ \vee \ \delta(a_i - (b_i - \theta_i)) + \delta c_i m_i$$
$$\frac{\delta(a_i - (b_i - \theta_i))}{1-\delta} - \frac{\delta(a_i - (b_i - \theta_i))(1-\delta)}{1-\delta} + \frac{\delta^2 q_G(\theta_i + c_i m_i)}{1-\delta} \ \vee \ \delta c_i m_i$$
$$\frac{\delta(a_i - (b_i - \theta_i)) - \delta(a_i - (b_i - \theta_i)) + \delta^2(a_i - (b_i - \theta_i))}{1-\delta} + \frac{\delta^2 q_G(\theta_i + c_i m_i)}{1-\delta} \ \vee \ \delta c_i m_i$$
$$\delta^2(a_i - (b_i - \theta_i) + q_G(\theta_i + c_i m_i)) \ \vee \ (1-\delta)\delta c_i m_i$$
$$\delta^2(a_i - b_i + \theta_i + q_G\theta_i) \ \vee \ \delta c_i m_i - \delta^2 c_i m_i - \delta^2 q_G c_i m_i$$
$$\delta^2(a_i - b_i + \theta_i + q_G\theta_i) \ \vee \ c_i m_i \delta(1 - \delta - \delta q_G)$$
$$\delta(a_i - b_i + \theta_i + q_G\theta_i) \ \vee \ c_i m_i(1 - \delta - \delta q_G)$$

The expressions cannot be concluded with certainty; all depends on the parameters. Consider some static cases. If δ is large enough (approaching 1), and $q_G = 1$, then the right-hand side of the expression is negative, and as long as the left-hand side is positive, it is always larger. This means that we can either have the technology developed at $t = 1$ but not at $t = 2$, or in both, but never at $t = 2$ and not at $t = 1$. Importantly, the left-hand side positively depends on θ_i, implying that larger damage increases the likelihood that the technology will be invested in at $t = 1$. In the extreme case of no local damage, $\theta_i = 0$, the left-hand side is also negative, and thus it is possible that investing later, at $t = 2$, is preferred to investing earlier, at $t = 1$. Alternatively, if future does not matter much, $\delta \to 0$, then the right-hand side can be larger than the left-hand side, and we can either have investment at both periods, or at $t = 2$ only.
A complete strategy is written below where "I" stands for "invest", "NI" – for "not invest", and they correspond to $t = 1$ and $t = 2$ respectively.

$$\begin{aligned}
\{\text{I,I}\} \text{ if } k_i &\leq \frac{\delta(a_i - (b_i - \theta_i)) + \delta^2 q_G(\theta_i + c_i m_i)}{1-\delta} \leq \delta(a_i - (b_i - \theta_i) + c_i m_i) \\
\text{or } k_i &\leq \delta(a_i - (b_i - \theta_i) + c_i m_i) \leq \frac{\delta(a_i - (b_i - \theta_i)) + \delta^2 q_G(\theta_i + c_i m_i)}{1-\delta}; \\
\{\text{I,NI}\} \text{ if } \delta(a_i - (b_i - \theta_i) + c_i m_i) &\leq k_i \leq \frac{\delta(a_i - (b_i - \theta_i)) + \delta^2 q_G(\theta_i + c_i m_i)}{1-\delta} \\
\{\text{NI,I}\} \text{ if } \frac{\delta(a_i - (b_i - \theta_i)) + \delta^2 q_G(\theta_i + c_i m_i)}{1-\delta} &\leq k_i \leq \delta(a_i - (b_i - \theta_i) + c_i m_i) \\
\{\text{NI,NI}\} \text{ if } k_i &> \frac{\delta(a_i - (b_i - \theta_i)) + \delta^2 q_G(\theta_i + c_i m_i)}{1-\delta} \geq \delta(a_i - (b_i - \theta_i) + c_i m_i) \\
\text{or } k_i &> \delta(a_i - (b_i - \theta_i) + c_i m_i) \geq \frac{\delta(a_i - (b_i - \theta_i)) + \delta^2 q_G(\theta_i + c_i m_i)}{1-\delta}
\end{aligned}$$

■

Without spillovers, a rational nation's behaviour is ambiguous and strongly depends on the parameters of the payoff function. Adding spillovers changes the dynamic because the first period borderline value increases due to the price paid in the second period or stays the same if no technology transfer occurs, but it never decreases. The parameters can be limited to exclude the case when there is investment at $t = 2$ but not at $t = 1$, but we do not believe it is a problematic conclusion. We can observe countries make pledges to be carbon-neutral by 2050 but not performing any meaningful actions at present to meet the pledge. With a small δ, this scenario can arise in the model. Even with a large δ, when the right-hand side is negative, with a very inefficient nascent green technology and an extremely efficient brown technology, $a_i - b_i << 0$, and small local damage, $\theta_i \longrightarrow 0$, we can still observe the cost of investment being too large to invest at $t = 1$, but not so for $t = 2$.

Proposition 3.1. *An equilibrium strategy at $t = 1$ for i is as follows.*

1. *Invest if* $k_i \leq \dfrac{\delta(a_i - (b_i - \theta_i)) + \delta^2 q_G(\theta_i + c_i m_i)}{1-\delta}$, *if nations j and k decide to invest at $t = 1$,*

2. *If nation j decides to invest at $t = 1$, while nation k does not,*

   2.1. Invest if $k_i \leq \dfrac{\delta(a_i - (b_i - \theta_i)) + \delta^2 q_G(\theta_i + c_i m_i)}{1-\delta}$, *if nation k would invest at $t = 2$,*

   2.2. Invest if $k_i \leq \dfrac{\delta(a_i - (b_i - \theta_i)) + \delta^2 q_G(\theta_i + c_i m_i) + \delta^2 c_i m_k}{1-\delta}$, *if nation k would not invest at $t = 2$,*

3. *If neither nation would invest at $t = 1$;*

   3.1. Invest if $k_i \leq \dfrac{\delta a_i + \delta(p_{ji} + p_{ki}) - \delta(b_i - \theta_i) + \delta^2 q_G(\theta_i + c_i m_i) + \delta^2 c_i m_k}{1-\delta}$, *if nation j would invest at $t = 2$, or*
   Invest if $k_i \leq \dfrac{\delta a_i + \delta(p_{ji} + p_{ki}) - \delta(b_i - \theta_i) + \delta^2 q_G(\theta_i + c_i m_i)}{1-\delta}$, *if both nations would invest at $t = 2$,*

   3.2. Invest if $k_i \leq \dfrac{\delta a_i + \delta(p_{ji} + p_{ki}) - \delta(b_i - \theta_i) + \delta^2 q_G(\theta_i + c_i m_i)}{1-\delta} + \dfrac{\delta^2 c_i(m_j + m_k)}{1-\delta}$, *if no one other nation would invest at $t = 2$.*

Proof. We begin by recognising that nations are symmetric in their payoff functions, but with different limits on k_i, and where one nation may find investment worthwhile, the other may not, so there are different action choices generated by the model.

Compare i's payoffs from investing at $t = 1$, $\pi_{i,1}(1)$, and at $t = 2$, $\pi_{i,2}(2)$.

$$[b_i - \theta_i - \delta\theta_i q_G - \delta^2 q_G^2(\theta_i + c_i m_i)] + [\delta a_i] + [\delta^2 a_i] - \delta^2 c_i G_{-i,3} - k_i + \delta(p_{ji} + p_{ki}) \vee$$
$$[b_i - \theta_i - \delta\theta_i q_G - \delta^2 q_G^2(\theta_i + c_i m_i)] + [\delta(b_i - \theta_i) - \delta^2 q_G(\theta_i + c_i m_i)] +$$
$$+ [\delta^2 a_i] - \delta^2 c_i G_{-i,3} - \delta k_i$$

Reduce the expression above to

$$\delta a_i - k_i(1-\delta) + \delta(p_{ji} + p_{ki}) - \delta^2 c_i G_{-i,3} \ \vee\ \delta(b_i - \theta_i) - \delta^2 q_G(\theta_i + c_i m_i) - \delta^2 c_i G_{-i,3}$$

Consider nation i that is contemplating the development of technology at $t = 1$. To make a decision, it is necessary to consider the strategies of other players. There are several scenarios to consider.

1. If both j and k develop their own technology at $t = 1$, then the global damage term is the same on LHS and RHS and equal to $G_{-i,3} = q_G^2(m_j g_{j,1} + m_k g_{k,1})$, thus the expression is further reduced to

$$\delta a_i - k_i(1-\delta) \ \vee \ \delta(b_i - \theta_i) - \delta^2 q_G(\theta_i + c_i m_i)$$

Therefore, i would prefer to invest in her own technology if and only if

$$k_i \le \frac{\delta(a_i - (b_i - \theta_i)) + \delta^2 q_G(\theta_i + c_i m_i)}{1-\delta} \tag{3.1}$$

2. If only j decides to invest but k does not, then at $t = 2$, k either finds it optimal to invest or not. If she does, then the global damage on the left and right-hand sides are once again the same and equal to $q_G^2(m_j g_{j,1} + m_k g_{k,1}) + q_G m_k g_{k,2}$, and i arrives at the following comparison:

$$\delta a_i - k_i(1-\delta) + \delta p_{ki} \ \vee \ \delta(b_i - \theta_i) - \delta^2 q_G(\theta_i + c_i m_i)$$

The only effect i's decision will have is if she sells her technology to k at $t = 2$. But same is true about j, so here, technological competition can be observed. According to Bertrand's model, the price would be driven to zero since the marginal cost of the technology is zero. This is where a treaty would be useful. It could set the price floor to protect investing nations from having to give the technology away for free. Nation k would be indifferent between purchasing technology from i or j, and i would expect to get $\delta\frac{p}{2}$. This term would enter the left-hand side only if $p \le k_k[\le \delta(a_k - (b_k - \theta_k - c_k m_k))]$, as a condition for k to prefer to buy rather than to invest [and to invest at $t = 2$ rather than not]. Without a treaty, the term does not matter, but the full condition on k_i in that case is written below.

$$k_i \le \frac{\delta(a_i - (b_i - \theta_i)) + \delta^2 q_G(\theta_i + c_i m_i) + \delta p_{ki}}{1-\delta} \tag{3.2}$$

Equation 3.2 is identical to Equation 3.1, because the price is driven to 0 by the competition between i and j in the absence of treaty, and k would invest at $t = 2$ anyway, so the condition of k_i is

$$k_i \le \frac{\delta(a_i - (b_i - \theta_i)) + \delta^2 q_G(\theta_i + c_i m_i)}{1-\delta} \tag{3.3}$$

