ISSN 1471-0498

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

LEADING THE WAY: COALITIONAL STABILITY IN TECHNOLOGICAL COOPERATION & SEQUENTIAL CLIMATE POLICY

Heinrich H. Nax & Thomas W.L. Norman

Number 585
December 2011

Manor Road Building, Oxford OX1 3UQ

Leading the Way: Coalitional Stability in Technological Cooperation & Sequential Climate Policy*

Heinrich H. Nax & Thomas W. L. Norman†

First version: October 14, 2011
This version: December 12, 2011

Abstract

The world's nations have yet to reach a truly effective treaty to control the emission of greenhouse gases. The importance of compatibility with private incentives of individual countries has been acknowledged (at least by game theorists) in designing climate policies for the post-Kyoto world. Individually incentive-compatible agreements, however, may still be spoilt if coalitional incentives to deviate as a group exist. As a first step toward understanding these incentives from a game-theoretic perspective, we propose a hybrid noncooperative–cooperative game theory model of coalition formation in technology collaboration. Serious coalitional instabilities inherent to the existing climate policy architectures are revealed. It turns out that coalitionally stable agreements are achieved via intermediate self-selecting subcoalitions. The sequence of coalitions forming and the size of the direct and spillover effects of R&D collaboration on countries' individual production technologies determine the effectiveness of the agreements to reduce carbon emissions. These coalitional group motives are already becoming important in the practice of climate change negotiations.

JEL classifications: C71, C73, D62, D86, F53, H87, Q54
Key words: Climate change policy, coalitions, cooperative game theory, environmental agreements, externalities, mechanism design, noncooperative game theory, R&D

*The authors would like to thank Roy Radner and Bard Harstad for discussions that helped in developing and improving the model of this paper. For correspondence, please contact Thomas Norman at Magdalen College, Oxford OX1 4AU, United Kingdom.
†University of Oxford.

1 Global reform through local integration

Agreements to cut greenhouse gas emissions should be "incentive-compatible", that is, consistent with countries' selfish pursuit of private interests. Otherwise, a treaty may either never be signed by all parties or risks being disrespected.[1] An example of such a treaty is the Kyoto protocol. Without an incentive-compatible policy architecture, it would require some form of legitimate global government to enforce, by legal means, emission reductions beyond countries' private interests. In the absence of such a world government, the world's nations have yet to agree on climate change policies that are based on voluntary collective action and, thus, promise to be truly effective. To achieve emission reductions in an economically efficient way, the importance of technology sharing and joint investments into research and development (R&D) has been noted. Only if renewable technology is substantially improved will countries voluntarily substitute their conventional fossil technologies in a meaningful way and, thus, cut their greenhouse gas emissions sufficiently.[2]

An important dimension of incentive compatibility, however, has been ignored in the game-theoretic study of climate change; coalitional stability. Beyond incentive compatibility at the individual level, a truly effective agreement must also be coalitionally stable. Otherwise, coalitions of countries may deviate jointly. For example, suppose (1) a global R&D agreement is formulated where all countries in the world share renewable energy technology and cut emissions, and (2) that agreement is individually incentive-compatible, that is, any potential individual defector who pollutes more than agreed or refuses to share its technology is worse-off. Given (1) and (2) hold, however, it may nonetheless be the case that a coalition of countries exists that prefers to deviate jointly. Given a reduction of fossil fuel consumption by the rest of the world, groups of countries that are fossil-rich and/ or have relatively more advanced renewable technology may be better-off by sharing technology only amongst themselves and increasing their own fossil use. For these coalitions, the decrease in fossil consumption by the rest of the world reduces its (opportunity) cost of consuming fossil themselves. Ceteris paribus, that incentive to deviate may outweigh the benefits from technology collaboration when the technology within that coalition is already sufficiently advanced. Such countries may just as well not collaborate with the rest of the world if it allows them to pollute more.

In practice, a tendency for negotiations to be lead via multinational interest groups and partial agreements proves that these coalitional incentives are very important. During the Durban Climate Change Conference (November/ December 2011), for instance, the discussion was mostly led by representatives of coalitions.[3] Existing coalitions like the European Union, the Alliance of Small Island States (AOSIS), the Coalition of Rainfor-

[1]Incentive compatibility as a key property to effective climate change policy has been stressed by a number of economists, including Barrett [2001], [2003], [2005], Dutta & Radner [2004], [2009], Aldy & Stavins [2007], [2009], Stavins [2010], Harstad [2011a], [2011b].

[2]The policies suggested by Barrett [2001], [2003], [2005], Dutta & Radner [2004], [2009], Harstad [2011a], [2011b] study this effect explicitly.

[3]See the daily programs of the Durban Climate Change Conference at unfccc.int and, for example, the pre-Durban report FCCC/AWGLCA/2011/9 of the UNFCCC Ad hoc Working Group on Long-term Cooperative Action under the Convention (AWG-LCA) from August 9, 2011.

est Nations, the African Group and the Group of Mountainous Landlocked Developing Countries pursued (multi)national interests through these institutions, whereas relatively few countries—most importantly the United States—continued to negotiate primarily as individual nations. Moreover, toward the end of the conference, new coalitional blocks began to form and statements were made on their multinational behalves, for example on behalf of the Group of 77 and China, on behalf of the Least Developed Countries, backing the EU's push for a firm emission reduction timetable after it got support from the African Group and AOSIS.

Toward a better game-theoretic understanding of these issues, in this paper, a formal strategic model of coalitional collaboration and individual energy investments is developed. The global emission of greenhouse gases is modelled as a direct consequence of these decisions. Countries differ with respect to their production technologies and geo-climatic locations that determine their exposure to the adverse effects of climate changes. These heterogeneities imply different incentives to cooperate and deviate. Standard assumptions are made with respect to the production technologies and costs. Each country makes the following two choices. First, countries form coalitions to collaborate in research and development (R&D). Second, given the coalition structure, countries privately make investments in conventional fossil and renewable-energy technologies without any binding commitments.

We evaluate the coalitional stability of the subgame perfect Nash equilibria (SPNE) of the game. The global welfare optimum is not a SPNE because, given any coalition structure, all countries have an inherent tendency to overpollute and underinvest in R&D. "Business as usual" with high pollution and low R&D is a SPNE but can be improved upon if the right coalitions form. A global agreement forming the grand coalition is not a SPNE if tied to emission quotas for all countries. Global agreements without emission quotas are SPNE but may result in insubstantial emission reductions. Furthermore, the global agreement is coalitionally unstable if groups of countries exist whose renewable energy technology is insufficiently improved by forming the grand coalition instead of a coalition amongst themselves.

