WORKING PAPERS

The value of commitment and delegation For the control of greenhouse gas emissions

Paul Pichler
Gerhard Sorger

June 2016

Working Paper No: 1604

DEPARTMENT OF ECONOMICS

UNIVERSITY OF VIENNA

All our working papers are available at: http://mailbox.univie.ac.at/papers.econ

The value of commitment and delegation for the control of greenhouse gas emissions*

Paul PICHLER[a,b] and **Gerhard SORGER**[a]

June 2016

Abstract: We analyze a stylized model of a world consisting of a large number of countries, which derive utility from energy consumption but suffer both from the emission of greenhouse gases (smog, black carbon, etc.) as well as from the external effects caused by climate change. The countries decide individually on investments in clean (i.e., emission free) technologies for energy production, whereas a supranational environmental authority decides for each country on the maximally permitted amount of emissions of greenhouse gases. We demonstrate that the authority faces a dynamic inconsistency problem that leads to welfare losses. Yet these welfare losses can be kept small if the mandate for the authority penalizes the local cost of emissions very heavily but puts little or no weight at all on the cost of climate change.

Journal of Economic Literature **classification codes:** F53, H87, O33, O44, Q43, Q54,

Key words: Climate change; supranational environmental authority; dynamic inconsistency; optimal delegation.

---

* The authors would like to thank Bård Harstad and Franz Wirl for helpful discussions and comments.

[a] Department of Economics, University of Vienna, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria

[b] Economic Studies Division, Austrian National Bank, Otto-Wagner-Platz 3, A-1011 Vienna, Austria

1 Introduction

Many scientists today consider climate change as the biggest threat humanity has ever faced. It will lead to severe disruptions of everyday life around the globe and it is an extremely difficult problem to tackle. The difficulties arise primarily for two reasons. First, climate change is due to strong negative external effects of economic activity (in particular energy production and transportation) which prevent its solution by market mechanisms.[1] Second, climate change is an inherently global phenomenon that can only be addressed on a global basis. To be effective, measures against climate change have to be implemented on a broad international scale, which requires difficult and protracted negotiations between many heterogeneous countries with very diverse goals; see, e.g., Barrett (2003) or Nordhaus (2013). In the past, such negotiations often had limited success or failed altogether.[2] Against this background, there have been pleas to create a supranational environmental authority (SEA) with the explicit mandate to fight climate change, and to delegate decision power over certain climate-relevant policies to this authority; see, e.g., Esty (1994), Runge (1994), Hempel (1996), or Yadav (2014).

This paper takes the suggestion to establish an SEA with the power to impose emission limits on the countries of the world as its starting point and analyzes a few issues that would have to be taken into account in order to make such an authority function well.[3] First we argue that, even if the SEA has global welfare as its mandate and all countries respect its policy prescriptions, the authority will face a serious commitment problem. To this end we show that the optimal policy of a benevolent SEA is in general dynamically inconsistent. Discretionary behavior of the authority causes under-investment in clean technologies, over-pollution and, accordingly, a welfare loss relative to the situation when it has commitment power. We then demonstrate that the dynamic inconsistency problem can be ameliorated by giving the SEA a goal that differs from the maximization of global welfare.[4] We find that an SEA that attaches *more* weight to

[1] Stern (2007, p. 1) refers to climate change as "the greatest example of market failure we have ever seen".

[2] Barrett (1994) argues that international environmental agreements need to be self-enforcing and, therefore, they are not likely to be signed by many countries unless the global benefits of cooperation are small.

[3] An alternative interpretation of our model is that of an environmental protection agency (EPA) which has the power to impose emission standards on firms on a national level. Due to the global nature of climate change, however, we prefer the to think of the authority as a supranational one.

[4] The idea of optimal delegation is borrowed from the literature on monetary policy, where dynamic incon-

the *local* costs of emissions (stemming from smog, polluted water, black carbon, etc.) than the individual countries themselves but *less or even zero* weight to the damage caused by the *global* stock of greenhouse gases, can implement a discretionary policy with significantly smaller welfare losses than a benevolent SEA. This seemingly counter-intuitive result obtains because an authority with these preferences provides stronger incentives for the individual countries to invest in clean technologies. We also prove that the optimal mandate of the authority depends on the initial level of global pollution: the higher is the stock of greenhouse gases, the more weight should be given to the local cost of emissions. This creates another dynamic inconsistency problem, this time on the side of the individual countries: even if they can commit to set up an SEA and to respect its policy prescriptions, they have a recurrent incentive to change the authority's mandate as the stock of greenhouse gases accumulates.

The framework in which we study the above mentioned issues is explained in section 2. It is a modification of the model used by Harstad (2012) and Harstad (2016), which was developed to analyze international environmental agreements.[5] Our model describes the world as consisting of many small countries, which derive utility from consuming energy and which suffer both from local emissions of greenhouse gases and from the consequences of climate change. Energy can be produced either by dirty technologies creating greenhouse gas emissions (e.g., power plants based on fossil fuels) or by clean technologies that do not generate such emissions (e.g., power plants using waterpower, solar energy, or wind power). In order to assess the value of commitment and delegation we start in section 3 by considering two extreme benchmarks: the first-best solution and business as usual (BAU). The first-best solution is the allocation that would be implemented by a social planner who makes all investment and emission decisions for all countries and seeks to maximize aggregate welfare in the world. BAU, on the other hand, describes the situation in which all individual countries pursue their own interests without any coordination among them.

sistency problems can be extenuated by augmenting the central bank's mandate by strong inflation aversion, as in Rogoff (1985), or by a desire for policy inertia, as in Woodford (2003). In the context of environmental policy, optimal delegation has been studied by Helm et al. (2004); see below.

[5] Battaglini and Harstad (2016) study participation and duration of environmental agreements in a closely related modelling framework.

Suppose now that there exists an SEA which sets the (maximum) emission levels, whereas the countries themselves decide on how much they invest into clean technologies. In section 4 we assume that the authority is *benevolent*, that is, it maximizes aggregate welfare in the world. In our setup it turns out that such an authority can implement the first-best solution provided it has *full commitment power*. In other words, by announcing the optimal emission levels at time 0 and by credibly committing to enforce these levels, the countries are given the correct incentives for investment in clean technologies. This property follows from the fact that the environmental externality operates solely through the stock of greenhouse gases, which does not depend on the technology stocks but only on the amount of emissions. The more interesting case, however, is when the authority lacks commitment power and acts under *discretion*. Maintaining the assumption of benevolence of the SEA, we show that it looses its ability to provide the correct incentives for investment. As a matter of fact, the countries choose the same technology stocks as under BAU, which fall short of the technology levels in the first-best solution. Given that less clean technology is available than in the first-best solution, it is not optimal for the SEA to impose the tight emission limits from the first-best solution. Taken together, these effects result in a welfare loss relative to the case of commitment.

In section 5 we drop the assumption of benevolence and turn to the question of *optimal delegation*. More specifically, we demonstrate that it can be advantageous to endow the authority with preferences that attach higher weight to the local costs of emissions rather than to the global costs of the stock of greenhouse gases in the atmosphere. The reason for this seemingly counter-intuitive result is that an authority with these preferences provides stronger incentives for the individual countries to invest in clean technologies which, in turn, mitigates the dynamic inconsistency problem of the authority and thereby moves the solution closer to the first-best allocation. One could also interpret our finding in the way that putting more weight on the cost of emissions (as compared to the cost of climate change) creates a more direct incentive for the authority to curb emissions.

Interestingly, the optimal relative weight of local costs (optimal with regard to global welfare) depends on the state of the model, i.e., on the initial stock of greenhouse gases in the atmosphere. A higher initial concentration of greenhouse gases would require a higher weight on local costs. This creates a new problem regarding the design of the SEA's mandate. Either one has to allow that the mandate is specified in terms of a *rule* that describes the goal of the SEA in state-dependent form, or there arises another *dynamic inconsistency* in the sense that the assembly of countries wants to change the mandate of the SEA as the concentration of greenhouse gases evolves. We address this issue in supplementary material contained in an online appendix. It turns out that the welfare loss due to this (second) dynamic inconsistency problem of the countries is very small relative to the one affecting the SEA (which we analyze in the main part of the paper).

It is worthwhile to discuss the role of local costs of greenhouse gas emissions both in reality and in our model. The most prominent greenhouse gas is carbon dioxide (CO_2), which is also the most plentiful after water vapour. CO_2 itself has no negative effects at the location where it is emitted but it stays in the atmosphere for very long and causes global warming (the global cost). However, many economic activities that generate CO_2 typically also lead to other emission that have short-lived but local effects, e.g., methane, hydrofluorocarbons, sulphur, or black carbon (soot). The policy prescription of our analysis is that endowing the SEA with a mandate that gives high weight to these local costs helps the authority to curb CO_2 emissions and their negative global consequences. In addition to that, it is known that restricting the short-lived greenhouse gas emissions would have noticeable direct effects on climate change; see, e.g., *The Economist* (2015ab).

Dynamic inconsistency of environmental policy has been studied in a variety of settings. Gersbach and Glazer (1999) point out that dynamic inconsistency creates a hold-up problem for firms, which can be solved by a introducing a market for emission permits. Petrakis and Xepapadeas (2003) study the location decisions of a monopolistic firm that is subject to emission taxes in its home country. Among other things they show that the government's commitment to an emission tax is not necessarily welfare improving. D'Amato and Dijkstra (2015) analyze firms' investment decisions in clean technologies under asymmetric information about costs of abatement with the new technology. They show that a regulator who sets emission taxes or emission permits cannot implement the first-best solution under commitment, but that this can be done with a time consistent discretionary policy. All of these papers, however, use a modelling framework that is very different from the one in the present paper, and they do not address the question of optimal delegation at all. The possibility of solving a dynamic inconsistency problem of environmental policy by delegation to a non-benevolent authority has been investigated by Helm et al. (2004). These authors study emission taxation of firms in a very simple, essentially static model and find that delegation to an "environmentalist policy maker" is beneficial to the society. The environmentalist policy maker attaches higher weight than the society to the disutility from pollution. Due to its simple structure, however, the model in Helm et al. (2004) does not capture the stock externality of greenhouse gas emissions at all, nor does it distinguish between local and global costs. The present paper, in contrast, analyzes the design of an SEA in a fully dynamic framework featuring the accumulation of greenhouse gases in the atmosphere and the resulting negative stock externality.

2 The economic environment

We consider a world consisting of a unit interval of infinitesimally small countries $i \in [0, 1]$. Energy can be produced either by a clean technology or by a dirty one. Only the latter generates greenhouse gas emissions. We denote by $R_{i,t}$ the stock of clean technology available in country i at time t and by $g_{i,t}$ the amount of emissions generated by this country in period t. We choose units of measurement in such a way that $R_{i,t} + g_{i,t}$ denotes the total amount of energy available to country i in period t. Furthermore, we denote by $r_{i,t}$ country i's investment in the clean technology in period t. This implies that

$$R_{i,t} = r_{i,t} + \rho R_{i,t-1}, \tag{1}$$

where $1 - \rho \in (0, 1)$ is the rate of depreciation of the technology stock. An initial value $R_{i,-1}$ is exogenously given for every country $i \in [0, 1]$.

The total stock of greenhouse gases in the atmosphere at time t is denoted by G_t. It is assumed that

$$G_t = \int_0^1 g_{i,t} \, \mathrm{d}i + \gamma G_{t-1}, \tag{2}$$

where $1 - \gamma \in (0, 1)$ is the natural decay rate of greenhouse gases. The initial stock G_{-1} is exogenously given.