If, on the other hand, k would not invest at $t = 2$, $k_k > \delta(a_k - (b_k - \theta_k - c_k m_k))$, then i can improve its payoff not only through selling the technology, but also by reducing global damage from $[q_G^2(m_j g_{j,1} + m_k g_{k,1}) + q_G m_k g_{k,2} + m_k g_{k,3}]$ to $[q_G^2(m_j g_{j,1} + m_k g_{k,1}) + q_G m_k g_{k,2}]$, if k acquires the technology at $t = 2$ and cuts its emissions at $t = 3$ as a result. In this case, i would compare

$$\begin{aligned} &\delta a_i - k_i(1-\delta) + \delta p_{ki} - \delta^2 c_i[q_G^2(m_j g_{j,1} + m_k g_{k,1}) + q_G m_k g_{k,2}] \\ \vee\ &\delta(b_i - \theta_i) - \delta^2 q_G(\theta_i + c_i m_i) - \delta^2 c_i[q_G^2(m_j g_{j,1} + m_k g_{k,1}) + q_G m_k g_{k,2} + m_k] \end{aligned} \tag{3.4}$$

$$\delta a_i - k_i(1-\delta) + \delta p_{ki} \vee\ \delta(b_i - \theta_i) - \delta^2 q_G(\theta_i + c_i m_i) - \delta^2 c_i m_k$$

and decide to invest if

$$k_i \leq \frac{\delta(a_i - (b_i - \theta_i)) + \delta^2 q_G(\theta_i + c_i m_i) + \delta p_{ki} + \delta^2 c_i m_k}{1-\delta}$$

Again, the payoff depends on how k decides whether to purchase the technology from i or j. The expression above is true if k buys from i. If k buys from j, then i's action is irrelevant for $G_{-i,3}$, and i is back to the expression below: invest at $t = 1$ if

$$k_i \leq \frac{\delta(a_i - (b_i - \theta_i)) + \delta^2 q_G(\theta_i + c_i m_i)}{1-\delta} \tag{3.5}$$

Generally, in equilibrium, as discussed before, the price would be driven to 0 and k would be indifferent who to buy from, as long as $\delta(a_i - (b_i - \theta_i - c_i m_i)) \geq 0$. If that's the case, the assumption about k's behaviour is irrelevant, i.e., k buys from i or from j when the price is the same. We can also assume that k randomizes by flipping a coin probability and say that i would expect with probability 0.5 that k buys from her and 0.5 that she buys from j. Then i's payoff consists of two equally likely terms if she invests:

$$\begin{aligned} &0.5[\delta a_i - k_i(1-\delta) + \delta \cdot 0 - \delta^2 c_i(q_G^2(m_j g_{j,1} + m_k g_{k,1}) + q_G m_k g_{k,2})] + \\ &+ 0.5[\delta a_i - k_i(1-\delta) - \delta^2 c_i(q_G^2(m_j g_{j,1} + m_k g_{k,1}) + q_G m_k g_{k,2})] = \\ &= \delta a_i - k_i(1-\delta) - \delta^2 c_i(q_G^2(m_j g_{j,1} + m_k g_{k,1}) + q_G m_k g_{k,2}) \end{aligned}$$

This is the same as the left-hand side in the Equation 3.4 with price set equal to 0. Thus, the expression for k_i is the same: invest at $t = 1$ if

$$k_i \leq \frac{\delta(a_i - (b_i - \theta_i)) + \delta^2 q_G(\theta_i + c_i m_i) + \delta^2 c_i m_k}{1-\delta} \tag{3.6}$$

This concludes the scenario of what if both i and j had invested at $t = 1$, while k had not and would not invest at $t = 2$ either, $k_k > \delta(a_k - (b_k - \theta_k - c_k m_k))$. It is evident that i's choice would be independent of others' decisions as long as others find it optimal to invest for themselves, with the only addition being the price of selling the technology, and even that disappears once there is more than one nation with the technology at $t = 2$.

3. If investing at $t = 1$ is not worth it for neither j nor k, then there are two cases to consider.

Investing at $t = 2$ is also not optimal, for one or both nations, i.e., $k_k > \delta(a_k - (b_k - \theta_k - c_k m_k))$ and/or $k_j > \delta(a_j - (b_j - \theta_j - c_j m_j))$. This is where i's decision is most affected by the externality-containing term and the price of the technology. Being the only country with the technology, it can choose the price for selling it and reap all of the benefits of doing so, while also reducing the externality term by $(m_j g_{j,3} + m_k g_{k,3})$, which enters the payoff as $-\delta^2 c_i(m_j g_{j,3} + m_k g_{k,3})$.

$$\begin{aligned}\delta a_i - k_i(1-\delta) + \delta(p_{ji} + p_{ki}) - \delta^2 c_i(q_G^2(m_j g_{j,1} + m_k g_{k,1}) + q_G(m_j g_{j,2} + m_k g_{k,2})) \vee \\ \vee\ \delta(b_i - \theta_i) - \delta^2 q_G(\theta_i + c_i m_i) - \delta^2 c_i(q_G^2(m_j g_{j,1} + m_k g_{k,1}) \\ + q_G(m_j g_{j,2} + m_k g_{k,2}) + (m_j g_{j,3} + m_k g_{k,3}))\end{aligned}$$

$$\delta a_i - k_i(1-\delta) + \delta(p_{ji} + p_{ki})\ \vee\ \delta(b_i - \theta_i) - \delta^2 q_G(\theta_i + c_i m_i) - \delta^2 c_i(m_j + m_k)$$

It appears that nation i would therefore invest at $t = 1$ if

$$k_i \leq \frac{\delta a_i + \delta(p_{ji} + p_{ki}) - \delta(b_i - \theta_i) + \delta^2 q_G(\theta_i + c_i m_i) + \delta^2 c_i(m_j + m_k)}{1-\delta} \tag{3.7}$$

Note that Equation 3.7 is true for when neither j nor k would invest at $t = 2$. But would they agree to purchase the technology? As long as $p_{ji} < \delta(a_j - b_j + \theta_j + c_j m_j)$, and since p_{ji} is a decision variable of i, as long as the right-hand side is positive, such price can be found.

If only one of them would invest at $t = 2$ (say j), and $p_{ji} < k_j$, then it would be reduced to

$$k_i \leq \frac{\delta a_i + \delta(p_{ji} + p_{ki}) - \delta(b_i - \theta_i) + \delta^2 q_G(\theta_i + c_i m_i) + \delta^2 c_i m_k}{1-\delta} \tag{3.8}$$

If both would invest at $t = 2$, and $p_{ji} < k_j$ and $p_{ki} < k_k$

$$k_i \leq \frac{\delta a_i + \delta(p_{ji} + p_{ki}) - \delta(b_i - \theta_i) + \delta^2 q_G(\theta_i + c_i m_i)}{1 - \delta} \tag{3.9}$$

The condition on the price is chosen by i strategically to maximize her payoff: it is the maximum price which is below investment cost, i.e., $p_{ji} = k_i - \varepsilon$, with $\varepsilon \to 0$.

The discussion above can be summarised as follows.

1. both j and k invest at $t = 1$; condition in Equation 3.1
2. j invests at $t = 1$, k does not;
   - 2.1. k invests at $t = 2$; condition in Equation 3.3
   - 2.2. k does not invest at $t = 2$; condition in Equation 3.6
3. neither invest at $t = 1$;
   - 3.1. either one of them or both invest at $t = 2$; condition in Equation 3.8 and Equation 3.9
   - 3.2. no one invests at $t = 2$; condition in Equation 3.7

■

Nation i's willingness to invest when allowed to trade increases compared to the critical level when no technology trade is introduce, but only in the case of no other nation investing at $t = 1$. We can arrange the critical values in the following order:

$$\begin{aligned} &\frac{\delta(a_i - (b_i - \theta_i)) + \delta^2 q_G(\theta_i + c_i m_i)}{1 - \delta} < \\ &\quad < \frac{\delta a_i - \delta(b_i - \theta_i) + \delta^2 q_G(\theta_i + c_i m_i) + \delta(p_{ji} + p_{ki})}{1 - \delta} < \\ &\quad < \frac{\delta a_i - \delta(b_i - \theta_i) + \delta^2 q_G(\theta_i + c_i m_i) + \delta(p_{ji} + p_{ki}) + \delta^2 c_i(m_j + m_k)}{1 - \delta} \end{aligned}$$

In such a simple linear model, the second period is easily solved because i's decision is only affected by i's own choices, and no spillovers can result from i's decision at $t = 2$. There, the threshold value for k_i, call it $\bar{k}_2$, is well and easily defined. However, this is not so for the first period. What is even more interesting, at $t = 1$, the threshold value $\bar{k}_1$ does not have a clear relationship with $\bar{k}_2$. Depending on the parameters, one can be larger than the other, so we can observe investment at $t = 2$ but not at $t = 1$, even though one may think that if investment takes place at all, it should be done as early as possible, because by switching to low-carbon earlier, a nation can enjoy the benefits of no pollution for longer. This result is due to defining a new technology as inefficient with $a_i < b_i$, having a weight in the global damage at $t = 3$ (larger country reduces the damage marginally more than a smaller country, but the size does not matter for personal pollution, so large own damage does not translate into large global damage if a country is small and the decision to remedy it would be a different decision than the one to fixing global damage), and local damage for how much a country is already suffering.

3.4 Comparative statics

To discuss the results of the Proposition 3.1 in more details, below we provide some interesting comparative statics.

Note that we will assume $q_G = 1$ from now on as it does not bear significance for the results.

3.4.1 No technological trade

Define the threshold for $t = 2$ as $\overline{k_{i,2}} = \delta(a_i - b_i + \theta_i + c_i m_i)$. The threshold for $t = 1$ depends on the timing of the investment decision, and is larger with spillovers, but we first consider the one which corresponds to no spillovers, $\overline{k_{i,1}} = \dfrac{\delta(a_i - b_i + \theta_i + \delta(\theta_i + c_i m_i))}{1-\delta}$. There is always global damage, $c_i \neq 0$ and $a_i < b_i$.

1. Small country and no local damage: $m_i \longrightarrow 0$ and $\theta_i \longrightarrow 0$ would never invest.

   $$\overline{k_{i,2}} = \delta(a_i - b_i)$$
   $$\overline{k_{i,1}} = \frac{\delta(a_i - b_i)}{1-\delta}$$

   Both values are negative for any δ, so unless the technology is subsidised, investment is never optimal. A small country who is not suffering from damage in the present has nothing to gain in terms of global damage.