Amongst the SPNE that are coalitionally stable, a certain class is of particular interest: when coalitions form amongst countries with similar resources and technology who are exposed to the same effects of climate change, all countries may agree to reduce their emissions. This may already lead to substantial emission reductions. Moreover, the formation of these local "clubs" facilitates larger coalitions forming afterwards that would initially have been unstable. As a result, subsequent agreements can be reached that lead to larger emission reductions than were previously negotiable. Following the technological progress in intermediate coalitions, the gains from collaboration improve and reduce incentives to deviate in the future. Subsequently, collaboration cascades through growing coalitions resulting in arrangements that are not only incentive compatible and coalitionally stable, but also achieve larger emission reductions than R&D mechanisms that are immediately negotiated at the global level. We view these results are a first step toward developing a better understanding of stable coalitions in climate change agreements and what effects these may have on global emission levels.

2 Climate policy architectures

The United Nations Framework Convention on Climate Change (UNFCCC) has two main policy targets in order of priority: (1) to reduce the harmful effects of greenhouse gas emissions, and (2) to cooperate toward development of renewable energy technology.[4] Much of the wellbeing of future generations may well depend on how successfully these will be implemented. Of course, the two policy targets are intimately interrelated because conventional fossil fuels (causing greenhouse gas emissions) and renewable energy technology (no emissions) are substitutes. A policy architecture aimed at either or both targets, therefore, can either curtail fossil fuel consumption by setting emission quotas and thus force countries into renewable technology, or promote renewable technology and thus foster substitution effects away from fossil fuels.

The most important agreement, the Kyoto protocol, is negotiated as a quota system with emission targets for participating countries for a five-year period from 2008-2012. Participating countries face emission quotas and build on the protocol's implementation strategies for renewable energy alternatives. Ever since the first meetings of the United Nations Conference on Environment and Development (UNCED) preceding the foundation of the UNFCCC, quota mechanisms have dominated the political debate because they address the main problem (emissions) directly, and because emissions are easier to verify than joint investments in renewable technology. This may be true for the policy makers perspective. When it comes to signing and adhering to these mechanisms, however, the countries have not played along quite as expected. As of 2011, one year before its expiry, 192 countries have voluntarily ratified the Kyoto protocol. Despite this seemingly overwhelming participation, emission reductions have been moderate. The reasons include the following major issues. First, the largest emitter of harmful greenhouse gases, the USA, has not ratified. Second, ratification of other main emitters has at most symbolic character; China and India negotiated no binding emission targets, other targets (including Russia's) are lax. Third, many targets are simply not met (e.g. Canada, Australia, New Zealand, Switzerland, Norway).[5]

In short, despite extensive negotiations about globally beneficial targets, most countries have essentially gone about emission policy as "business as usual", resulting in exactly the kind of continued high emissions that pre-Kyoto meetings feared with too low investments in renewable technology. The basic dilemma of the protocol is that globally beneficial measures are incentive-incompatible at the individual country level: countries that expect to be worse-off by committing to certain targets either opt not to ratify at all (e.g. the USA), or, when they ratify, to set overly modest targets (e.g. Russia) or no binding targets at all (e.g. China). Countries who set proper targets but fail to meet these (e.g. Australia), moreover, may subsequently choose to disrespect the protocol's implementation mechanisms at little cost.

Now, serious reductions are needed soon or else climatic effects may get increasingly difficult and costly to address.[6] Unfortunately, the kind of global climate "Marshall

[4] See the ultimate objective list, formulated in the Kyoto protocol [1998] under Article 2.

[5] See, for example, the reports of the European Climate Change Programme (ECCP).

[6] The Copenhagen accord (2009) contains emission targets for 2020.

plan" that may seem necessary is largely non-implementable in a global market economy without a world government.[7] Substantial research has gone into ways of adapting the Kyoto protocol to market mechanisms.[8] It turns out that mechanisms in the spirit of the Kyoto protocol, that is, emissions-trading markets and direct global tax schemes for carbon consumption are incentive-incompatible in much the same way.[9]

Toward a more effective post-Kyoto agreement, theorists have over the last years come to the conclusion that renewable technology needs to be more seriously promoted to foster substantial substitution away from fossil fuels in an incentive-compatible way. The Cancun agreement (2010) already contains mechanism tools for how, in addition to the 2020 emission targets (Copenhagen accord), the necessary knowledge transfer may be facilitated using strategies such as joint investment in R&D, knowledge and resource sharing, joint energy networks, etc. Barrett [2001], [2003], [2005] proposes the first formal climate change policy architecture of this type. Unfortunately, based on global R&D collaboration alone, the implied resulting emission reductions may not be very large. Dutta & Radner [2004], [2009], Harstad [2011a], [2011b], Datta & Somanathan [2011], Bosetti & Frankel [2011] explore their nature further and see how larger emissions may be achieved. It turns out R&D agreements come with their own incentive issues because certain (high-tech) countries may prefer not to share their technology. Also, since investments are harder to verify than emissions and because certain sets of countries may have an incentive to delay or fake their contributions, global agreements based on participation quotas may again be spoilt.[10] All of these incentive issued have been noticed and studied but not overcome.

There is another dimension of incentive compatibility that theorists have yet to consider; coalitional stability. Because any climate change agreement is based on cooperative contracts that individual countries voluntarily sign, these contracts should also be robust to coalitional renegotiations. Otherwise, coalitions may deviate jointly. Even without deviation at the individual level, therefore, groups of countries may be able to negotiate superior arrangements by writing subcoalitional agreements before. With emission targets, for example, coalitions of countries (e.g. the European Union) may circumvent tougher individual targets by formulating laxer coalitional targets with internal trading markets that are also difficult to evaluate. Coalitions need to be modelled more explicitly.[11]

In this paper, we propose a model where countries set their quantities privately and chose with whom to share their technology. It is shown that sequential agreements born in small local clubs amongst countries with similar technology ultimately integrate more effectively: compared to agreements where all countries negotiate collaboration

[7]Schelling [2002] discusses the architecture of such a plan, noting in particular the problems of participation and compliance.

[8]See Aldy, Barrett & Stavins [2003] for a discussion of earlier proposals.

[9]This is true for improved emissions-trading markets as, for example, proposed by Aldy, Orszag & Stiglitz [2001] and Bradford [2002], and also holds for global carbon tax schemes, as proposed by Cooper [1998] and Nordhaus [1998].

[10]See Harstad [2011b].

[11]In Dutta & Radner [2004], [2009], the EU is considered a single player and not in terms of its individual constituting member countries. A good assumption for their purpose, such an assumption could not be made in general.

at once, countries can be shown to agree to more when they cooperate with countries with similar incentives first. This leads to technology and productivity increases that facilitate subsequent coalition formation. Ultimately, emissions are cut further and global welfare is increased compared to agreements that are immediately negotiated at the global level.