The utility derived by country i during period t is measured by

$$U(G_t, R_{i,t}, g_{i,t}, r_{i,t}) = B(R_{i,t} + g_{i,t}) - C(g_{i,t}) - D(G_t) - kr_{i,t},$$

where $B(R_{i,t} + g_{i,t})$ is the benefit derived from energy consumption, $C(g_{i,t})$ measures the cost of (local) emissions generated by country i, $D(G_t)$ denotes the cost generated by the (global) stock of greenhouse gases, and k is the constant unit cost of investment. Note that all countries have the same utility function but that the countries can differ from each other with respect to their initial technology levels. We assume that $B : \mathbb{R} \mapsto \mathbb{R}$ is concave and smooth, that $C : \mathbb{R} \mapsto \mathbb{R}$ and $D : \mathbb{R} \mapsto \mathbb{R}$ are strictly convex and smooth, and that k is a positive constant. Denoting the common time-preference factor of all countries by $\beta \in (0, 1)$, we can write total discounted welfare derived by country i as

$$\sum_{t=0}^{+\infty} \beta^t U(G_t, R_{i,t}, g_{i,t}, r_{i,t}) \tag{3}$$

and aggregate world-welfare as

$$\sum_{t=0}^{+\infty} \beta^t \int_0^1 U(G_t, R_{i,t}, g_{i,t}, r_{i,t}) \,\mathrm{d}i. \tag{4}$$

The setup described above is very similar to that used by Harstad (2012, 2016). In particular, we use an additively separable utility function and a linear investment cost function. This specification ensures high analytical tractability. The only two differences between the model described above and that in Harstad (2012, 2016) are (i) that we assume a continuum of countries whereas Harstad (2012, 2016) considers a world consisting of finitely many countries and (ii) that the countries in our model face costs of local emissions (described by the function C) which are absent in Harstad (2012, 2016). Besides the fact that the emission of greenhouse gases typically generates local costs (as has been argued in the introduction), there is also a technical reason why we include a local cost of emission. Contrary to the framework studied by Harstad (2012, 2016), in our model there is no strategic interaction between the countries because each one is infinitesimally small. If the global cost of climate change were the only negative consequence of emissions, all countries would like to increase their emissions and, consequently, their benefits from energy consumption without bound, because the global cost of climate change is a purely external effect. The local costs of emitting greenhouse gases serves as a limiting factor to energy production and thereby ensures that an equilibrium exists. Finally, note that from section 4 onwards, we will introduce a supranational environmental authority, which is also absent from Harstad (2012, 2016).

3 Benchmarks

In this section we describe two benchmarks that will be useful for assessing the value of commitment and delegation in the main part of the paper. The first benchmark describes business as usual, a situation in which all countries pursue their own interests and there is no coordination among them whatsoever. For the second benchmark we assume that there exists a social planner who chooses allocations in such a way that aggregate welfare is maximized. These two benchmarks correspond to no internalisation of the environmental externality and to full internalisation, respectively.

3.1 Business as usual

Suppose that all countries act separately in their own interest. The countries do neither coordinate their actions nor is there any institution that facilitates such a coordination. Since every country is infinitesimally small, it disregards its influence on the global stock of greenhouse gases. In other words, it neglects (2) and treats the stocks of greenhouse gases $(G_t)_{t=0}^{+\infty}$ as exogenous to its own decision problem. We refer to the resulting equilibrium as business as usual (BAU-equilibrium).

Definition 1 A BAU-equilibrium consists of a global sequence $(G_t)_{t=0}^{+\infty}$ and individual sequences $\{(R_{i,t}, g_{i,t}, r_{i,t})_{t=0}^{+\infty} \mid i \in [0,1]\}$ such that the following two conditions are satisfied:
(i) Given $(G_t)_{t=0}^{+\infty}$ it holds for all $i \in [0,1]$ that $(R_{i,t}, g_{i,t}, r_{i,t})_{t=0}^{+\infty}$ maximizes (3) subject to (1).
(ii) Equation (2) holds for all $t \geq 0$.

Lemma 1 *The unique BAU-equilibrium is given by*

$$g_{i,t} = g^{\mathrm{BAU}}, \tag{5}$$

$$R_{i,t} = R^{\mathrm{BAU}}, \tag{6}$$

$$G_t = \frac{g^{\mathrm{BAU}}}{1-\gamma} + \left(G_{-1} - \frac{g^{\mathrm{BAU}}}{1-\gamma}\right)\gamma^{t+1} \tag{7}$$

for all $i \in [0,1]$ *and all* $t \geq 0$*, where* g^{BAU} *and* R^{BAU} *are determined by*

$$B'(R^{\mathrm{BAU}} + g^{\mathrm{BAU}}) = C'(g^{\mathrm{BAU}}) = (1-\beta\rho)k.$$

All formal proofs are relegated to the appendix. The above lemma shows that in the BAU-equilibrium all countries, irrespective of their initial technology levels, choose the same emission rates and technology stocks from period $t = 0$ onwards. The global stock of greenhouse gases evolves according to equation (7), which implies that it converges to the steady state G^{BAU} determined by

$$C'((1-\gamma)G^{\mathrm{BAU}}) = (1-\beta\rho)k. \tag{8}$$

3.2 First-best solution

Now suppose that a social planner makes the investment and emission decisions for all countries in order to maximize world welfare as given by (4). We call the resulting solution the first-best solution.

Definition 2 The first-best solution consists of a global sequence $(G_t)_{t=0}^{+\infty}$ and individual sequences $\{(R_{i,t}, g_{i,t}, r_{i,t})_{t=0}^{+\infty} \,|\, i \in [0,1]\}$ which together maximize (4) subject to (1)-(2).

Lemma 2 *The first-best solution is characterized by equations (1)-(2) along with the first-order conditions*

$$\begin{aligned} g_{i,t} &= g_t, && (9)\\ R_{i,t} &= R_t, && (10)\\ B'(R_t + g_t) &= (1-\beta\rho)k, && (11)\\ D'(G_t) &= (1-\beta\gamma)(1-\beta\rho)k - C'(g_t) + \beta\gamma C'(g_{t+1}) && (12) \end{aligned}$$

for all $i \in [0,1]$ *and all* $t \geq 0$ *as well as the transversality condition*

$$\lim_{t\to\infty} \beta^t[(1-\beta\rho)k - C'(g_t)]G_t = 0. \tag{13}$$

Comparing the first-best solution with the BAU-equilibrium, we see that total energy consumption $R_{i,t} + g_{i,t}$ by every country $i \in [0, 1]$ and in any given period $t \geq 0$ is the same in both solutions. The only difference in the two solutions consists therefore in the way how this amount of energy is produced. In the BAU-equilibrium the technology level R^{BAU} is installed in period 0 and maintained forever, whereas in the first-best solution the sequences $(g_{i,t})_{t=0}^{+\infty}$ and $(R_{i,t})_{t=0}^{+\infty}$ are not constant. Their evolution together with that of $(G_t)_{t=0}^{+\infty}$ is determined by equations (2) and (11)-(13). The stock of greenhouse gases converges to G^{FB} defined by

$$D'(G^{\mathrm{FB}}) = (1 - \beta\gamma)[(1 - \beta\rho)k - C'((1 - \gamma)G^{\mathrm{FB}})].$$

Comparing this equation to (8) and using strict convexity of C and D it follows that G^{FB} is smaller than G^{BAU}. This implies of course that the steady state emission rate g^{FB} in the first-best solution is smaller than g^{BAU} and, consequently, that the steady state technology stock R^{FB} in the first-best solution must be larger than R^{BAU}. To summarize, full internalisation of the external effects of greenhouse gas emission does not change the benefit from energy consumption for any country but it speeds up the replacement of the dirty technology by the clean one.

3.3 The linear-quadratic case

Throughout the paper we will illustrate our results by considering a version of the model in which the benefit and cost functions are given by

$$B(y) = -\frac{(y - z)^2}{2}\,,\ C(g) = \frac{cg^2}{2}\,,\ D(G) = \frac{dG^2}{2}, \tag{14}$$

where c and d are positive parameters and z is a bliss point for energy consumption. These functional forms are borrowed from Harstad (2016), except for $C(g)$ which is absent in his analysis. The quadratic specifications allow us to compute the two benchmark solutions easily. For example, the steady state values in the BAU-equilibrium are given by

$$G^{\mathrm{BAU}} = \frac{(1 - \beta\rho)k}{(1 - \gamma)c},$$
$$R^{\mathrm{BAU}} = z - \frac{(1 - \beta\rho)(1 + c)k}{c},$$

whereas the corresponding values in the first-best solution are

$$G^{\mathrm{FB}} = \frac{(1-\beta\gamma)(1-\beta\rho)k}{d+(1-\gamma)(1-\beta\gamma)c},$$
$$R^{\mathrm{FB}} = z - \frac{[d+(1-\gamma)(1-\beta\gamma)(1+c)](1-\beta\rho)k}{d+(1-\gamma)(1-\beta\gamma)c}.$$

As has been mentioned before, it holds that $G^{\mathrm{FB}} < G^{\mathrm{BAU}}$ and $R^{\mathrm{BAU}} < R^{\mathrm{FB}}$.

Example 1 Assume the functional forms stated in (14) with the following parameters: $k = 1$, $z = 1$, $\beta = 0.96$, $\rho = 0.9$, $\gamma = 0.9$, $c = 1$, and $d = 0.1$. Under these assumptions, the steady state in the BAU-equilibrium is given by $G^{\mathrm{BAU}} = 1.36$ and $R^{\mathrm{BAU}} = 0.728$, whereas the steady state in the first-best solution is $G^{\mathrm{FB}} = 0.1628$ and $R^{\mathrm{FB}} = 0.8477$.

4 Delegation to a benevolent authority

In this section we turn to the main investigation of this paper by introducing a supranational environmental authority (SEA), to which all countries $i \in [0,1]$ delegate their decisions on the emission levels $(g_{i,t})_{t=0}^{+\infty}$. The countries keep control over their investments $(r_{i,t})_{t=0}^{+\infty}$ and seek to maximize their individual welfare given in (3). We distinguish between the case where the SEA has commitment power so that it can announce and implement the entire family of sequences $\{(g_{i,t})_{t=0}^{+\infty} \,|\, i \in [0,1]\}$ at the outset of period 0, and the case where it lacks such commitment power. In both cases we assume that the SEA is benevolent in the sense that it seeks to maximize aggregate welfare as given in (4).[6]

4.1 Commitment

In this subsection we assume that the SEA has commitment power and announces the emission levels for all countries and all periods at the outset of period 0. The countries act as followers. We start by determining the best response of an arbitrary country $i \in [0,1]$ to the fixed family of sequences $\{(g_{j,t})_{t=0}^{+\infty} \,|\, j \in [0,1]\}$.

[6]We shall drop the assumption of benevolence in section 5.

Lemma 3 Let $(G_t)_{t=0}^{+\infty}$ and $\{(g_{j,t})_{t=0}^{+\infty} \,|\, j \in [0,1]\}$ be given sequences such that condition (2) holds. For every $i \in [0,1]$, country i's best response is characterized by

$$B'(R_{i,t} + g_{i,t}) = (1 - \beta\rho)k \text{ for all } t \geq 0. \tag{15}$$

Using the above characterization of best responses, we can define an equilibrium under commitment. Our first main result demonstrates that such an equilibrium coincides with the first-best solution.

Definition 3 An equilibrium under commitment consists of a global sequence $(G_t)_{t=0}^{+\infty}$ and individual sequences $\{(R_{i,t}, g_{i,t}, r_{i,t})_{t=0}^{+\infty} \,|\, i \in [0,1]\}$ which maximize (4) subject to (1), (2), and (15).

Theorem 1 *An equilibrium under commitment coincides with the first-best solution.*

The above result follows from the observation that in the case where the authority commits to any family of sequences of emissions $\{(g_{j,t})_{t=0}^{+\infty} \,|\, j \in [0,1]\}$, the countries choose corresponding efficient investment levels. This is reflected by the coincidence of equations (11) and (15). On a more intuitive level, one can say that the negative external effect operates solely via the emissions and the resulting stock of greenhouse gases and, hence, by setting the emission levels at the socially optimal values, the first-best solution can be achieved. The problem with this solution, however, is that it is not dynamically consistent. By announcing low emission levels for the future, the SEA provides the correct incentives for the countries to invest into clean technologies. But once these technologies are implemented, the authority itself faces an incentive to allow higher emission levels in order to provide more energy consumption to the countries. If the SEA lacks full commitment power, it will therefore not want to stick to its announced policy. We now turn to the analysis of this case.