2. Large country with no local damage: $m_i > 0$ and $\theta_i \longrightarrow 0$. It would never invest for $c_i m_i < b_i - a_i$; invest at $t = 2$ only if $\delta < \delta^*$ and $\overline{k_{i,1}} < k_i, < \overline{k_{i,2}}$; invest at $t = 1$ if $\delta > \delta^*$ and $k_i < \overline{k_{i,1}}$, and some other cases that are not of

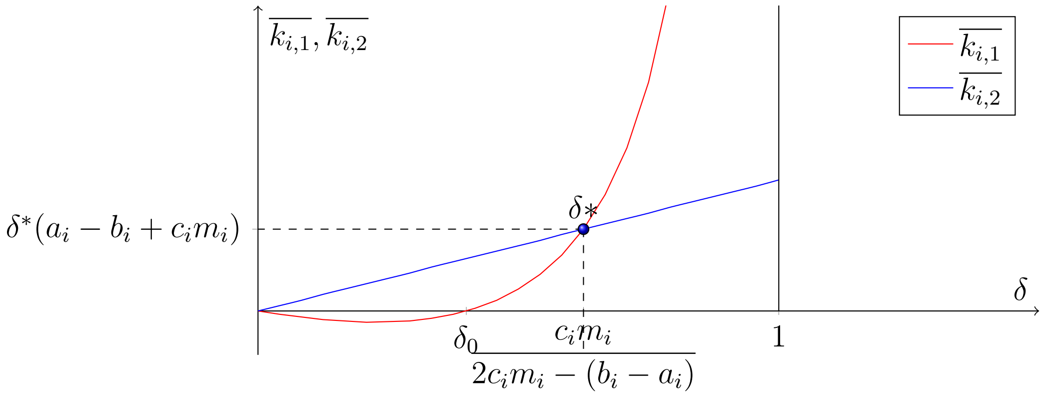

FIGURE 3.1: For $\delta < \delta^*$, there exist such values of k_i, that $\overline{k_{i,1}} < k_i < \overline{k_{i,2}}$, implying that for a range of investment cost, only investing at $t = 2$ is attractive, not at $t = 1$. For large δ, however, investing earlier is preferred.

particular interest.

$$\overline{k_{i,2}} = \delta(a_i - b_i + c_i m_i)$$
$$\overline{k_{i,1}} = \frac{\delta(a_i - b_i + \delta c_i m_i)}{1 - \delta}$$

The first expression is a linear function of δ which ranges from 0 to $(a_i - b_i + c_i m_i)$, and depending on the relation between $c_i m_i$ and $b_i - a_i$, it slopes positively or negatively. But so does $\overline{k_{i,1}}$, with the only difference being in its shape. See Figure 3.1 for $c_i m_i > b_i - a_i$. When this does not hold, both thresholds are negative and i never invests. The critical value of the discount factor that changes the relation between investment thresholds is

$$\delta^* = \frac{c_i m_i}{2c_i m_i - (b_i - a_i)}$$

For $\delta < \delta^*$, there exist such values of k_i, that $\overline{k_{i,1}} < k_i < \overline{k_{i,2}}$, implying that for a range of investment cost, only investing at $t = 2$ is attractive, not at $t = 1$. For large δ, however, investing earlier is preferred. Unlike a small country, a large country can appreciably affect global damage, and even if it does not suffer from local pollution, it can still choose to invest, if the future damage is significant and the discount factor is small.

3. Small country with local damage: $m_i \longrightarrow 0$ and $\theta_i > 0$. It would never invest if $b_i - a_i > 2\theta_i$; invest at $t = 1$ if $k_i < \overline{k_{i,1}}$ and $b_i - a_i < \theta_i$, or if $\delta > \dfrac{b_i - \theta_i - a_i}{\theta_i}$.

In this case, i would never prefer investment at $t = 2$ over investment at $t = 1$.

$$\overline{k_{i,2}} = \delta(a_i - b_i + \theta_i)$$
$$\overline{k_{i,1}} = \frac{\delta(a_i - b_i + \theta_i + \delta\theta_i)}{1 - \delta}$$

Note that both can slope upward or downward without restriction. If both are downward sloping, which happens when $b_i - a_i > 2\theta_i$, i would never invest. If both are upward sloping, $b_i - a_i < \theta_i$, the threshold for $t = 1$ trumps the threshold for $t = 2$, and if investment takes place because k_i is low enough, it necessarily takes place at $t = 1$: $k_i < \overline{k_{i,1}}$. Finally, when the thresholds slope in different directions, $2\theta_i > b_i - a_i > \theta_i$, $\overline{k_{i,2}}$ is always negative, while $\overline{k_{i,1}}$ is negative until δ reaches $\delta_0 = \dfrac{b_i - \theta_i - a_i}{\theta_i} > 0$, and only for $\delta > \delta_0$ can we have $k_i < \overline{k_{i,1}}$, so that investment would happen at $t = 1$ or never.

4. Large country with local damage: $m_i > 0$ and $\theta_i > 0$. This case is not significantly different from the one before, but it involves many subcases. New thresholds are:

$$\overline{k_{i,2}} = \delta(a_i - b_i + \theta_i + c_i m_i)$$
$$\overline{k_{i,1}} = \frac{\delta((a_i - b_i + \theta_i) + \delta(\theta_i + c_i m_i))}{1 - \delta}$$

Similar reasoning tells us that the points where different threshold values cross each other and where $\overline{k_{i,1}}$ crosses the zero line are:

$$\delta^* = \frac{c_i m_i}{2(c_i m_i + \theta_i) - (b_i - a_i)}$$
$$\delta_0 = \frac{b_i - a_i - \theta_i}{c_i m_i + \theta_i}$$

A lot is defined by $a_i - b_i + \theta_i$ versus 0.

(a) If $a_i - b_i + \theta_i > 0$, then up to δ^*, $\overline{k_{i,2}}$ trumps $\overline{k_{i,1}}$, and should k_i happen to fall between the thresholds, investment would be made at $t = 2$. For $\delta > \delta^*$, should investment be made, it would be made at $t = 1$ (for $k_i < \overline{k_{i,1}}$). Refer Figure 3.2.

(b) If $a_i - b_i + \theta_i < 0$, and $c_i m_i + 2\theta_i > b_i - a_i$ (both conditions merge into $c_i m_i + 2\theta_i > b_i - a_i > \theta_i$), $\overline{k_{i,2}}$ can slope both ways.

i. If it slopes upwards, $c_i m_i + \theta_i > b_i - a_i$, the scenario is the same as in Figure 3.1.

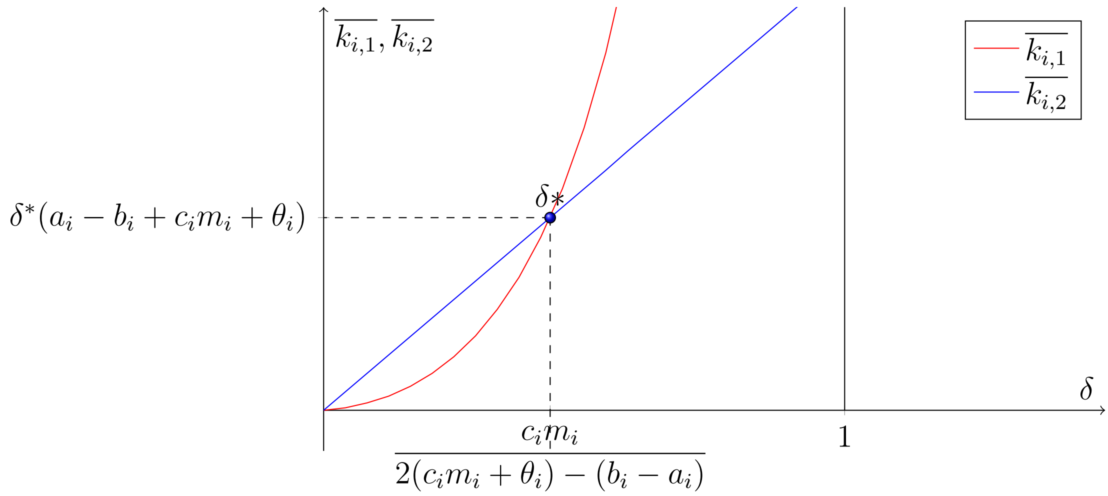

FIGURE 3.2: For $\delta < \delta^*$, there exist such values of k_i, that $\overline{k_{i,1}} < k_i < \overline{k_{i,2}}$, implying that for a range of investment cost, only investing at $t = 2$ is attractive, not at $t = 1$. For δ larger than δ^*, however, investing at $t = 1$ is preferred.

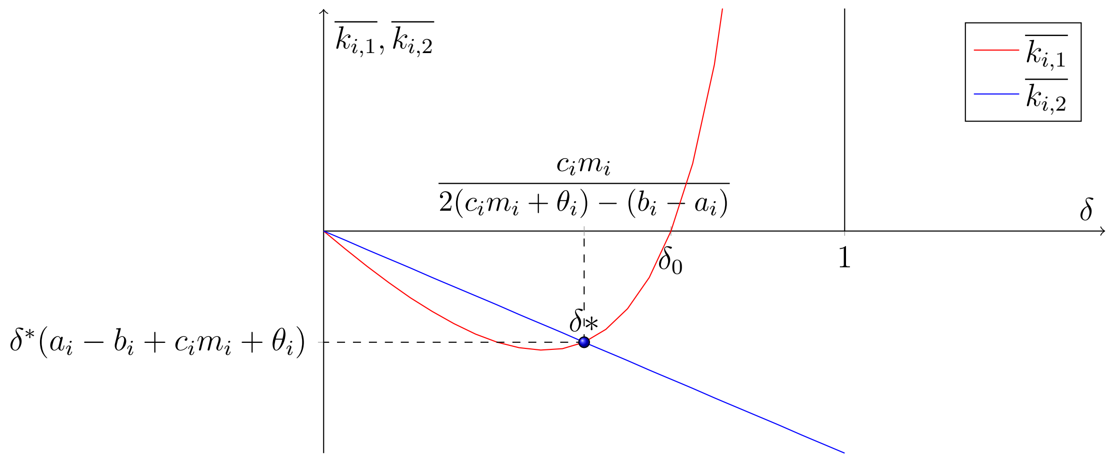

FIGURE 3.3: For $\delta < \delta_0$, investment will never happen. For the discount factor above δ_0, investment may happen at $t = 1$ if $k_i < \overline{k_{i,1}}$.

ii. If it slopes downwards, $c_i m_i + \theta_i < b_i - a_i$, see Figure 3.3. Obviously, investment will not take place at $t = 2$, but it might at $t = 1$, provided that $\delta > \delta_0 = \frac{b_i - a_i - \theta_i}{c_i m_i + \theta_i}$ and $k_i < \overline{k_{i,1}}$.

(c) If $a_i - b_i + \theta_i < 0$, and $c_i m_i + 2\theta_i < b_i - a_i$ (both conditions merge into $b_i - a_i > c_i m_i + 2\theta_i > \theta_i$), $\overline{k_{i,1}}$ slopes downwards and is always negative. These conditions imply that $b_i - a_i > c_i m_i + \theta_i$, so $\overline{k_{i,2}}$ also slopes downwards and is always below zero. There will be no investment in this case.

As can be seen from the scenarios above, for large discount factor, a nation will invest early on. This is consistent with standard folk theorems. For a discount factor close to one, any nation will invest at $t = 1$ with certainty. However, in some instances and under certain parameter specifications, a nation will either prefer to wait and invest at $t = 2$, or to never invest at all. A small country which is not suffering from immediate local pollution will never invest; a small country who does suffer will either invest early or never; but a large country (who does or does not suffer at present) may do either depending on the cost of investment and other parameters: it may never invest, it may wait and invest at $t = 2$, or it may invest early on, at $t = 1$.

