# Mirage on the Horizon: Geoengineering and Carbon Taxation Without Commitment*

Daron Acemoglu
MIT

Will Rafey
MIT

March 6, 2018

**Abstract**

We show that, in a model without commitment to future policies, geoengineering breakthroughs can have adverse environmental and welfare effects because they change the (equilibrium) carbon taxes. In our model, energy producers emit carbon, which creates a negative environmental externality, and may decide to switch to cleaner technology. A benevolent social planner sets carbon taxes without commitment. Higher future carbon taxes both reduce emissions given technology and encourage energy producers to switch to cleaner technology. Geoengineering advances, which reduce the negative environmental effects of the existing stock of carbon, decrease future carbon taxes and thus discourage private investments in conventional clean technology. We characterize the conditions under which these advances diminish—rather than improve—environmental quality and welfare.

JEL Classification: Q01, Q4, Q54, Q55, Q58, O30, O31, O33, C65
Keywords: carbon tax, geoengineering, environment, innovation, time inconsistency

*We thank Marco Battaglini, Patrick Behrer, Steve Cicala, Bård Harstad, Gustavo Joaquim, Antoine Levy, Torben Mideksa, William Nordhaus, Alp Simsek, James Stock, Ludwig Straub, Massimo Tavoni, Gernot Wagner, Martin Weitzman, and seminar participants at the U.C. Berkeley Advanced Workshop in Climate Economics, the Harvard Environmental Economics Lunch, the Harvard Solar Geoengineering Research Program, the MIT Joint Program on the Science and Policy of Global Change, and the MIT Macro Lunch for comments and useful suggestions. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 1122374. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Rafey also acknowledges financial support from the Becker Friedman Institute at the University of Chicago and the MIT Environmental Solutions Initiative.

# 1 Introduction

There is increasing recognition that a transition to cleaner technology has to be the bedrock of future reductions in carbon emissions.[1] While there have been important advances in "conventional" clean technologies, such as wind and solar, and economic research supports the notion that carbon taxes and other subsidies can contribute to the adoption of these technologies,[2] some experts and policy-makers instead pin their hopes on geoengineering breakthroughs, such as large-scale carbon sequestration, ocean fertilization, and solar radiation management (for example, Keith, 2013, Flannery, 2015, and Morton, 2015). Although such breakthroughs, if realized, could enable the global economy to achieve lower environmental damages without high carbon taxes, there are concerns that the prospect of geoengineering may delay or undermine other policy responses to climate change. As the Intergovernmental Panel on Climate Change (IPCC) concludes in its most recent report,

> One of the most prominent arguments against geoengineering suggests that geoengineering research activities might hamper mitigation efforts... which presumes that geoengineering should not be considered an acceptable substitute for mitigation. The central idea is that research increases the prospect of geoengineering being regarded as a serious alternative to emission reduction. (IPCC, 2014, p. 219)

In this paper, we provide a new and complementary reason why the prospect of geoengineering may, paradoxically, lead to worse environmental outcomes. In addition to incorporating geoengineering, our model features two plausible modifications relative to the simplest model of Pigovian carbon taxation. First, we introduce a conventional clean technology, which firms can adopt in order to reduce emissions when faced with a future carbon tax. Consistent with much of the evidence in the area of innovation, we assume that the development or adoption of cleaner conventional technologies today will make it cheaper to adopt them in the future.[3] Second, we assume that

[1]On anthropogenic climate change, see, e.g., Mann *et al.* (2017) on the contribution of human activity to recent European heat waves, Yan *et al.* (2016) on the intensification of tropical cyclones, Kopp *et al.* (2016) on new models of sea level rise, and Cook *et al.* (2016) on scientific consensus. On economic costs of climate change, see Dell *et al.* (2012) on growth, Hsiang *et al.* (2013) on conflict, and Costinot *et al.* (2016) on agriculture, as well as Greenstone *et al.* (2013) on calculating the social cost of carbon. For recent macroeconomic analyses of climate change, see Hassler *et al.* (2016), Nordhaus (2008, 2014), Desmet and Rossi-Hansberg (2015), and Golosov *et al.* (2014).

[2]We refer to wind, solar and geothermal technologies and to energy-saving incremental improvements, which firms themselves develop or invest in, as "conventional" technologies to distinguish them from the less-tested geoengineering technologies (which are likely to be developed by other entities).

[3]This type of externality arises naturally in almost all models of endogenous technology, including the quality ladder models of Aghion and Howitt (1992), Grossman and Helpman (1991) and Klette and Kortum (2004), as well as many of the horizontal innovation models, such as Romer (1990) and Jones (1995). See Acemoglu (2007) for a discussion. Aghion and Griffith (2005) and Akcigit and Kerr (forthcoming) provide evidence consistent with this type of externality in general, while Aghion *et al.* (2016) provide evidence for it in the context of the adoption of cleaner technologies in the automobile industry and Gillingham and Bollinger (2014) provide evidence in the context of solar installations.

policy is chosen by a social planner without the ability to commit to future policies. That policy-making is potentially "time inconsistent"—both because future decision-makers may be different than the current one and because even the same decision-maker may wish to revise policy plans and deviate from promises made in the past—has long been emphasized in many areas of economics.[4] In environmental economics, even if it has not been studied extensively, its importance has been recognized.[5] For example, in reviewing several economic frameworks for climate policy evaluation, William Nordhaus observes that

> [N]one of these approaches touch on the structure of actual intertemporal decision-making, in which this generation cannot decide for or tie the hands of future generations. Instead, each generation is in the position of one member of a relay team, handing off the baton of capital to the next generation, and hoping that future generations behave sensibly and avoid catastrophic choices by dropping or destroying the baton.... but this is largely uncharted territory in economic growth theory. (Nordhaus, 2007, p. 693)

The core of our argument is that time-inconsistency—beyond its general import—qualitatively changes the positive and normative implications of new technologies. A natural reference point for the carbon tax in a model with harmful carbon emissions is the Pigovian benchmark (where the carbon tax equals the marginal damage from one more unit of carbon). However, when the (social) planner would also like to encourage a transition to cleaner (conventional) technology, she would like to deviate from the Pigovian benchmark and set a higher tax rate to encourage more rapid technology adoption.[6] But in a world without commitment to future policies, firms will anticipate that any promised taxes above the Pigovian level will be revised, and underinvest in clean technology.

It is into this setting that we introduce the prospect of geoengineering. For clarity, and in

[4] Throughout, we use the term "time-inconsistency" in the spirit of Kydland and Prescott (1977) and Calvo (1978) to signify that Bellman's (1957) theorem of sequential optimality fails even for a standard, additively separable exponentially-discounted objective function because the constraint facing the decision-maker changes over time (here, due to decisions made by other agents). For examples in the context of fiscal policy, see Chari and Kehoe (1990), Benhabib and Rustichini (1997), Phelan and Stacchetti (2001), and Klein *et al.* (2008).

[5] Several noteworthy about-turns in environmental policy in the OECD illustrate the relevance of time-inconsistency concerns in this area. Major political reversals include Canada's withdrawal from the Kyoto Accord in 2011, Australia's repeal of its carbon tax in 2014, and the U.S.'s rollback of the Clean Power Plan and repudiation of the Paris Accord in 2017. Examples of regulators revising energy pricing schemes in response to technological change include the Spanish solar-feed-in-tariff, where the government reneged on solar subsidies after the production costs fell by more than regulators anticipated, and the U.K.'s decision in late 2011 to cut solar subsidies under the 2008 Energy Act by 55%.

[6] As emphasized in Acemoglu *et al.* (2012, 2016), the presence of clean technology innovations or investments necessitates augmenting carbon taxes with direct subsidies for the development or adoption of these technologies. In our model, in contrast, we assume that the planner uses a carbon tax to encourage these investments. This is motivated by practical difficulties in using subsidies to costlessly encourage a complete transition to clean technology (see, e.g., Borenstein, 2012, pp. 79–80). We show in Proposition 6 how our qualitative results generalize to the case in which the social planner can also subsidize clean technology.

line with the IPCC's own taxonomy, we distinguish between two different types of geoengineering technologies (recognizing that some real-world technologies are a mixture of these two types): type I technologies, *carbon removal*, which correspond to a rightward shift of the environmental damage function (or, equivalently, reduce the effective stock of carbon that enters the damage function by a constant amount), and type II technologies, *climate adaptation* or *solar radiation management*, which reduce marginal damages from carbon in the atmosphere.[7] Examples of type I technologies include all forms of large-scale carbon dioxide removal, like mass afforestation, biochar, ambient air capture, and ocean fertilization.[8] Examples of type II technologies include solar radiation management, such as albedo enhancement, space reflectors, or stratospheric aerosols.[9] While some type II geoengineering technologies appear to be the most empirically relevant, due to their low predicted costs, type I technologies may also experience breakthroughs.

Though each type of geoengineering technology has somewhat different implications, they both work in a similar manner. Without an equilibrium response, geoengineering of either type reduces future damages, and thus future Pigovian carbon taxes. But when future Pigovian carbon taxes determine today's investment in conventional clean technologies, we show that the advent of geoengineering technologies *increases* underinvestment in these socially valuable technologies.

More specifically, we demonstrate that type I geoengineering technologies reduce investment in conventional clean technology so much that overall damages remain at the same level as before geoengineering. Intuitively, in our model to restore incentives for the adoption of conventional clean technologies, the Pigovian tax (marginal damage of carbon) needs to be at a certain level. With type I technologies, when the marginal value of damages remains the same so does the level of overall damages. Interestingly, even though overall damages remain constant, welfare may decline because the problem of underinvestment in cleaner technologies becomes more severe with the geoengineering advances. More ominously, we show that type II geoengineering technologies may actually lead to greater damages (depending on an elasticity condition for the damage function) and are more likely to reduce welfare.

The reason why geoengineering technologies backfire in our model is very different from those emphasized in previous discussions, which focus on potential downsides of the prospect of geoengineering because major geoengineering breakthroughs may not be realized or may create new, unrelated environmental risks. Instead, our framework identifies potential inefficiencies from geo-

[7]See IPCC (2014, pp. 484–489): "Two categories of geoengineering are generally distinguished. Removal of GHGs, in particular carbon dioxide termed 'carbon dioxide removal' or CDR, would reduce atmospheric GHG concentrations.. . . 'Solar radiation management' or SRM technologies aim to increase the reflection of sunlight to cool the planet and do not fall within the usual definitions of mitigation and adaptation."

[8]See, for example, Lenton and Vaughan (2009) for a scientific assessment of these technologies for global climate dynamics and Socolow *et al.* (2011) and Lackner *et al.* (2012) for cost estimates for direct carbon removal.

[9]See, for example, the analysis in Rasch *et al.* (2008) and the survey in National Research Council (2015). On low predicted costs, see McClellan *et al.* (2012) and Keith *et al.* (2017).

engineering that arise precisely because the breakthroughs *will be* realized.

We first develop these ideas in the simplest setting, which is a static world with ex ante identical firms. Each firm first undertakes a costly investment to switch to a cleaner production technology anticipating the future carbon tax and any geoengineering breakthroughs. A benevolent planner sets the carbon tax after these conventional clean technology investments are made, but before production decisions. Production decisions create emissions, which contribute to the stock of carbon in the atmosphere, and a convex (social) damage function determines the welfare costs from this stock of carbon. The key technological externality—that clean technology investments make future clean technology cheaper—arises from a simple premise: a fraction of firms are replaced by new entrants, and if they have invested in the clean technology, the new entrant can inherit this improvement. This externality implies that the planner would like to choose a carbon tax rate above the Pigovian level, but the aforementioned time-inconsistency problem means that she cannot deviate from Pigovian taxes, leading to underinvestment in the conventional clean technology.

What simplifies the analysis of this setting is that there exists a unique level of the carbon tax that satisfies the technology IC—making the ex ante identical firms indifferent between investing in the clean technology and not. Provided that it is optimal to have some firms invest in the clean technology, the stock of carbon in the atmosphere has to adjust in order to satisfy the technology IC. In this light, the implications of various different types of geoengineering technologies become straightforward. A type I geoengineering technology, for example, shifts the damage function to reduce the level of the Pigovian carbon tax at a given stock of carbon in the atmosphere. But at this lower level of carbon tax, the technology IC is violated. To restore IC, the stock of carbon in the atmosphere must increase to offset the benefits from geoengineering. The logic for type II technologies is similar, except that in this case following geoengineering, the overall level of damages increase not to their original level but to restore the marginal value of damages to their original level. Depending on the elasticity of the damage function, this may involve an increase in the level of damages relative to the benchmark without geoengineering.

After expositing our main ideas in a transparent manner in this static model, we move to a continuous-time model of endogenous technological change with quality ladders. This model is useful for micro-founding the technological externality introduced above and demonstrating that the results discussed in the previous two paragraphs do not depend on assuming a static setting.

In our dynamic model, each active firm operates the best available technology in a given energy-related activity, and is stochastically replaced by a new entrant that builds and improves upon its productivity. The key technology externality emerges from the assumption that firms face nonzero probabilities of replacement. We characterize the dynamic equilibrium with a time-inconsistent planner in this setting. Though Pigovian taxes become more complicated (because they take into account future damages), we show that the results in the unique balanced growth path (BGP) are qualitatively identical to those we obtained in the static setting (and that the dynamic equilibrium converges to the BGP).

We also discuss how relaxing several of the assumptions adopted for simplicity and transparency does not affect our qualitative conclusions. These include introducing heterogeneity across firms in terms of their costs of switching to clean technology, allowing policy to imperfectly internalize climate externalities, introducing direct subsidies to clean technology, making geoengineering advances stochastic, and modifying how environmental externalities impact utility or productivity.

Our work relates to the small but growing literature on clean technology investments and innovations. In addition to Acemoglu *et al.* (2012, 2016) and Aghion *et al.* (2016), which have been mentioned above, Bovenberg and Smulders (1995, 1996), Goulder and Mathai (2000), Goulder and Schneider (1999), Grimaud *et al.* (2011), Hartley *et al.* (2016), Hassler *et al.* (2012), Newell *et al.* (1999), Popp (2002, 2004), and van der Zwaan *et al.* (2002) also discuss endogenous technology in the context of environmental policy and climate change.