4.2 Discretion

Consider the situation where the SEA cannot make a binding commitment. To formalize this idea, we assume that the authority makes a consistent plan in the sense of Strotz (1955-1956).[7] This means that the SEA consists of an infinite sequence of selves, one for each period. In every period t the current self of the SEA decides on $\{g_{i,t} \,|\, i \in [0,1]\}$ only. But in doing this, it correctly anticipates how its choice of the emission rates affects the investment behavior of the individual countries and the future emission levels determined by later selves of the authority. Formally, the different selves of the authority play a game among each other. We solve this game for a stationary Markov-perfect Nash equilibrium.[8]

A stationary Markov-perfect equilibrium can be defined most easily in recursive form using recursive notation. Consider any given period t. Variables dated t will be written without a subscript, e.g., G, R_i, g_i, and r_i. Variables of the preceding period $t-1$ will be denoted by the subscript $_-$, e.g., G_-, $R_{i,-}$, $g_{i,-}$, and $r_{i,-}$. At the start of a given period the aggregate states G_- and $\tilde{R}_- = \{R_{i,-} \,|\, i \in [0,1]\}$ are known to all countries and to the SEA, where $\tilde{R}_-$ is interpreted as the distribution of technology levels.[9] As for the timing, we follow Harstad (2012) and assume that every time period is split into two stages. In the first stage all countries invest into clean technologies (investment stage) and in the second stage the authority sets the emission levels and the countries consume (emission stage). Note that the emission stage begins once the investment costs are sunk and that the SEA knows the countries' technology stocks at the end of the investment stage, $\tilde{R} = \{R_i \,|\, i \in [0,1]\}$, when it has to decide on the emission levels $\{g_i \,|\, i \in [0,1]\}$.

---

[7]This approach is often called the sophisticated solution according to the multiple-selves model; see, e.g., chapter 6 in Sorger (2015) for a recent textbook presentation.

[8]Here, stationarity is used synonymously to symmetry: it means that every self of the SEA uses the same strategy. Markov-perfection is ensured by allowing only strategies that depend solely on the payoff-relevant state variables and by using backward induction for defining the equilibrium.

[9]By referring to these variables as 'aggregate states' we indicate that no individual country can affect them by unilateral actions. In particular, country i can decide on its own technology level R_i, but it cannot alter the entire distribution $\tilde{R}$. It will turn out that the equilibrium strategies are independent of the distribution $\tilde{R}_-$ and that the equilibrium value functions depend on $\tilde{R}_-$ only via the aggregate (or, equivalently, the average) technology level $\int_0^1 R_{i,-}\,\mathrm{d}i$; see lemma 4 below.

We restrict attention to those stationary Markovian strategies $\mathbf{g}$ of the SEA according to which the authority's choice of the emission level for country i depends only on the aggregate state variables G_- and $\tilde{R}$ as well as on country i's own technology stock R_i. We write such a strategy in the form $g_i = \mathbf{g}(G_-, \tilde{R}, R_i)$.[10]

Before we can define an equilibrium under discretion, we need to explain some preliminary considerations. To begin with, we need to find out how the countries react to a given strategy $\mathbf{g}$ of the SEA. Suppose therefore that all selves of the SEA play the strategy $\mathbf{g}$. It follows from (2) that the stock of greenhouse gases evolves according to

$$G = \mathcal{G}(G_-, \tilde{R}), \tag{16}$$

where

$$\mathcal{G}(G_-, \tilde{R}) = \int_0^1 \mathbf{g}(G_-, \tilde{R}, R_i)\,\mathrm{d}i + \gamma G_-. \tag{17}$$

Now suppose that we are at the beginning of period t and consider an arbitrary country i that has technology stock $R_{i,t-1}$. Since country i is infinitesimally small, it takes $(G_s)_{s=t-1}^{+\infty}$ and $(\tilde{R}_s)_{s=t-1}^{+\infty}$ as given and maximizes

$$\begin{aligned}
&\sum_{s=t}^{+\infty} \beta^{s-t}\Big[B(R_{i,s} + \mathbf{g}(G_{s-1}, \tilde{R}_s, R_{i,s})) - C(\mathbf{g}(G_{s-1}, \tilde{R}_s, R_{i,s})) - D(G_s) - k(R_{i,s} - \rho R_{i,s-1})\Big] \\
= \; & k\rho R_{i,t-1} \\
&+ \sum_{s=t}^{+\infty} \beta^{s-t}\Big[B(R_{i,s} + \mathbf{g}(G_{s-1}, \tilde{R}_s, R_{i,s})) - C(\mathbf{g}(G_{s-1}, \tilde{R}_s, R_{i,s})) - D(G_s) - (1-\beta\rho)kR_{i,s}\Big]
\end{aligned}$$

by choosing $(R_{i,s})_{s=t}^{+\infty}$. This is equivalent to setting $R_{i,t} = \mathbf{R}(G_{t-1}, \tilde{R}_t)$, where

$$\mathbf{R}(G_-, \tilde{R}) = \operatorname{argmax}\{B(x + \mathbf{g}(G_-, \tilde{R}, x)) - C(\mathbf{g}(G_-, \tilde{R}, x)) - (1-\beta\rho)kx \mid x \in \mathbb{R}\}. \tag{18}$$

We proceed under the assumption that the maximization problem in (18) has a unique solution, i.e., that $\mathbf{R}$ is a single-valued function.[11] This implies in particular that, independently of the

[10]In general, stationary Markovian strategies would be of the form $g_i = \mathbf{g}(G_-, \tilde{R}, i)$. An example of a strategy that is ruled out in our analysis is $g_i = aG_- + bR_{i/2}$, where a and b are constants. In this example, the emission of country i depends explicitly on the stock of technology available in country $j = i/2$.

[11]Whether the problem in (18) has a solution and, if so, whether this solution is unique depends among other things on the properties of the strategy $\mathbf{g}$. In the linear-quadratic example from section 3.3 the strategy $\mathbf{g}$ is a linear function and concavity of B together with strict convexity of C implies that there exists a unique solution.

form of the distribution $\tilde{R}_{-}$, all countries $i \in [0,1]$ choose the same technology level $\mathbf{R}(G_{-}, \tilde{R})$ so that the distribution $\tilde{R}$ is degenerate. This feature is due to the linear investment cost. The distribution $\tilde{R}$ can of course depend on G_{-} and we therefore write

$$\tilde{R} = \mathcal{R}(G_{-}). \tag{19}$$

By definition, the mapping $\mathcal{R}$ must satisfy

$$\mathcal{R}(G_{-})_i = \mathbf{R}(G_{-}, \mathcal{R}(G_{-})) \text{ for all } i \in [0,1]. \tag{20}$$

We assume that $\mathcal{R}(G_{-})$ is the unique fixed point of equation (20), which is indeed true in the linear-quadratic case. Equations (16) and (19) together form the law of motion of the aggregate states G and $\tilde{R}$ corresponding to the given strategy $\mathbf{g}$. One can combine these two equations to obtain

$$G = \mathcal{G}_0(G_{-}) \tag{21}$$

where

$$\mathcal{G}_0(G_{-}) = \mathcal{G}(G_{-}, \mathcal{R}(G_{-})). \tag{22}$$

This shows that the dynamics of the greenhouse gas stock can be separated from the accumulation of technology.

Now that we have determined the reaction of the countries to a fixed strategy $\mathbf{g}$ and have described the induced dynamics of the aggregate state variables, we can evaluate the strategy $\mathbf{g}$ according to the SEA's welfare measure. Let us denote by $V(G_{-}, \tilde{R}_{-})$ the value of the strategy $\mathbf{g}$ for a self of the authority that takes over control in a situation when the aggregate states are given by G_{-} and $\tilde{R}_{-}$. The value function must satisfy the recursive equation

$$\begin{aligned} V(G_{-}, \tilde{R}_{-}) = \int_0^1 \Big\{ & B(\mathcal{R}(G_{-})_i + \mathbf{g}(G_{-}, \mathcal{R}(G_{-}), \mathcal{R}(G_{-})_i)) - C(\mathbf{g}(G_{-}, \mathcal{R}(G_{-}), \mathcal{R}(G_{-})_i)) \\ & - D(\mathcal{G}_0(G_{-})) - k[\mathcal{R}(G_{-})_i - \rho R_{i,-}] \Big\} \, \mathrm{d}i + \beta V(\mathcal{G}_0(G_{-}), \mathcal{R}(G_{-})). \end{aligned}$$

It remains to determine the equilibrium policy function $\mathbf{g}$. To this end, we have to find out how off-equilibrium choices of the emission levels $\{g_i \,|\, i \in [0,1]\}$ affect the countries' investments and future technology stocks under the assumption that all future emission levels are chosen according to the equilibrium policy function $\mathbf{g}$. Consider a self that takes over control when the aggregate states are G_- and $\tilde{R}_-$. This self rationally anticipates that all its successors play the equilibrium strategy $\mathbf{g}$. Since the countries make their investment decisions before the SEA sets the emission levels, the self in charge can react to off-equilibrium choices of the countries' technology stocks. Thus the self under consideration needs to find the optimal emission levels $\{g_i \,|\, i \in [0,1]\}$ for given aggregate states G_- and $\tilde{R}$. The individual countries, on the other hand, move before the SEA and, hence, their only rational expectation about the authority's behavior is the SEA's equilibrium strategy. Formally, even under off-equilibrium behavior of the authority, every country chooses its technology level such that (18)-(20) hold. To summarize, the self under consideration chooses the emission levels $\{g_i \,|\, i \in [0,1]\}$ so as to maximize

$$\int_0^1 [B(R_i + g_i) - C(g_i) - D(G) - k(R_i - \rho R_{i,-})] \, \mathrm{d}i + \beta V(G, \tilde{R})$$

subject to (2). Dropping terms that do not depend on the emission levels and using (2) to eliminate G, it follows that the mapping $i \mapsto g_i = \mathbf{g}(G_-, \tilde{R}, R_i)$ maximizes the functional

$$\int_0^1 \left[B(R_i + g_i) - C(g_i) - D\left(\int_0^1 g_j \, \mathrm{d}j + \gamma G_- \right) \right] \mathrm{d}i + \beta V\left(\int_0^1 g_j \, \mathrm{d}j + \gamma G_-, \tilde{R} \right). \qquad (23)$$

We are finally ready to define an equilibrium under discretion.

Definition 4 An equilibrium under discretion consists of functions $(\mathbf{g}, \mathbf{R}, \mathcal{G}, \mathcal{G}_0, \mathcal{R}, V)$ such that conditions (17)-(18), (20), and (22)-(23) hold and such that, for all G_- and all $\tilde{R}$, the mapping $i \mapsto g_i = \mathbf{g}(G_-, \tilde{R}, R_i)$ maximizes (23).

It is easy to verify that the first-best solution does not qualify as an equilibrium under discretion, as stated in the following theorem.

Theorem 2 *The first-best solution cannot be implemented as an equilibrium under discretion.*

The separability and linearity assumptions made in section 2 imply a number of useful properties of an equilibrium under discretion which we state in the following lemma. To this end, we denote by $\bar{R}_-$ the average technology stock of the distribution $\tilde{R}_-$, that is, $\bar{R}_- = \int_0^1 R_{i,-} \, \mathrm{d}i$.

Lemma 4 **(a)** *The authority's equilibrium value function V is of the form*

$$V(G_-, \tilde{R}_-) = V_0(G_-) + k\rho \bar{R}_-.$$

(b) *The equilibrium strategy* $\mathbf{g}$ *is of the form*

$$\mathbf{g}(G_-, \tilde{R}, R_i) = \mathbf{g}_0(G_-, R_i).$$

(c) *The function* $\mathbf{R}$ *is of the form*

$$\mathbf{R}(G_-, \tilde{R}) = \mathbf{R}_0(G_-)$$

and $\mathcal{R}(G_-)_i = \mathbf{R}_0(G_-)$ *holds for all* $i \in [0, 1]$.

(d) *The functions* V_0, $\mathbf{g}_0$, *and* $\mathbf{R}_0$ *from parts (a)-(c) satisfy*

$$\begin{aligned} V_0(G_-) \quad = \quad & B(\mathbf{R}_0(G_-) + \mathbf{g}_0(G_-, \mathbf{R}_0(G_-))) - C(\mathbf{g}_0(G_-, \mathbf{R}_0(G_-))) \\ & -D(\mathcal{G}_0(G_-)) - (1 - \beta\rho)k\mathbf{R}_0(G_-) + \beta V_0(\mathcal{G}_0(G_-)) \end{aligned} \tag{24}$$

and

$$B'(R_i + \mathbf{g}_0(G_-, R_i)) - C'(\mathbf{g}_0(G_-, R_i)) = D'\left(\mathcal{G}_0(G_-)\right) - \beta V_0'\left(\mathcal{G}_0(G_-)\right) \tag{25}$$

for all G_- *and all* R_i.

An interesting property of the equilibrium emission strategy $\mathbf{g}_0$ is that its sensitivity with respect to the technology level R_i is independent of how the stock of greenhouse gases affects country i's utility (i.e., it is independent of the form of the cost function D) whereas it does depend on how the local emissions by country i impact its own welfare (i.e., it depends on the form of the cost function C). This observation, which will become relevant in section 5 below, is formalized in the following lemma.