Our results complement the game-theoretic analyses of emission targets and quota trade that have not yet considered coalition formation. More broadly, it offers a novel hybrid cooperative-noncooperative approach compared to the existing game-theoretic models of climate change. As is usual in traditional game theory, this literature can be sorted into cooperative and noncooperative models. On the cooperative side, models exist that either take the coalition structure as given and then provide bounds for the stability of quota transfers, or where, for a given transfer market, stable coalition structures are found.[12] We complement the cooperative strand of climate change game theory models with a non-cooperative argument for which and why coalitions are likely to form. We explain what kinds of emission cuts result from countries privately setting investments and emissions. In the noncooperative strand of the literature, a series of recent papers has explored the incentive compatibility of partial contracts on emissions (Dutta & Radner [2004], [2009], Harstad [2011a], [2011b]). Dutta & Radner [2004], [2009] show that emission quotas can be implemented from now to infinity as Nash equilibria of an emission game over an infinite investment horizon provided countries care sufficiently about the long run. Harstad [2011a], [2011b] adds to this a study of the timing and commitment issues beyond the short run in more complex settings. Of particular interest for us his result that policies facilitating technological investment, which may be easier to legislate in practice, can be used to substitute tougher emission legislation, which may be harder to enforce (Harstad [2011b]). Sacrificing the detailed dynamic timing analysis, we complement this strand of models with a simple game to identify the most effective coalition formation process to understand what kind of outcomes are implemented not only as mutual best replies but also as coalitionally stable agreements. It is an avenue for further research to combine these two strands of game-theoretic analyses in a more complex and dynamic economy. Mechanism designers in practice need to tie these two types of mechanisms (emission quotas and R&D collaboration) further together.

Outside the climate change literature, our model relates to the R&D literature and to some models in international trade. In the R&D literature, the mixed incentives to cooperate in R&D coalitions are analysed, for example, in Amir & Wooders [1997]. Of particular interest is their result that heterogeneity may render joint ventures unstable, whereas joint ventures are stable for more homogeneous partners. This is in line with our findings. Our results also relate to findings concerning coalition formation in international trade. In that literature, an important hypothesis views regional agreements as stepping stones for global agreements (Levy & Srinivasan [1996]).[13] Regional trade unions may create trade liberalisation with positive externalities that ultimately reduce the barriers for global trade agreements.

[12] See Finus [2008] for an overview.

[13] The other side argues that regional agreements are stumbling blocks that will lead to trade wars between rival trade unions, creating negative externalities that prevent the formation of global agreements (see Levy & Srinivasan [1996] for a discussion).

3 Renewable energy game

We first introduce a static two-stage game where a population of countries, $N = \{1, ..., n\}$, first form technology coalitions and then make simultaneous investment choices on their energy mix of traditional fossil consumption and alternative renewable energy. Fossil fuels incur a costly climate-change externality on the environment. The effects of climate change vary according to countries' geographic locations and climatic conditions. Regarding renewable energy collaboration, we assume existence of positive knowledge spillovers between countries that make cooperative efforts to improve their energy industries jointly.

Each investment in either technology represents a discounted stream of consumptions over an infinite horizon, determining future production levels and methods. The resulting utility represents the discounted present value of production minus costs. Given a coalition structure, we assume that this value can be represented by a single real-valued number. For that to be a valid assumption, we implicitly assume that discount factors and uncertainties are taken to be constant over time and that countries, whenever deciding to enter a coalition, base their decisions on the relative present values alone and do not consider further coalitional rearrangements in future periods.

When solving the game, we assume subgame perfection at the individual level: given the coalition structure that is agreed, all quantities are set as best replies. Foreseeing that, individual countries will not enter coalitions that make them worse off. Instead, coalitions of countries form that make all members better off.

The coalition formation game results in a coalition structure ρ, that is, a partition of N. Given ρ, each state i decides to invest f_i in fossil fuels and r_i in renewable energy. The total fossil fuel consumption is $F = \sum_{i \in N} f_i$. Coalitions, $C \subseteq N$, of states may form to share renewable technology and infrastructure. The total renewable-energy level in any C is $R_C = \sum_{i \in C} r_i$. Fossils have a global negative externality effect on all players; renewable-energy consumption has a positive externality effect within coalitions. Given ρ, all states set $\{(f_i, r_i)\}_{i \in N}$, which implies state i's utility as $U_i = U_i(\rho, \{(f_j, r_j)\}_{j \in N}) \;\; \forall i$.

Players. As a proxy for each country's geographic location and climatic conditions, each player $i \in N$ has a **latitudinal factor** $L_i \in \{0, 1, ..., 90\}$, representing the country's mean distance to the closest pole. For example, states have $L_P = 0$ on either pole and $L_E = 90$ on the Equator, state i at $(70°N, 25°E)$ has $L_i = 90 - 70 = 20$.

The production function. We assume the same production function for all states: $P_i = P(f_i + r_i) \;\; \forall i$. Returns to scale in both production inputs diminish; we assume $\frac{\partial P}{\partial f_i} = \frac{\partial P}{\partial r_i} = P'(f_i + r_i) > 0$, $\frac{\partial^2 P}{\partial f_i^2} = \frac{\partial^2 P}{\partial r_i^2} = P''(f_i + r_i) < 0$. f_i and r_i are perfect substitutes; the total energy level, $f_i + r_i$, independent of inner allocation, determines the state's output.

Costs. States' costs of production are additively separable in the two energy inputs: $C_i = C_i^f + C_i^r \;\; \forall i$.

Fossil fuel costs. The cost of fossil fuels is determined by the same function for all states, depending on the latitudinal factor: $C_i^f = C^f(f_i, F, L_i) \;\; \forall i$. The cost of fossil fuels increases in i's own consumption f_i, in total fossil-fuel consumption F as a consequence of demand pressure on fossil resource prices, and in the state's latitudinal location L_i as a consequence of global climate change. The negative externality of F on i rises with L_i; indeed, for all levels of inputs $x \in \{f_i, F, L_i\}$, we assume $\frac{\partial C^f}{\partial x} > 0$, $\frac{\partial^2 C^f}{\partial x^2} > 0$, $\frac{\partial^2 C^f}{\partial F \partial L_i} > 0$.

Renewable energy costs. The cost of renewable energy depends on each country's renewable technology: $C_i^r = C_i^r(r_i, R_C) \quad \forall i$. Heterogeneities amongst the agents are expressed through C_i^r. Note this also allows to express heterogeneities with respect to P and C^f in terms of opportunity costs. For each country, the cost of renewable energy increases in i's own r_i with $\frac{\partial C_i^r}{\partial r_i} > 0$, $\frac{\partial^2 C_i^r}{\partial r_i^2} > 0$. Cost decreases in R_C as members of a non-singleton coalition, C, share renewable technology and infrastructure. We assume increasing cost cuts from cooperation: $\frac{\partial^2 C_i^r}{\partial R_C^2} < 0$.

3.1 Technology coalitions

Before energy investment choices are made countries from population $N = \{1, ..., n\}$ form coalitions $C \subseteq N$ to collaborate on their renewable energy R&D. Each player $i \in N$ specifies one coalition $C_i \subseteq N$. Given the vector of specified coalitions, $\{C_1, ..., C_n\}$, all $C \subseteq N$ form where $C_i = C$ for all $i \in C$. All other countries remain singletons.