Our work is also related to a very small literature on environmental policy without commitment (surveyed in Karp and Newbery, 1993). Laffont and Tirole (1996a,b) study pollution permit markets and innovation in a two-period setting with asymmetric information, where the regulator cannot commit and the usual hold-up problem arises. Harstad (2012) and Harstad and Battaglini (2016) study incentives to invest in clean technologies in the presence of multiple regulators without commitment, and show there will be underinvestment because additional investments reduce each regulator's bargaining position against the others. Also notable is Harstad (2016) who analyzes environmental policy and technology choices under hyperbolic discounting. We are not aware of any papers that model or note how with time-inconsistency, improvements in technologies can lead to declines in welfare.[10]

Finally, there are some recent papers on geoengineering in the economics literature. Most relevant to our work are Barrett (2008), Weitzman (2015), and Moreno-Cruz (2015), who focus on the international political economy dimensions of geoengineering technologies to study the risks of unilateral geoengineering when the technology imposes externalities on other countries. Separately, Emmerling and Tavoni (2018) explore the optimal combination of conventional technology and stochastic geoengineering from the viewpoint of a social planner, using an integrated assessment model that assumes full commitment.

The remainder of the paper is organized as follows. Section 2 introduces our model and char-

[10]This point is related to but distinct from the general idea that equilibrium responses may undo or weaken benefits from technological improvements. For example, Peltzman (1975) has observed that adjustment in driver behavior may undo the benefits from automobile safety devices such as seatbelts. Kousky *et al.* (2006) and Boustan *et al.* (2012) make a similar point in the context of flood levees, private investments and migration decisions in areas prone to natural disaster.

acterizes the equilibrium. Section 3 extends our results to an infinite-horizon setting in continuous time. Section 4 concludes. Appendix A contains proofs omitted from the text, while online Appendix B contains proofs for the infinite-horizon model and demonstrates the robustness of our main results to various variations on our model and its assumptions.

# 2 Baseline Model

In this section, we introduce our baseline static model. In the next section, we consider a dynamic model which provides a clearer microfoundation for some of the assumptions used in this section and fits more naturally with existing economic models of climate change, but still delivers essentially identical results.

## 2.1 Production and Environmental Damages

We consider an economy consisting of a range of energy-related activities, represented by the continuum $[0, 1]$. For simplicity, we take the output of these activities to be perfectly substitutable. Initially, firm $i$ controls the production technology for activity $i \in [0, 1]$, and by using $k_i$ units of the final good as inputs, it can produce

$$f_d(k_i)$$

units of output. The production function $f_d$ is assumed to be twice continuously differentiable, increasing and concave with the usual Inada conditions to ensure interior solutions (i.e., $\lim_{k\to 0} f_d'(k) = \infty$ and $\lim_{k\to\infty} f_d'(k) = 0$). Since all activities are perfectly substitutable, firms will act competitively and we choose the price of energy output as numéraire (normalizing it to 1).[11]

As indicated by the subscript "$d$,"the initial production technology is "dirty," and generates $k_i$ units of carbon emissions. By incurring a cost $\Gamma > 0$, each firm can upgrade to a (conventional) cleaner technology that produces

$$f_c(k_i)$$

units of output (where $f_c$ is also twice continuously differentiable, increasing and concave, and satisfies $\lim_{k\to 0} f_c'(k) = \infty$ and $\lim_{k\to\infty} f_c'(k) = 0$), but only $\gamma k_i$ units of carbon (where $\gamma < 1$). We think of clean firms as switching to a technology that produces energy output from cleaner sources such as wind or solar energy, or upgrading their existing plant's efficiency to reduce emissions per unit of energy production.

We also assume that in each activity $i \in [0, 1]$, a new entrant replaces the incumbent firm with probability $\lambda \in [0, 1)$. If the incumbent has already transitioned to clean technology, the entrant

[11] Though different activities are perfect substitutes, the Inada conditions on the production functions ensure that all of them will be produced in equilibrium.

inherits it.[12] The entrant has access to the same production technology as the incumbent ($f_d$ if there has not been a transition to clean technology, and $f_c$ if there has been such a transition).[13]

Given these assumptions, denoting the fraction of activities that have switched to clean technology by $q$, total emissions in the economy can be written as

$$E = q\gamma k_c + (1-q)k_d, \tag{1}$$

where $k_c$ is the equilibrium production level of clean technology and $k_d$ is the equilibrium production level of dirty technology (here we are using the fact that both entrants and incumbents will choose the same level of investment given their technology). The presence of the term $\gamma < 1$ captures the fact that input usage by clean firms creates lower emissions.

Finally, we assume that the stock of carbon in the atmosphere is given by

$$S = (1-\delta)S_0 + E, \tag{2}$$

where $S_0 \geq 0$ is the initial level of carbon, and $\delta \in [0,1]$ denotes "depreciation" of this stock of carbon (for example, by absorption by oceans and forest cover). We choose this formulation to create continuity with the dynamic model in the next section. The damages from carbon in the atmosphere are denoted by

$$D(S;\xi,\upsilon), \tag{3}$$

where $D$ is an increasing, twice continuously differentiable and strictly convex function, and for now, damages are taken to be additive, and the parameters $\xi$ and $\upsilon$ will be used to model the effects of other types of geoengineering advances on environmental damages. For now, we suppress these parameters, writing environmental damages simply as $D(S)$.

## 2.2 Carbon Tax and Production Decisions

Firms pay a carbon tax of $\tau$ per unit of their emissions. Thus the profit maximization problems of the two types of firms can be written as

$$\begin{aligned} \pi_d(\tau) &= \max_{k\geq 0} f_d(k) - (1+\tau)k \\ &= f_d(k_d(\tau)) - (1+\tau)k_d(\tau), \end{aligned}$$

and

[12] Nothing in our qualitative results below change if we instead assume that the entrant could use the clean technology at some cost $\Gamma_{\text{entrant}} < \Gamma$. Our specification can be viewed as the special case with $\Gamma_{\text{entrant}} = 0$, adopted for simplicity. An incumbent's clean investment may create lower costs for an entrant if, for example, there exists any learning-by-doing in that activity, or if imperfect patent protection (or, in the dynamic model, quality-ladder-location-specific investment) prevents an exiting incumbent from recouping the entire value of their clean investment. For evidence on these types of spillovers, see footnote 3.

[13] This structure of entrants replacing incumbents will be further micro-founded in the context of the dynamic model in the next section.

$$\begin{aligned}\pi_c(\tau) &= \max_{k\geq 0} f_c(k) - (1+\gamma\tau)k \\ &= f_c(k_c(\tau)) - (1+\gamma\tau)k_c(\tau),\end{aligned}$$

where $k_d(\tau)$ is defined as the profit-maximizing level of input choice for a dirty firm, and $k_c(\tau)$ is the profit-maximizing level of input choice for a clean firm.

The difference between the profit-maximization problem of the two types of firms stems from the difference in their production functions and—more crucially for our focus—from the fact that clean firms pollute less per unit of input (i.e., $\gamma < 1$). That clean firms pollute less per unit of input does not, however, guarantee that their overall emissions are less than that of dirty firms, since they may choose higher levels of input usage. This possibility, first noted by Jevons (1866), may lead to greater overall emissions by clean firms. Our next assumption ensures that this is not the case.

**Assumption 1 (No Jevons)** *For all $\tau \geq 0$, we have*

$$\Lambda(\tau) \equiv k_d(\tau) - \gamma k_c(\tau) > 0.$$

This assumption is not restrictive and is automatically satisfied when $\gamma = 0$ or $f_c = f_d$.

## 2.3 Clean Technology Decisions

The difference in profits between a clean and a dirty firm can be written as

$$\begin{aligned}\Psi(\tau) &= \pi_c(\tau) - \pi_d(\tau) \\ &= [f_c(k_c(\tau)) - (1+\gamma\tau)k_c(\tau)] - [f_d(k_d(\tau)) - (1+\tau)k_d(\tau)] \\ &= [f_c(k_c(\tau)) - k_c(\tau)] - [f_d(k_d(\tau)) - k_d(\tau)] + \tau\Lambda(\tau),\end{aligned} \tag{4}$$

where $\Lambda(\tau)$ is the change in emissions from switching to a clean technology defined in Assumption 1.

Recall that firms make their investment to switch to clean technology before they know whether they will be replaced by a new firm, and enjoy the additional profits from clean technology, $\Psi(\tau)$, only if they are not thus replaced (an event of probability $1-\lambda$). Consequently, a firm will find it (privately) optimal to switch to clean technology only if the condition

$$(1-\lambda)\Psi(\tau) \geq \Gamma$$

is satisfied. Our key results will follow from the interplay between the effect of various geoengineering technologies and the equilibrium tax rate implied by this Technology IC constraint.

In what follows, we denote the fraction of firms that switch to clean technology by $q$. The following lemma is immediate (proof omitted):

**Lemma 1 (Incentive Compatible Technology Choice)**

$$\begin{cases} \Psi(\tau) > \frac{\Gamma}{1-\lambda} \implies q = 1 \\ \Psi(\tau) = \frac{\Gamma}{1-\lambda} \implies q \in [0,1] \\ \Psi(\tau) < \frac{\Gamma}{1-\lambda} \implies q = 0. \end{cases} \tag{Technology IC}$$

Note that when Technology IC holds exactly, i.e.,

$$\Psi(\tau) = \frac{\Gamma}{1-\lambda}, \tag{5}$$

any fraction of firms switching to clean technology is privately optimal. Conversely, when this equality does not hold, either all firms or no firms will make the switch to clean technology. Since we show that $\Psi(\tau)$ is increasing in the next lemma, (5) defines a unique carbon tax rate, which we denote by $\hat{\tau}$.

The following lemma shows that higher taxes increase the incentives to switch to clean technology (proof omitted).

**Lemma 2 (Carbon Tax and Technology IC)** *Suppose Assumption 1 holds. Then*

$$\frac{d\Psi(\tau)}{d\tau} = \Lambda(\tau) > 0.$$

The result that a small increase in the carbon tax affects (Technology IC) only through $\Lambda(\tau)$ follows from the Envelope Theorem, or simply from using the fact that both clean and dirty firms are choosing profit-maximizing levels of input usage. That this effect is positive is a consequence of Assumption 1. This result greatly simplifies our analysis by ensuring that the function $\Psi$ is monotone.

## 2.4 The Planner's Problem

The (social) planner maximizes utilitarian welfare. Imposing, without loss of any generality, that all dirty (clean) firms choose the same level of inputs, welfare can be written as

$$\begin{aligned} W &= (1-q)[f_d(k_d) - k_d] + q[f_c(k_c) - k_c] - q\Gamma - D(S) \\ &= (1-q)\pi_d + q\pi_c + (k_d - q\Lambda)\tau - q\Gamma - D(S), \end{aligned} \tag{6}$$

where, as in Assumption 1, we write $\Lambda = k_d - \gamma k_c > 0$.

There are three important observations. First, differently from private firms, the planner cares about the actual cost of inputs, and not about the taxes; this can be seen by the presence of the term $\Lambda$. Second, she also cares about the externality from emissions, as captured by the term $D(S)$. Third, the probability that a current producer is replaced by a new entrant, $\lambda$, which was important for private decisions to invest in clean technology, does not feature in this objective function because the new entrant will be able to produce with the same technology.

Until Section 2.11, we assume that the planner has access to a single instrument—a carbon tax, $\tau$.

### 2.5 Timing of Events

The key assumption, already highlighted in the Introduction, is that of *the lack of commitment* to future policies, which induces *time-inconsistency.* Namely, the planner is not able to choose, and commit to, the carbon tax sequence ahead of all other decisions. In the static model, we incorporate this feature with the following timing of events:

- All firms simultaneously make their technology decisions.
- Firms that will be replaced by new entrants are revealed.
- The planner chooses the carbon tax, $\tau$.
- Given the carbon tax $\tau$, all firms simultaneously choose their input levels.

### 2.6 Equilibrium

Given the above description, a (subgame perfect) equilibrium can be defined as tuple $(q^*, \tau^*, k_d^*, k_c^*)$ such that

- Given $q^*$, $\tau^*$ maximizes $W$ as in (6);
- $q^*$ satisfies (Technology IC);
- Given $\tau^*$, $k_d^*$ and $k_c^*$ maximize, respectively, $\pi_d$ and $\pi_c$.

Since the maximization problem of both clean and dirty firms is strictly concave, the equilibrium will always feature the same level of inputs for a given type of firm, denoted respectively by $k_d(\tau)$ and $k_c(\tau)$ as defined above. Then, once $q^*$ and $\tau^*$ are determined, the level of emissions can be computed from equation (1) as $E(\tau^*, q^*)$, and the level of stock of carbon in the atmosphere from equation (2) as $S(\tau^*, q^*)$. In view of this, we summarize the equilibrium simply by $(\tau^*, q^*)$, corresponding to the level of carbon tax and fraction of firms switching to clean technology.

### 2.7 Pigovian Carbon Taxes

A first implication of the timing of events adopted here (which incorporates the time-inconsistency feature mentioned above) is that the carbon tax will always be Pigovian—it will equal the marginal damage created by one more unit of emissions. This structure of Pigovian taxation contrasts with the case in which the planner can commit to carbon taxes, as we will see later.

More formally, we have:

**Proposition 1** *In equilibrium, the carbon tax is given as*

$$\tau^* = D'(S(\tau^*, q^*)). \tag{7}$$

This result follows straightforwardly by differentiating the planner's objective function, (6). The Pigovian tax given in (7) will play a central role throughout the paper.

### 2.8 Characterization of Equilibrium

In the rest of the analysis, we impose the following assumption, which ensures the existence of an *interior equilibrium*, meaning one in which some firms switch to clean technology, while others do not.[14]

**Assumption 2 (Conditions for Interior Equilibrium)** *We have*

$$\frac{\Gamma}{1-\lambda} \in (\Psi(\underline{\tau}), \Psi(\overline{\tau}))$$

*where* $\underline{\tau} = D'\left((1-\delta)S_0 + \gamma k_c(\underline{\tau})\right)$ *and* $\overline{\tau} = D'\left((1-\delta)S_0 + k_d(\overline{\tau})\right)$.

This assumption ensures that condition (5) holds and the equilibrium is interior. It implies that when all firms make the switch to clean technology, the stock of carbon is low enough that the planner chooses a relatively low level of carbon tax (the one given by $\underline{\tau}$ in this assumption), and when no firm makes the switch, the stock of carbon is high enough that the planner chooses a relatively high level of carbon tax (the one given by $\overline{\tau}$ in this assumption).