Lemma 5 *Suppose that the strategy* $\mathbf{g}_0$ *is differentiable. Then it holds that* $\partial\mathbf{g}_0(G_-, R_i)/(\partial R_i) \in (-1, 0)$ *and that this partial derivative is independent of the form of the function* D.

4.3 The linear-quadratic case

In this subsection we determine an equilibrium under discretion in the linear-quadratic environment introduced in section 3.3.

Theorem 3 *Consider the linear-quadratic specification from section 3.3. There exists an equilibrium under discretion in which the policy functions take the forms*

$$\mathbf{g}_0(G_-, R_i) = h_1 + h_G G_- + h_R R_i \tag{26}$$

and

$$\mathbf{R}_0(G_-) = R^{\text{BAU}}. \tag{27}$$

The coefficients h_1, h_G, and h_R are given by

$$h_1 = \frac{(1+c)(z+\beta v_G) + (d-\beta v_{GG})R^{\text{BAU}}}{(1+c)(1+c+d-\beta v_{GG})}, \tag{28}$$

$$h_G = -\frac{(d-\beta v_{GG})\gamma}{1+c+d-\beta v_{GG}}, \tag{29}$$

$$h_R = -\frac{1}{1+c}, \tag{30}$$

where

$$v_{GG} = \frac{1+c+d-\beta(1+c)\gamma^2 - \sqrt{[1+c+d-\beta(1+c)\gamma^2]^2 + 4(1+c)\beta d\gamma^2}}{2\beta}, \tag{31}$$

$$v_G = \frac{\gamma(d-\beta v_{GG})R^{\text{BAU}}}{(1+c)(1-\beta\gamma) + d - \beta v_{GG}}. \tag{32}$$

In this equilibrium, the stock of greenhouse gases converges monotonically to

$$G^{\text{DIS}} = \frac{(1+c)(1-\beta\gamma)(1-\beta\rho)k}{c[d+(1-\gamma)(1-\beta\gamma)(1+c)]}. \tag{33}$$

The equilibrium described in the above theorem has a number of interesting properties that we would like to point out. First of all, all countries choose the technology stock R^{BAU} in period 0 and maintain that stock forever. This shows that, if the SEA lacks commitment power, it is unable to induce the countries to invest more than in the BAU-equilibrium. As can be seen from the proof of the theorem, this property is due to the fact that the equilibrium strategy $\mathbf{g}_0$ of the authority has the reaction coefficient $h_R = -1/(1+c)$ as stated in (30). This, in turn, has the consequence that the countries play the constant investment strategy $\mathbf{R}_0$ described in (27). Given that in equilibrium the countries' investments are not responsive to any changes in the stock of greenhouse gases, the SEA can improve welfare only by curtailing emission rates relative to business as usual. As shown in the theorem, this results in the long-run stock of greenhouse gases to be equal to G^{DIS} stated in (33). It is easy to see that

$$G^{\text{FB}} < G^{\text{DIS}} < G^{\text{BAU}}$$

holds whenever d is strictly positive.[12]

Example 2 Assume the same functional forms and parameters as in example 1. The steady state in a discretionary equilibrium with a benevolent SEA that has commitment power coincides with the first-best steady state, $G^{\text{FB}} = 0.1628$ and $R^{\text{FB}} = 0.8477$. By contrast, the steady state in the equilibrium where emission decisions are delegated to a discretionary and benevolent SEA is given by $G^{\text{DIS}} = 0.2908$ and $R^{\text{DIS}} = 0.7280$.

Figure 1 expands on example 2 by plotting the welfare functions under business as usual, commitment, and discretion over the domain $G_{-} \in [0, 0.45]$.[13] As can be seen, delegation to a benevolent SEA noticeably improves welfare relative to the BAU equilibrium even when the authority lacks commitment.

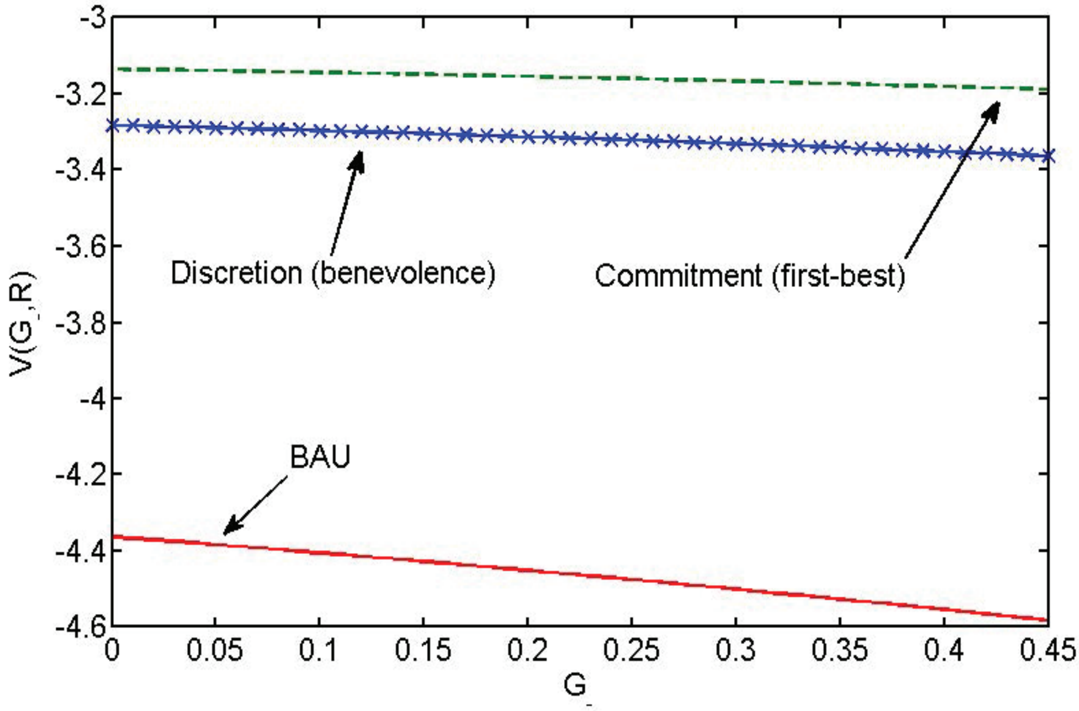
5 Delegation to a non-benevolent authority

The delegation of emission control to an SEA can improve aggregate welfare relative to business as usual because the authority can internalize the external spillover effects of greenhouse gas emissions. However, as we have seen in theorem 3, the authority cannot accelerate the conversion from fossil fuel based-energy production to emission free technologies if it lacks commitment power. As will be shown in the present section, the dynamic inconsistency problem at the root of this result can be mitigated by endowing the SEA with an objective functional that differs from aggregate welfare (4). The idea of optimal delegation is borrowed from the literature on monetary policy, where dynamic inconsistency problems can be extenuated by augmenting the central bank's mandate by strong inflation aversion, as in Rogoff (1985), or by a desire for policy inertia, as in Woodford (2003).

[12] If d would be equal to 0, the steady state values G^{BAU}, G^{FB}, and G^{DIS} would coincide.

[13] This choice of the domain is motivated by the property that the steady states under commitment and under discretion lie well within this interval.

Figure 1: Welfare comparisons



5.1 Equilibrium definition

Let us for example assume that the objective functional of the SEA is of the form

$$\sum_{t=0}^{+\infty} \beta^t \int_0^1 \hat{U}(G_t, R_{i,t}, g_{i,t}, r_{i,t})\,\mathrm{d}i,$$

where

$$\hat{U}(G_t, R_{i,t}, g_{i,t}, r_{i,t}) = B(R_{i,t} + g_{i,t}) - kr_{i,t} - \hat{C}(g_{i,t}) - \hat{D}(G_t)$$

and where $\hat{C}$ and $\hat{D}$ are strictly concave cost functions that may be different from the functions C and D describing the preferences of the countries. Of course, one could consider much more general mandates for the authority but the proposed form will be sufficient to make our point. We are particularly interested in the question of whether the SEA should attach more weight to the cost of local emissions $g_{i,t}$ or to the cost of the global stock of greenhouse gases G_t.

An equilibrium under discretion in this situation can be defined along the same lines as in the previous section. The only difference to the previous arguments is that one needs to take into account that the SEA evaluates its policy using the functions $\hat{C}$ and $\hat{D}$. Analogously to section 4.2 we can therefore obtain the following equilibrium conditions.[14]

Definition 5 An equilibrium under discretion consists of functions $(\mathbf{g}_0, \mathbf{R}_0, \mathcal{G}, \mathcal{G}_0, \mathcal{R}, \hat{V}_0)$ such that the following conditions hold:

$$\mathcal{G}(G_-, \tilde{R}) = \int_0^1 \mathbf{g}_0(G_-, R_i)\, \mathrm{d}i + \gamma G_-, \tag{34}$$

$$\mathbf{R}_0(G_-) = \mathrm{argmax}\{B(x + \mathbf{g}_0(G_-, x)) - C(\mathbf{g}_0(G_-, x)) - (1 - \beta\rho)kx \mid x \in \mathbb{R}\}, \tag{35}$$

$$\mathcal{R}(G_-)_i = \mathbf{R}_0(G_-) \text{ for all } i, \tag{36}$$

$$\mathcal{G}_0(G_-) = \mathcal{G}(G_-, \mathcal{R}(G_-)), \tag{37}$$

$$\begin{aligned} \hat{V}_0(G_-) = {} & B(\mathbf{R}_0(G_-) + \mathbf{g}_0(G_-, \mathbf{R}_0(G_-))) - \hat{C}(\mathbf{g}_0(G_-, \mathbf{R}_0(G_-))) \\ & - \hat{D}(\mathcal{G}_0(G_-)) - (1 - \beta\rho)k\mathbf{R}_0(G_-) + \beta\hat{V}_0(\mathcal{G}_0(G_-)), \end{aligned} \tag{38}$$

$$B'(R_i + \mathbf{g}_0(G_-, R_i)) - \hat{C}'(\mathbf{g}_0(G_-, R_i)) = \hat{D}'(\mathcal{G}_0(G_-)) - \beta\hat{V}_0'(\mathcal{G}_0(G_-)). \tag{39}$$

Note that the function $\hat{V}$ defined by $\hat{V}(G_-, \tilde{R}_-) = \hat{V}_0(G_-) + k\rho\bar{R}_-$ is the equilibrium value function of the authority. Aggregate welfare according to (4) is given by $V(G_-, \tilde{R}_-) = V_0(G_-) + k\rho\bar{R}_-$, where V_0 solves the recursive equation

$$\begin{aligned} V_0(G_-) = {} & B(\mathbf{R}_0(G_-) + \mathbf{g}_0(G_-, \mathbf{R}_0(G_-))) - C(\mathbf{g}_0(G_-, \mathbf{R}_0(G_-))) \\ & - D(\mathcal{G}_0(G_-)) - (1 - \beta\rho)k\mathbf{R}_0(G_-) + \beta V_0(\mathcal{G}_0(G_-)). \end{aligned} \tag{40}$$

We would like to find out how the SEA's cost functions $\hat{C}$ and $\hat{D}$ must be chosen so that aggregate welfare is improved relative to the equilibrium under a benevolent SEA. In the following subsection, we derive a few analytical results for the linear-quadratic example from section 3.3.

5.2 The linear-quadratic case

We consider the same specifications as introduced in section 3.3 and assume furthermore that

$$\hat{C}(g) = \frac{\hat{c}g^2}{2}\,,\ \hat{D}(G) = \frac{\hat{d}G^2}{2},$$

[14] Note, in particular, that a result analogous to lemma 4 holds also in the present situation.

where $\hat{c}$ and $\hat{d}$ are positive parameters.

On first thought, one may think that the value of $\hat{c}$ is of subordinate importance but that the authority should be given a high value of $\hat{d}$. After all, the external effect operates via the global stock of greenhouse gases, and it is tempting to suggest that this stock must be highly penalized. The following analysis, however, proves that this conjecture is wrong and shows that the parameter $\hat{c}$ of the cost of local emissions plays the key role.

We start by deriving an equilibrium under discretion in the case where $\hat{c} = c$ holds and where $\hat{d}$ is an arbitrary positive number.