3.2 Technology investments

The NTU game. Given any partition ρ of N, the fixed population of states, $N = \{1, ..., n\}$, plays the cooperative renewable energy game, $G(v, N)$, a nontransferable utility game in partition function form. For any ρ, a characteristic function exists such that, $\forall i$, $v(\,\cdot\,; \rho) : \; i \to \mathbb{R}$.

Given ρ, we assume quantities are decided upon privately and as best-replies, $\{(f_i^\rho, r_i^\rho)\}_{i \in N}$, such that $U_i^\rho \in \arg\max_{(f_i, r_i)} U_i(\rho, (f_j, r_i); \{(f_j^\rho, r_j^\rho)\}_{j \neq i}) \quad \forall i$. Given any ρ, each $v(i; \rho)$ is defined by the above best reply: for all i, $v(i; \rho) = U_i^\rho = (\rho, \{(f_j^\rho, r_j^\rho)\}_{j \in N})$.

4 Benchmark solutions

In this section, we analyse a number of benchmark solutions. The analysis in this section shows what kind of agreements cannot be reached as coalitionally stable outcomes when the countries play the renewable energy game and form coalitions.

4.1 World government

We begin our formal analysis with the benchmark of what a utilitarian world government would do if it could dictate coalition structures and each country's investment and emission choices. Such a government maximises $\sum_{i \in N} U_i$ by forming the grand coalition and choosing globally optimal levels of fossil fuel and renewable energy investment for each country. The positive externality effects on the cost of renewable energies and the negative externality effects on the cost of fossil fuels are fully endogenised, thus solving the tragedy of the commons.

Given the grand coalition, $\rho = N$, the first-order derivatives are

$$\frac{\partial \sum_{i \in N} U_i}{\partial f_i} = \frac{\partial P}{\partial f_i} - \frac{\partial C^f}{\partial f_i} - \sum_{j \in N} \frac{\partial C^f}{\partial F}, \tag{1}$$

$$\frac{\partial \sum_{i \in N} U_i}{\partial r_i} = \frac{\partial P}{\partial r_i} - \frac{\partial C_i^r}{\partial r_i} - \sum_{j \in N} \frac{\partial C_j^r}{\partial R_N}. \tag{2}$$

Evaluated at zero for the optimal levels of $\{(f_i^{WG}, r_i^{WG})\}_{i \in N}$ for all countries, the players' marginal individual utilities are

$$\frac{\partial U_i}{\partial f_i} = \frac{\partial P}{\partial f_i} - \frac{\partial C^f}{\partial f_i} - \frac{\partial C^f}{\partial F} = \sum_{j \neq i} \frac{\partial C^f}{\partial F} > 0, \tag{3}$$

$$\frac{\partial U_i}{\partial r_i} = \frac{\partial P}{\partial r_i} - \frac{\partial C_i^r}{\partial r_i} - \frac{\partial C_i^r}{\partial R_N} = \sum_{j \neq i} \frac{\partial C_j^r}{\partial R_N} < 0. \tag{4}$$

Further, given distinct states i and j, (1) implies that

$$P'(f_i + r_i) - P'(f_j + r_j) = \frac{\partial C^f}{\partial f_i} - \frac{\partial C^f}{\partial f_j}. \tag{5}$$

Similarly, (2) implies that

$$P'(f_i + r_i) - P'(f_j + r_j) = \frac{\partial C_i^r}{\partial r_i} - \frac{\partial C_j^r}{\partial r_j}. \tag{6}$$

Hence,

$$\frac{\partial C^f}{\partial f_i} - \frac{\partial C^f}{\partial f_j} = \frac{\partial C_i^r}{\partial r_i} - \frac{\partial C_j^r}{\partial r_j}. \tag{7}$$

$\{(f_i^{WG}, r_i^{WG})\}_{i \in N}$ must solve these various equations simultaneously. The following comparative statics are noted:

If the latitude and the renewable technology of both countries is the same, i.e. $L_i = L_j$ and $C_i^r \equiv C_j^r$, then $f_i^{WG} = f_j^{WG}$ and $r_i^{WG} = r_j^{WG}$; otherwise if, for example, $f_i^{WG} < f_j^{WG}$, then $r_i^{WG} < r_j^{WG}$ by (7), but the RHS of (5) is negative and the LHS positive - a contradiction.

When $L_i > L_j$ and $C_i^r \leq C_j^r$, we must have $f_i^{WG} < f_j^{WG}$. Suppose otherwise; $f_i^{WG} \geq f_j^{WG}$ and the LHS of (7) is positive, implying $r_i^{WG} > r_j^{WG}$, but then the RHS of (5) is positive and the LHS negative. Similarly, if $C_i^r < C_j^r$ and $L_i \geq L_j$, then we must have $r_i^{WG} > r_j^{WG}$. Suppose otherwise; $r_i^{WG} \leq r_j^{WG}$ and the RHS of (7) is negative, implying $f_i^{WG} < f_j^{WG}$, but then the RHS of (6) is negative and the LHS positive.

To summarize, the socially optimal world government outcome can be described as follows. First, controlling for renewable technology, the closer a country is to the equator (greater L_i), the smaller its fossil consumption. Second, controlling for latitude, the better a country's renewable technology, the higher its renewable energy consumption. Third, each country's marginal utility of fossils is positive, implying that countries have an incentive to increase fossil consumption - the world government keeps the negative effects of climate change down to a social optimum. Fourth, each country's marginal utility of renewable energy is negative, implying that countries have an incentive to reduce their renewable energy investment - the world government keeps the technology level up to a social optimum. The optimum is obviously not a Nash equilibrium because each country has a unilateral incentive to set different quantities.

4.2 Status quo or business as usual

In reality, however, there is no world government. As a result, there is little reason to expect the above outcome or any other outcome that maximises a different social welfare function. If instead no coalitions form at all and each country sets its own quantity, each state i takes $F - f_i$ and $R_C = r_i$ as given. Compared to the world optimum, the result is worldwide overinvestment in fossil fuels and underinvestment in renewables because the externalities are not endogenised. Unfortunately, this describes the status quo, that is, business as usual.

Given no coalitions, $\rho = \{(1), ..., (n)\}$, the first-order derivates are

$$\frac{\partial U_i}{\partial f_i} = \frac{\partial P}{\partial f_i} - \frac{\partial C^f}{\partial f_i} - \frac{\partial C^f}{\partial F}, \tag{8}$$

$$\frac{\partial U_i}{\partial r_i} = \frac{\partial P}{\partial r_i} - \frac{\partial C_i^r}{\partial r_i}, \tag{9}$$

evaluated at zero for the optimal levels of $\{(f_i^{NA}, r_i^{NA})\}_{i \in N}$ for all countries: each country sets (f_i^{NA}, r_i^{NA}) as a best reply to all others' $\{(f_j^{NA}, r_j^{NA})\}_{j \neq i}$. Compared to the FOCs of the world-government case, note that the fossil externality is not endogenised and that the knowledge spillover effect on the cost of renewable energy is gone.