When this assumption does not hold, there exists a unique equilibrium in which all firms switch to the clean technology or no firm switches to the clean technology, and in neither case do we have interesting comparative statics of investment in clean technology (small changes in parameters will not impact clean technology decisions). Thus Assumption 2 restricts the analysis to the interesting subset of the parameter space, where the equilibrium is interior. This is also empirically reasonable—in practice, only a limited fraction of energy producers have made the transition to clean technology, and there exist marginal clean investment decisions that will be impacted by future carbon taxes.

The next proposition characterizes the unique interior equilibrium.

[14]This equilibrium can also be labeled "asymmetric" because some ex-ante identical firms switch to clean technology, while others do not. We show in Appendix A2, however, that asymmetry is not the important feature, and similar results obtain when firms are heterogeneous ex ante in terms of their cost of switching to clean technology. The important feature, thus, is that in our equilibrium, the transition to clean technology is not complete.

**Proposition 2 (Interior Equilibrium)** *Suppose Assumptions 1 and 2 hold. Then there exists a unique equilibrium given by* $(\tau^*, q^*) = (\hat{\tau}, \hat{q})$, *where* $(\hat{\tau}, \hat{q}) = (D'(S(\hat{\tau}, \hat{q})), \hat{q})$. *This equilibrium is interior in the sense that the fraction of firms switching to clean technology* $\hat{q}$ *is strictly between 0 and 1.*

**Proof.** See Appendix A. ■

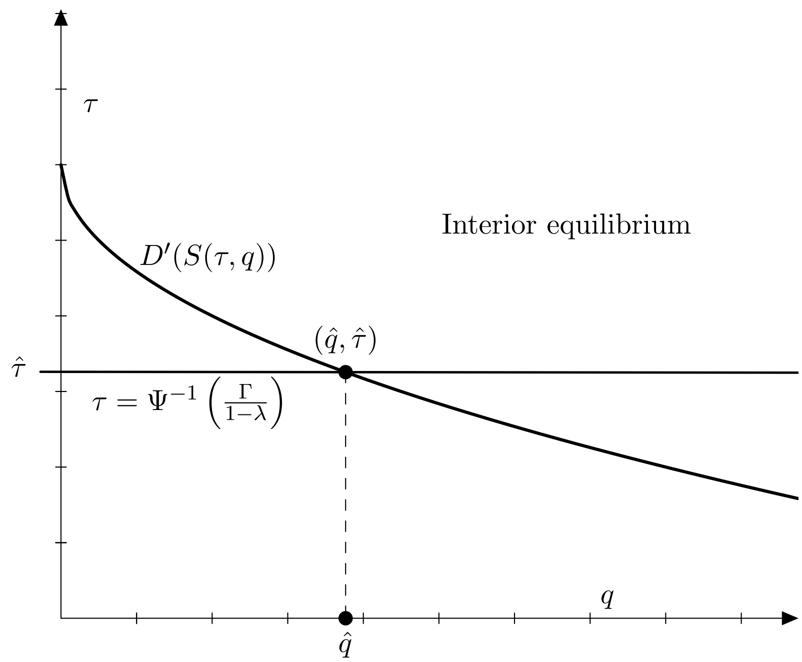


Figure 1: Unique interior equilibrium $(\hat{\tau}, \hat{q})$.

The first noteworthy result in this proposition is the uniqueness of an interior equilibrium. The reason why the equilibrium is interior and only a fraction of firms switch to the clean technology relates to the main economic force in our model. Firms, at the margin, switch to clean technology because of the carbon tax. The higher the carbon tax, the more inclined they are to make this transition. However, the carbon tax is determined by the planner after the technology decisions are made and will be lower when more firms have made the switch to clean technology—and herein lies the time-inconsistency problem. In particular, as already emphasized, in an interior equilibrium (5) needs to hold as equality. This implies that the carbon tax needs to take a specific value, $\hat{\tau}$, as given in Proposition 2 and represented by the horizontal line in Figure 1. Given the convexity of damages in (3), the Pigovian carbon tax the social planner will set is increasing in the stock of carbon and hence decreasing in $\hat{q}$, guaranteeing uniqueness. But, for $\hat{\tau}$ to emerge as the planner's choice, the stock of carbon in the atmosphere needs to take a specific value, $S(\hat{\tau}, \hat{q})$, and exactly $\hat{q}$ fraction of firms need to switch to clean technology. If more firms than $\hat{q}$ were to switch to clean technology, there would be less carbon in the atmosphere than $S(\hat{\tau}, \hat{q})$, and consequently, the planner would choose a lower carbon tax than $\hat{\tau}$, violating (5). Likewise, if fewer firms than $\hat{q}$ made the switch, the carbon tax rate would be higher than $\hat{\tau}$, once again violating (5).

In addition to the existence of a unique interior equilibrium, the most important conclusion of Proposition 2 is that the level of carbon taxes will be Pigovian. This is dictated by the timing of events. At the time the planner sets the tax rate, technology decisions have already been made—in view of the fact that the planner cannot commit to carbon taxes ex ante. Without an influence on technology decisions, there is no reason for the planner to deviate from the Pigovian benchmark.

This contrasts with what the planner would have preferred if she could commit to the carbon tax, as we show next.

### 2.9 Second-Best

In this subsection, we briefly contrast the equilibrium with the "second-best" allocation where the planner can commit to carbon taxes in advance of the technology decisions of energy firms (but still cannot dictate input choices and technology decisions, hence the label "second-best"). This comparison will highlight the implications of time-inconsistency in our model.

Suppose that the planner sets a carbon tax rate $\tau$, and commits to it, before the technology decisions of firms.[15] The next proposition shows that as long as $\lambda > 0$ the planner deviates from the Pigovian tax and induces more firms to switch to the clean technology than in the case without commitment.

**Proposition 3 (Second-best)** *1. Suppose $\lambda > 0$. Then the planner commits to a carbon tax $\tau^{SB} = \hat{\tau} > D'(S^{SB})$, and the equilibrium fraction of firms that switch to clean technology is $q^{SB} > \hat{q}$, where $S^{SB}$ is the stock of carbon in the second-best allocation (with commitment).*

*2. Suppose $\lambda = 0$. Then the planner commits to a carbon tax $\tau^{SB} = \hat{\tau} = D'(S^{SB})$, and the equilibrium fraction of firms that switch to clean technology is $q^{SB} = \hat{q}$.*

**Proof.** See Appendix A. ■

The first part of this proposition shows that, provided that $\lambda > 0$, the planner would like to deviate from Pigovian taxation. Recall that Pigovian taxation implies $\tau^{SB} = D'(S^{SB})$, whereas the planner would like to commit to a tax $\tau^{SB} > D'(S^{SB})$.[16] This is because when $\lambda > 0$, there is underinvestment in clean technology, because firms do not take into account the benefit they

[15] When we endow the planner with commitment power, if the planner commits to the (unique) incentive-compatible tax $\hat{\tau}$, then technically any $q \in [0, 1]$ may be an equilibrium. In the spirit of general mechanism design or principal-agent problems, we impose the incentive-compatibility constraints and let the planner or the principal choose her favorite allocation consistent with incentive-compatibility.

[16] Note, however, that we still have $\tau^{SB} = \hat{\tau}$, since the planner cannot control investments in clean technology and thus has to satisfy (Technology IC).

create for others who will build on their clean technology investments. As a result, in the second-best allocation where she cannot directly control technology investments but still can commit to a tax, the planner would like to encourage greater investment in clean technology by setting higher carbon taxes than the Pigovian benchmark, in order to induce more firms to switch to the clean technology. However, without commitment, the planner cannot achieve a non-Pigovian carbon tax, and the equilibrium always involves too little investment in clean technology, i.e., $\hat{q} < q^{SB}$.

The second part of the proposition highlights the role of $\lambda > 0$. When $\lambda = 0$, firms fully internalize the benefits from a switch to clean technology. In this case, setting the right price of carbon—i.e., the Pigovian tax—is sufficient to induce the right level of technology investment, and thus the planner has no reason to resort to a non-Pigovian carbon tax.

One consequence of Proposition 3 is that, when $\lambda > 0$ as we assume to be the case throughout the rest of the analysis, there is too little investment in clean technology and too much carbon in the atmosphere. Thus any further increase in the stock of carbon has a first-order negative impact on welfare.

### 2.10 The Effects of Geoengineering

We next study the implications of geoengineering technologies on equilibrium carbon taxes, investment in clean technologies, environmental damages and welfare. By geoengineering technologies, we refer to technological advances that reduce the damages from a given stock of carbon and are operated by the government or some other entity (but not the firms themselves). We distinguish between two different types of geoengineering technologies, which we first enumerate and motivate. We then analyze their implications separately. Actual geoengineering breakthroughs may combine features from these two types, but it is useful for our purposes to exposit their implications separately.

To incorporate each type of geoengineering, let us make the role of the different parameters explicit as follows, writing

$$D(S;\xi,\upsilon) = (1-\upsilon)\tilde{D}((1-\delta)S_0 - \xi + E),$$

where $\tilde{D}$ is a base damage function, and changes in the parameters $\xi \geq 0$ and $\upsilon \in [0,1)$ each shift the environmental damage function.

More specifically, the first type of geoengineering technology, which we refer to as *carbon removal* or *geoengineering technology of type I*, corresponds to an increase in $\xi$, and thus leads to a parallel rightward shift of the environmental damage function as shown in the left panel of Figure 2. In practice, this corresponds to large-scale carbon sequestration schemes that capture carbon from the air, such as permanent afforestation or algae blooms.

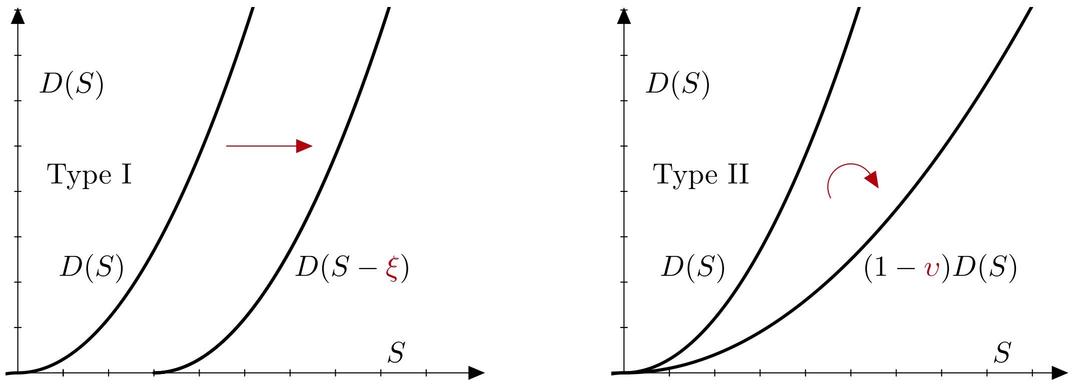


Figure 2: Geoengineering of type I ("carbon removal") and type II ("climate adaptation" or "solar radiation management").

The second type of geoengineering technology, *climate adaptation*, *solar radiation management*, or *geoengineering technology of type II*, corresponds to an increase in $\upsilon$, a proportional rightward shift or rotation of the environmental damage function as illustrated in the right panel of Figure 2. We interpret this class as representing a range of technologies related to solar radiation management, aiming to slow temperature rise at a given emissions level. The most well-known example is the injection of sulfur dioxide into the stratosphere—suggested most notably by the Nobel prize-winning chemist Paul Crutzen—in order to reduce surface temperatures. Less esoteric geoengineering solutions that reduce marginal damages via various technological adaptations fall within this category as well.

In addition to the way in which we map the technical aspects of climate engineering into climate damages, there are two other features of the way that we model geoengineering that matter for our results. First, we conceive of geoengineering advances as *exogenous*, in the sense that they are not consequences of decisions made by agents within the model. Second, it is also important for our results that these advances are small (hence the infinitesimal changes in the propositions). Specifically, geoengineering improvements cannot be so large as to shift the economy out of an interior equilibrium into a world in which clean technology investments no longer respond at the margin to carbon taxes.

We now show that, in our framework, both types of geoengineering technologies do, to some extent, backfire, and they may increase emissions and even reduce welfare.

**Proposition 4 (Implications of Geoengineering Technologies of Type I)** *Suppose that Assumptions 1 and 2 hold. Consider a geoengineering technology improvement of type I that increases $\xi$ by a small amount $d\xi$. Then we have*

- $d\hat{\tau}/d\xi = 0$ *(there is no effect on the equilibrium carbon tax).*

- $d\hat{q} = -\frac{1}{\Lambda} d\xi$ *(investment in clean technology declines).*
- $dE = d\xi > 0$ *(emissions increase, through lower $\hat{q}$).*
- $dD/d\xi = 0$ *(environmental damages remain constant).*
- $dW/d\xi < 0$ *if and only if* $\lambda(\pi_c - \pi_d) > \Lambda\tau$ *(welfare may decline).*

**Proof.** See Appendix A. ■

The key economic force driving the result in Proposition 4 is that even after the geoengineering advances, the technology IC (5) still pins down the carbon tax rate at $\hat{\tau}$. This is because with a small change in $\xi$, Assumption 2 will continue to hold and the equilibrium has to be interior. This in turn requires that the social planner still prefers to set the carbon tax at $\hat{\tau}$, which is only possible if the marginal environmental damage remains constant. Since geoengineering shifts the damage function rightward by $d\xi$, the total stock of carbon must increase by $d\xi$. This happens by fewer firms making the switch to clean technology. This is visually illustrated in Figure 3. Geoengineering shifts the curve representing marginal damages rightwards as shown by the red curve. If there was no change in investment in conventional clean technology, marginal and overall damages would both decline. But in equilibrium, marginal damages have to remain constant, and as shown by the arrows along the red curve, the adjustment involves a reduction in investment in conventional clean technology, which increases emissions and the total stock of carbon and restores marginal damages to the same level that prevailed before geoengineering.