Lemma 6 *Assume that $\hat{c} = c$ and $\hat{d} > 0$. There exists an equilibrium under discretion in which the functions $\mathbf{g}_0$ and $\hat{V}_0$ have the forms*

$$\mathbf{g}_0(G_-, R_i) = \hat{h}_1 + \hat{h}_G G_- + \hat{h}_R R_i, \tag{41}$$

and

$$\hat{V}_0(G_-) = \hat{v}_1 + \hat{v}_G G_- + \frac{\hat{v}_{GG}}{2}(G_-)^2 \tag{42}$$

and in which (27) holds. The coefficients $\hat{h}_1$, $\hat{h}_G$, $\hat{h}_R$, $\hat{v}_1$, $\hat{v}_G$, and $\hat{v}_{GG}$ are described by the same formulas as the corresponding coefficients h_1, h_G, h_R, v_1, v_G, and v_{GG} in theorem 3 except that the parameter d is replaced by $\hat{d}$ wherever it appears. The stock of greenhouse gases converges monotonically to

$$\hat{G}^{\mathrm{DIS}} = \frac{(1-\beta\gamma)(1-\beta\rho)(1+c)k}{c[\hat{d} + (1-\gamma)(1-\beta\gamma)(1+c)]}. \tag{43}$$

It will be convenient to refer to the pair $(\hat{c}, \hat{d})$ as the *type* of the SEA. The benevolent authority, for example, is of type (c, d). The following theorem demonstrates that it is not possible to reduce the welfare loss caused by discretionary behavior by variations of $\hat{d}$ alone (i.e., by keeping $\hat{c}$ equal to c).

Theorem 4 *Let $\hat{d} > 0$ be arbitrarily given. Aggregate welfare in the discretionary equilibrium with a non-benevolent SEA of type $(c, \hat{d})$ is smaller than or equal to aggregate welfare in the discretionary equilibrium with a benevolent SEA of type (c, d) (which is described in theorem 3).*

The intuitive explanation for this result is that it is not the externality per se that causes the discretionary equilibrium to have poor welfare, but the dynamic inconsistency. The dynamic inconsistency, in turn, can only be mitigated by changing the incentives for the countries to invest in green technologies. As we have seen in lemma 5, these incentives are independent of the cost function D (or $\hat{D}$, respectively) but do depend on C (or $\hat{C}$, respectively). What, then, is the role of the parameter $\hat{d}$ in the present example? It is obvious from equation (43) that by varying $\hat{d}$ between 0 and $+\infty$ one can generate any steady state greenhouse gas stock between 0 and

$$G^{\mathrm{BAU}} = \frac{(1-\beta\rho)k}{(1-\gamma)c}.$$

For example, by choosing $\hat{d} = (1+c)d/c > d$ one obtains the first-best steady state G^{FB}. One could also try to find that value of $\hat{d}$ which maximizes per-period welfare in steady state. To this end, recall from lemma 6 that independently of the value of $\hat{d}$ it holds in the discretionary equilibrium for all $i \in [0,1]$ and all $t \geq 0$ that $R_{i,t} = R^{\mathrm{BAU}}$. Furthermore, in steady state we must have $g_{i,t} = (1-\gamma)G$. When one maximizes per-period welfare across those steady states for which $R_{i,t} = R^{\mathrm{BAU}}$ holds, one must therefore choose G in such a way that

$$B(R^{\mathrm{BAU}} + (1-\gamma)G) - C((1-\gamma)G) - D(G)$$

is maximized. One might call this a restricted Golden Rule, where the word 'restricted' indicates that $R_{i,t} = R^{\mathrm{BAU}}$ is imposed. Solving the first-order condition for this problem yields

$$G = \frac{(1-\gamma)(z - R^{\mathrm{BAU}})}{d + (1-\gamma)^2(1+c)},$$

which can be achieved by setting $\hat{d} = (1-\beta\gamma)d/(1-\gamma) > d$.

To summarize, by fixing the value of $\hat{c}$ at c and varying the cost parameter $\hat{d}$ only, one can steer the stock of greenhouse gases towards its first-best steady state or other desirable values (like the Golden Rule stock) but one cannot achieve higher welfare than the benevolent SEA would implement. We must therefore also consider variations of the cost parameter for local emissions $\hat{c}$. Whereas the general procedure to compute the equilibrium remains the same as in lemma 6, the algebra becomes more cumbersome. In particular, it is no longer the case that (27) holds. We therefore relegate the relevant formulas to the appendix and proceed with an illustrative example.

Example 3 Assume the same parameters as in example 1. Under this assumption, the optimal type of the discretionary SEA, i.e., the pair of parameters $(\hat{c}, \hat{d})$ that maximizes aggregate welfare over the domain $G_- \in [0, 0.45]$ is given by $\hat{c} = 10.48$ and $\hat{d} = 0$. The steady state in a discretionary equilibrium with an authority of type $(\hat{c}, \hat{d}) = (10.48, 0)$ is given by $\hat{G}^{DIS} = 0.1409$ and $\hat{R}^{DIS} = 0.8382$.

Surprisingly, optimal delegation under discretion requires the authority to place a large weight on local emission costs and *zero* weight on the global costs of climate change. While this result may seem counter-intuitive at first, the underlying mechanism is readily seen. Preferences that put a large weight on local pollution costs endow the SEA with an incentive to punish countries that underinvest in green technologies, which in turn alleviates the SEA's time inconsistency problem. A large weight on local pollution costs, in turn, makes it optimal to put relatively little weight on global pollution costs in order not to distort the optimal trade-off between green and dirty energy consumption too much. Delegation of emission policies to an 'optimally designed' discretionary SEA eliminates almost the entire welfare loss resulting from lack of commitment, as is visualized in figure 2.

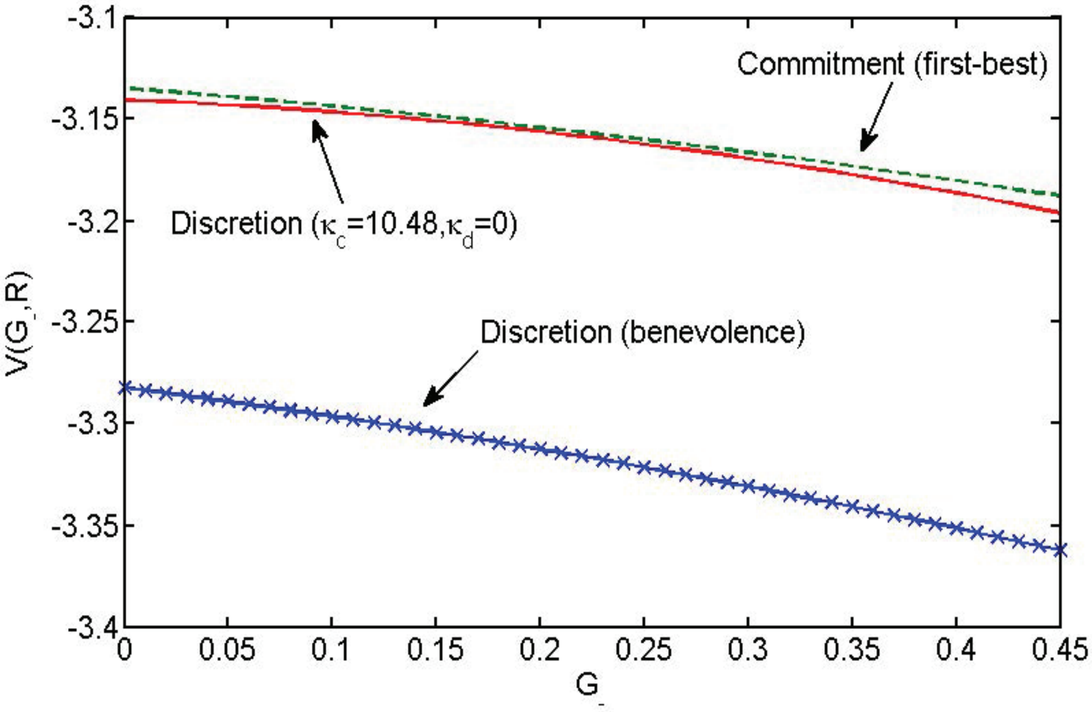
Example 4 Assume the same parameters as in example 1. Assume furthermore that the initial state is given by the steady state of the BAU-equilibrium, $G_- = 1.36$ and $R_- = 0.728$. Under these assumptions, optimal delegation under discretion requires an authority with type $(\hat{c}, \hat{d}) = (8.4010, 0.44131)$. The resulting steady state is given by $\hat{G}^{DIS} = 0.0401$ and $\hat{R}^{DIS} = 0.8442$. By contrast, if the initial state is the first-best steady state, $G_- = 0.1628$ and $R_- = 0.8477$, the optimal type of the authority is $(\hat{c}, \hat{d}) = (9.7927, 0)$ and the resulting steady state features $\hat{G}^{DIS} = 0.1515$ and $\hat{R}^{DIS} = 0.8365$.

Example 4 illustrates that the optimal type of the SEA (i.e., the optimal choice of preference parameters $\hat{c}$ and $\hat{d}$) in our model depends on the initial state (G_-, R_-). For any given state, the optimal preference parameters are given by

$$\text{argmin}_{\hat{c},\hat{d}} \left| V^{FB}(G_-, R_-) - V^{DIS}(G_-, R_- \mid \hat{c}, \hat{d}) \right| . \tag{44}$$

Figure 3 displays the optimal type as a function of G_-, while holding R_- constant. It shows that when G_- is not too large, the optimal choice is to set $\hat{c} > c$ and $\hat{d} = 0$, in line with our

Figure 2: Welfare comparison: commitment versus discretion

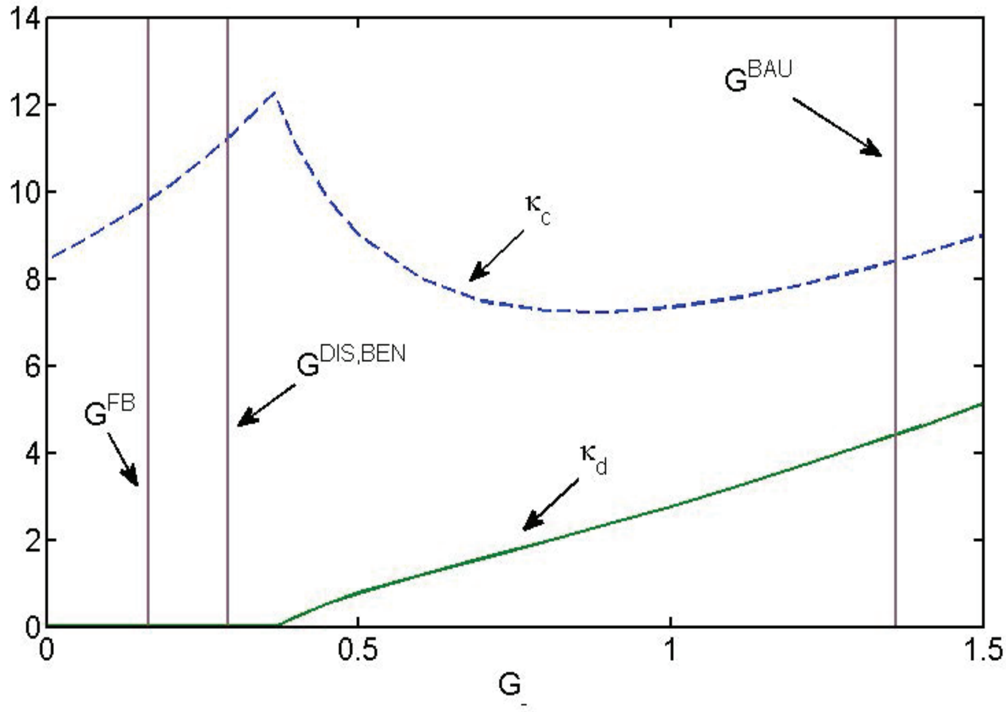


Note: $\kappa_c = \hat{c}/c$ and $\kappa_d = \hat{d}/d$.

discussion of example 3. If, however, the initial stock of greenhouse gases is very large relative to the steady state, then it is optimal to put a positive weight on the global costs of climate change, $\hat{d} > 0$.