Given distinct states i and j, (8) implies that

$$P'(f_i + r_i) - P'(f_j + r_j) = \frac{\partial C^f}{\partial f_i} - \frac{\partial C^f}{\partial f_j}. \tag{10}$$

Similarly, (9) implies that

$$P'(f_i + r_i) - P'(f_j + r_j) = \frac{\partial C_i^r}{\partial r_i} - \frac{\partial C_j^r}{\partial r_j}. \tag{11}$$

Since these are the same as (5) and (6), the comparative statics from the world government case continue to hold. For any level of (f_i, r_i), the RHS of (8) is larger than the RHS of (1), and the RHS of (9) is smaller than the RHS of (2).

The no agreement outcome can be described as follows. First, controlling for renewable technology, the closer a country is to the equator (greater L_i), the smaller its fossil consumption. Second, controlling for latitude, the better a country's renewable technology, the higher its renewable energy consumption. Third, each country's marginal utility of fossils is zero, implying that countries overall use more fossil fuel than optimal. Fourth, each country's marginal utility of renewable energy is zero, implying that countries overall use less renewable energy than optimal.

Because no country can form a coalition on its own and sets is individual quantities as best replies to the others, the no agreement benchmark is a Nash equilibrium. However, we will prove in Section Five that this outcome is coalitionally unstable because coalitions of players exist with joint and unilateral incentives to share R&D. This makes us optimistic that, with the right coalitions forming, a stable improvement on the status quo is possible.

4.3 Kyoto quotas

The basic idea of the Kyoto protocol is (1) to set for each country a series of carbon emission targets for future periods so as to reduce the overall emissions by a global target, and (2) to promote collaboration on renewable energy. Expressed by one period, $K < F^{NA}$, the total cut in fossil consumption necessary to achieve this reduction when all countries collaborate, is supposed to be shared amongst the countries according to a collection of quotas, $(K_i)_{i \in N}$, so that $\sum_{i \in N} K_i = F^{NA} - K$. Suppose N forms.

Kyoto targets.
For each $i \in N$, a Kyoto target K_i is set such that fossil consumption is $f_i \leq K_i$.

Given the constraint, $f_i \leq K_i$, each country may then choose a constrained optimum input, (f_i^K, r_i^K). Overall fossil consumption will be $F^K \leq F^{NA} - K$. Note that, as a result of the overall cut, individual fossil consumption of some given level f_i will be cheaper: for any $K > 0$, $C(f_i; L_i, F^{NA}) > C(f_i; L_i, F^{NA} - K)$ because $\frac{\partial C^f}{\partial F} > 0$.

For the constrained-optimal level of fossil consumption, this implies directly that

$$f_i^K \begin{cases} = K_i & \text{if } f_i^{NA} > K_i \\ \in (f_i^{NA}, K_i] & \text{if } f_i^{NA} \leq K_i. \end{cases}$$

For the constrained-optimal level of renewable consumption, this also implies

$$r_i^K \begin{cases} > r_i^{NA} & \text{if } f_i^{NA} > K_i \\ < r_i^{NA} & \text{if } f_i^{NA} \leq K_i. \end{cases}$$

To see the latter, note that - whilst (8) need no longer equal zero - (9) still implies that

$$P'(f_i + r_i) = \frac{\partial C_i^r}{\partial r_i}. \tag{12}$$

Hence, if $f_i^{NA} > K_i = f_i^K$ and $r_i^K \leq r_i^{NA}$, the LHS of (12) is strictly increased by Kyoto targets whilst the RHS is weakly decreased, a contradiction. And if $f_i^{NA} < f_i^K \leq K_i$ and $r_i^K \geq r_i^{NA}$, the LHS of (12) is strictly decreased by Kyoto targets whilst the RHS is weakly increased, a contradiction.

Globally, except for (just) achieving the targeted reduction of total fossil consumption at the expense of countries with $f_i^{NA} > K_i$, the welfare and growth effects of such quotas are ambiguous compared to the status quo. Most importantly, total investment in renewable technology may rise or fall. If it falls by too much, total welfare may actually decrease for particular sets of preferences. Total welfare is only sure to increase as a result of effective Kyoto targets when it also causes a global increase in investment in renewable technology.

Moreover, any individual country with $f_i^{NA} > K_i$ will either not voluntarily ratify the agreement or disrespect his individual target, since $U_i^{NA}(f_i^{NA}, r_i^{NA}) > U_i^K(f_i^K, r_i^K) \;\; \forall i : f_i^{NA} > K_i$. Ratification of and adherence to a global Kyoto protocol is not a Nash equilibrium unless countries can be seriously constrained in setting their individual quantities. The list of reasons why any given country has incentives that are incompatible with its Kyoto targets include that the country is fossil-rich, does not care enough about the future, is little affected by climate change and/ or has little renewable technology.

4.4 A global R&D agreement

Because Kyoto-type targets are not compatible with the incentives of certain countries, agreements have been considered where countries share renewable technology and set their individual consumption levels freely as mutual best replies.[14] If the effect on renewable technology is large enough, this may already significantly reduce the burden of global climate change due to overuse of fossil fuels compared to the no agreement case. The total welfare effects of such agreements are unambiguously positive compared to the status quo. With the new technology in place, it would be in the best interests of certain countries to cut their fossil-fuel emissions. Sufficiently fossil-rich countries sufficiently far away from the equator or countries with very poor renewable technology, however, may actually consume more fossil as a result of falling overall demand and prices. If this latter effect is too large, this may still cause failure to achieve the global emission-cut target. Furthermore, coalitional instabilities may spoil these arrangements.

Given that the grand coalition forms, the partials are

$$\frac{\partial U_i}{\partial f_i} = \frac{\partial P}{\partial f_i} - \frac{\partial C^f}{\partial f_i} - \frac{\partial C^f}{\partial F}, \tag{13}$$

$$\frac{\partial U_i}{\partial r_i} = \frac{\partial P}{\partial r_i} - \frac{\partial C_i^r}{\partial r_i} - \frac{\partial C_i^r}{\partial R_N}, \tag{14}$$

evaluated at zero for the optimal levels of $\{(f_i^N, r_i^N)\}_{i \in N}$ for all i: each country sets (f_i^N, r_i^N) as a best reply to all others' $\{(f_j^N, r_j^N)\}_{j \neq i}$. This gives the same FOCs for fossil

[14]The EU's efforts to establish a common energy policy with subsidies and feed-in tariffs is an example of such a treaty.

fuel as in the NA case, and renewable energy FOCs similar to the NA case but with a small positive spillover effect from technology transfer.

Once again, given distinct states i and j, (13) implies that

$$P'(f_i + r_i) - P'(f_j + r_j) = \frac{\partial C^f}{\partial f_i} - \frac{\partial C^f}{\partial f_j}, \tag{15}$$

and (14) implies that

$$P'(f_i + r_i) - P'(f_j + r_j) = \frac{\partial C_i^r}{\partial r_i} - \frac{\partial C_j^r}{\partial r_j}, \tag{16}$$

and the comparative statics continue to hold. For any level of (f_i, r_i), the RHS of (13) lies in between the RHSs of (3) and (8), the RHS of (2) in between the RHSs of (4) and (8).