The effects of this type of geoengineering advance on welfare are ambiguous because of two competing forces. On the one hand, since $S$ remains constant and $q$ declines, society saves the costs of switching to clean technology. If investment in clean technology were optimal (which happens when $\lambda = 0$), this would be its sole impact because reductions in investment in clean technology would only have second-order welfare costs. Thus in this case, despite the increase in emissions, welfare would go up. However, because $\lambda > 0$, investments in clean technology are distorted, and a further reduction in the fraction of firms making the switch to clean technology creates a first-order welfare loss. Put differently, the benefit from investment in clean technology is not only the reduction in emissions, but also the fact that $\pi_c = f_c(k_c(\hat{\tau})) - (1 + \gamma\hat{\tau})k_c(\hat{\tau})$ may be greater than $\pi_d = f_d(k_d(\hat{\tau})) - (1 + \hat{\tau})k_d(\hat{\tau})$. A reduction in $q$ implies that this gain is forgone, which can outweigh the cost savings from lower investments. The condition for welfare to diminish as a result of a geoengineering advance of type I in the last part of the proposition indeed requires that $\lambda$ and $\pi_c - \pi_d$ are sufficiently large to compensate for the fixed cost savings. In fact, a large value of $\lambda$, by creating a larger wedge between the planner's objective function and private incentives to switch to clean technology, is sufficient to ensure that welfare declines as a result of this type of geoengineering advance.

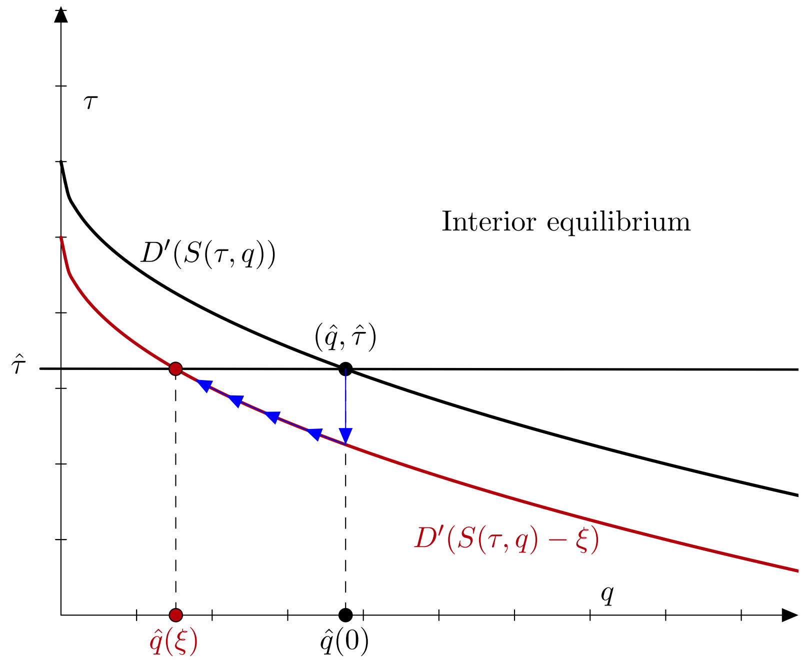


Figure 3: Equilibrium response of clean technology after type I geoengineering.

We also note that, if instead of a small increase in $\xi$, we consider a large increase, we may violate Assumption 2, and if so, the planner may wish to deviate from (5), forgoing any investment in clean technology. In this case, damages may again decline and welfare may increase. A similar caveat applies to the type II geoengineering technology discussed next.

**Remark 1 (Carbon Leakage)** Though our focus is on geoengineering technologies, Proposition 4 holds identically in a different setting. Suppose that our model applies to a specific country (say the United States), and another country (say China) reduces its emissions by an amount $d\xi > 0$. This reduction in the global carbon stock would reduce the Pigovian tax of the domestic government, violating (5). To restore this constraint, emissions by domestic firms increase, again through reduced investments in clean technology.

The implications of geoengineering technologies of type II are broadly similar but slightly more involved because they can also lead to greater overall damages.

**Proposition 5 (Implications of Geoengineering Technologies of Type II)** *Suppose that Assumptions 1 and 2 hold. Consider a geoengineering technology improvement of type II that increases $\upsilon$ by a small amount $d\upsilon > 0$, and let $\eta = \hat{S}D''(\hat{S})/D'(\hat{S})$ be the elasticity of the marginal damage function (where $\hat{S} = S(\hat{\tau}, \hat{q})$). Then we have*

- $d\hat{\tau}/d\upsilon = 0$ *(there is no effect on the equilibrium carbon tax).*

- $dS/dv > 0$ *(the total stock of carbon increases).*

- $d\hat{q}/dv < 0$ *(investment in clean technology declines).*

- $dE/dv > 0$ *(emissions increase, through lower $\hat{q}$).*

- $dD/dv > 0$ *if and only if $\eta \leq \eta^*$ (environmental damage increases if the damage function is not too convex), where $\eta^* \geq 1$.*

- $dW/dv < 0$ *if and only if*
$$\eta \leq \eta^{\mathrm{II}}(\lambda) \equiv a\lambda \left( \frac{\pi_c - \pi_d}{\Lambda \tau} \right)$$
*(welfare declines if the damage function is not too convex), where $a \equiv SD'(S)/D(S) > 1$.*

**Proof.** See Appendix A. ■

As in Proposition 4, the results of Proposition 5 are a consequence of the fact that to sustain an interior clean technology adoption rate, the carbon tax needs to remain at $\hat{\tau}$, and this necessitates an increase in emissions. In the case of a type I geoengineering improvement, emissions increased in such a way as to keep the total stock of carbon in the atmosphere and overall environmental damages constant. With a type II advance, emissions must again increase to keep the marginal damage constant, but this might involve a higher level of overall damages. In particular, if the elasticity of the marginal damage function, $\eta$, is high, marginal damages can change significantly without a large change in the level of the stock of carbon. In this case, the direct environmental benefit from geoengineering dominates the equilibrium decline in clean technology, and environmental damages fall. Conversely, if $\eta$ is low (in particular, less than some $\eta^*$), to restore marginal damages to their initial value and thereby sustain the Pigovian tax at $\hat{\tau}$, the stock of carbon needs to change by a large amount, which translates into an increase in overall environmental damages. This is the case illustrated in the bottom panel of Figure 4.[17]

The effects on welfare are once again ambiguous for similar reasons to those discussed above. But provided that the elasticity of the marginal damage function $\eta$ is sufficiently low (in this case less than $\eta^{\mathrm{II}}$) and $\lambda > 0$, the negative effect of distorting investment in clean technology dominates savings from the transition costs $q\Gamma$, and overall welfare declines (in fact, as in the previous case, a sufficiently large $\lambda$ ensures that welfare always declines). Conversely, a sufficiently elastic marginal damage function or a sufficiently low $\lambda$ will make welfare increase as a result of an improvement in geoengineering technologies of type II. But the reason why these two conditions rule out negative

[17]The figure is drawn for a quadratic $D$ function. We show in Appendix A that when $D$ is quadratic, the condition $\eta \leq \eta^*$ is always satisfied, and thus environmental damages always increase as a result of a geoengineering technology improvement of type II. In this case, the condition $\lambda \geq \lambda^* \equiv \frac{1}{2}\left(\frac{\Lambda\tau}{\pi_c - \pi_d}\right)$ is necessary and sufficient for welfare to decline (note that $\lambda^* < 1/2$).

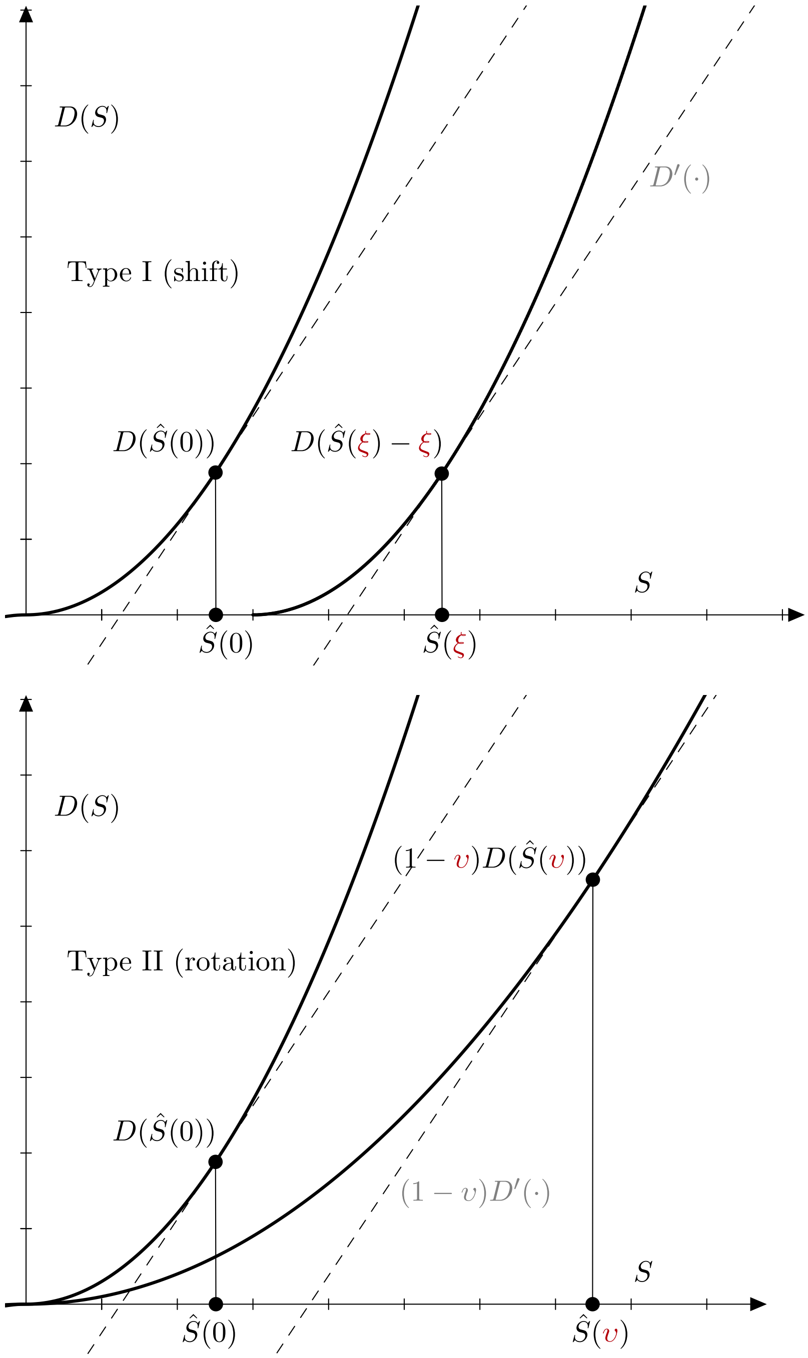


Figure 4: Environmental damage before and after type I ("carbon removal") and type II ("climate adaptation" or "solar radiation management") geoengineering. The top panel depicts the increase in the stock of carbon necessary to keep marginal damages constant with a type I technology, and the bottom panel does the same for a type II technology.

welfare effects are somewhat different: when $\eta$ is sufficiently high, overall environmental damages decline by a large amount; when $\lambda$ is very small, private investment in clean technology is nearly optimal, and a further reduction in these investments only has second-order welfare costs.

Together, Propositions 4 and 5 are the main results of our static model: the negative equilibrium response of clean technology entirely offsets the environmental benefits of geoengineering improvements. In all interior equilibria, geoengineering technologies that remove carbon directly from the atmosphere do not affect overall environmental damages (Proposition 4), while geoengineering technologies that flatten the damage function sometimes increase environmental damages (Proposition 5). When $\lambda$ is sufficiently large, both geoengineering improvements will reduce welfare.

### 2.11 Extensions

Many of the assumptions adopted so far are for simplicity and transparency. We now briefly discuss how several of them can be relaxed. Throughout, we assume that slight variations of Assumptions 1 and 2 (adapted to the extended environment) continue to hold, but do not state them formally to conserve space.

**Direct subsidies to clean technology.** We have so far assumed that the social planner has access to a single policy instrument, the carbon tax. This is not central to our results as long as the social planner cannot perfectly control investments in clean technology. Suppose, for illustration, that she can subsidize a proportion $\sigma \in [0, 1]$ of each firm's investments (fixed costs $\Gamma$) in clean technology. We capture the fact that the planner cannot perfectly control these clean technology investments by assuming that these subsidies are not pure transfers and there is a social cost of $\sigma\omega\Gamma$ in terms of the final good, where $\omega > 0$ parameterizes the extent of inefficiency from the clean technology subsidies. In this case the qualitative results in Propositions 3–5 continue to apply. In particular, we can summarize the results in this case with the following proposition.

**Proposition 6 (Direct Subsidies for Clean Technology)** *Suppose that in addition to the carbon tax, the social planner has access to a clean technology subsidy whereby a fraction $\sigma \in [0, 1]$ of clean investment costs is subsidized at a social cost $\sigma\omega\Gamma$ in terms of the final good. Then*

- *If $\omega > 0$, the planner still prefers to commit to a tax above the Pigovian benchmark when $\lambda > 0$.*
- *$dW/d\xi < 0 \iff \tilde{\lambda}(\pi_c - \pi_d) > \Lambda\tau$.*
- *$dW/dv < 0 \iff \eta \leq \eta^{\mathrm{II}}(\tilde{\lambda})$, where $\tilde{\lambda} \equiv \frac{\lambda-\sigma-(1-\lambda)\sigma\omega}{1-\sigma}$.*

**Proof.** See Appendix A. ■

This result is intuitive. As long as the planner cannot perfectly control clean technology decisions—in particular because clean technology subsidies are socially costly (the case where $\omega > 0$)—she would prefer a carbon tax above the Pigovian level to encourage additional clean technology investments. The commitment problem, however, prevents this, and the logic of our results above apply and yield the same insights. However, our results do not extend to the case in which $\omega = 0$, where the planner can perfectly control clean technology investments without any social distortions. In this case, she has no reason to deviate from Pigovian taxation, the technology externality emphasized above is no longer present, and geoengineering improves welfare because the equilibrium is constrained efficient.

**Political economy considerations.** Our model takes a rather charitable view of policy-makers, assuming that they fully internalize environmental externalities. In practice, many regulators and politicians appear to be far away from this ideal benchmark. For example, they may be captured by special interest groups or receive campaign contributions that influence their policy agendas. In several simple political economy settings, such behavior can be modeled by assuming that the policy-maker maximizes a weighted social welfare function, with greater weights on groups capable of lobbying or making campaign contributions (e.g., Grossman and Helpman, 1994). In our setting, this corresponds to the policy-maker having an objective function that assigns a lower weight to environmental damage. The next proposition shows that our qualitative results remain unchanged in this case, except that negative welfare consequences of geoengineering become more likely.