The conclusion that a small country would never invest if the private present damage is not too large is consistent with the results of other studies. For instance, in a stage game with a threshold that marks a climate catastrophe, non-pivotal regions are more likely to free-ride (Emmerling et al. 2021); similarly, in a model with a possibility of carbon leakage, an environmental policy must be implemented by a larger region to be globally effective (Sanna-Randaccio et al. 2017) (and technology transfer must take place). It is however inconsistent with real world observations of countries like Sweden, Norway, Denmark, Switzerland and other small developed nations who are leading the world's decarbonization. In a model where technological trade is not allowed, this pattern of investment cannot be observed. Due to the simplistic nature of the model, the only concern nations have when contemplating green investments is avoided environmental damage in the absence of spillovers, and a small nation does not have enough environmental damage to appropriate if it stops polluting. The model does not account for such potentially decision-affecting factors as public preferences, wealth accumulation, or historical responsibility. Having incorporated these would have resulted in qualitatively different predictions.

On the other hand, it is conceivable that these nations are seeing the consequences of changing climate already at present, and are therefore classed as small countries with local damage in the terminology of the model. Climate change is becoming obvious to more and more people every day, and a nation with a high level of public consciousness may start to link observed environmental changes to global warming, even if at present they are not particularly damaging. For northern regions like Sweden and Norway, it is a common notion that climate change may have a short-term productivity-enhancing effect due to longer growing season and more land becoming arable as permafrost retreats. But if a median voter is environmentally conscious to realise that these benefits are short-lived and only serve as an indication of serious environmental changes, they are more likely to be viewed as negative. Reinterpreting present local damage in the model as *perceived* local damage can bridge the gap between the model and reality in explaining why countries like Norway, Sweden, Denmark are heavily investing in climate change mitigation and adaptation – they are perceiving environmental changes as damages already at present.

A note on discounting. It is a fairly standard approach to measuring the cost of climate change in continuous time, to use exponential discounting where δ comes as a power on the base of natural logarithm. A baseline for many other studies, Stern et al. (2006), uses a near-zero (utility) discount rate of 0.1%, which is interpreted as a probability of not surviving a catastrophic event in a given year, so that survival until year 2100 is very likely, with a probability of 90.5%. Other studies argue for a larger discount rate: Nordhaus (2007) employs a discount rate of 4.3% in his research; [3] Grijalva et al. (2014) empirically measure a (constant) discount rate of 4.9%, but note that an altogether different discounting model, one offered by Loewenstein and Prelec (1992), with a hyperbolic discounting and a decreasing discount rate, offers a better fit for their data. The 100-years-from-now discounting rate from a specification suggested by Loewenstein and Prelec (1992) approaches the one used by Stern et al. (2006), but this is not a flat discount rate, and hence policy conclusions would be different in the two papers.

Exponential discounting in discrete time is used by Harstad in all of his studies (2012, 2016, 2016, 2019). Compared to continuous time models, larger discount factor means more weight is allocated to future consumption. This is a notational difference to keep in mind when saying "higher"/"lower" discount factor. When discussing the discount rate, we will be referring to Harstad's definition. In his research, he does not assume any specific magnitude of the discount factor.

At a policy design level, there is much debate as to what that rate should be. Ethically, it should be large. Empirically, there are many ways to find it, some better than others. Using capital markets to establish it may not be the best idea in the context of climate change; using revealed preferences is complicated because the time horizon is so large. We could agree with the previous research in that the discount factor is large, but then any stringent policy would be justified. There would be no issue of procrastination and everything would be tackled at the earliest.

However, nothing is tackled, global warming is spiralling out of control, and a significant number of people are in fact against governments spending mtaxpayers' money on green projects. One would think that if the threat of global warming was serious enough, and the discount factor was large, everybody would want to switch to low-carbon as soon as possible, but it is clearly not the case. Hence, either the threat is not serious enough, or the discount factor too small. We therefore assume a reasonably small discount factor, between 0 and $\delta*$, to explore the most interesting

[3] These not exactly the same discount rates due to differences in the underlying models, but Stern's discount rate is still much lower even when the differences are accounted for; see Goulder and Williams (2012).

cases that we see in real life: small poorer countries transitioning to low carbon before larger and more developed countries do; nations pledging to become net zero by and not implementing the policies consistent with those pledges (they may be doing so later, which would be consistent with not investing at $t = 1$ while doing so at $t = 2$). Note that for a large country, current local pollution does not change its behaviour drastically, and we may observe similar dynamics whether or not there is damage at present.

3.4.2 Technology spillovers

But this discussion is based on the absence of knowledge exchange and technological trade between the nations. If countries are allowed to sell their technology, their incentives to invest are increased. Recall the threshold values when there are spillovers.

$$\overline{k_{i,1}} = \frac{\delta(a_i - b_i + \theta_i) + \delta^2(\theta_i + c_i m_i)}{1-\delta} \text{ if both others invest at } t = 1, \text{ or only one does, and the other would at } t = 2 \quad (3.10)$$

$$\overline{k_{i,1}} = \frac{\delta(a_i - b_i + \theta_i) + \delta^2(\theta_i + c_i m_i) + \delta^2 c_i m_k}{1-\delta} \text{ if one invests at } t = 1, \text{ and the other would not invest at all} \quad (3.11)$$

$$\overline{k_{i,1}} = \frac{\delta(a_i - b_i + \theta_i) + \delta(p_{ji} + p_{ki}) + \delta^2(\theta_i + c_i m_i)}{1-\delta} \text{ if neither invests at } t = 1, \text{ but both would at } t = 2 \quad (3.12)$$

$$\overline{k_{i,1}} = \frac{\delta(a_i - b_i + \theta_i) + \delta(p_{ji} + p_{ki}) + \delta^2(\theta_i + c_i m_i) + \delta^2 c_i(m_j + m_k)}{1-\delta} \text{ if neither invest at } t = 1, \text{ and would not at } t = 2 \quad (3.13)$$

$$\overline{k_{i,2}} = \delta(a_i - b_i + \theta_i + c_i m_i) \quad (3.14)$$

As per our model, the price p_{ij} is chosen by each selling nation. There is no assumption about the bargaining power, since there is no bargain over the technology; nation i benefits from selling its technology to nation j and in optimum, the

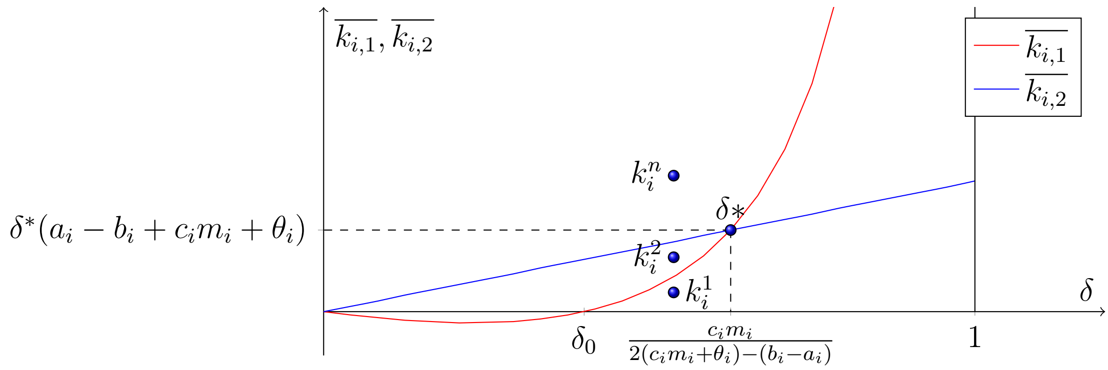

Figure 3.4: Different investment costs that would induce investment to take place at different periods: never for $k_i^n > \delta(a_i - b_i + c_i m_i + \theta_i) > \dfrac{\delta(a_i - b_i + \theta_i) + \delta^2(c_i m_i + \theta_i)}{1-\delta}$; at $t = 2$ for $\delta(a_i - b_i + c_i m_i + \theta_i) > k_i^2 > \dfrac{\delta(a_i - b_i + \theta_i) + \delta^2(c_i m_i + \theta_i)}{1-\delta}$; at $t = 1$ for $\delta(a_i - b_i + c_i m_i + \theta_i) > \dfrac{\delta(a_i - b_i + \theta_i) + \delta^2(c_i m_i + \theta_i)}{1-\delta} > k_i^1$. The discount factor is the same, $\delta < \delta^*$, to observe the scenario in which investment does not take place at $t = 1$ but does at $t = 2$.

price is set low enough so that nation i would purchase it, but high enough so that nation i would be convinced to invest. Ideally, it would want to sell at a price just shy of the exact cost of j's investment, k_j, which means the higher the cost, the higher the price, and the higher i's threshold for own investment. But all nations think the same way. So who will invest first?

Let's start from Equation 3.10. If nations j and k invest at $t = 1$, and δ is such that $k_i > \dfrac{\delta(a_i - b_i + \theta_i) + \delta^2(\theta_i + c_i m_i)}{1-\delta}$ (provided $\delta < \delta^*$), then nation i will not invest at $t = 1$. At $t = 2$, it compares k_i and $\delta(a_i - b_i + \theta_i + c_i m_i)$. This is case 4 above with Figure 3.2 and Figure 3.3. Note that the cases 4(a) and 4(b(i)) are very similar and can be considered as one, while cases 4(b(ii)) and 4(c) do not add substantial value to the analysis, so I will focus on 4(a): both thresholds are non-negative and upward-sloping.

Every outcome is possible in terms of the timing of the investment; see Figure 3.4. For a given δ, investment cost can happen to be one of the three depicted options (there are, of course, more, but they do not add value to the analysis and are not of particular interest to us). In the absence of technological trade, the thresholds are fixed and each country behaves according to the position of its k_i.