The above finding points to a further complication of the optimal specification of an SEA's mandate. If the countries establish an SEA with optimal type $(\hat{c}, \hat{d})$ at time 0, when the stock of greenhouse gases is equal to G_{-1}, they will face an incentive to revise this mandate when the greenhouse gas concentration changes in the course of time. One can think of two possible ways to approach this situation. One possibility is to allow for state-dependent mandates, that is, for SEA types $(\hat{c}(G_{-}), \hat{D}(G_{-}))$. A SEA which is endowed with such a mandate would of course anticipate that it has to use different evaluations of the respective local and global costs if the greenhouse gas concentration changes. The strategies derived above would therefore no longer be an equilibrium. Another possible approach would be to maintain the assumption of state-independent mandates $(\hat{c}, \hat{d})$ but to take the incentives of the countries for a revision of

Figure 3: Welfare comparison: commitment versus discretion



Note: $\kappa_c = \hat{c}/c$ and $\kappa_d = \hat{d}/d$.

the mandate into account. In other words, one would have to address yet another dynamic inconsistency problem, this time one that pertains to the decision over the mandate of the SEA made by the countries. We further elaborate on this issue in the online appendix.

6 Concluding remarks

Dynamic inconsistency is ubiquitous in economic policy, and environmental policy is no exception. In the present paper we have pointed out that such a problem would also occur if a supranational environmental authority were established in order to control the emission of greenhouse gases in a world consisting of many individual countries. We have used a very simple and stylized, yet fully dynamic model to make our point. The main result is that the welfare loss that results from lack of commitment power can be kept very small by endowing the authority with a mandate to heavily penalize the local costs of emissions and to attach only little weight to the cost of the resulting (global) climate change.

Appendix

Proof of lemma 1

Consider an arbitrary country $i \in [0,1]$. Since this country takes $(G_t)_{t=0}^{+\infty}$ as given, one can drop the additively separated terms $-\beta^t D(G_t)$ from its objective functional (3) and it follows from (1) that country i maximizes

$$\begin{aligned} & \sum_{t=0}^{+\infty} \beta^t \left[B(R_{i,t} + g_{i,t}) - C(g_{i,t}) - k(R_{i,t} - \rho R_{i,t-1}) \right] \\ = \quad & k\rho R_{i,-1} + \sum_{t=0}^{+\infty} \beta^t \left[B(R_{i,t} + g_{i,t}) - C(g_{i,t}) - (1 - \beta\rho) k R_{i,t} \right] \end{aligned}$$

with respect to $(R_{i,t}, g_{i,t})_{t=0}^{+\infty}$. Obviously, this is the case if and only if

$$B'(R_{i,t} + g_{i,t}) = C'(g_{i,t}) = (1 - \beta\rho)k \tag{45}$$

holds for all $i \in [0,1]$ and all $t \geq 0$. Since the latter condition does neither explicitly depend on the country index i nor on time t, we obtain (5)-(6). Finally, by combining (2) and (5) one obtains (7).

Proof of lemma 2

Using (1) we can rewrite (4) as

$$\begin{aligned} & \sum_{t=0}^{+\infty} \beta^t \int_0^1 \left[B(R_{i,t} + g_{i,t}) - C(g_{i,t}) - D(G_t) - k(R_{i,t} - \rho R_{i,t-1}) \right] \mathrm{d}i \\ = \quad & k\rho \int_0^1 R_{i,-1} \,\mathrm{d}i + \sum_{t=0}^{+\infty} \beta^t \int_0^1 \left[B(R_{i,t} + g_{i,t}) - C(g_{i,t}) - D(G_t) - (1 - \beta\rho) k R_{i,t} \right] \mathrm{d}i. \end{aligned}$$

This expression is maximized with respect to the variables $R_{i,t}$ for all $i \in [0,1]$ and all $t \geq 0$ if and only if

$$B'(R_{i,t} + g_{i,t}) = (1 - \beta\rho)k, \tag{46}$$

which implies $B(R_{i,t}+g_{i,t})-(1-\beta\rho)kR_{i,t} = B(A)-(1-\beta\rho)k(A-g_{i,t})$, where $A=(B')^{-1}((1-\beta\rho)k)$. It follows therefore that the family of sequences $\{(g_{i,t})_{t=0}^{+\infty} \,|\, i\in[0,1]\}$ has to be chosen so as to maximize

$$\sum_{t=0}^{+\infty}\beta^t\int_0^1[(1-\beta\rho)kg_{i,t}-C(g_{i,t})-D(G_t)]\,\mathrm{d}i \tag{47}$$

subject to (2). Since the emission rates $g_{i,t}$ enter the constraint (2) only via the integral $\int_0^1 g_{i,t}\,\mathrm{d}i$ and since the cost function C is strictly convex, it follows that for every fixed $t\geq 0$ all countries' emission levels $g_{i,t}$ must be equal. Denoting the common value by g_t it follows from (46) that (9)-(11) hold. The Lagrangian function of the optimization problem is therefore

$$L=\sum_{t=0}^{+\infty}\beta^t\left\{(1-\beta\rho)kg_t-C(g_t)-D(G_t)+\lambda_t[g_t+\gamma G_{t-1}-G_t]\right\},$$

where λ_t denotes the Lagrange multiplier corresponding to constraint (2). Writing down the necessary and sufficient first-order conditions plus transversality condition and eliminating the Lagrange multipliers from these conditions yields (12)-(13).

Proof of lemma 3

Given $(G_t)_{t=0}^{+\infty}$ and $\{(g_{j,t})_{t=0}^{+\infty} \,|\, j\in[0,1]\}$, the maximization of (3) with respect to $(R_{i,t})_{t=0}^{+\infty}$ is equivalent to the maximization of

$$\sum_{t=0}^{+\infty}\beta^t\left[B(R_{i,t}+g_{i,t})-k(R_{i,t}-\rho R_{i,t-1})\right]=k\rho R_{i,-1}+\sum_{t=0}^{+\infty}\beta^t[B(R_{i,t}+g_{i,t})-(1-\beta\rho)kR_{i,t}]$$

with respect to $(R_{i,t})_{t=0}^{+\infty}$. Obviously, the latter expression is maximized if and only if (15) is satisfied.

Proof of theorem 1

Condition (15) coincides with condition (46). Using the same arguments as in the proof of lemma 2 it can be seen that the authority's optimization problem is equivalent to the maximization of (47) subject to (2). Since this is the same problem that a social planner would solve, the proof of the theorem is complete.

Proof of theorem 2

The first-order condition associated with problem (18) requires that in any equilibrium under discretion it holds for all $i \in [0,1]$ that

$$\begin{aligned} 0 &= B'(R_i + \mathbf{g}(G_-, \tilde{R}, R_i)) - (1-\beta\rho)k \\ &\quad + \left[B'(R_i + \mathbf{g}(G_-, \tilde{R}, R_i)) - C'(\mathbf{g}(G_-, \tilde{R}, R_i))\right] \frac{\partial \mathbf{g}(G_-, \tilde{R}, R_i)}{R_i}. \end{aligned} \tag{48}$$

Now suppose that the first-best solution is an equilibrium under discretion. Combining the above condition with (11) one obtains

$$0 = \left[B'(R_i + \mathbf{g}(G_-, \tilde{R}, R_i)) - C'(\mathbf{g}(G_-, \tilde{R}, R_i))\right] \frac{\partial \mathbf{g}(G_-, \tilde{R}, R_i)}{R_i}.$$

We claim that this cannot be the case. Suppose first that $\partial \mathbf{g}(G_-, \tilde{R}, R_i)/R_i = 0$. In this case, the authority's equilibrium policy would assign the same emission level to all countries independent of their technology level R_i, that is, there would exist an emission level g such that the optimal solution of problem (23) satisfies $g_i = g$ for all $i \in [0,1]$. Holding the total amount of emissions $\int_0^1 g_j \,\mathrm{d}j$ constant at the optimal level g, a reallocation of emission rights from countries with larger technology stocks to those with smaller ones would improve world welfare because B is concave. Formally, this can be seen from the first-order condition for the maximization of (23) subject to the constraint $\int_0^1 g_j \,\mathrm{d}j = g$. As a matter of fact, this condition is

$$B'(R_i + g) - C'(g) + \lambda = 0,$$

where λ is the Lagrange multiplier of the constraint. Obviously, this equation cannot hold for all $i \in [0,1]$ if there exist countries with different technology levels.

Now assume that $B'(R_i + \mathbf{g}(G_-, \tilde{R}, R_i)) - C'(\mathbf{g}(G_-, \tilde{R}, R_i)) = 0$ is satisfied along the first-best equilibrium path. Then it follows from equations (11)-(12) that $D'(G_t) = 0$ holds along this path, which cannot be true as D is a strictly convex function. This completes the proof.

Proof of lemma 4

The result in (a) follows immediately from equation (23). Using the result from part (a), the result in (b) follows easily from equation (23) and the result in (c) follows from (18) and (20).

Due to the results in (a) and (b), equation (23) simplifies to (24). A necessary and sufficient first-order condition for the maximization of (23) with respect to $\{g_i \,|\, i \in [0,1]\}$ is

$$\begin{aligned} & B'(R_i + \mathbf{g}_0(G_-, R_i)) - C'(\mathbf{g}_0(G_-, R_i)) \\ = \; & D'\left(\int_0^1 \mathbf{g}_0(G_-, R_j)\,\mathrm{d}j + \gamma G_-\right) - \beta V_0'\left(\int_0^1 \mathbf{g}_0(G_-, R_j)\,\mathrm{d}j + \gamma G_-\right) \end{aligned}$$

for all $i \in [0,1]$, where we have utilized the results from parts (a) and (b). Applying part (b) once more and combining it with (17), (19), and (22) we obtain (25).

Proof of lemma 5

Differentiating equation (25) with respect to R_i it follows that

$$B''(R_i + \mathbf{g}_0(G_-, R_i))[1 + \partial \mathbf{g}_0(G_-, R_i)/\partial R_i] = C''(\mathbf{g}_0(G_-, R_i))\partial \mathbf{g}_0(G_-, R_i)/\partial R_i,$$

which yields

$$\frac{\partial \mathbf{g}_0(G_-, R_i)}{\partial R_i} = \frac{B''(R_i + \mathbf{g}_0(G_-, R_i))}{C''(\mathbf{g}_0(G_-, R_i)) - B''(R_i + \mathbf{g}_0(G_-, R_i))}.$$

Since B is concave and C is strictly convex, the above formula proves the first claim. Since the above formula does not involve the function D, the second claim holds as well.

Proof of lemma 6

Starting from the conjectures stated in (41)-(42) one can proceed analogously to the proof of theorem 3 to find that

$$\begin{aligned} & \mathbf{g}_0(G_-, R_i) = \frac{z - R_i - \hat{d}\mathcal{G}_0(G_-) + \beta[\hat{v}_G + \hat{v}_{GG}\mathcal{G}_0(G_-)]}{1+c}, \\ & \hat{h}_R = -\frac{1}{1+c}, \\ & \mathbf{R}_0(G_-) = R^{\mathrm{BAU}}, \\ & \mathcal{G}_0(G_-) = \hat{h}_1 + (\hat{h}_G + \gamma)G_- - \frac{R^{\mathrm{BAU}}}{1+c}. \end{aligned}$$

It is not difficult to see that there is no structural difference between the situation analyzed in theorem 3 and the present case. That is, all we need to do is replace the parameter d wherever it appears by $\hat{d}$. This completes the proof.

Proof of theorem 3

Due to the linear-quadratic structure of the problem, we guess that the equilibrium strategy $\mathbf{g}$ is linear and that the equilibrium value function V is quadratic in G_-. According to lemma 4 this means that (26) and

$$V_0(G_-) = v_1 + v_G G_- + \frac{v_{GG}}{2}(G_-)^2$$

must hold, where h_1, h_G, h_R, v_1, v_G, and v_{GG} are undetermined coefficients. Solving equation (25) for $\mathbf{g}_0(G_-, R_i)$ yields

$$\mathbf{g}_0(G_-, R_i) = \frac{z - R_i - d\mathcal{G}_0(G_-) + \beta[v_G + v_{GG}\,\mathcal{G}_0(G_-)]}{1+c}, \tag{49}$$

which proves (30). Using (30) it is straightforward to solve (18) and we obtain

$$\mathbf{R}_0(G_-) = z - \frac{1+c}{c}(1-\beta\rho)k = R^{\mathrm{BAU}},$$

which proves (27). Condition (20) implies that the function $\mathcal{R}$ is constant as well. Its constant value is the degenerate distribution with

$$\mathcal{R}(G_-)_i = R^{\mathrm{BAU}} \tag{50}$$

for all $i \in [0,1]$. From (17) and (26) we obtain

$$\mathcal{G}(G_-, \tilde{R}) = h_1 + (h_G + \gamma)G_- - \frac{\bar{R}}{1+c}$$

and from (22) and (50) it follows therefore that

$$\mathcal{G}_0(G_-) = h_1 + (h_G + \gamma)G_- - \frac{R^{\mathrm{BAU}}}{1+c}. \tag{51}$$

Substituting this into (49) and equating the coefficients of the powers of G_- on both sides yields two linear equations for h_1 and h_G. These equations have a solution if and only if

$$v_{GG} \neq \frac{1+c+d}{\beta}. \tag{52}$$

If (52) holds, the unique solution is given by (28)-(29).