**Proposition 7 ($\beta$-Benevolence)** *Suppose that the policy-maker values only $\beta \in (0,1)$ of environmental damages and thus maximizes*

$$W(\beta) = (1-q)[f_d(k_d) - k_d] + q[f_c(k_c) - k_c - \Gamma] - \beta D(S) \tag{6′}$$

*rather than (6)—but true welfare is still given by (6). Then Proposition 4 and Proposition 5 hold, except that in the latter case welfare declines if*

$$\eta \leq a\left(\beta\lambda\frac{\pi_c - \pi_d}{\Lambda\tau} + 1 - \beta\right) < \eta^{\mathrm{II}}. \tag{8}$$

**Proof.** See Appendix A. ■

Intuitively, because the policy-maker undervalues environmental damages, the equilibrium is further away from the second-best, making it more likely that the decline in clean investment resulting from geoengineering reduces welfare.

**Heterogeneous costs of clean investment.** Our analysis is simplified by the fact that the (Technology IC) constraint holds with equality both before and after the arrival of geoengineering technologies. This feature relies on all firms having the same cost of switching to clean technology.

In Appendix B we show that our qualitative results generalize to the case in which these costs are heterogeneous. Suppose, in particular, that firm $i$'s cost of switching to clean technology is $\Gamma_i = \Gamma + \chi_i$, where $\chi_i$ is mean zero and distributed according to a probability density $h$. We then establish that—provided that the density $h$ is sufficiently concentrated around zero—all of our qualitative results, including the potential negative welfare consequences of geoengineering, continue to hold. Intuitively, the fact that $h$ is concentrated around zero ensures that the social planner does not find it beneficial to change the carbon tax by too much after the arrival of the geoengineering technology, in which case all of the insights emphasized above apply in the same fashion.

**Stochastic arrival of geoengineering technologies.** Our analysis also presumed for simplicity that a geoengineering advance would arrive with certainty and that this was fully recognized by all agents. In practice, as our discussion in the Introduction emphasized, there is considerable uncertainty about whether and when large-scale geoengineering will be feasible. We can incorporate this feature by assuming that both the agents in the economy and the social planner expect the geoengineering technology to arrive with some probability $\kappa \in (0, 1)$. We show in Appendix B that all of our results generalize immediately to this case, but we also gain an additional insight. We see that the worst outcomes in terms of welfare and environmental damages are realized when geoengineering is expected to succeed with high probability—substantially lowering investment in clean technology—but then fails. In this case, our model delivers a specific channel for the general concerns noted by the IPCC in the Introduction.

**Endogenous geoengineering.** It is also straightforward to see that all of our results apply with endogenous geoengineering, meaning that the social planner has access to a technology to generate possibly stochastic geoengineering advances and cannot commit to not deploying this technology. In the no-commitment equilibrium, the social planner chooses her optimal geoengineering investment after clean technology investments are made, and then the equilibrium is very similar to the one with stochastic geoengineering technologies given the resulting equilibrium probability of geoengineering success.

**Different formulations of environmental damages.** The additive formulation of environmental damages simplified our analysis, but it is also unnecessary. In Appendix B, we show that our qualitative results are unaffected if damages affect productivity as in Nordhaus (1991, 2008) and Golosov *et al.* (2014), or affect utility in a non-additively separable manner.

## 3 Dynamic Model

We now extend our static model to a dynamic economy where production decisions are made continuously, firms enter and exit, and technological quality and the stock of carbon accumulate over time. Our model is constructed to mimic both the structure of our static setup and the quality-ladder models of Aghion and Howitt (1992) and Grossman and Helpman (1991) as closely as possible. The quality-ladder structure enables us to endogenize the replacement probability $\lambda$ as the flow rate of creative destruction. After deriving the unique balanced growth path (BGP) and characterizing the structure of the dynamic equilibrium, we show that the effects of geoengineering technologies on the BGP are essentially identical to those derived in the static model.

### 3.1 Production, Entry and Environmental Damages

As in the static model, we consider an economy with a unique final energy good, produced by a continuum of perfectly substitutable activities indexed by $[0,1]$. We think of each activity $i \in [0,1]$ as being produced at a site $i \in [0,1]$ dedicated to that activity and each site can house only one firm (so that the number of active firms will be constant as in the static model). Time $t$ is infinite and discrete, of length $\Delta > 0$. In what follows, we simplify the exposition by taking $\Delta \to 0$ to work directly with differential equations. The production technology differs from the static model only in that the productivity of each activity depends on where it is located on a quality ladder, denoted by $n_{it} \in \mathbb{N}$ for activity $i$ at time $t$. This productivity applies both to dirty and clean technologies. If there has not been a switch to clean technology in activity $i$, then the firm with the best technology in this line at time $t$ will be active in site $i$ and produce

$$A^{n_{it}} f_d(k_{it}), \tag{9}$$

where $k_{it} \equiv K_{it}/A^{n_{it}}$ is "normalized investment," $K_{it}$ is investment (again in terms of the final good), $A = 1 + \alpha > 1$ so that each higher rung on the quality ladder secures a proportional improvement in productivity, and we continue to make the same assumptions on $f_d$ ($f_d' > 0$, $f_d'' < 0$, and the Inada conditions).[18] We also assume that the dirty production technology emits $K_{it}$ units of carbon given investment $K_{it}$.

If, on the other hand, activity $i$ has switched to clean technology, the firm with the best technology for this activity at time $t$ has access to the production technology

$$A^{n_{it}} f_c(k_{it}), \tag{10}$$

where again the same assumptions as in the static model apply to $f_c$, and as before, clean technology emits $\gamma K_{it}$ units of carbon when the level of investment is $K_{it}$, where $\gamma < 1$. Consequently, total emissions at time $t$ are

$$E_t = \int_0^1 \left[ \mathbf{1}_{\{i \text{ is dirty}\}} K_{it} + \gamma \mathbf{1}_{\{i \text{ is clean}\}} K_{it} \right] di. \tag{11}$$

[18] As in the static model, the Inada conditions imply that, despite productivity differences across activities, all activities will produce positive output. Moreover, since each site can house a single firm, only the firm with the best technology in that activity will produce.

An important feature of our formulation is that even though productivity varies across activities, the level of normalized investment will only differ between dirty and clean activities, and we thus denote it by $k_{dt}$ and $k_{ct}$ respectively for dirty and clean technologies at time $t$. Consequently, total emissions can also be expressed as

$$E_t = (1 - q_t)k_{dt}\mathbb{E}\left[A^{n_{it}} \middle|\ i \text{ is dirty}\right] + q_t\gamma k_{ct}\mathbb{E}\left[A^{n_{it}} \middle|\ i \text{ is clean}\right],$$

where $q_t$ denotes the aggregate fraction of clean firms at time $t$.

The dynamics of the stock of carbon in the atmosphere, which we write directly in differential form since we focus on $\Delta \to 0$, are given as

$$\dot{S}_t = \frac{E_t}{A_t} - \delta S_t, \tag{12}$$

where $S_0 \geq 0$, $\delta > 0$ is the environmental regeneration rate, and environmental damages are

$$A_t D(S_t; \xi, \upsilon),$$

where

$$A_t \equiv \int_0^1 A^{n_{it}} di \tag{13}$$

is the average productivity of the economy at time $t$,

$$D(S; \xi, \upsilon) \equiv (1 - \upsilon)\tilde{D}(S - \xi) \tag{14}$$

as in the static model, and $\tilde{D}(\cdot)$ is increasing, strictly concave, and twice continuously differentiable in the stock of carbon $S$. We set the geoengineering parameters as $(\xi, \upsilon) = 0$ and omit them from our notation until the final subsection of this section. Note that damages are multiplied by average productivity, while emissions are divided by average productivity. This formulation captures the fact that when the productivity or consumption level of the economy is higher, a given stock of carbon in the atmosphere will have more negative productivity or disutility implications, while ensuring that damages grow at the same rate as the economy.

Finally, we assume that the economy is inhabited by a representative household, who discounts the future at the exponential rate $\rho > 0$. In the text we simplify the analysis (and keep it as close as possible to the static model) by assuming that this household obtains linear flow utility (more general utility functions are discussed in Appendix B). Thus the objective function of the household at time $t$ is

$$\sum_{s=t}^{\infty} \left[C_{t+\Delta(s-t)} - A_{t+\Delta(s-t)}D(S_{t+\Delta(s-t)}; \xi, \upsilon)\right] e^{-\rho\Delta(s-t)},$$

where $C_s$ is consumption at time $s$. Once again, taking the limit $\Delta \to 0$, we work with the continuous-time equivalent,

$$\int_t^{\infty} \left[C_s - A_s D(S_s; \xi, \upsilon)\right] e^{-\rho(s-t)} ds. \tag{15}$$

The switch from dirty to clean technology has a fixed cost of $A^{n_{it}}\Gamma > 0$ in terms of the final good for activity $i$ with productivity $A^{n_{it}}$, and is incurred only once for each activity (because once an activity switches to clean technology, all future productivity improvements build on the existing clean technology in that activity or site). This formulation, which makes the cost of switching to clean technology proportional to productivity, ensures that the incentives to switch to clean technology remain independent of an activity's productivity.

Productivity improvements take place in a manner analogous to the standard quality-ladder models. Specifically, potential entrants invest in research and development (R&D) in order to improve over existing products. R&D uses a scarce input, say scientists, which has an inelastic supply of $Z > 0$.[19] We also assume that R&D is undirected, meaning that entrants decide their R&D effort, but cannot choose which activity they are researching and are randomly matched to one of the activities in $[0, 1]$. A successful innovation for activity $i$ currently with productivity $A^{n_{it}}$ enables the entrant to replace the incumbent producer of this activity with a new technology with productivity $A^{n_{it}+1}$. Let us denote R&D effort (scientists hired) at time $t$ by $z_t$. Then the (Poisson) arrival rate of a successful innovation is

$$\lambda_t = \varphi z_t, \tag{16}$$

where $\varphi > 0$. The cost of R&D effort of $z_t$ is $z_t w_t$, where $w_t$ denotes the equilibrium wage for scientists. This wage is determined from the market-clearing condition for scientists given by

$$z_t = Z \quad \text{for all } t \geq 0. \tag{17}$$

This naturally ensures that in equilibrium

$$\lambda_t = \lambda \equiv \varphi Z.$$

Taking into account the expenditures on switching to clean technology, the resource constraint of the economy implies that consumption at time $t$ is given as

$$C_t = \int_0^1 A^{n_{it}}[f_i(k_{it}) - k_{it} - \mathbf{1}(t = \inf\{t \geq 0 : q_{it} = 1\})\Gamma]di,$$

which integrates over the output levels of different activities and then subtracts the costs of investment in clean technology (where $\mathbf{1}(t = \inf\{t \geq 0 : q_{it} = 1\})$ is the indicator function for the time at which activity $i$ switches to clean technology and incurs the fixed cost $A^{n_{it}}\Gamma$).

[19] This formulation with an inelastic supply of scientists ensures that the overall growth rate of the economy will be insensitive to the rate of carbon taxation. We view this as a desirable benchmark property, since otherwise the planner would have an incentive to manipulate carbon taxes in order to affect the long-run growth rate.

## 3.2 Carbon Tax and Production Decisions

As in the static model, there is a carbon tax of $\tau_t$ at time $t$. Profits of dirty and clean firms can be written, respectively, as

$$\Pi_{idt} = \max_{k \geq 0} A^{n_{it}} \left[ f_d(k) - (1+\tau_t)k \right] = A^{n_{it}} \left[ f_d(k_d(\tau_t)) - (1+\tau_t) k_d(\tau_t) \right] \tag{18}$$

and

$$\Pi_{ict} = \max_{k \geq 0} A^{n_{it}} \left[ f_c(k) - (1+\gamma\tau_t)k \right] = A^{n_{it}} \left[ f_c(k_c(\tau_t)) - (1+\gamma\tau_t) k_c(\tau_t) \right], \tag{19}$$

where $k_c(\tau_t)$ and $k_d(\tau_t)$ are then defined as the optimal input decisions for dirty and clean firms respectively. We use $\pi_j(\tau_t) \equiv \Pi_{ijt}/A^{n_{it}}$ to denote normalized profits of activity $j \in \{c, d\}$ at time $t$.

We next write the value functions of firms with clean and dirty technologies as a function of their productivity. At time $t$, a clean incumbent with productivity $A^n$ has (expected) net present discounted value given by the usual dynamic programming recursion (provided that this value is a differentiable function of time):

$$r_t V_{ct}(n) = A^n \pi_c(\tau_t) + \dot{V}_{ct}(n) - \lambda V_{ct}(n).$$

Intuitively, the firm receives a "dividend" of $A^n \pi_c(\tau_t)$ on its asset of $V_{ct}(n)$, but also recognizes that this asset may change value, captured by the term $\dot{V}_{ct}(n)$, and may entirely disappear because of creative destruction coming from improvements by entrants, which takes place at the Poisson rate $\lambda_t$ and will make the incumbent lose the asset entirely. This stream of profits is then discounted at the interest rate $r_t$. Because the household's preferences are linear, the interest rate is always equal to the discount rate, i.e.,

$$r_t = \rho,$$

and thus this expected net present discounted value can be expressed as

$$V_{ct}(n) = A^n \int_t^\infty \pi_c(\tau_s) e^{-(\rho+\lambda)(s-t)} ds, \tag{20}$$

which is just the discounted integral of flow profits $\pi_c(\tau_s)$ over time, adjusted for the baseline productivity of the firm and the Poisson rate $\lambda$ of arrival of creative destruction.

The expected net present discounted value of dirty firms is similar, except that they can choose whether to switch to clean technology at a cost $A^{n_{it}}\Gamma > 0$,

$$V_{dt}(n) = \max \left\{ V_{ct}(n) - A^n \Gamma, \frac{A^n \pi_d(\tau_t) + \dot{V}_{dt}(n)}{\rho + \lambda} \right\}. \tag{21}$$

The max operator takes care of the choice to switch to clean technology, while the second part is the dynamic programming recursion rearranged (with $r_t = \rho$ imposed).