However, allowing for trade changes everything. The plot of the basic thresholds compared with the ones augmented by the spillovers is in Figure 3.5. If nation i starts with investment cost k_i located between the green and red lines, above the blue, in the absence of technology spillovers, it would never invest. But if both

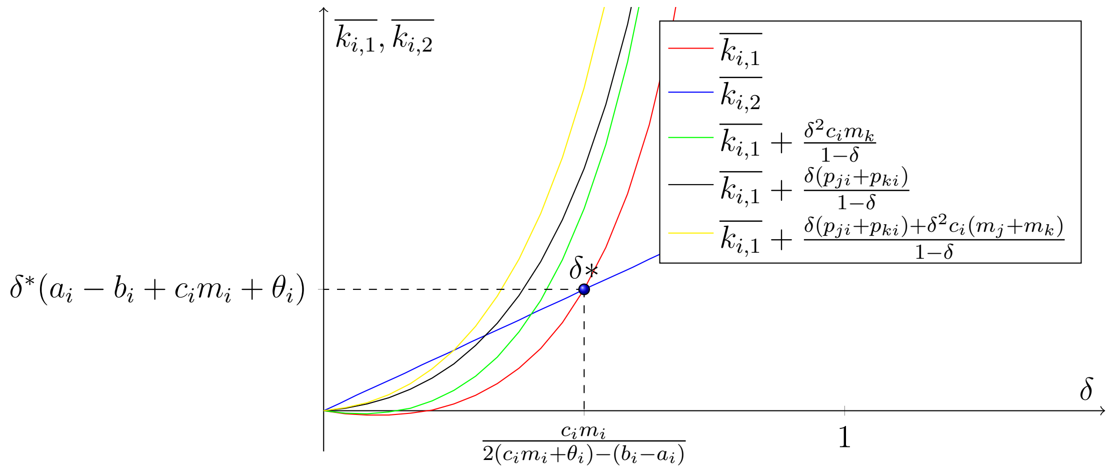

FIGURE 3.5: Different thresholds for different scenarios: (1) basic case with no spillovers, (2) the case which assumes that j invests at $t = 1$ but k never invests, (3) the case which assumes that neither j nor k invest at $t = 1$ but do at $t = 2$, (4) the case which assumes that both j and k never invest. The thresholds are depicted as a function of δ. Note that some of them are exogenous, such as c_i and m_j, while some are endogenously decided by i to maximize the payoff, such as p_{ji} and p_{ki}. It is assumed that j and k are not small nations; otherwise, the whole benefit only comes through the price of technology and not through reduction of global emissions.

other nations start in the same position, meaning they would never invest either, the threshold for nation i shifts to the left and turns into the yellow line, and so it now decides that investing at $t = 1$ is optimal.

But the other two nations see the same picture, and for them it becomes also optimal to invest at $t = 1$, and no one gains anything from such overinvestment. To avoid thinking that j knows that k knows that j knows ad infinitum, the model could be reformulated into a Bayesian game with some priors on the distribution of types. However, note that the type here would be defined with respect to the price of technology, and since this is a decision variable that is fully determined by the investing nation, it is hard to define a prior distribution over it. The game is Bayesian, however, with respect to the uncertainty surrounding the local damage type θ_i .

The possibility of trade is alone capable of inducing investment where there was previously none. Let us look at the cases in Table 3.1. The blue cells are of interest. There, no nation has an incentive to invest at $t = 1$, but may reconsider if the technology can be traded.

- Case D, $k_i = k_i^n$, $k_j = k_j^n$, $k_k = k_k^2$. The case is easy on the face of it, however, we can't say that $k_k < k_j$, because every country has its own set of parameters, and nation k may actually have a much larger investment

$k_k = k_k^2$		Nation j		
		$k_j = k_j^1$	$k_j = k_j^2$	$k_j = k_j^n$
	$k_i = k_i^1$	$(1, 1, 2)$	$(1, 2, 2)$	$(1, 2, 2)$
Nation i	$k_i = k_i^2$	$(2, 1, 2)$	A	B
	$k_i = k_i^n$	$(2, 1, 2)$	C	D

$k_k = k_k^n$		Nation j		
		$k_j = k_j^1$	$k_j = k_j^2$	$k_j = k_j^n$
	$k_i = k_i^1$	$(1, 1, 2)$	$(1, 2, 2)$	$(1, 2, 2)$
Nation i	$k_i = k_i^2$	$(2, 1, 2)$	E	F
	$k_i = k_i^n$	$(2, 1, 2)$	G	H

TABLE 3.1: An array showing what investment decisions nations would make, (i, j, k) respectively, when the investment costs are as specified (nation k's cost are in the corner of each matrix, specified for the whole matrix). The blue coloured numbers in the tables are the decisions that would have been made differently were the spillovers absent. The blue coloured cells are the most interesting scenarios, where no one would want to invest at $t = 1$ but could reconsider when the technology can be sold. The third table, for $k_k = k_k^1$, is not considered because k would always invest at $t = 1$ and sell the technology at an appropriate price when other nations would not want to invest themselves.

cost, but still benefit from implementing the technology, whereas nation j can have tiny investment cost, but not benefit from the transition in any way. The fair price of technology can be established by considering the incentive compatibility constraints for each nation and finding the set of parameters where they all hold. Take Figure 3.4 for reference.

$$\begin{cases} k_i > \delta(a_i - b_i + c_i m_i + \theta_i) > \dfrac{\delta(a_i - b_i + \theta_i) + \delta^2(c_i m_i + \theta_i)}{1-\delta} \\ k_j > \delta(a_j - b_j + c_j m_j + \theta_j) > \dfrac{\delta(a_j - b_j + \theta_j) + \delta^2(c_j m_j + \theta_j)}{1-\delta} \\ \delta(a_k - b_k + c_k m_k + \theta_k) > k_k > \dfrac{\delta(a_k - b_k + \theta_k) + \delta^2(c_k m_k + \theta_k)}{1-\delta} \end{cases}$$

To invest at $t = 1$, k's investment cost needs to be below the new threshold,

$$\frac{\delta(a_k - b_k + \theta_k) + \delta(p_{jk} + p_{ik}) + \delta^2(\theta_k + c_k m_k) + \delta^2 c_k(m_j + m_i)}{1-\delta}$$

which exceeds the old threshold by $\dfrac{\delta(p_{jk} + p_{ik}) + \delta^2 c_k(m_j + m_i)}{1-\delta}$. The second half of the expression is fixed, while the first half is under full control of nation k. Both j and i have to agree to purchase the technology, which they will do if

$$\begin{cases} k_i > \delta(a_i - b_i + c_i m_i + \theta_i) \geq p_{ik} > \dfrac{\delta(a_i - b_i + \theta_i) + \delta^2(c_i m_i + \theta_i)}{1-\delta} \\ k_j > \delta(a_j - b_j + c_j m_j + \theta_j) \geq p_{jk} > \dfrac{\delta(a_j - b_j + \theta_j) + \delta^2(c_j m_j + \theta_j)}{1-\delta} \end{cases}$$

Assuming benevolent nations (i.e., when indifferent, choose a socially superior option), the prices will be exactly equal to the $t = 2$ threshold:

$$\begin{cases} \delta(a_i - b_i + c_i m_i + \theta_i) = p_{ik} \\ \delta(a_j - b_j + c_j m_j + \theta_j) = p_{jk} \end{cases}$$

Taking this to the difference between k's thresholds, we get

$$\frac{\delta^2((a_i - b_i + c_i m_i + \theta_i) + (a_j - b_j + c_j m_j + \theta_j)) + \delta^2 c_k(m_j + m_i)}{1-\delta}$$

which, added to the threshold, has to reverse the order of the thresholds for nation k,

$$\frac{\delta(a_k - b_k + \theta_k) + \delta^2\left(\theta_k + c_k m_k + a_i - b_i + c_i m_i + \theta_i + a_j - b_j + c_j m_j + \theta_j\right)}{1-\delta} +$$
$$+ \frac{\delta^2 c_k(m_j + m_i)}{1-\delta} > \delta(a_k - b_k + c_k m_k + \theta_k)$$
$$(a_k - b_k + \theta_k) + \delta(\theta_k + c_k m_k + a_i - b_i + c_i m_i + \theta_i + a_j - b_j + c_j m_j + \theta_j + c_k(m_j + m_i)) >$$
$$> a_k - b_k + c_k m_k + \theta_k - \delta(a_k - b_k + c_k m_k + \theta_k)$$
$$\delta\left(\theta_k + c_k m_k + (a_i - b_i + c_i m_i + \theta_i) + (a_j - b_j + c_j m_j + \theta_j) + c_k(m_j + m_i)\right) >$$
$$> c_k m_k - \delta(a_k - b_k + c_k m_k + \theta_k)$$
$$\delta(\theta_k + c_k m_k + c_k m_j + c_k m_i) - c_k m_k + \delta(a_i - b_i + c_i m_i + \theta_i) + \delta(a_j - b_j + c_j m_j + \theta_j) >$$
$$> -\delta(a_k - b_k + c_k m_k + \theta_k)$$

With the restrictions on the parameters imposed above (so that the thresholds only slope upward), the last expression is true as long as

$$\delta(\theta_k + c_k m_j + c_k m_i) + c_k m_k(\delta - 1) > 0 \qquad (3.15)$$

If nation k does not suffer from local damage, and the other two nations are small relative to nation k, then the inequality would not hold. It is therefore necessary that either both countries j and i are large, or that country k suffers from its own pollution, or both. Assuming condition in Equation 3.15 holds, nation k would decide to invest at $t = 1$ instead of $t = 2$ and will sell its technology to nations i and j at a price $p_{ik} = \delta(a_i - b_i + c_i m_i + \theta_i)$ and $p_{jk} = \delta(a_j - b_j + c_j m_j + \theta_j)$ respectively at $t = 2$, swaying them away from never investing.

- Cases G and F are identical, with nations j and i investing at $t = 1$, respectively. Thus, the corresponding necessary conditions are

$$\delta(\theta_j + c_j m_k + c_j m_i) + c_j m_j(\delta - 1) > 0 \tag{3.16}$$

$$\delta(\theta_i + c_i m_j + c_i m_k) + c_i m_i(\delta - 1) > 0 \tag{3.17}$$

and the corresponding optimal prices are

$$\begin{cases} p_{ij} = \delta(a_i - b_i + c_i m_i + \theta_i) \\ p_{kj} = \delta(a_k - b_k + c_k m_j + \theta_j) \end{cases}$$

and

$$\begin{cases} p_{ji} = \delta(a_j - b_j + c_j m_j + \theta_j) \\ p_{ki} = \delta(a_k - b_k + c_k m_j + \theta_j) \end{cases}$$

- Cases C, B, and E are also mirroring each other; if suffices to consider one of them; for example, case E: without spillovers, nations i and j would invest at $t = 2$, k would never invest. The initial setup is

$$\begin{cases} \delta(a_i - b_i + c_i m_i + \theta_i) > k_i > \dfrac{\delta(a_i - b_i + \theta_i) + \delta^2(c_i m_i + \theta_i)}{1 - \delta} \\ \delta(a_j - b_j + c_j m_j + \theta_j) > k_j > \dfrac{\delta(a_j - b_j + \theta_j) + \delta^2(c_j m_j + \theta_j)}{1 - \delta} \\ k_k > \delta(a_k - b_k + c_k m_k + \theta_k) > \dfrac{\delta(a_k - b_k + \theta_k) + \delta^2(c_k m_k + \theta_k)}{1 - \delta} \end{cases}$$

It is evident that nations i and j need a smaller push to tip them over to invest at $t = 1$ than k does. From doing so, they would gain $\delta p_{ki} + \delta^2 c_i m_k$ or $\delta p_{kj} + \delta^2 c_j m_k$, respectively. Also note that the nation that decides to invest earlier, gains in payoff because itself now pollutes one period less than before. Consider nation i, for example. It would gain $\delta p_{ki} + \delta^2 c_i m_k$ if it sells the technology to nation k, but it can also sell the technology to nation j and gain δp_{ji}. From its own minimised pollution, it gains $\delta\theta_i\delta^2 c_i m_i$, while it loses $-k_i + \delta k_i$ from the investment undertaken now at $t = 1$ instead of $t = 2$. The net gain is thus $\delta p_{ki} + \delta^2 c_i m_k + \delta p_{ji} + \delta^2 c_i m_i - k_i + \delta k_i$, and it is positive if and only if $p_{ki} + p_{ji} > k_i \frac{1-\delta}{\delta} - \delta c_i(m_k + m_i) - \theta_i$. It is left to check the incentive compatibility constraints for nations j and k to see whether such prices exist that satisfy all national constraints. Note that a similar constraint would be true for nation j to move first. A quick look at the inequalities suggests that the nation who would be the one to invest early will not only be the one whose investment cost is lower, but also the one which suffers more from local pollution and who is larger in size.