The grand coalition outcome can be described as follows. First, controlling for renewable technology, the closer a country is to the equator (greater L_i), the smaller its fossil consumption. Second, controlling for latitude, the better a country's renewable technology, the higher its renewable consumption. Third, each country's marginal utility of fossils is zero, implying that countries overall use more fossil fuel than optimal but less than in the no-agreement case. Fourth, each country's marginal utility of renewables in the grand coalition is zero, implying that countries overall produce less renewable energy than optimal, though more than in the no agreement case.

However, if the agreement is tied to emission reductions, some countries may deviate because they actually prefer to increase their consumption of fossils as a result of an agreement being reached in the rest of society. A polar country with $L_i = 0$ (unaffected by climate change) and without any renewable technology, for example, has $U_i = P(f_i) - C^f(f_i, F)$. Given any ρ forms, such a country maximises utility by setting $\frac{\partial U_i}{\partial f_i} = P'(f_i) - \frac{\partial C^f}{\partial f_i} - \frac{\partial C^f}{\partial F}$ to zero at f_i^ρ given $\sum_{j \neq i} f_i^\rho$, implying that the optimal consumption of fossils rises as $\sum_{j \neq i} f_i^\rho$ falls. Such a country contributes nothing to the positive externality on renewable technology, and acts in the same way whether included in a coalition or not. Similarly, consider a country i with very poor renewable technology that sets (f_i^{NA}, r_i^{NA}) in the no agreement case such that f_i^{NA} is very high and r_i^{NA} very low. When N forms, i may again set a very low r_i^N but now a significantly higher f_i^N as a result of the overall fall in total fossil demand of the others. Compared to the Kyoto target, $F^{NA} - K$, F^N may still be too high: it is possible that $F^{NA} > F^N > F^{NA} - K$. Furthermore, countries for which $\frac{\partial U_i}{\partial f_i} = P'(f_i + r_i^N) - \frac{\partial C^f}{\partial f_i} - \frac{\partial C^f}{\partial F} > 0 \;\; \forall f_i \leq K_i$, will not voluntarily ratify or obey any imposed quota. Their incentive is to increase fossil consumption.

Without quotas, such a global agreement is a Nash equilibrium because no individual country is better off by setting different quantities and no country prefers to lose the positive effect of forming the knowledge-transfer coalition. Whether the outcome is actually coalitionally stable, however, depends on the precise specifications. Note that if, for any subcoalition $S \subset N$, a profitable deviation exists such that $v(S; \rho) > \sum_{i \in S} v(i; N)$ for all $\rho : S \subset \rho$, the above global agreement may break down. In the next section, we will show that these global agreements are generally unstable.

5 Stable R&D coalition formation

The R&D treaty we describe at the end of the last section has two main problems. First, it may not be core-stable. Second, even if it is stable, it may result in less than sufficient emission reductions due to the fact that some countries continue to overuse fossil resources. In this section, we show that, ceteris paribus, "similar" countries will collaborate locally and reduce their fossil consumption in the absence of global agreements. Thereafter, global agreements are easier to reach in subsequent steps of coalition formation and will result in greater emission reductions than are achievable when a global agreement is sought first. This has significant implications for mechanism design: (1) Clubs of similar countries should form first. (2) Once countries are integrated in local clubs, global agreements result in stable agreements with superior outcomes.

5.1 Leading the way

We sort countries by "pools" of countries with the same latitudinal factors and the same renewable energy technologies.

Country Pool. *Pool $P \subset N$ consists of all i such that $L_i = L$ and $C_i^r = C^r$ for all $i \in P$.*

We now show that, in any such pool, core-stable local coalitions ("clubs") exist. These clubs can be achieved by subgame perfect Nash equilibrium play of the renewable energy game.

Proposition 1. *Suppose that, within any pool P, two agents form a coalition. If their production does not fall too much, their renewable-energy costs fall sufficiently and other countries do not raise their fossil investment too much, then there exists a coalition $M \subseteq P$ such that the subgame of $G(v, N)$ over population M has a nonempty core.*

Proof. In the following proof, for any $C \subseteq S \subseteq N$, the expression $v_{\rho_{N\setminus S}}(C)$ is shorthand for $\max_{\rho_{S\setminus C}} v(C; \rho_{S\setminus C}, \rho_{N\setminus S})$. To prove nonemptiness of the core of M, we will show convexity of $v_{\rho_{N\setminus S}}$.[15]

First note that, by symmetry, $f_i^S = f^S$ and $r_i^S = r^S$ for all i belonging to some $S \subseteq P$. For convexity of $v_{\rho_{N\setminus M}}$, we require that, $\forall S \subseteq T \subseteq M\setminus\{i\}$, $\forall i \in M$,

$$v_{\rho_{N\setminus M}}(S \cup \{i\}) - v_{\rho_{N\setminus M}}(S) \leq v_{\rho_{N\setminus M}}(T \cup \{i\}) - v_{\rho_{N\setminus M}}(T)$$
$$\Leftrightarrow \quad (|S|+1)\left(P\left(f^{S\cup\{i\}}, r^{S\cup\{i\}}\right) - C^f\left(f^{S\cup\{i\}}, F^{S\cup\{i\}}, L\right) - C^r\left(r^{S\cup\{i\}}, R_{S\cup\{i\}}\right)\right)$$

[15] Indeed, a variety of definitions of the core of partition function games applies, depending on the conjecture of each $C \subset S$ regarding $\rho_{S\setminus C}$ (Hafalir [2007]). If, given $v_{\rho_{N\setminus S}}(C) = \max_{\rho_{S\setminus C}} v(C; \rho_{S\setminus C}, \rho_{N\setminus S})$ for all $C \subset S$, the core is nonempty, however, all these cores will also be nonempty.

$$
\begin{aligned}
&-|S|\left(P\left(f^S,r^S\right)-C^f\left(f^S,F^S,L\right)-C^r\left(r^S,R_S\right)\right)\leq\\
&(|T|+1)\left(P\left(f^{T\cup\{i\}},r^{T\cup\{i\}}\right)-C^f\left(f^{T\cup\{i\}},F^{T\cup\{i\}},L\right)-C^r\left(r^{T\cup\{i\}},R_{T\cup\{i\}}\right)\right)\\
&-|T|\left(P\left(f^T,r^T\right)-C^f\left(f^T,F^T,L\right)-C^r\left(r^T,R_T\right)\right).
\end{aligned}
$$