Equations (20) and (21) show that $V_{jt}(n)/A^n$ is independent of $n$ for $j \in \{c, d\}$, and we thus define $v_{jt} \equiv V_{jt}(n)/A^n$ as the normalized value function.

### 3.3 Clean Technology and R&D Decisions

Equation (21) immediately gives us the equivalent of (Technology IC) in the static model. Firms are happy to switch to clean technology only when the maximization operator in this expression picks the first term, or, put in terms of normalized value functions, when

$$v_{dt} = v_{ct} - \Gamma. \tag{22}$$

This binding constraint will play an analogous role to (5) in the static model, and implies the following form for incentive-compatible technology choice $q_t$:

$$\begin{cases} v_{dt} = v_{ct} - \Gamma & \Longrightarrow \; q_t \in [0,1] \\ v_{dt} > v_{ct} - \Gamma & \Longrightarrow \; q_t = 0 \end{cases} \tag{Dynamic Technology IC}$$

which closely resembles its analogue in the static model.[20]

Next, using the characterization of the value functions in the previous subsection, we derive equilibrium R&D decisions. Since potential entrants have access to the R&D technology given by (16), equilibrium requires the following free-entry condition to hold with complementary slackness

$$\varphi \int_0^1 \left[ q_t V_{ct}(n_{it}+1) + (1-q_t) V_{dt}(n_{it}+1) \right] di - w_t = 0,$$

where $V_{jt}(n)$ for $j \in \{c, d\}$ are the expected value functions defined in (20) and (21), $w_t$ is the equilibrium wage for scientists, and the integral reflects the fact that R&D is undirected and may lead to an improvement over a clean or dirty technology. Using the definition of normalized value functions, the free-entry condition can be simplified to the following form

$$q_t v_{ct} + (1-q_t) v_{dt} = \frac{w_t}{\varphi A_t}. \tag{23}$$

At each $t$, the wage for scientists, $w_t$, adjusts to satisfy (23) (so $z_t = Z$).

### 3.4 Planner's Problem

As in the static model, the (social) planner is benevolent, and therefore maximizes the same objective as the representative household, (15). She will seek to achieve this objective by choosing a sequence of carbon taxes, $(\tau_t)_{t \geq 0}$. We also continue to assume that the planner does not have access to a commitment technology, so the sequence of carbon taxes can be revised at any $t$. As in the static model, the planner's preferred allocation differs from that of the firms in two ways. First, firms do not internalize the environmental damage they create (except through the carbon taxes

[20] Unlike the static condition (Technology IC), however, there is no case in which $v_{dt} < v_{ct} - \Gamma$, since $v_{dt} = \max\{v_{ct} - \Gamma, (\rho+\lambda)^{-1}(\pi_{dt} + \dot{v}_{dt})\}$ implies that $v_{dt} \geq v_{ct} - \Gamma$ for all $t \geq 0$. Naturally, the equilibrium involves $q_t = 1$ when the max operator always strictly picks the first term in (21). We provide conditions for this not to be the case in equilibrium in Assumption $2'$ below.

that the planner imposes). Second, they fail to internalize the positive externality that they create for future producers of the same activity when they switch to clean technology. This externality is again proportional to the likelihood of replacement, $\lambda$.

### 3.5 Definition of Equilibrium

We focus on Markovian equilibria where no agent can condition its strategy at $t$ on the history of play except through the state variables $(S_t, q_t, \{n_{it}\}_{i\in[0,1]})$. This focus on Markovian equilibria is motivated by our main interest, which is to understand the implications of lack of commitment to future carbon taxes. In an infinite-horizon setup, non-Markovian equilibria may sometimes mimic commitment policies.[21]

A *dynamic (Markov) equilibrium*, or an equilibrium for short, is given by a path of technology choices, taxes, input decisions, wages for scientists, and stock of carbon $\{(q_t^*)_{t\geq 0}, (\tau_t^*)_{t\geq 0}, (k_{dt}^*)_{t\geq 0}, (k_{ct}^*)_{t\geq 0}, (w_t^*)_{t\geq 0}, (S_t^*)_{t\geq 0}\}$,

- Given $(q_t^*)_{t\geq 0}$, carbon taxes $(\tau_t^*)_{t\geq 0}$ maximize household utility (15) at each $t \geq 0$,
- Given $(\tau_t^*)_{t\geq 0}$, clean technology decisions $(q_t^*)_{t\geq 0}$ satisfy (Dynamic Technology IC),
- Given $\tau_t^*$, input choices $k_{dt}^*$ and $k_{ct}^*$ maximize, respectively, $\pi_{dt}$ and $\pi_{ct}$ in (18) and (19), for all $t \geq 0$,
- Given $(\tau_t^*)_{t\geq 0}$ and $(q_t^*)_{t\geq 0}$, the equilibrium R&D intensity $z_t$ and wages $w_t$ satisfy labor market clearing (17) and free entry (23) for each $t \geq 0$.

The equilibrium has a block recursive structure whereby the remaining variables can be determined from $(\tau_t^*)_{t\geq 0}$ and $(q_t^*)_{t\geq 0}$. In view of this, we use the shorthand of referring to an equilibrium as $(\tau_t^*, q_t^*)_{t\geq 0}$.

We also define a *Balanced Growth Path Equilibrium (BGP)* as an equilibrium in which $(\tau_t^*, q_t^*) = (\hat{\tau}, \hat{q})$ for all $t$, so that aggregate output $A_t$ grows at a constant rate given by

$$g \equiv \alpha\lambda = \alpha\varphi Z,$$

where the presence of the term $\alpha = A - 1$ follows from the properties of the Poisson process.[22] We will also see that in a BGP, $S_t = \hat{S}$ for all $t$. When this causes no confusion, we will also include $\hat{S}$ in the definition of a BGP (or $S_t^*$ in the definition of an equilibrium).

[21] In our setup, this would take the form of the social planner expecting worse actions from the firms following a lower-than-promised carbon tax. Though such schemes are not always feasible, they nevertheless complicate the analysis.

[22] Each $n_{it}$ is a sample path of a Poisson process with intensity $\lambda t$, so $\int_0^1 A^{n_{it}} di$ corresponds to the expectation of $A^{N_t} = \exp(N_t \log A)$, where $N_t \sim \text{Pois}(\lambda t)$ so that $\mathbb{E}[e^{\phi N_t}] = \exp(\lambda t(e^{\phi} - 1))$ for any $\phi \in \mathbb{R}$.

### 3.6 Farsighted Pigovian Taxes

To characterize the equilibrium tax sequence, we start by determining the evolution of marginal environmental damages or, equivalently, the *shadow price of carbon emissions*, which will give us the dynamic equivalent of Pigovian taxation (or what we will call "farsighted Pigovian taxes"). Consider the Hamiltonian corresponding to the planner's maximization problem, in (15), subject to the evolution of the stock of carbon given in (12),

$$H_t(K_t, S_t) = C_t - A_t D(S_t) - \mu_t \left[E_t/A_t - \delta S_t\right], \tag{24}$$

where $\mu_t$ is the costate variable associated with the stock of carbon in the atmosphere.[23] Since emissions are divided by average productivity, $A_t$, the shadow value of carbon emissions is given by

$$p_t = \mu_t/A_t. \tag{25}$$

Since the Hamiltonian is concave, the necessary and (with the usual transversality condition) sufficient first-order condition for optimality is

$$\partial H_t/\partial S = \dot{\mu}_t - \rho\mu_t,$$

which yields a simple form for the shadow price of carbon emissions provided that the planner's maximization problem in (24) is well-behaved (in particular has a finite value). The next assumption ensures this:

**Assumption 3 (Growth)**

$$g \equiv \alpha\varphi Z < \rho + \delta.$$

Under this assumption, we have:

**Lemma 3 (Shadow Cost of Carbon)** *Suppose Assumption 3 holds. Then, along any optimal path,*

$$\dot{p}_t = -D'(S_t) + (\delta + \rho - \alpha\lambda)\, p_t \tag{26}$$

*and thus*

$$p_t = \int_t^\infty D'(S_s) e^{-(\delta+\rho-\alpha\lambda)(s-t)} ds, \tag{27}$$

*for all $t \geq 0$.*

[23] The full maximization problem would also need to impose constraints for the evolution of the states of clean technology, $q_t$, and average productivity in the economy, $A_t$, but as these constraints do not change the expression for the shadow price of carbon, we omit them from our exposition in the text.

**Proof.** See Appendix B. ■

We refer to the tax trajectory implied by (27) as "farsighted Pigovian." This terminology emphasizes that this tax sequence is a direct generalization of our static Pigovian tax. The generalization accounts for the fact that emissions create damages not only today but at all future dates, which means that the shadow price of carbon emissions must incorporate the discounted cost of these future damages.

Our next result shows that equilibrium taxes—due to the lack of commitment of the planner—must equal the farsighted Pigovian taxes characterized in (27), at least once clean technology converges.[24]

**Proposition 8 (Pigovian Best-response)** *There exists $T < \infty$ such that equilibrium taxes are given by*

$$\tau_t = p_t \tag{28}$$

*for all $t \geq T$.*

**Proof.** See Appendix B. ■

Proposition 8 shows that, despite the complicated dependence of clean technology and R&D decisions on the entire tax trajectory, equilibrium carbon taxes take a simple form. In fact, in (27), these taxes only depend on the evolution of the stock of carbon in the atmosphere $(S_t)_{t\geq 0}$. The key to understanding this result is that absent technology choices, the (farsighted) Pigovian taxes are optimal (with or without commitment), and the lack of commitment, combined with the Markovian restriction, precludes the planner from choosing a tax sequence that is ex post distortionary (different from Pigou), once the transition to cleaner technology is complete (either with $q_t = 1$ or $q_t = \hat{q} < 1$). This transition is completed within some finite time $T$, enabling us to use backward induction to prove the proposition.[25]

**Remark 2 (Counterexample to pure Pigovian taxes)** Proposition 8 establishes that $\tau_t = p_t$ for all $t \geq T$. In addition, we can prove that $\tau_t \leq p_t$ for all $t \geq 0$. But there might be some circumstances in which the social planner prefers to set a tax rate strictly less than the Pigovian one in the interval $[0, T]$ in order to increases future Pigovian taxes and encourage a faster switch to clean technology. We analyze the conditions under which this possibility could arise in Appendix B, but also prove that such a counterexample is possible only if $\lambda$ is very high (in fact, so high that all geoengineering technologies are strictly welfare reducing.)

[24] The main technical detail, showing that equilibrium clean technology indeed always converges in finite time, is stated and proven as Lemma B1 in the Appendix.

[25] This result is reminiscent of the generic time-inconsistency result of Calvo (1978). It is also simplified since $\Delta \to 0$, which removes the possibility of choosing a distortionary tax today in order to affect behavior until the taxes are adjusted tomorrow.

We will see later that, if she could commit, the planner would prefer to deviate from this Pigovian tax scheme.

### 3.7 Characterization of Equilibrium

To characterize the dynamic equilibrium, we impose dynamic analogues of Assumptions 1 and 2, which will again rule out Jevons' paradox and guarantee an "interior" equilibrium.

**Assumption 1′ (Dynamic No Jevons)** *For all $t \geq 0$ and all $\tau \geq 0$, we have*

$$\Lambda(\tau_t) \equiv k_d(\tau_t) - \gamma k_c(\tau_t) > 0.$$

This assumption enables us to develop another parallel with the static model. Analogously with (4), let us define

$$\begin{aligned} \Psi(\tau_t) &\equiv \pi_c(\tau_t) - \pi_d(\tau_t) \\ &= f_c(k_c(\tau_t)) - k_c(\tau_t) - (f_d(k_d(\tau_t)) - k_d(\tau_t)) + \Lambda(\tau_t)\tau_t \end{aligned} \tag{29}$$

as the difference in (normalized) profits between clean and dirty technologies at carbon tax $\tau_t$. Recall that in the static model, Lemma 2 ensured that $\Psi'(\tau) = \Lambda > 0$. Here, we similarly have $\Psi'(\tau_t) = \Lambda(\tau_t) > 0$ by Assumption 1′. Moreover, in an interior BGP where $(\tau_t, q_t) = (\hat{\tau}, \hat{q})$ with $\hat{q} \in (0, 1)$, we obtain a simplified form of (Dynamic Technology IC),

$$v_{dt} - v_{ct} = \frac{\Psi(\hat{\tau})}{\rho + \lambda} = \Gamma. \tag{30}$$

Here the first equality exploits the fact that when $\hat{q} < 1$, $v_{dt}$ is equal to the discounted stream of profits from dirty technology, and that profits, taxes and the creative destruction rate, $\lambda$, are constant, and the second equality follows from (22).

**Assumption 2′ (Conditions for Dynamic Interior Equilibrium)** *Let the initial carbon stock be $S_0$. Then for all $t \geq 0$,*

$$\Gamma \in \left( \int_t^\infty \Psi(\underline{\tau}_s) e^{-(\rho+\lambda)(s-t)} ds, \ \int_t^\infty \Psi(\overline{\tau}_s) e^{-(\rho+\lambda)(s-t)} ds \right)$$

*where*

$$\overline{\tau}_t = \int_t^\infty D' \left( S_0 e^{-\delta s} + \int_0^s k_d(\overline{\tau}_\nu) e^{-\delta(s-\nu)} d\nu \right) e^{-(\delta+\rho-\alpha\lambda)(s-t)} ds$$

*and*

$$\underline{\tau}_t = \int_t^\infty D' \left( S_0 e^{-\delta s} + \int_0^t k_d(\underline{\tau}_\nu) e^{-\delta(t-\nu)} d\nu + \gamma \int_t^s k_c(\underline{\tau}_\nu) e^{-\delta(s-\nu)} d\nu \right) e^{-(\delta+\rho-\alpha\lambda)(s-t)} ds.$$

Although notationally cumbersome, this assumption has an identical interpretation as its static counterpart, Assumption 2. Specifically, it ensures that the cost of switching to clean technology is neither too high nor too low—and the relevant thresholds depend on the farsighted Pigovian taxes and R&D intensities that will prevail when no firm ever switches to clean technology, $(\overline{\tau}_t)_{t\geq 0}$, or all firms switch to clean technology, $(\underline{\tau}_s)_{s\geq t}$. As in its static analogue, Assumption 2, the conditions in Assumption $2'$ depend on the initial stock of carbon, because this determines the entire path of Pigovian taxes.