When technology spillovers are added to the comparison, essentially anything is possible. It is still more likely that a larger country with more significant present environmental damage invests in green technology before a small nation does, but even a small well off nation can find investment lucrative if it can sell the technology to large emitters who would not have reduced their own emissions otherwise.

While being the second (after the local damage) driver of green investment in our model, the possibility of profiting from selling green technology may not be of great importance to nations in real life. Existing body of research does not permit a definitive conclusion to be made. However, even assuming some other factors unaccounted for in the model play a larger role in national decision making, investment patterns of reach unaffected nations like Sweden and Norway can still be observed in our model where technological trade and own present environmental damage are the only driving forces.

That being said, it is worth keeping in mind that the present analysis is conducted under the assumption that nations invest in mitigation technology. Analysis for other types of technology may well lead to different results.

3.5 Discussion and conclusion

While being incomplete in many ways, the analysis introduced above sheds some light on the interaction between technological trade and opens the way to answer many more questions in a model that is not typical of current literature. Taking national self-interest and small discount rate (future matters less than it ethically and philosophically should) at par value, the non-cooperative behaviour generates interesting dynamics between the nations. The model suggests that while local damage is an important determinant of early investment, it is far from being the only one. In fact, the size of the nation matters, and a small suffering nation is unlikely to invest in mitigation technology because it stands little to gain from its own abatement. It may however change its mind if other polluting nations are large but unwilling to invest for themselves. There is a complex parameter interaction at play, and almost any outcome can be observed as optimal depending on parameter specification. This is good and bad news, because, while global emission reduction can arise as a result of technological trade even when initially no nation was interested in investing early, but in terms of policy implications, it is hard to know which nation would be the first to do so. The analysis suggests small nations are unlikely to invest in mitigation technology whether or not they suffer from global warming already, and that the gaze should be turned towards large nations who preferably already suffer at present from self-generated pollution, such as China. However, if large nations cannot be convinced to invest, a different equilibrium where small nations invest instead can be conceived. A treaty would then be a coordination device that specifies the technology price to avoid over-investment, as well as a mechanism to protect intellectual property rights to ensure that investing nations receive the fees for their technology to encourage investment.

The next step in the analysis would be to ensure the incentive compatibility constraints hold for the nations who are to purchase the technology, to describe the complete optimal strategy profile. Once that is done, the mitigation technology scenario can be considered complete. There are then two more to consider: the cases of adaptation technology and of carbon capture and storage. After that, the uncertainty can be introduced to local damages by making θ_i be unknown at the time of decision making at $t = 1$. At $t = 2$, of course, every nation will become aware of its own damage, but in a game with only three periods, if no nation has invested at $t = 1$, there can be no spillovers and thus no path to global emissions reduction. Uncertainty may then completely negate promising predictions of the deterministic model and lead the world to unmitigated climate change. This result would be consistent with empirical findings by Barrett and Dannenberg (2014) and Dannenberg et al. (2015).

There are multiple other interesting extensions described throughout the paper that can also be considered. The simplicity of the model allows for the solution in a closed form, while not limiting the versatility and applicability of the results. Its linearity is also not a disadvantage, since there is no differentiation and the core results would not be altered by assuming a different functional form to national payoffs. Extending the model to infinity would unlikely yield some qualitatively varying conclusion, but could be an interesting exercise to perform. Considering more than three nations could potentially introduce new dynamics to their interactions, but having three nations already possesses substantial advantage over two, so it remains to be seen what insights n nations would have for the conclusions of the paper.

From the analysis performed in this chapter, it appears that the damage from climate change has to be experienced in order to observe at least some emissions reduction. To observe more than some, such damage would have to be done strategically, to the nations who can afford to invest in green technology and also emit more greenhouse gases than many others. The failure of Kyoto over not being able to get the US and China on board seems to also be the culprit of a treaty based on technology. In any case, unless the solution to climate change can be found in some magical green technology, there is no treaty the world can sign that will lead to sizable global emissions reduction. All that remains is to wait for such technology to be developed and pray that the environment is more resilient than all our models suggest.

Bibliography

Acemoglu, D., F. A. Gallego, and J. A. Robinson (2014). Institutions, Human Capital, and Development. *Annual Review of Economics 6*, 875–912.

Acemoglu, D., S. Johnson, and J. A. Robinson (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review 91*(5), 1369–1401.

Acemoglu, D., S. Johnson, and J. A. Robinson (2002). Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution. *The Quarterly Journal of Economics 117*(4), 1231–1294.

Acemoglu, D. and J. A. Robinson (2000a). Political Losers as a Barrier to Economic Development. *American Economic Review 90*(2), 126–130.

Acemoglu, D. and J. A. Robinson (2000b). Political losers as a barrier to economic development. *American Economic Review 90*(2), 126–130.

Acemoglu, D. and J. A. Robinson (2013). *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*. Broadway Business.

Aklin, M. and M. Mildenberger (2020). Prisoners of the wrong dilemma: why distributive conflict, not collective action, characterizes the politics of climate change. *Global Environmental Politics 20*(4), 4–27.

Alesina, A. and G. Tabellini (1990). A Positive Theory of Fiscal Deficits and Government Debt. *The Review of Economic Studies 57*(3), 403–414.

Amador, M. (2003). A Political Economy Model of Sovereign Debt Repayment. *Manuscript, Stanford University*.

Aoyama, N. and E. C. D. Silva (2022). Endogenous abatement technology agreements under environmental regulation. *Games 13*(2), 32.

Archsmith, J., A. Kendall, and D. Rapson (2015). From cradle to junkyard: assessing the life cycle greenhouse gas benefits of electric vehicles. *Research in Transportation Economics 52*, 72–90.

Arrhenius, S. and E. S. Holden (1897). On the influence of carbonic acid in the air upon the temperature of the earth. *Publications of the Astronomical Society of the Pacific 9*(54), 14–24.

Arthur, W. B. (1989). Competing Technologies, Increasing Returns, and Lock-in by Historical Events. *The Economic Journal 99*(394), 116–131.

Barrett, S. (1994). Self-Enforcing International Environmental Agreements. *Oxford Economics Papers 46*, 878–894.

Barrett, S. (2001). International Cooperation for Sale. *European Economic Review 45*(10), 1835–1850.

Barrett, S. (2005a). *Environment and Statecraft: The Strategy of Environmental Treaty-Making.* Oxford University Press.

Barrett, S. (2005b). *Environment and Statecraft: The Strategy of Environmental Treaty-Making.* Oxford University Press.

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

Barrett, S. (2013). Climate Treaties and Approaching Catastrophes. *Journal of Environmental Economics and Management 66*(2), 235–250.

Barrett, S. and A. Dannenberg (2012). Climate negotiations under scientific uncertainty. *Proceedings of the National Academy of Sciences 109*(43), 17372–17376.

Barrett, S. and A. Dannenberg (2014). Sensitivity of Collective Action to Uncertainty about Climate Tipping Points. *Nature Climate Change 4*(1), 36.

Barrett, S. and A. Dannenberg (2016). An experimental investigation into 'pledge and review'in climate negotiations. *Climatic Change 138*(1), 339–351.

Barrett, S. and A. Dannenberg (2017). Tipping versus cooperating to supply a public good. *Journal of the European Economic Association 15*(4), 910–941.

Barrett, S. and R. Stavins (2003). Increasing Participation and Compliance in International Climate Change Agreements. *International Environmental Agreements 3*(4), 349–376.

Battaglini, M. and B. Harstad (2016). Participation and Duration of Environmental Agreements. *Journal of Political Economy 124*(1), 160–204.

Battaglini, M. and B. Harstad (2020). The political economy of weak treaties. *Journal of Political Economy 128*(2), 544–590.

Baum, C. M., S. Low, and B. K. Sovacool (2022). Between the sun and us: Expert perceptions on the innovation, policy, and deep uncertainties of space-based solar geoengineering. *Renewable and Sustainable Energy Reviews 158*, 112179.

Bayer, P. and J. Urpelainen (2013). External sources of clean technology: Evidence from the clean development mechanism. *The Review of International Organizations 8*(1), 81–109.

Beccherle, J. and J. Tirole (2011). Regional initiatives and the cost of delaying binding climate change agreements. *Journal of Public Economics 95*(11-12), 1339–1348.

Beise, M. (2004). Lead markets: country-specific drivers of the global diffusion of innovations. *Research policy 33*(6-7), 997–1018.

Beise, M. and T. Cleff (2004). Assessing the lead market potential of countries for innovation projects. *Journal of International management 10*(4), 453–477.

Beiser-McGrath, L. F. and T. Bernauer (2019). Commitment failures are unlikely to undermine public support for the paris agreement. *Nature climate change 9*(3), 248–252.

Bento, A., D. Kaffine, K. Roth, and M. Zaragoza-Watkins (2014). The effects of regulation in the presence of multiple unpriced externalities: Evidence from the transportation sector. *American Economic Journal: Economic Policy 6*(3), 1–29.

Beresteanu, A. and S. Li (2011). Gasoline prices, government support, and the demand for hybrid vehicles in the united states. *International Economic Review 52*(1), 161–182.

Bernauer, T., L. Dong, L. F. McGrath, I. Shaymerdenova, and H. Zhang (2016). Unilateral or reciprocal climate policy? experimental evidence from china. *Politics and Governance 4*(3), 152–171.

Bernauer, T. and R. Gampfer (2015). How robust is public support for unilateral climate policy? *Environmental Science & Policy 54*, 316–330.

Bernstein, S. and M. Hoffmann (2018). The politics of decarbonization and the catalytic impact of subnational climate experiments. *Policy Sciences 51*(2), 189–211.

Biermann, F., J. Oomen, A. Gupta, S. H. Ali, K. Conca, M. A. Hajer, P. Kashwan, L. J. Kotzé, M. Leach, D. Messner, et al. (2022). Solar geoengineering: The case for an international non-use agreement. *Wiley Interdisciplinary Reviews: Climate Change 13*(3), e754.

Binger, B. R. and E. Hoffman (1989). Institutional Persistence and Change: The Question of Efficiency. *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft 145*(1), 67–84.

Blumsack, S., C. Samaras, and P. Hines (2008). Long-term electric system investments to support plug-in hybrid electric vehicles. In *2008 IEEE Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century*, pp. 1–6. IEEE.

Boute, A. and A. Zhikharev (2019). Vested interests as driver of the clean energy transition: Evidence from russia's solar energy policy. *Energy Policy 133*, 110910.