Supposing that $T=S\cup\{k\}$ for an arbitrary $k\in P$, this becomes

$$
\begin{aligned}
&|T|\left(P\left(f^T,r^T\right)-C^f\left(f^T,F^T,L\right)-C^r\left(r^T,R_T\right)\right)\\
&-|S|\left(P\left(f^S,r^S\right)-C^f\left(f^S,F^S,L\right)-C^r\left(r^S,R_S\right)\right)\leq\\
&(|T|+1)\left(P\left(f^{T\cup\{i\}},r^{T\cup\{i\}}\right)-C^f\left(f^{T\cup\{i\}},F^{T\cup\{i\}},L\right)-C^r\left(r^{T\cup\{i\}},R_{T\cup\{i\}}\right)\right)\\
&-|T|\left(P\left(f^T,r^T\right)-C^f\left(f^T,F^T,L\right)-C^r\left(r^T,R_T\right)\right)\\
\Leftrightarrow\quad &(|T|+1)\left(P\left(f^{T\cup\{i\}},r^{T\cup\{i\}}\right)-C^f\left(f^{T\cup\{i\}},F^{T\cup\{i\}},L\right)-C^r\left(r^{T\cup\{i\}},R_{T\cup\{i\}}\right)\right)\\
&+|S|\left(P\left(f^S,r^S\right)-C^f\left(f^S,F^S,L\right)-C^r\left(r^S,R_S\right)\right)\\
&-2|T|\left(P\left(f^T,r^T\right)-C^f\left(f^T,F^T,L\right)-C^r\left(r^T,R_T\right)\right)\geq 0\\
\Leftrightarrow\quad &(|T|+1)\Big(\left(P\left(f^{T\cup\{i\}},r^{T\cup\{i\}}\right)-P\left(f^T,r^T\right)\right)\\
&+\left(C^f\left(f^T,F^T,L\right)-C^f\left(f^{T\cup\{i\}},F^{T\cup\{i\}},L\right)\right)+\left(C^r\left(r^T,R_T\right)-C^r\left(r^{T\cup\{i\}},R_{T\cup\{i\}}\right)\right)\Big)\geq\\
&(|T|-1)\Big(\left(P\left(f^T,r^T\right)-P\left(f^S,r^S\right)\right)\\
&+\left(C^f\left(f^S,F^S,L\right)-C^f\left(f^T,F^T,L\right)\right)+\left(C^r\left(r^S,R_S\right)-C^r\left(r^T,R_T\right)\right)\Big).
\end{aligned}
$$

If $S=\emptyset$, the RHS is zero. Since $f^T>f^{T\cup\{i\}}$ and $F^T>F^{T\cup\{i\}}$, the middle term of the LHS is positive. The expression will then hold if, as a result of i joining T, production does not fall too much, renewable-energy costs fall sufficiently and fossil-fuel costs do not rise too much. If $S\neq\emptyset$, the condition requires some combination of production not falling too much more, and fossil-fuel and renewable-energy costs not changing too differently, as (the larger) T gets an additional member versus when S does. Other things equal, the condition becomes harder to satisfy the larger the coalition T. Until the condition fails, however, we can apply induction to achieve convexity on the relevant coalition M. □

Proposition 1 proves that clubs in all of the pools in society can form stably at the same time. The resulting partition ρ' contains a number of these clubs $M\subseteq P$. M may include some or all of the members of P. All members in the club now have better renewable technology and invest more in renewable energy. Overall, the use of fossil fuels falls.

5.2 Follow-up coalitions

All $C\subset N$ that form as technology clubs are stable. Now, with the population consisting of $N'=\rho'$ and $G(v,N')$ accordingly defined, further coalitions may form by playing the same coalition formation game on N' again. Individual nations continue to set best-reply quantities as a response to the coalitions that form. Now clubs of countries in the pools of countries with the same latitude and technology have formed. Absent clubs, as argued in earlier sections, the game overall for N may not be convex. With clubs, however, the renewable technology overall is now improved, which increases the chance for convexity of larger coalitions in N'. We now show that these coalitions may extend in a core-stable way beyond existing clubs and pools if sufficient technology gain is achieved in the clubs.

Proposition 2. *Suppose that two clubs from different pools form a coalition. If the clubs are sufficiently large, their production does not fall too much, their renewable-energy costs fall sufficiently and other countries do not raise their fossil investment too much, then there exists a coalition $M \subseteq N'$ such that the subgame of $G(v, N')$ over population M has a nonempty core.*

Proof. For convenience, we drop the superscripts and write N for N'. We need some more notation: as before, for any $C \subseteq S \subseteq N$, the expression $v_{\rho_{N\setminus S}}(C)$ is shorthand for $\max_{\rho_{S\setminus C}} v(C; \rho_{S\setminus C}, \rho_{N\setminus S})$. To prove nonemptiness of the core of M, we will show convexity of $v_{\rho_{N\setminus S}}$. But to do so we also need to define, for any $C \subseteq S \subseteq N$, the expressions $M^C_{\rho_{N\setminus S}}(i)$ and $m^C_{\rho_{N\setminus S}}(i)$ as shorthand for $\max_{i\in C} v(i; \rho_{S\setminus C}, \rho_{N\setminus S})$ and $\min_{i\in C} v(i; \rho_{S\setminus C}, \rho_{N\setminus S})$ respectively.

For convexity of $v_{\rho_{N\setminus M}}$, we require that, $\forall S \subseteq T \subseteq M\setminus\{i\}$, $\forall i \in M$,

$$\begin{aligned}
v_{\rho_{N\setminus M}}(S \cup \{i\}) - v_{\rho_{N\setminus M}}(S) &\leq v_{\rho_{N\setminus M}}(T \cup \{i\}) - v_{\rho_{N\setminus M}}(T) \\
\Leftarrow (|S|+1)(M^{S\cup i}_{\rho_{N\setminus S}}(i) - m^S_{\rho_{N\setminus S}}(i)) &\leq |S|(m^{T\cup i}_{\rho_{N\setminus S}}(i) - M^T_{\rho_{N\setminus S}}(i))
\end{aligned}$$

Let $f^S_{\min} \triangleq \min_{i\in S} f^S_i$, $f^S_{\max} \triangleq \max_{i\in S} f^S_i$, $r^S_{\min} \triangleq \min_{i\in S} r^S_i$, $r^S_{\max} \triangleq \max_{i\in S} r^S_i$, $L^S_{\min} \triangleq \min_{i\in S} L_i$, and $L^S_{\max} \triangleq \max_{i\in S} L_i$.