We start by characterizing the BGP in which $(\tau_t, q_t) = (\hat{\tau}, \hat{q})$ for all $t$, which also ensures that the stock of carbon in the atmosphere converges to some finite $\hat{S}$. From (12), this limiting value of the stock of carbon must satisfy

$$\hat{q}\gamma k_c(\hat{\tau}) + (1-\hat{q})k_d(\hat{\tau}) = \delta\hat{S}. \tag{31}$$

Using (27) and (28), the stationary Pigovian tax $\hat{\tau}$ is given by

$$\hat{\tau} = \frac{D'(\hat{S})}{\delta+\rho-\alpha\lambda}. \tag{32}$$

These two equations together with (30) determine $(\hat{S}, \hat{\tau}, \hat{q})$. The next proposition establishes that such a BGP exists and is unique.

**Proposition 9 (Existence, Uniqueness of the Balanced Growth Path)** *Suppose Assumptions $1'$, $2'$, and 3 hold. Then there exists a unique BGP where $(S_t, \tau_t, q_t) = (\hat{S}, \hat{\tau}, \hat{q})$, and $(\hat{S}, \hat{\tau}, \hat{q})$ is the unique solution to equations (30), (31), and (32).*

**Proof.** See Appendix B. ■

The existence of a BGP $(\hat{S}, \hat{\tau}, \hat{q})$ follows from the equations and arguments proceeding the proposition. The uniqueness of this BGP is a consequence of the fact that the BGP farsighted Pigovian tax $\hat{\tau}$ is a decreasing function of $\hat{q}$. Once the incentive-compatible carbon tax, $\hat{\tau}$, is pinned down by equation (30), there exists a unique $\hat{q}$ that solves (32). These two variables then yield a unique value of $\hat{S}$.

A noteworthy feature of the unique BGP is that, as in our static model, $\hat{q} \in (0,1)$ and the equilibrium is "interior." This, in particular, ensures that in the BGP, (30) holds, which restricts the value of the BGP carbon tax to $\hat{\tau}$. The next proposition shows that every equilibrium converges to the BGP equilibrium in Proposition 9, and does so by some $T < \infty$.

**Proposition 10 (Interior Dynamic Equilibrium)** *Suppose Assumptions $1'$, $2'$, and 3 hold. Then the unique dynamic equilibrium takes the following form. There exists a $T < \infty$ such that:*

1. *for all $t \in [0, T)$, $\tau_t$ and $S_t$ grow continuously and $q_t = 0$.*

*2. for all $t \geq T$, $(S_t, q_t, \tau_t) = (\hat{S}, \hat{q}, \hat{\tau})$, where $(\hat{S}, \hat{q}, \hat{\tau})$ is given in Proposition 9.*

**Proof.** See Appendix B. ■

Figure 5 illustrates the shape of the dynamic equilibrium. The stock of carbon is always nondecreasing, and smoothly increasing until the economy reaches the BGP. Therefore marginal environmental damages and Pigovian taxes also increase until they reach their constant BGP level $\hat{\tau}$. As the Pigovian tax grows, clean technology incentives also increase—eventually (by monotonicity of $\Psi(\tau_t)$) reaching the value for which (30) holds, at which point clean technology leaps from zero to $\hat{q}$.

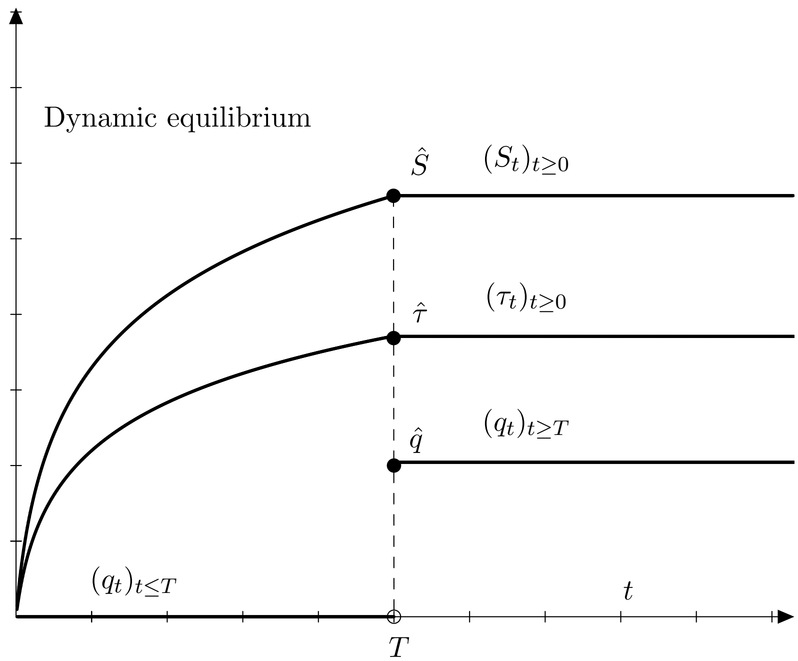


Figure 5: Time paths of the carbon stock $(S_t)_{t\geq 0}$, optimal taxes $(\tau_t)_{t\geq 0}$, and clean technology $(q_t)_{t\geq 0}$ in the dynamic equilibrium.

The proof of Proposition 10 is provided in Appendix B. Here we give some intuition. Proposition 9 established that the BGP has to be "interior"—if all activities eventually switched to clean technology, the subsequent carbon taxes would be too low to make such a switch optimal, whereas if no activity switches to clean technology, the stock of carbon and thus future carbon taxes would be sufficiently high to incentivize investment in clean technology. Proposition 10 then shows how we get to this BGP. Initially, with a lower stock of carbon in the atmosphere than the BGP value, the marginal damage of carbon emissions is low, so Pigovian taxes are also low, and consequently the transition path involves faster growth of emissions than in the BGP. When the stock of carbon reaches $\hat{S}$, the fraction of firms that have already transitioned to clean technology must be exactly the BGP value, $\hat{q}$, to sustain the (stationary) Pigovian tax sequence that maintains the dynamic technology IC, (30), so that we have $\tau_t = \hat{\tau}$ for all $t \geq T$.

### 3.8 Second-best

We noted above that, as in the static model, if she could commit, the planner would set a carbon tax sequence different than the Pigovian one. In this subsection, we prove this claim. As in Proposition 3 in our static analysis, the next result shows that whenever $\lambda > 0$, the second-best deviates from Pigovian taxation. The main differences are that the condition that $\lambda > 0$ is now automatically satisfied in any BGP with productivity growth (provided that $Z > 0$). Second-best carbon taxes, $\tau_t^{SB}$, exceed Pigovian ones (are greater than the shadow price of carbon emissions, $p_t$), and induce more firms to switch to clean technology. In contrast, if $\lambda = 0$ so that there is no growth in productivity in this economy, second-best and Pigovian taxes coincide.

**Proposition 11 (Dynamic Second-best)** *1. Suppose that $Z > 0$ (which ensures that $\lambda > 0$). Then the planner commits to a carbon tax $\tau_t^{SB} \geq p_t^{SB}$ for all $t \geq 0$, with $\tau_t^{SB} > p_t^{SB}$ for some $t \geq 0$, and the equilibrium fraction of firms that switch to clean technology converges to $q^{SB} > \hat{q}$.*

*2. Suppose that $Z = 0$ (so that $\lambda = 0$). Then for all $t \geq 0$, $\tau_t^{SB} = p_t$ and the equilibrium fraction of firms that switch to clean technology converges to $q_t^{SB} = \hat{q}$.*

**Proof.** See Appendix B. ■

### 3.9 Geoengineering

We next consider the implications of geoengineering breakthroughs on dynamic carbon taxation, environmental damages and welfare. We focus on the BGP derived in Proposition 9, and show that the results are essentially identical to the effects of geoengineering in the static model, derived in Section 2.10. We again distinguish between the two types of geoengineering advances, captured by the parameters $\xi$ and $\upsilon$ in the general damage function $(1-\upsilon)D(S_t - \xi)$.

**Proposition 12 (Dynamic Implications of Type I Geoengineering Technologies)** *Suppose that Assumptions 1′, 2′, and 3 hold, and the economy's unique BGP is given by $(\hat{S}, \hat{q}, \hat{\tau})$. Consider a geoengineering technology improvement of type I that increases $\xi$ by a small amount $d\xi > 0$. Then:*

- $d\hat{\tau}/d\xi = 0$ *(taxes do not change).*
- $d\hat{S}/d\xi = 1$ *(the stock of carbon increases).*
- $d\hat{q}/d\xi = -\delta/\Lambda < 0$ *(clean technology falls).*

- $dW/d\xi < 0 \iff \frac{\lambda}{\rho+\lambda}(\hat{\pi}_c - \hat{\pi}_d) > \Lambda\hat{\tau}$ *(welfare may decline).*

**Proof.** See Appendix B. ∎

This proposition shows that any geoengineering advance of type I results in conclusions similar to Proposition 4—in the BGP, the stock of carbon in the atmosphere increases and welfare (in the BGP) may even decline if there is a sufficiently strong response of investment in clean technology.

The next proposition gives the dynamic analogues of Proposition 5. Geoengineering reduces the BGP carbon tax, and welfare may decline.

**Proposition 13 (Dynamic Implications of Type II Geoengineering Technologies)** *Suppose that Assumptions 1′, 2′, and 3 hold, and the economy's unique BGP is given by* $(\hat{S}, \hat{q}, \hat{\tau})$. *Consider a geoengineering technology improvement of type II that increases* $\upsilon$ *by a small amount* $d\upsilon > 0$, *and let* $\eta = \hat{S}D''(\hat{S})/D'(\hat{S})$ *be the elasticity of the marginal damage function. Then*

- $d\hat{\tau}/d\upsilon = 0$ *(taxes do not change).*
- $d\hat{S}/d\upsilon = \frac{D'(\hat{S})}{(1-\upsilon)D''(\hat{S})} > 0$ *(the stock of carbon increases).*
- $d\hat{q}/d\upsilon < -\frac{\delta\hat{S}}{(1-\upsilon)\Lambda}\frac{1}{\eta} < 0$ *(clean technology declines).*
- $dW/d\upsilon < 0 \iff \eta < \eta^{\mathrm{II}}(\lambda)$, *where*

$$\eta^{\mathrm{II}}(\lambda) \equiv a\left(\delta(\rho - g + \delta)\frac{\lambda}{\rho+\lambda}\frac{\hat{\pi}_c - \hat{\pi}_d}{\Lambda\hat{\tau}} + \frac{\delta}{\rho - g + \delta}\right)$$

  *and* $a \equiv \hat{S}D'(\hat{S})/D(\hat{S}) > 1$ *(welfare may decline).*

**Proof.** See Appendix B. ∎

We note in addition that the conditions for welfare to decline as a result of a geoengineering advance of type II are again very similar to those we have obtained in the static model in Proposition 5. In particular, as in the static model, if $\lambda$ is sufficiently large, welfare declines following both types of geoengineering.

# 4 Conclusions

Many scientists and policymakers are pinning their hopes on major geoengineering advances to stem damages from the rapidly-rising concentration of carbon in the atmosphere. Others, on the other hand, have worried that the prospect of geoengineering advances may jeopardize more conventional solutions to our environmental maladies, most notably the necessary increases in carbon taxes. Many of these concerns center around the possibility that the promise of geoengineering solutions may not materialize, or that geoengineering may have harmful side effects. In this paper, we have proposed an alternative perspective on the possible dark side of geoengineering. We have argued, theoretically, that geoengineering may damage the environment and welfare precisely because it is expected to and will materialize (or at least do so with a high probability). At the center of our argument is the possibility that the expectation of geoengineering makes future carbon taxes non-credible (because once geoengineering advances have been made, the damage from carbon emissions is reduced), which will discourage current investments in conventional cleaner technology (where our emphasis on "conventional" is to distinguish it from geoengineering technologies).

To advance this argument, we have developed a model of an endogenous transition to clean technology with policy-making without commitment. Both of these elements are relatively new in the environmental literature and important for our argument. Though transition to various types of clean technology (including wind, solar, and geothermal) is generally seen as a bedrock of any reduction in the pace of buildup of carbon in the atmosphere, there are relatively few analyses of this process in the economics literature (see the references in the Introduction). The modeling of the transition to conventional clean technology is critical for understanding the potential adverse effects of geoengineering, because it is these types of investments that may be discouraged if future carbon taxes are expected to be low. Lack of commitment to future policies in general and carbon taxes in particular is also an evident reality, but most economic analyses of environmental policy have stayed away from the time-inconsistency issues that arise in the absence of such commitment. It plays a pivotal role in our setting because it is this lack of commitment that makes it impossible for future carbon taxes to remain high when geoengineering advances materialize.

We start with a static model in which existing energy producers can undertake costly investments to switch to clean technology and once these technology investments are made, a benevolent planner sets the carbon tax. Lack of commitment to policies means that the planner cannot deviate from the Pigovian carbon tax once technology investments are sunk. But because such investments create a positive externality—for other firms that can build on them—the planner would have preferred to commit to a carbon tax greater than the Pigovian level, had this been possible. Furthermore, we restrict attention to parameters such that the equilibrium is "interior" where some firms switch to clean technology, while others do not. In the static model, an interior equilibrium arises only if the carbon tax that they anticipate takes a specific value.

We then introduce geoengineering breakthroughs into this framework. For simplicity, we distinguish between two different types of geoengineering advances. Type I, which corresponds to various technologies aiming at carbon removal from the atmosphere, shifts the damage function from the stock of carbon in the atmosphere rightwards in a parallel fashion—and is thus equivalent to a decline in the effective stock of carbon. If no economic decisions changed following this type of geoengineering breakthrough, the marginal and overall damages would decline, leading to lower carbon taxes in the future and higher welfare. But anticipating a lower rate of carbon taxation, all firms would then abandon their investments in clean technology. This would increase emissions and the stock of carbon. Provided that the geoengineering breakthrough is not so large as to destroy the interior equilibrium, we must then have sufficiently higher emissions so that marginal damages are restored to their pre-geoengineering level and energy producers are once again incentivized to invest in conventional clean technology. In this case, therefore, overall environmental damages remain constant despite the geoengineering breakthrough, and overall welfare may decrease. With geoengineering advances of type II, which correspond to climate adaptation technologies including solar radiation management, environmental damages decline proportionately. In this case, we show that similar reasoning leads to an increase or decrease in the overall environmental damages depending on the elasticity of the marginal damage function. Intuitively, the stock of carbon in the atmosphere has to increase so that the marginal environmental damage and future carbon taxes do not decline, and depending on the aforementioned elasticity, this may necessitate a large or a small increase in the stock of carbon in the atmosphere (the greater is the elasticity, the larger is the requisite change in the stock of carbon). As a result, welfare may again decrease. Overall, both types of geoengineering breakthroughs generate countervailing negative effects, and may make the problem of reducing and controlling carbon emissions much more difficult.