Bracking, S. (2015). The anti-politics of climate finance: the creation and performativity of the green climate fund. *Antipode 47*(2), 281–302.

Brandt, U. S. and N. Nannerup (2013). Unilateral actions as signals of high damage costs: distorting pre-negotiations emissions in international environmental problems. *Environmental economics 4*(2), 31–41.

Brunner, S., C. Flachsland, and R. Marschinski (2012). Credible commitment in carbon policy. *Climate Policy 12*(2), 255–271.

Buen, J. (2006). Danish and norwegian wind industry: The relationship between policy instruments, innovation and diffusion. *Energy policy 34*(18), 3887–3897.

Calel, R. and A. Dechezleprêtre (2016). Environmental policy and directed technological change: evidence from the european carbon market. *Review of economics and statistics 98*(1), 173–191.

Carattini, S., S. Levin, and A. Tavoni (2019). Cooperation in the climate commons. *Review of Environmental Economics and Policy 13*(2), 227–247.

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

Chander, P. (2017). Subgame-perfect cooperative agreements in a dynamic game of climate change. *Journal of Environmental Economics and Management 84*, 173–188.

Chander, P. and H. Tulkens (2006a). The core of an economy with multilateral environmental externalities. In *Public goods, environmental externalities and fiscal competition*, pp. 153–175. Springer.

Chander, P. and H. Tulkens (2006b). A core-theoretic solution for the design of cooperative agreements on transfrontier pollution. In *Public goods, environmental externalities and fiscal competition*, pp. 176–193. Springer.

Chatterji, S. and S. Ghosal (2009). Technology, Unilateral Commitments and Cumulative Emissions Reduction. *CESifo Economic Studies 55*(2), 286–305.

Clinton, B. C. and D. C. Steinberg (2019). Providing the spark: Impact of financial incentives on battery electric vehicle adoption. *Journal of Environmental Economics and Management 98*, 102255.

Coate, S. and S. Morris (1999). Policy Persistence. *American Economic Review 89*(5), 1327–1336.

Dannenberg, A., A. Loschel, G. Paolacci, C. Reif, and A. Tavoni (2015). On the Provision of Public Goods with Probabilistic and Ambiguous Thresholds. *Environmental and Resource Economics 61*(3), 365–383.

Diamond, J. M. (1997). *Guns, Germs and Steel: a Short History of Everybody for the Last 13,000 Years.* London: Jonathan Cape.

Dockner, E. J., N. Van Long, and G. Sorger (1996). Analysis of Nash Equilibria in a Class of Capital Accumulation Games. *Journal of Economic Dynamics and Control 20*(6-7), 1209–1235.

Dollar, D. and A. Kraay (2003). Institutions, Trade, and Growth. *Journal of Monetary Economics 50*(1), 133–162.

Dutta, B., S. Ghosal, and D. Ray (2005). Farsighted Network Formation. *Journal of Economic Theory 122*(2), 143–164.

Dutta, P. K. and R. Radner (2009). A Strategic Analysis of Global Warming: Theory and Some Numbers. *Journal of Economic Behavior & Organization 71*(2), 187–209.

Elsayyad, M. and F. Morath (2016). Technology transfers for climate change. *International Economic Review 57*(3), 1057–1084.

Emmerling, J., U. Kornek, V. Bosetti, and K. Lessmann (2021). Climate thresholds and heterogeneous regions: Implications for coalition formation. *The Review of International Organizations 16*, 293–316.

Espín-Sánchez, J.-A. (2017). Institutional Inertia: Persistent Inefficient Institutions in Spain. *The Journal of Economic History 77*(3), 692–723.

Fernandez, R. and D. Rodrik (1991). Resistance to reform: Status quo bias in the presence of individual-specific uncertainty. *The American economic review 5*(81), 1146–1155.

Finus, M. and P. Pintassilgo (2013). The role of uncertainty and learning for the success of international climate agreements. *Journal of Public Economics 103*, 29–43.

Frankel, J. A. and D. H. Romer (1999). Does Trade Cause Growth? *American Economic Review 89*(3), 379–399.

Galiani, S., I. Torre, and G. Torrens (2019). International organizations and the political economy of reforms. *Journal of International Economics 121*, 103249.

Gallagher, K. S. and E. Muehlegger (2011). Giving green to get green? incentives and consumer adoption of hybrid vehicle technology. *Journal of Environmental Economics and management 61*(1), 1–15.

Geels, F. W., B. K. Sovacool, T. Schwanen, and S. Sorrell (2017). Sociotechnical transitions for deep decarbonization. *Science 357*(6357), 1242–1244.

Glaeser, E. L., R. La Porta, F. Lopez-de Silanes, and A. Shleifer (2004). Do Institutions Cause Growth? *Journal of Economic Growth 9*(3), 271–303.

Góes, C. (2016). Institutions and growth: A gmm/iv panel var approach. *Economics Letters 138*, 85–91.

Gollier, C. and J. Tirole (2015). Negotiating effective institutions against climate change. *Economics of Energy & Environmental Policy 4*(2), 5–28.

Golombek, R. and M. Hoel (2004). Unilateral Emission Reductions and Cross-Country Technology Spillovers. *The BE Journal of Economic Analysis & Policy 3*(2), 1–27.

Gosens, J. and Y. Lu (2013). From lagging to leading? technological innovation systems in emerging economies and the case of chinese wind power. *Energy Policy 60*, 234–250.

Goulder, L. H. and I. Williams, Roberton C (2012, August). The choice of discount rate for climate change policy evaluation. Working Paper 18301, National Bureau of Economic Research.

Graff Zivin, J. S., M. J. Kotchen, and E. T. Mansur (2014). Spatial and temporal heterogeneity of marginal emissions: Implications for electric cars and other electricity-shifting policies. *Journal of Economic Behavior & Organization 107*, 248–268.

Grafström, J. and Å. Lindman (2017). Invention, innovation and diffusion in the european wind power sector. *Technological Forecasting and Social Change 114*, 179–191.

Grijalva, T. C., J. L. Lusk, and W. D. Shaw (2014). Discounting the distant future: An experimental investigation. *Environmental and Resource Economics 59*, 39–63.

Guo, J.-X. and Y. Fan (2017). Optimal abatement technology adoption based upon learning-by-doing with spillover effect. *Journal of Cleaner Production 143*, 539–548.

Haas, P. M. (1992). Banning Chlorofluorocarbons: Epistemic Community Efforts to Protect Stratospheric Ozone. *International organization 46*(1), 187–224.

Halimanjaya, A. and E. Papyrakis (2015). Donor characteristics and the allocation of aid to climate mitigation finance. *Climate Change Economics 6*(03), 1550014.

Hall, R. E. and C. I. Jones (1999). Why do Some Countries Produce So Much More Output Per Worker than Others? *The Quarterly Journal of Economics 114*(1), 83–116.

Harstad, B. (2012). Climate Contracts: A Game of Emissions, Investments, Negotiations, and Renegotiations. *Review of Economic Studies 79*(4), 1527–1557.

Harstad, B. (2016). The Dynamics of Climate Agreements. *Journal of the European Economic Association 14*(3), 719–752.

Harstad, B. (2020). Technology and time inconsistency. *Journal of Political Economy 128*(7), 2653–2689.

Harstad, B. (2023). Pledge-and-review bargaining: from kyoto to paris. *The Economic Journal 133*(651), 1181–1216.

Harstad, B., F. Lancia, and A. Russo (2019). Compliance Technology and Self-enforcing Agreements. *Journal of the European Economic Association 17*(1), 1–29.

Hawkins, T. R., B. Singh, G. Majeau-Bettez, and A. H. Strømman (2013). Comparative environmental life cycle assessment of conventional and electric vehicles. *Journal of Industrial Ecology 17*(1), 53–64.

Heuson, C., W. Peters, R. Schwarze, and A.-K. Topp (2015). Investment and adaptation as commitment devices in climate politics. *Environmental and resource economics 62*, 769–790.

Hoel, M. (1992). International environment conventions: the case of uniform reductions of emissions. *Environmental and Resource Economics 2*, 141–159.

Hoel, M. and A. De Zeeuw (2010). Can a Focus on Breakthrough Technologies Improve the Performance of International Environmental Agreements? *Environmental and Resource Economics 47*, 395–406.

Holland, S. P., E. T. Mansur, N. Z. Muller, and A. J. Yates (2016). Are there environmental benefits from driving electric vehicles? The importance of local factors. *American Economic Review 106*(12), 3700–3729.

Holtsmark, B. and A. Skonhoft (2014). The norwegian support and subsidy policy of electric cars. should it be adopted by other countries? *Environmental science & policy 42*, 160–168.

Jacobsson, S. and V. Lauber (2006). The politics and policy of energy system transformation—explaining the german diffusion of renewable energy technology. *Energy policy 34*(3), 256–276.

Ji, S., C. R. Cherry, M. J. Bechle, Y. Wu, and J. D. Marshall (2012). Electric vehicles in china: emissions and health impacts. *Environmental science & technology 46*(4), 2018–2024.

Jiao, J., G. Jiang, and R. Yang (2018). Impact of r&d technology spillovers on carbon emissions between china's regions. *Structural Change and Economic Dynamics 47*, 35–45.

Jones, B. F. and B. A. Olken (2005). Do Leaders Matter? National Leadership and Growth since World War II. *The Quarterly Journal of Economics 120*(3), 835–864.

Kalkuhl, M., O. Edenhofer, and K. Lessmann (2012). Learning or lock-in: Optimal technology policies to support mitigation. *Resource and Energy Economics 34*(1), 1–23.

Kemp, R., J. Schot, and R. Hoogma (1998). Regime shifts to sustainability through processes of niche formation: the approach of strategic niche management. *Technology analysis & strategic management 10*(2), 175–198.

Keohane, R. O. and M. Oppenheimer (2016). Paris: Beyond the climate dead end through pledge and review? *Politics and Governance 4*(3), 142–151.

Kim, J. D. and M. Rahimi (2014). Future energy loads for a large-scale adoption of electric vehicles in the city of los angeles: Impacts on greenhouse gas (ghg) emissions. *Energy Policy 73*, 620–630.

Klenow, P. J. and A. Rodriguez-Clare (2005). Externalities and growth. *Handbook of economic growth 1*, 817–861.

Knack, S. and P. Keefer (1995). Institutions and economic performance: Cross-country tests using alternative institutional measures. *Economics and Politics 7*(3), 207–227.

Kolstad, C. and A. Ulph (2008). Learning and International Environmental Agreements. *Climatic Change 89*(1-2), 125–141.

Kolstad, C. D. (2007). Systematic Uncertainty in Self-enforcing International Environmental Agreements. *Journal of Environmental Economics and Management 53*(1), 68–79.

Kolstad, C. D. and A. Ulph (2011). Uncertainty, Learning and Heterogeneity in International Environmental Agreements. *Environmental and Resource Economics 50*(3), 389–403.