It remains to determine the coefficients v_1, v_G, and v_{GG}. To this end, we evaluate (24). Using the above results, this equation can be written as

$$
\begin{aligned}
& v_1 + v_G G_- + \frac{v_{GG}}{2}(G_-)^2 \qquad (53)\\
= & -\frac{1}{2}\left(h_1 + h_G G_- + \frac{cR^{\mathrm{BAU}}}{1+c}\right)^2 - \frac{c}{2}\left(h_1 + h_G G_- - \frac{R^{\mathrm{BAU}}}{1+c}\right)^2 \\
& -\frac{d}{2}\left[h_1 + (h_G+\gamma)G_- - \frac{R^{\mathrm{BAU}}}{1+c}\right]^2 - (1-\beta\rho)kR^{\mathrm{BAU}} \\
& +\beta\left\{v_1 + v_G\left[h_1 + (h_G+\gamma)G_- - \frac{R^{\mathrm{BAU}}}{1+c}\right] + \frac{v_{GG}}{2}\left[h_1 + (h_G+\gamma)G_- - \frac{R^{\mathrm{BAU}}}{1+c}\right]^2\right\}.
\end{aligned}
$$

Equating the coefficients of G_- on both sides of equation (53) and replacing h_1 and h_G by the values from (28)-(29) one obtains (32). Equating the coefficients of G_-^2 on both sides of (53) yields

$$v_{GG} = -\frac{(1+c)h_G^2 + d(h_G+\gamma)^2}{1-\beta(h_G+\gamma)^2}.$$

Substituting for h_G from (29) and noting (52), it follows that

$$\beta v_{GG}^2 - [1+c+d-\beta(1+c)\gamma^2]v_{GG} - (1+c)d\gamma^2 = 0.$$

This quadratic equation has one negative and one positive root. Since the equilibrium value function must be concave, we consider only the negative root, which is given by (31). Since v_{GG} is negative, condition (52) is trivially satisfied.

Next, we compute the steady state values G^{DIS}, R^{DIS}, and g^{DIS} that correspond to the above equilibrium. We already know that $R^{\mathrm{DIS}} = R^{\mathrm{BAU}}$. Using the fact that $\mathbf{R}_0$ is a constant function and that $\partial \mathbf{g}_0(G_-, R_i)/(\partial G_-) = h_G$ and $\partial \mathcal{G}_0(G_-)/(\partial G_-) = h_G + \gamma$, it follows from (24) by differentiating and evaluating at the steady state that

$$V_0'(G^{\mathrm{DIS}}) = [B'(R^{\mathrm{BAU}} + g^{\mathrm{DIS}}) - C'(g^{\mathrm{DIS}})]h_G - [D'(G^{\mathrm{DIS}}) - \beta V_0'(G^{\mathrm{DIS}})](h_G+\gamma).$$

Furthermore, evaluation of (25) at the steady state leads to

$$B'(R^{\mathrm{BAU}} + g^{\mathrm{DIS}}) - C'(g^{\mathrm{DIS}}) = D'(G^{\mathrm{DIS}}) - \beta V_0'(G^{\mathrm{DIS}}).$$

Solving the last two equations for $V_0'(G^{\mathrm{DIS}})$ and $B'(R^{\mathrm{BAU}} + g^{\mathrm{DIS}}) - C'(g^{\mathrm{DIS}})$ one obtains in particular

$$B'(R^{\mathrm{BAU}} + g^{\mathrm{DIS}}) - C'(g^{\mathrm{DIS}}) = \frac{D'(G^{\mathrm{DIS}})}{1-\beta\gamma}.$$

Using the functional forms of B, C, and D as well as $g^{\mathrm{DIS}} = \gamma G^{\mathrm{DIS}}$, this equation can be written as

$$-R^{\mathrm{BAU}} - (1-\gamma)(1+c)G^{\mathrm{DIS}} + z = \frac{dG^{\mathrm{DIS}}}{1-\beta\gamma},$$

from which we obtain (33).

Finally, it has to be shown that the greenhouse gas stock converges monotonically to G^{DIS}. Because of (21) and (51) it is sufficient to verify that $0 < h_G + \gamma < 1$ holds. Using (29) we obtain

$$h_G + \gamma = \frac{\gamma(1+c)}{1+c+d-\beta v_{GG}}.$$

It follows immediately from this expression, from $\gamma < 1$, and from $v_{GG} < 0$ that $0 < h_G + \gamma < 1$ is satisfied. This completes the proof of the theorem.

Proof of theorem 4

We know from lemma 6 that $\hat{c} = c$ implies that $\mathbf{R}_0(G_-) = R^{\mathrm{BAU}}$. This result is independent of the value of $\hat{d}$. In order to prove theorem 4 it is therefore sufficient to show that the maximization of aggregate welfare (4) subject to (2) and $R_{i,t} = R^{\mathrm{BAU}}$ for all $i \in [0,1]$ and all $t \geq 0$ leads to the same emission levels as the equilibrium under a benevolent authority described in theorem 3.

The maximization of (4) subject to (2) and $R_{i,t} = R^{\mathrm{BAU}}$ for all $i \in [0,1]$ and all $t \geq 0$ is equivalent to the maximization of

$$\sum_{t=0}^{+\infty} \beta^t \int_0^1 [B(R^{\mathrm{BAU}} + g_{i,t}) - C(g_{i,t}) - D(G_t)]\,\mathrm{d}i$$

subject to (2). Concavity of B and strict convexity of C imply that $g_{i,t} = g_t$ holds for all $i \in [0,1]$. The necessary and sufficient first-order optimality conditions are

$$G_t = g_t + \gamma G_{t-1},$$
$$D'(G_t) + C'(g_t) - B'(R^{\mathrm{BAU}} + g_t) = \beta\gamma[C'(g_{t+1}) - B'(R^{\mathrm{BAU}} + g_{t+1})],$$
$$\lim_{t\to+\infty} \beta^t [C'(g_t) - B'(R^{\mathrm{BAU}} + g_t)]G_t = 0.$$

Using the specification of the functions B, C, and D, the first two of these conditions can be rewritten as

$$\beta\gamma(1+c)G_{t+1} - [d + (1+\beta\gamma^2)(1+c)]G_t + \gamma(1+c)G_{t-1} = (1-\beta\gamma)(R^{\mathrm{BAU}} - z).$$

Using the definition of R^{BAU} it is easy to see that the unique fixed point of this difference equation is G^{DIS} as defined in (33). The eigenvalues of the difference equation are

$$\lambda_{1,2} = \frac{(1+\beta\gamma^2)(1+c) + d \pm \sqrt{[(1+\beta\gamma^2)(1+c)+d]^2 - 4\beta\gamma^2(1+c)^2}}{2\beta\gamma(1+c)}.$$

One can show that the smaller one of these two eigenvalues coincides with $h_G + \gamma$ where h_G is specified in (29). We also know from the proof of theorem 3 that this value is an element of the interval $(0,1)$. Hence, it follows that the optimal state trajectory of the optimization problem under consideration satisfies

$$G_t = (h_G + \gamma)G_{t-1} + (1 - h_G - \gamma)G^{\mathrm{DIS}}.$$

Since this difference equation coincides with the aggregate law of motion $G_t = \mathcal{G}_0(G_{t-1})$ from theorem 3, the proof of the present theorem is complete.

Formulas for the case $\hat{c} \neq c$

Consider the example from subsection 5.2 with $\hat{c} \neq c$. Equation (39) implies that

$$\mathbf{g}_0(G_-, R_i) = \hat{h}_1 + \hat{h}_G G_- + \hat{h}_R R_i = \frac{z - R_i - \hat{d}\mathcal{G}_0(G_-) + \beta[\hat{v}_G + \hat{v}_{GG}\mathcal{G}_0(G_-)]}{1+\hat{c}}, \tag{54}$$

such that

$$\hat{h}_R = -\frac{1}{1+\hat{c}}.$$

Equation (34) therefore gives

$$\mathcal{G}(G_-, \tilde{R}) = \hat{h}_1 + (\hat{h}_G + \gamma)G_- - \frac{1}{1+\hat{c}}\bar{R}. \tag{55}$$

Solving condition (35) we obtain

$$\mathbf{R}_0(G_-) = \frac{(1+\hat{c})[\hat{h}_1(c-\hat{c}) - (1-\beta\rho)k(1+\hat{c}) + \hat{c}z]}{c+\hat{c}^2} + \frac{(1+\hat{c})\hat{h}_G(c-\hat{c})}{c+\hat{c}^2}G_-. \tag{56}$$

Combining (36), (37), (55), and (56) it follows that

$$\begin{aligned} \mathcal{G}_0(G_-) &= \hat{h}_1 + (\hat{h}_G + \gamma)G_- - \frac{1}{1+\hat{c}}\mathbf{R}_0(G_-) \\ &= \frac{\hat{h}_1(\hat{c}+\hat{c}^2) + (1-\beta\rho)k(1+\hat{c}) - \hat{c}z}{c+\hat{c}^2} + \left(\gamma + \frac{\hat{c}+\hat{c}^2}{c+\hat{c}^2}\hat{h}_G\right)G_-. \end{aligned} \tag{57}$$

Substituting (57) into (54) and comparing the coefficients of G_- on both sides, we can solve for $\hat{h}_1$ and $\hat{h}_G$ in terms of $\hat{v}_G$ and $\hat{v}_{GG}$, which yields

$$\hat{h}_1 = \frac{(c+\hat{c}^2)z + \hat{d}[\hat{c}z - (1+\hat{c})(1-\beta\rho)k] + \beta[(c+\hat{c}^2)\hat{v}_G + (k+\hat{c}k-\hat{c}z)\hat{v}_{GG}] - \beta^2(1+\hat{c})k\rho\hat{v}_{GG}}{(1+\hat{c})[c+\hat{c}(\hat{c}+\hat{d}-\beta\hat{v}_{GG})]},$$

$$\hat{h}_G = -\frac{(c+\hat{c}^2)\gamma(\hat{d}-\beta\hat{v}_{GG})}{(1+\hat{c})[c+\hat{c}(\hat{c}+\hat{d}-\beta\hat{v}_{GG})]}.$$

Combining these two results with (56) and (57) we can express $\mathbf{R}_0(G_-)$, $\mathbf{g}_0(G_-, \mathbf{R}_0(G_-))$, and $\mathcal{G}_0(G_-)$ in terms of the variable G_- and the undetermined coefficients $\hat{v}_G$ and $\hat{v}_{GG}$. If we substitute these expressions into equation (38) and compare the coefficients of equal powers of G_- on both sides of the equation, we obtain three non-linear equations for the coefficients $\hat{v}_1$, $\hat{v}_G$, and $\hat{v}_{GG}$. Once we have solved these equations, we have determined an equilibrium under discretion, i.e., the authority's policy function $\mathbf{g}_0$ and value function $\hat{V}_0$. One can then solve equation (40) to determine aggregate welfare in this equilibrium.

References

[1] Barrett, S., "Self-enforcing international environmental agreements", *Oxford Economic Papers* **46** (1994), 879-894.

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

[3] Battaglini, M., and Harstad, B., "Participation and duration of environmental agreements" *Journal of Political Economy* **124** (2016), 160-204.

[4] D'Amato, A., and Dijkstra, B.R., "Technology choice and environmental regulation under asymmetric information", *Resource and Energy Economics* **41** (2015), 224-247.

[5] Esty, D., *Greening the GATT: Trade, Environment, and the Future*, Institute for International Economics, 1994.

[6] Gersbach, H., and Glazer, A., "Markets and regulatory hold-up problems, *Journal of Environmental Economics and Management* **37** (1999), 151-164.

[7] Harstad, B., "Climate contracts: A game of emissions, investments, negotiations, and renegotiations", *Review of Economic Studies* **79** (2012) 1527-1557.

[8] Harstad, B., "The dynamics of climate agreements", *Journal of the European Economic Association* **14** (2016) 719-752.

[9] Helm, D., Hepburn, C., and Mash, R., "Time-inconsistent environmental policy and optimal delegation", Discussion Paper 175, Oxford University Department of Economics, 2004.