Then, the above equations hold if

$$\begin{aligned}
(|S|+1)\left(P\left(f^{S\cup\{i\}}_{\max}, r^{S\cup\{i\}}_{\max}\right) - C^f\left(f^{S\cup\{i\}}_{\min}, F^{S\cup\{i\}}, L_{\min}\right) - C^r\left(r^{S\cup\{i\}}_{\min}, R_{S\cup\{i\}}\right)\right) \\
-|S|\left(P\left(f^S_{\min}, r^S_{\min}\right) - C^f\left(f^S_{\max}, F^S, L_{\max}\right) - C^r\left(r^S_{\max}, R_S\right)\right) \leq \\
(|T|+1)\left(P\left(f^{T\cup\{i\}}_{\min}, r^{T\cup\{i\}}_{\min}\right) - C^f\left(f^{T\cup\{i\}}_{\max}, F^{T\cup\{i\}}, L_{\max}\right) - C^r\left(r^{T\cup\{i\}}_{\max}, R_{T\cup\{i\}}\right)\right) \\
-|T|\left(P\left(f^T_{\max}, r^T_{\max}\right) - C^f\left(f^T_{\min}, F^T, L_{\min}\right) - C^r\left(r^T_{\min}, R_T\right)\right).
\end{aligned}$$

Supposing that $T = S \cup \{k\}$ for an arbitrary $k \in M$, this becomes

$$\begin{aligned}
\Leftrightarrow \quad (|T|+1)\left(P\left(f^{T\cup\{i\}}_{\min}, r^{T\cup\{i\}}_{\min}\right) - C^f\left(f^{T\cup\{i\}}_{\max}, F^{T\cup\{i\}}, L_{\max}\right) - C^r\left(r^{T\cup\{i\}}_{\max}, R_{T\cup\{i\}}\right)\right) \\
+|S|\left(P\left(f^S_{\min}, r^S_{\min}\right) - C^f\left(f^S_{\max}, F^S, L_{\max}\right) - C^r\left(r^S_{\max}, R_S\right)\right) \\
-2|T|\left(P\left(f^T_{\max}, r^T_{\max}\right) - C^f\left(f^T_{\min}, F^T, L_{\min}\right) - C^r\left(r^T_{\min}, R_T\right)\right) \geq 0 \\
\Leftrightarrow \quad (|T|+1)\Big(\left(P\left(f^{T\cup\{i\}}_{\min}, r^{T\cup\{i\}}_{\min}\right) - P\left(f^T_{\max}, r^T_{\max}\right)\right) \\
+\left(C^f\left(f^T_{\min}, F^T, L_{\min}\right) - C^f\left(f^{T\cup\{i\}}_{\max}, F^{T\cup\{i\}}, L_{\max}\right)\right) + \left(C^r\left(r^T_{\min}, R_T\right) - C^r\left(r^{T\cup\{i\}}_{\max}, R_{T\cup\{i\}}\right)\right)\Big) \geq
\end{aligned}$$

$$(|T|-1)\Big(\left(P\left(f^T_{\max},r^T_{\max}\right)-P\left(f^S_{\min},r^S_{\min}\right)\right)$$
$$+\left(C^f\left(f^S_{\max},F^S,L_{\max}\right)-C^f\left(f^T_{\min},F^T,L_{\min}\right)\right)+\left(C^r\left(r^S_{\max},R_S\right)-C^r\left(r^T_{\min},R_T\right)\right)\Big).$$

Without heterogeneity, this is the same condition as we had in the proof of Proposition 1, but the presence of heterogeneity may now cause it to fail where previously it held. However, sufficiently large clubs can compensate for this by increasing the difference between the third pair of terms on each side, since $\partial^2 C^r_i/\partial R^2_C < 0$ by assumption. □

Thus, it may well be the case that, absent clubs, countries' renewable energy technologies are insufficiently good to deliver coalitionally stable global agreements. In the presence of sufficiently technology-improving clubs, however, such agreements exist and follow-up agreements may be possible.

References

J. E. Aldy, S. Barrett & R. N. Stavins, "Thirteen Plus One: A Comparison of Global Climate Policy Architectures," *Climate Policy 3*, 373-397, 2003.

J. E. Aldy, P. R. Orszag, & J. E. Stiglitz, "Climate Change: An Agenda for Global Collective Action," *The Timing of Climate Change Policies; conference working paper*, Pew Center on Global Climate Change, 2001.

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

J. E. Aldy & R. N. Stavins (eds.), *Post-Kyoto International Climate Policy*, Cambridge University Press, 2009.

J.-C. Altamirano-Cabrera & M. Finus, "Permit Trading and Stability of International Climate Agreements," *Journal of Applied Economics 9*, 19-48, 2006.

R. Amir & J. Wooders, "Cooperation vs. Competition in R&D," *Zeitschrift fur Nationalokonomie 67*, 63-73, 1998.

S. Barrett, "Towards a Better Climate Treaty," *World Economics 3*, 35-45, 2001.

S. Barrett, *Environment and Statecraft*, Oxford University Press, 2003.

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

V. Bosetti & J. Frankel, "Sustainable Cooperation in Global Climate Policy: Specific Formulas and Emission Targets to Build on Copenhagen and Cancun," *Discussion Paper 2011-46,* Harvard Project on Climate Agreements, Belfer Center for Science and International Affairs, Harvard Kennedy School, 2011.

D. F. Bradford, "Improving on Kyoto: Greenhouse Gas Control as the Purchase of a Global Public Good," *Working paper series*, Princeton University, 2002.

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

R. Cooper, "Toward a Real Treaty on Global Warming," *Foreign Affairs 77*, 66-79, 1998.

A. Datta & E. Somanathan, "Climate Policy and Innovation in the Absence of Commitment," *Discussion Paper 2011-45,* Harvard Project on International Climate Agreements, Belfer Center for Science and International Affairs, Harvard Kennedy School, 2011.

P. K. Dutta & R. Radner, "Self-Enforcing Climate Change Treaties," *Proceedings of the National Academy of Sciences of the United States of America 101,* 5174-5179, 2004.

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

M. Finus, "Game Theoretic Research on the Design of International Environmental Agreements: Insights, Critical Remarks and Future Challenges," *International Review of Environmental and Resource Economics 2,* 29-67, 2008.

I. Hafalir, "Efficiency in Coalition Games with Externalities," *Games and Economic Behavior 61,* 242-258, 2007.

B. Harstad, "Climate Contracts: A Theory of Emissions, Investments, Negotiations and Renegotiations," mimeo, Northwestern University, 2011a.

B. Harstad, "The Dynamics of Climate Agreements," mimeo, Northwestern University, 2011b.

P. Levy & T. N. Srinivasan, "Regionalism and the (Dis)Advantage of Dispute-Settlement Access," *AEA Papers and Proceedings 86*, 93-98, 1996.

H. H. Nax, "Cooperation in Multiple Spheres of Interaction," *Economics Series Working Papers*, University of Oxford, 2008.

W. D. Nordhaus, "Is the Kyoto Protocol a Dead Duck? Are There Any Live Ducks Around? Comparison of Alternative Global Tradable Emissions Regimes," *Working Paper, Department of Economics,* Yale University, 1998.

T. C. Schelling, "What Makes Greenhouse Sense? Time to Rethink the Kyoto Protocol," *Foreign Affairs 81*, 2-9, 2002.

R. N. Stavins, "The Problem of the Commons: Still Unsettled After 100 Years," *Discussion Paper 2010-43*, Harvard Project on International Climate Agreements, Belfer Center for Science and International Affairs, Harvard Kennedy School, 2010.