Langniß, O. and L. Neij (2004). National and international learning with wind power. *Energy & Environment 15*(2), 175–185.

Lee, C.-C., S. Qin, and Y. Li (2022). Does industrial robot application promote green technology innovation in the manufacturing industry? *Technological Forecasting and Social Change 183*, 121893.

Lenton, T. M. (2011). Early warning of climate tipping points. *Nature climate change 1*(4), 201–209.

Li, S., L. Tong, J. Xing, and Y. Zhou (2017). The market for electric vehicles: indirect network effects and policy design. *Journal of the Association of Environmental and Resource Economists 4*(1), 89–133.

Li, S. and C. Wu (2023). Abatement technology innovation modelling with spillover effects under different market structures. *Journal of Industrial and Management Optimization 19*(7), 5163–5182.

Lindman, Å. and P. Söderholm (2012). Wind power learning rates: A conceptual review and meta-analysis. *Energy economics 34*(3), 754–761.

Loewenstein, G. and D. Prelec (1992). Anomalies in intertemporal choice: Evidence and an interpretation. *The Quarterly Journal of Economics 107*(2), 573–597.

Marechal, K. and N. Lazaric (2010). Overcoming inertia: insights from evolutionary economics into improved energy and climate policies. *Climate Policy 10*(1), 103–119.

Markard, J., R. Raven, and B. Truffer (2012). Sustainability transitions: An emerging field of research and its prospects. *Research policy 41*(6), 955–967.

Martimort, D. and W. Sand-Zantman (2016). A mechanism design approach to climate-change agreements. *Journal of the European Economic Association 14*(3), 669–718.

Marx, K. (1852). The Eighteenth Brumaire of Louis Bonaparte. October 29, 2018, from `https://www.marxists.org/archive/marx/works/1852/18th-brumaire/ch01.htm`.

Mauro, P. (1995). Corruption and growth. *The quarterly journal of economics 110*(3), 681–712.

McCarthy, R. and C. Yang (2010). Determining marginal electricity for near-term plug-in and fuel cell vehicle demands in california: Impacts on vehicle greenhouse gas emissions. *Journal of Power Sources 195*(7), 2099–2109.

McDonald, M. (2023). Geoengineering, climate change and ecological security. *Environmental Politics 32*(4), 565–585.

Meadowcroft, J. (2011). Engaging with the politics of sustainability transitions. *Environmental Innovation and Societal Transitions 1*(1), 70–75.

Meckling, J., N. Kelsey, E. Biber, and J. Zysman (2015). Winning coalitions for climate policy. *Science 349*(6253), 1170–1171.

Michalek, J. J., M. Chester, P. Jaramillo, C. Samaras, C.-S. N. Shiau, and L. B. Lave (2011). Valuation of plug-in vehicle life-cycle air emissions and oil displacement benefits. *Proceedings of the National Academy of Sciences 108*(40), 16554–16558.

Morrisette, P. M. (1989). The Evolution of Policy Responses to Stratospheric Ozone Depletion. *Natural Resources Journal 29*(3), 793–820.

Muttitt, G. and S. Kartha (2020). Equity, climate justice and fossil fuel extraction: principles for a managed phase out. *Climate Policy 20*(8), 1024–1042.

Na, S.-l. and H. S. Shin (1998). International environmental agreements under uncertainty. *Oxford Economic Papers 50*(2), 173–185.

Neumayer, E. (2005). Is the allocation of food aid free from donor interest bias? *The Journal of Development Studies 41*(3), 394–411.

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

Nordhaus, W. (2019). Climate change: The ultimate challenge for economics. *American Economic Review 109*(6), 1991–2014.

Nordhaus, W. D. (1993a). Optimal greenhouse-gas reductions and tax policy in the" dice" model. *The American Economic Review 83*(2), 313–317.

Nordhaus, W. D. (1993b). Rolling the 'dice': an optimal transition path for controlling greenhouse gases. *Resource and Energy Economics 15*(1), 27–50.

Nordhaus, W. D. (2007). A review of the stern review on the economics of climate change. *Journal of economic literature 45*(3), 686–702.

Nordhaus, W. D. and Z. Yang (1996). A Regional Dynamic General-Equilibrium Model of Alternative Climate-Change Strategies. *The American Economic Review 86*(4), 741–765.

North, D. C. (1991). Institutions. *Journal of Economic Perspectives 5*(1), 97–112.

North, D. C. and B. R. Weingast (1989). Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-century England. *The Journal of Economic History 49*(4), 803–832.

Oye, K. A. and J. H. Maxwell (1994). Self-interest and environmental management. *Journal of Theoretical Politics 6*(4), 593–624.

Panhans, M., L. Lavric, and N. Hanley (2017). The effects of electricity costs on firm re-location decisions: insights for the pollution havens hypothesis? *Environmental and Resource Economics 68*, 893–914.

Parry, I. W. (1998). A second-best analysis of environmental subsidies. *International Tax and Public Finance 5*(2), 153–170.

Pickering, J. and P. Mitchell (2017). What drives national support for multilateral climate finance? international and domestic influences on australia's shifting stance. *International Environmental Agreements: Politics, Law and Economics 17*(1), 107–125.

Putnam, R. D. (1988). Diplomacy and Domestic Politics: The Logic of Two-Level Games. *International organization 42*(3), 427–460.

Quitzow, R. (2015). Dynamics of a policy-driven market: The co-evolution of technological innovation systems for solar photovoltaics in china and germany. *Environmental Innovation and Societal Transitions 17*, 126–148.

Richardson, K., W. Steffen, W. Lucht, J. Bendtsen, S. E. Cornell, J. F. Donges, M. Drüke, I. Fetzer, G. Bala, W. von Bloh, G. Feulner, S. Fiedler, D. Gerten, T. Gleeson, M. Hofmann, W. Huiskamp, M. Kummu, C. Mohan, D. Nogués-Bravo, S. Petri, M. Porkka, S. Rahmstorf, S. Schaphoff, K. Thonicke, A. Tobian, V. Virkki, L. Wang-Erlandsson, L. Weber, and J. Rockström (2023). Earth beyond six of nine planetary boundaries. *Science Advances 9*(37), eadh2458.

Rodrik, D., A. Subramanian, and F. Trebbi (2004). Institutions Rule: the Primacy of Institutions over Geography and Integration in Economic Development. *Journal of Economic Growth 9*(2), 131–165.

Román, M. (2010). Governing from the middle: the c40 cities leadership group. *Corporate Governance: International Journal of Business in Society 10*(1), 73–84.

Rosenberg, N., L. E. Birdzell, and G. W. Mitchell (1986). *How the West Grew Rich*. Popular Prakashan Mumbai.

Roth, M. B., P. J. Adams, P. Jaramillo, and N. Z. Muller (2022). Policy spillovers, technological lock-in, and efficiency gains from regional pollution taxes in the us. *Energy and Climate Change 3*, 100077.

Sagar, A. D. and B. Van der Zwaan (2006). Technological innovation in the energy sector: R&d, deployment, and learning-by-doing. *Energy policy 34*(17), 2601–2608.

Sanna-Randaccio, F., R. Sestini, and O. Tarola (2017). Unilateral climate policy and foreign direct investment with firm and country heterogeneity. *Environmental and Resource Economics 67*, 379–401.

Segura-Cayuela, R. (2006). Inefficient Policies, Inefficient Institutions and Trade. working paper.

Shapley, L. S. and M. Shubik (1969). On the Core of an Economic System with Externalities. *The American Economic Review 59*(4), 678–684.

Shleifer, A. and R. W. Vishny (1993). Corruption. *The quarterly journal of economics 108*(3), 599–617.

Singh, E. M. (2012). Structural Change, Vested Interests, and Scandinavian Energy Policy-Making: Why Wind Power Struggles in Norway and not in Denmark. *The Open Renewable Energy Journal 5*(1), 19–31.

Smith, A. and D. Stewart (1963). *An Inquiry into the Nature and Causes of the Wealth of Nations*, Volume 1. Wiley Online Library.

Sokoloff, K. L. and S. L. Engerman (2000). Institutions, Factor Endowments, and Paths of Development in the New World. *Journal of Economic Perspectives 14*(3), 217–232.

Steffen, W., K. Richardson, J. Rockström, S. E. Cornell, I. Fetzer, E. M. Bennett, R. Biggs, S. R. Carpenter, W. De Vries, C. A. De Wit, et al. (2015). Planetary boundaries: Guiding human development on a changing planet. *Science 347*(6223), 1259855.

Stern, N., S. Peters, V. Bakhshi, A. Bowen, C. Cameron, S. Catovsky, D. Crane, S. Cruickshank, S. Dietz, N. Edmonson, et al. (2006). *Stern Review: The Economics of Climate Change*, Volume 30. HM treasury London.

Strunz, S. (2014). The German energy transition as a regime shift. *Ecological Economics 100*, 150–158.

Tarui, N. and S. Polasky (2005). Environmental regulation with technology adoption, learning and strategic behavior. *Journal of Environmental Economics and Management 50*(3), 447–467.

Thaler, R. H. and S. Benartzi (2004). Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving. *Journal of Political Economy 112*(S1), S164–S187.

Ulph, A. (2004). Stable International Environmental Agreements with a Stock Pollutant, Uncertainty and Learning. *Journal of Risk and Uncertainty 29*(1), 53–73.

Ulph, A. and D. Ulph (1996). Who Gains from Learning about Global Warming? In *Economics of Atmospheric Pollution*, pp. 31–67. Springer.

Ulph, A. and D. Ulph (2007). Climate change—environmental and technology policies in a strategic context. *Environmental and Resource Economics 37*, 159–180.

Unruh, G. C. (2000a). Understanding carbon lock-in. *Energy policy 28*(12), 817–830.

Unruh, G. C. (2000b). Understanding carbon lock-in. *Energy policy 28*(12), 817–830.

Unruh, G. C. (2002). Escaping carbon lock-in. *Energy policy 30*(4), 317–325.

Urpelainen, J. (2014). Sinking costs to increase participation: technology deployment agreements enhance climate cooperation. *Environmental Economics and Policy Studies 16*, 229–240.

Wang, H., H. Cui, and Q. Zhao (2021). Effect of green technology innovation on green total factor productivity in china: Evidence from spatial durbin model analysis. *Journal of Cleaner Production 288*, 125624.

Watkins, L., R. Aitken, and D. Mather (2016). Conscientious consumers: a relationship between moral foundations, political orientation and sustainable consumption. *Journal of Cleaner Production 134*, 137–146.

Wooliscroft, B., A. Ganglmair-Wooliscroft, and A. Noone (2014). The hierarchy of ethical consumption behavior: The case of new zealand. *Journal of Macromarketing 34*(1), 57–72.

Zavaleta, A. (2016). Climate change and breakthrough technologies: The role of markets. *Environmental and Resource Economics 64*, 597–617.

Zhou, R., X. Bi, and K. Segerson (2020). Evaluating voluntary environmental programs with spillover effects. *Journal of the Association of Environmental and Resource Economists 7*(1), 145–180.