[10] Hempel, L. *Environmental Governance: The Global Challenge*, Island Press, 1995.

[11] Nordhaus, W., *The Climate Casino*, Yale University Press, 2013.

[12] Petrakis, E., and Xepapadeas, A., "Location decisions of a polluting firm and the time consistency of environmental policy", *Resource and Energy Economics* **25** (2003), 197-214.

[13] Rogoff, K., "The optimal degree of commitment to an intermediate monetary target", *Quarterly Journal of Economics* **100** (1985) 1169-1189.

[14] Runge, W., *Freer Trade, Protected Environment: Balancing Trade Liberalization and Environmental Interests*, Council on Foreign Relations Press, 1994.

[15] Sorger, G., *Dynamic Economic Analysis: Deterministic Models in Discrete Time*, Cambridge University Press, 2015.

[16] Stern, N., *The Economics of Climate Change: The Stern Review*, Cambridge University Press, 2007.

[17] Strotz, R.H., "Myopia and inconsistency in dynamic utility maximization", *Review of Economic Studies* **23** (1955-1956), 165-180.

[18] *The Economist*, "Short-lived climate pollutants: low-hanging dirt", October 3rd, 2015a, p. 56.

[19] *The Economist*, "The way forward: second-best solutions", Special report: Climate change, November 28th, 2015b, p. 15-16.

[20] Woodford, M., "Optimal interest-rate smoothing", *The Review of Economic Studies* **70** (2003) 861-886.

[21] Yadav, R., "Climate Diplomacy: Do We Need a Supranational Authority", *The Politic*, November 1, 2014.

7 Online appendix

In this appendix we assume that the countries decide on the SEA's mandate in a discretionary fashion, a situation one could call *discretionary delegation*. At the start of every time period, the countries jointly choose the mandate $(\hat{c}, \hat{d})$ for the current authority. Subsequently they make their individual investment decisions and, finally, the SEA chooses the emission levels assigned to all countries. To simplify the exposition we restrict attention to the linear quadratic case.

In the environment under consideration, the mandate given to the authority will be a function of the stock of greenhouse gases, G_-. We refer to this function as a *mandate rule* and use the notation $(\hat{c}, \hat{d}) = M(G_-)$.[15] To determine the equilibrium mandate rule we proceed in three steps. First, we define equilibrium for a given arbitrary mandate rule M, which allows us to determine the welfare levels of the countries and the authority conditional on the mandate rule M being in place forever. Second, we determine the optimal current mandate $m = (\hat{c}, \hat{d})$ under the assumption that future mandates are determined by the rule M. Since the optimal current mandate depends on the current stock of greenhouse gases G_-, this step defines an optimal current mandate rule for any given future mandate rule, i.e., $\tilde{M}(G_-; M)$. Finally, we define the equilibrium mandate rule M^* as the fixed-point satisfying $\tilde{M}(G_-; M^*) = M^*(G_-)$.[16]

The equilibrium under an arbitrary mandate rule M can be characterized in analogy to definition 5 so that we do not present further details here. Let us denote the value functions of the SEA and the countries in this equilibrium by

$$\hat{V}(G_-, \tilde{R}_-; M) = \hat{V}_0(G_-; M) + k\rho\bar{R}_-,$$
$$V(G_-, \tilde{R}_-; M) = V_0(G_-; M) + k\rho\bar{R}_-,$$

where the notation we use makes explicit that the value functions depend on the mandate rule in place.

[15] By taking the limit of a finite horizon economy when the time horizon approaches infinity, it can be shown that in equilibrium the mandate rule is independent of $\tilde{R}_-$.

[16] Note that the equilibrium mandate rule derived in this way is time-consistent since the countries do not have an incentive to change the rule today if they perceive that this rule is in place from tomorrow onwards.

Let us now consider the optimal emission policies by a self of the SEA that is endowed with the mandate $m = (\hat{c}, \hat{d})$ and perceives future mandates to be determined by the rule M. This self chooses the emission levels $\{g_i \,|\, i \in [0,1]\}$ so as to maximize

$$\int_0^1 \left[-\frac{1}{2}(R_i + g_i - z)^2 - \frac{\hat{c}}{2} g_i^2 - \frac{\hat{d}}{2} G^2 \right] \mathrm{d}i + \beta \hat{V}_0(G; M)$$

subject to (2). Let the solution to this maximization problem be denoted by $\mathbf{g}_i(G_-, R_i; m, M)$, $i \in [0,1]$. Given the current mandate m and the policies $\mathbf{g}_i(G_-, R_i; m, M)$, The countries' optimal investment decisions can be derived as

$$\mathbf{R}(G_-; m, M) = \mathrm{argmax}_R \{ -\frac{1}{2}(R + \mathbf{g}_i(G_-, R; m, M) - z)^2 - \frac{c}{2} \mathbf{g}_i(G_-, R; m, M)^2 - (1-\beta\rho) k R \}.$$

Note that, as before, all countries choose the same technology level, independent of their initial stock $R_{i,-}$. Hence all countries will also receive the same emission quota,

$$\mathbf{g}(G_-; m, M) = \mathbf{g}_i(G_-, \mathbf{R}(G_-; m, M); m, M).$$

Aggregate world welfare, conditional on the current mandate m and the future mandate rule M, can be expressed as

$$\begin{aligned} V_0(G_-; m, M) + k\rho\bar{R}_- \;=\; & -\frac{1}{2}[\mathbf{R}(G_-; m, M) + \mathbf{g}(G_-; m, M) - z]^2 - \frac{c}{2}\mathbf{g}(G_-; m, M)^2 \\ & -\frac{d}{2}[\mathbf{g}(G_-; m, M) + \gamma G_-]^2 - k[\mathbf{R}(G_-; m, M) - \rho\bar{R}_-] \\ & +\beta V_0(\mathbf{g}(G_-; m, M) + \gamma G_-; M) + \beta k\rho \mathbf{R}(G_-; m, M). \end{aligned}$$

The optimal current mandate rule given the future mandate rule M can thus be computed as

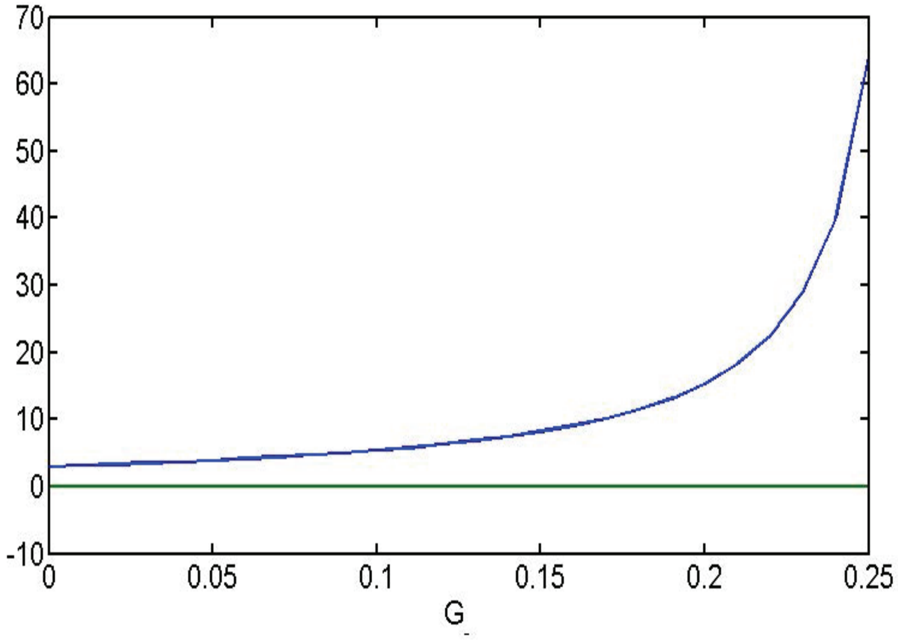
$$\tilde{M}(G_-; M) = \arg\max_{(\hat{c}, \hat{d})} V_0(G_-; (\hat{c}, \hat{d}), M).$$

Finally, the equilibrium mandate rule is determined as the fixed-point satisfying $M^*(G_-) = \tilde{M}(G_-; M^*)$.

Example 5 Assume the same parameters as in example 1. The steady state in the equilibrium with discretionary delegation is given by $G^{DD} = 0.1485$ and $R^{DD} = 0.8342$. In this steady state, the mandate given to the supranational environmental authority is $\hat{c} = 8.0784$, $\hat{d} = 0$.

Figure 4 expands on example 5 by illustrating the mandate rule in the neighborhood of the steady state. It shows that the optimal mandate rule for $\hat{c}$ is strongly increasing in the stock of greenhouse gases G_{-}, whereas the the optimal mandate rule for $\hat{d}$ is flat at $\hat{d} = 0$. Figure 5 compares world welfare under discretionary delegation (blue dashed line) to the

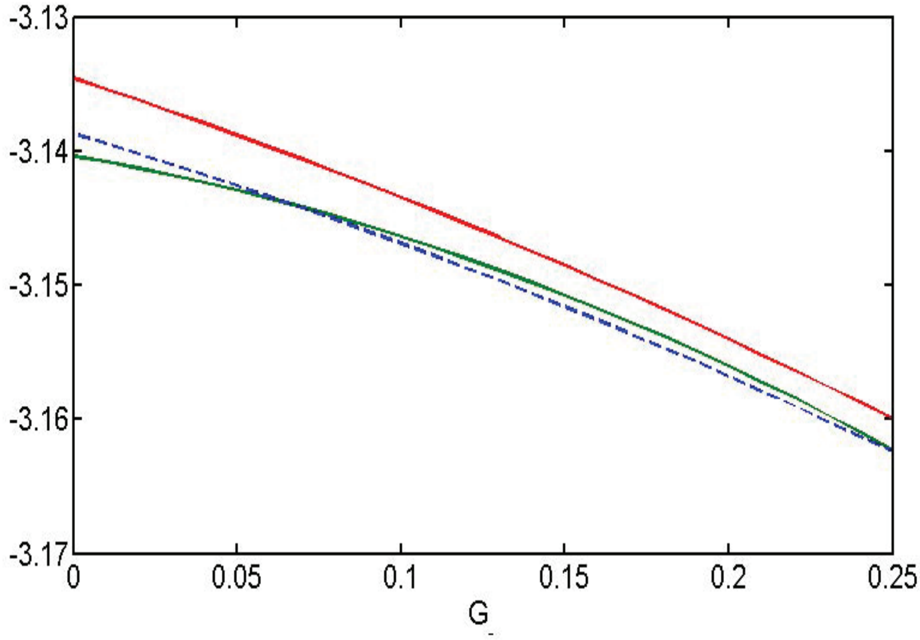
Figure 4: Discretionary delegation: mandate



Note: optimal mandate rule for $\hat{c}$ (blue line) and $\hat{d}$ (green line).

welfare levels in the first best (red line) and under optimal delgation to a non-benevolent authority with a time-invariant mandate (green line). Quantitatively, differences between the welfare levels under discretionary delegation and optimal time-invariant delegation are very small. Interestingly, the welfare functions intersect. This is possible because there are two mechanisms at work that are responsible for the welfare differences. First, in our environment with discretionary delegation the mandate is allowed to depend on the state of the economy. This greater flexibility compared to the case of a constant, time-invariant mandate leads to welfare gains. Second, the countries cannot commit over time to an optimal mandate, which leads to welfare losses relative to the case where countries commit to a constant, time-invariant mandate. These two opposing effects determine the welfare differences that we observe in figure 5. Finally, figure 6 shows the adjustment process to

Figure 5: Discretionary delegation: welfare



Note: World welfare under discretionary delegation (blue dashed line) to the welfare levels in the first best (red line) and under optimal delgation to a non-benevolent authority with a time-invariant mandate (green line).

the steady state, starting from a relatively large initial stock of greenhouse gases equal to $G_{-} = 0.25$. Under commitment and discretionary delegation, greenhouse gases are reduced noticeably faster compared to optimal delegation to a non-benevolent, discretionary authority with a time-invariant mandate. The stock of green technologies adjusts gradually under commitment and discretionary delegation, whereas it jumps to the steady state level instantaneously under optimal delegation to a non-benevolent, discretionary authority. The latter property is due to the fact that the optimal mandate prescribes $\hat{d} = 0$ in the region of the state space we consider, which implies in our model that both g and R are constant and independent of G_{-}.

Figure 6: Adjustment to the steady state