We show that the general insights are not dependent on specific assumptions made for tractability and clarity in our model. Adding heterogeneity or changing the way in which damages are modeled does not change our qualitative conclusions. More importantly, similar results apply in the context of a dynamic model in which the stock of carbon in the atmosphere and technology evolve gradually. In this dynamic model, the positive externalities from switching to clean technology have a more compelling microfoundation: technological progress takes the form of firms ascending a quality ladder, and investments for switching to a clean technology enable further improvements on that ladder to build on the foundations laid by this clean technology. We show that the BGP equilibrium in this dynamic model has a very similar structure to our static equilibrium, and the effects of the two types of geoengineering breakthroughs are essentially identical to what we described in the previous paragraph.

We see this paper as a first step both in the investigation of the implications of policy-making without commitment in the context of environmental policies and in the study of the consequences of geoengineering. In addition to considering richer menus of different technologies for reducing carbon emissions and combating climate change, future theoretical work could consider direct competition between firms using clean and dirty technologies (see Acemoglu *et al.*, 2016, for one attempt in this direction). A major element missing from our analysis is the interaction between different countries and jurisdictions, which would require political economy considerations in addition to the issues of policy-making without commitment. Perhaps even more important is to provide empirical evidence on the two-way interactions between technology and policy—how current and future policy affects investments in clean technology, and how new technologies impact future policies.

Finally, while the paper's analysis has been positive, its results have normative implications. First, much of the conversation over climate engineering thus far has centered on scientific assessments of the probabilities that geoengineering will succeed or create adverse environmental risks. We suggest that the result of this research may *itself* affect economic equilibria by impacting investments in conventional clean technology, and in this sense our model provides a note of caution for geoengineering policy-makers. Second, by stressing the costs of the policy-makers' inability to commit to future carbon taxes, our results highlight that there are additional benefits from efficient subsidies for clean technology (which would remove the excessive reliance on carbon taxes to incentivize innovation) and from new commitment devices in the context of environmental policy.[26]

# Appendix A

## A1 Omitted proofs

**Proof of Proposition 2 (Interior Equilibrium).** (*Interiority*). From Proposition 1, $\hat{\tau} = D'(\cdot;\hat{q})$. Assumption 2, which imposes that $(1-\lambda)\Psi(\underline{\tau}) < \Gamma$ and $(1-\lambda)\Psi(\overline{\tau}) > \Gamma$, then implies that neither $q = 0$ nor $q = 1$ are subgame perfect equilibria.

(*Existence and uniqueness*). The private gain from switching to clean technology, $\Psi(\tau)$, is continuous in $\tau$, so the intermediate value theorem gives existence of a point $\hat{\tau}$ such that $(1-\lambda)\Psi(\hat{\tau}) = \Gamma$. Since $\Psi(\tau)$ is increasing (from Lemma 2), $\hat{\tau}$ is unique. Moreover, because $D'' > 0$, the Pigovian tax,

$$\tau = D'((1-\delta)S_0 - q\Lambda(\tau) + k_d(\tau)),$$

is decreasing in $q$. Consequently $\hat{q}$ is also unique. ■

**Proof of Proposition 3 (Second-best).** The derivative of welfare with respect to $q$ is

$$\frac{\partial W}{\partial q} = f_c(k_c) - k_c - \Gamma - (f_d(k_d) - k_d) - \frac{\partial S}{\partial q} D'(S)$$

which, using $\partial S/\partial q = \Lambda$ and the fact that in the interior, $f_c(k_c) - k_c - (f_d(k_d) - k_d) + \Lambda\tau = \frac{\Gamma}{1-\lambda} = \Gamma - \frac{\lambda}{1-\lambda}\Gamma$, becomes

$$\frac{\partial W}{\partial q} = \frac{\lambda}{1-\lambda}\Gamma - \Lambda\tau + \Lambda D'(S). \tag{A1}$$

[26] For example, enforceable carbon price floors, such as the auction reserve price in the California cap-and-trade mechanism under AB 32 (see Borenstein *et al.*, 2015) may help to provide some medium-term commitment in the context of carbon markets. Laffont and Tirole (1996a) also discuss a number of more complicated options contracts that can be used to mimic commitment in dynamic pollution permit markets.

At $\tau = D'(S)$, (A1) is positive, implying that $\tau^{SB} > D'(S)$ yields strictly higher welfare than $\tau = D'(S)$ if and only if $\lambda > 0$. ■

**Proof of Proposition 4 (Type I Geoengineering).** (*Taxes, damages do not change*). In an interior equilibrium,

$$\hat{\tau} = \frac{1}{\Lambda}\left[\frac{\Gamma}{1-\lambda} - f_c(k_c) + k_c + f_d(k_d) - k_d\right]$$

and the RHS is invariant to a level shift in $S_0$, so $d\hat{\tau} = 0$. If $\hat{\tau} = D'(S)$, then $dS = 0$, which implies that $-\Lambda d\hat{q} = d\xi$.

(*Welfare*). We can calculate the total derivative of welfare, $W = q(f_c(k_c) - k_c - \Gamma) + (1 - q)(f_d(k_d) - k_d) - D((1-\delta)S_0 - \xi + E)$, with respect to $\xi$ as

$$\begin{aligned} \frac{dW}{d\xi} &= \left[q(f_c'(k_c) - 1)\frac{dk_c}{d\tau} + (1-q)(f_d'(k_d) - 1)\frac{dk_d}{d\tau}\right]\frac{d\hat{\tau}}{d\xi} \\ &\quad + \left[f_c(k_c) - k_c - \Gamma - (f_d(k_d) - k_d)\right]\frac{d\hat{q}}{d\xi} + D' - \frac{dE}{d\xi}D' \\ &= \left[f_c(k_c) - k_c - \Gamma - (f_d(k_d) - k_d)\right]\frac{d\hat{q}}{d\xi} + D' - \frac{dE}{d\xi}D' \\ &= \left[f_c(k_c) - k_c - \Gamma - (f_d(k_d) - k_d)\right]\frac{d\hat{q}}{d\xi} \end{aligned} \tag{A2}$$

where the second line uses $d\hat{\tau}/d\xi = 0$ and the third uses $dE/d\xi = 1$. (And $dE/d\xi = 1$ confirms $dD/d\xi = 0$). Using (5), (A2) simplifies to

$$\frac{dW}{d\xi} = \left[\lambda\left(f_c(k_c) - k_c - (f_d(k_d) - k_d)\right) - (1-\lambda)\Lambda\tau\right]\frac{d\hat{q}}{d\xi}.$$

Using $d\hat{q}/d\xi = -1/\Lambda$, and $\pi_c - \pi_d - \Lambda\tau = f_c(k_c) - k_c - (f_d(k_d) - k_d)$, we conclude that

$$\frac{dW}{d\xi} < 0 \iff \lambda(\pi_c - \pi_d) < \Lambda\tau.$$

■

**Proof of Proposition 5 (Type II Geoengineering).** (*I. Taxes*). As in the proof of Proposition 4, only $d\hat{\tau}/d\upsilon = 0$ sustains IC.

(*II. Environmental damage*). Differentiating total environmental damage, $(1-\upsilon)D(S)$, with respect to $\upsilon$, we obtain

$$\frac{dD(\cdot)}{d\upsilon} = D(S) + (1-\upsilon)D'(S)\frac{dS}{d\upsilon}. \tag{A3}$$

To calculate $dS/d\upsilon$, note that because $d\hat{\tau}/d\upsilon = 0$, we can differentiate

$$(1-\upsilon)D'(S) = \hat{\tau}$$

with respect to $\upsilon$ to obtain

$$-D'(S) + \frac{dS}{d\upsilon}(1-\upsilon)D''(\cdot) = 0 \implies \frac{dS}{d\upsilon} = \frac{1}{1-\upsilon}\frac{D'(S)}{D''}.$$

The total effect in (A3) then becomes

$$\begin{aligned}\frac{dD(\cdot)}{dv} &= -D(S) + \frac{1}{1-v}\frac{D'(S)}{D''}(1-v)D'(S) \\ &= -D(S) + \frac{1}{\eta}\cdot SD'(S),\end{aligned} \tag{A4}$$

where $\eta \equiv SD''(S)/D'(S)$ is the relative curvature of $D(\cdot)$ at $S$. By convexity ($D'' \geq 0$), the quantity $D(S)$ is bounded above by $SD'(S)$, so letting $\eta^* \equiv SD'(S)/D(S) > 1$ we have

$$\eta < \eta^* \iff dD/dv > 0.$$

(*III. Welfare*) Aggregate welfare changes with $v$ according to

$$\begin{aligned}\frac{dW}{dv} &= \frac{\partial}{\partial q}\left[q[f_c(k_c)-k_c] + (1-q)[f_d(k_d)-k_d] - q\Gamma\right]\frac{dq}{dv} - \frac{dD(\cdot)}{dv} \\ &= [f_c(k_c) - k_c - [f_d(k_d)-k_d] - \Gamma]\frac{dq}{dv} - \frac{dD(\cdot)}{dv} \\ &= \left[\frac{\lambda}{1-\lambda}\Gamma - \Lambda\tau\right]\frac{dq}{dv} - \frac{dD(\cdot)}{dv},\end{aligned}$$

where the last substitution follows from (5).

Differentiating the tax invariance condition $D'(S) = (1-v)D'(S)$ as before, noting that $S$ can adjust only through $q$, and that $\partial E/\partial q = -\Lambda$, we obtain

$$\frac{dq}{dv} = \left[\frac{\partial E}{\partial q}\right]^{-1}\frac{dS}{dv} = -\frac{1}{\Lambda(1-v)}\frac{D'(S)}{D''}, \tag{A5}$$

or equivalently,

$$\frac{dq}{dv} = -\frac{1}{\Lambda(1-v)}\frac{S}{\eta}.$$

As $\tau = (1-v)D'(S)$, using this expression for $dq/dv$ above gives

$$\left[\frac{\lambda}{1-\lambda}\Gamma - \Lambda\tau\right]\frac{dq}{dv} = -\frac{\lambda}{1-\lambda}\Gamma\cdot\frac{1}{\Lambda(1-v)}\frac{S}{\eta} + \frac{1}{\eta}SD'(S). \tag{A6}$$

From above, the total effect on environmental damage is

$$-\frac{dD(\cdot)}{dv} = D(S) - \frac{1}{\eta}SD'(S). \tag{A7}$$

The last term in each of the previous two expressions cancels when summed, and we obtain

$$\frac{dW}{dv} = D(S) - \frac{\lambda}{1-\lambda}\Gamma\cdot\frac{1}{\Lambda(1-v)}\frac{S}{\eta}.$$

From (5), we have

$$\Gamma = (1-\lambda)(\pi_c - \pi_d), \tag{A8}$$

and multiplying both sides by $\Lambda(1-\upsilon)D'(S)/D(S)$, we obtain

$$\frac{dW}{d\upsilon} < 0 \iff \Lambda\tau - \lambda(\pi_c - \pi_d)\frac{1}{\eta}\frac{SD'(S)}{D(S)} < 0$$

and letting $a \equiv SD'(S)/D(S) > 1$ (where the inequality follows from the strict convexity of $D(\cdot)$), we conclude that

$$\eta < \eta^{\mathrm{II}}(\lambda) \equiv a\lambda\left(\frac{\pi_c - \pi_d}{\Lambda\tau}\right)$$

characterizes the family of damage functions for which $dW/d\upsilon < 0$. ■

**Proof of Claim in Footnote 17 (Quadratic Damages).** (*Damages always increase*). If $D$ is quadratic, then the approximation

$$D(S) \approx SD' - \frac{1}{2}S^2 D''$$

is exact, so that $D/SD' = 1 - \eta/2$. By (A4), $dD/d\upsilon > 0 \iff -D(S) + \eta^{-1}SD'(S)$, so

$$dD/d\upsilon > 0 \iff -1 + \eta/2 + 1/\eta > 0,$$

or $dD/d\upsilon > 0 \iff \eta^2/2 - \eta + 1 > 0$. But $\eta^2/2 - \eta + 1$ is a polynomial with only imaginary roots, and is thus always positive.

(*Welfare*). Under the assumption that $D$ is quadratic, $a = 1 - \eta/2$ and from (A4), we conclude that the condition

$$\eta(1-\eta/2) - \lambda\left(\frac{\pi_c - \pi_d}{\Lambda\tau}\right) < 0$$

characterizes the region for which $dW/d\upsilon < 0$. The resulting polynomial has only imaginary roots when

$$\lambda\left(\frac{\pi_c - \pi_d}{\Lambda\tau}\right) > \frac{1}{2}$$

which is precisely the condition that $\lambda \geq \lambda^*$. ■

**Proof of Proposition 6 (Direct Subsidies).** Welfare in this extended model is given by

$$W = (1-q)[f_d(k_d) - k_d] + q[f_c(k_c) - k_c - (1+\sigma\omega)\Gamma] - D(S),$$

the new binding IC constraint becomes

$$\Psi(\tau) = \frac{1-\sigma}{1-\lambda}\cdot\Gamma, \tag{A9}$$

and the first-order condition for $\sigma$ is

$$[f_c(k_c) - k_c - (f_d(k_d) - k_d) + \Lambda\tau - (1+\sigma\omega)\Gamma]\frac{dq}{d\sigma} - q\omega\Gamma = 0. \tag{A10}$$

Note also that we can obtain

$$\frac{dq}{d\sigma} = \frac{\Gamma}{1-\lambda}\frac{1}{\Lambda D''(S)} \tag{A11}$$

by differentiating (A9) totally with respect to $\sigma$ (noting that $\tau$ depends on $\sigma$) and using $\Psi'(\tau) = \Lambda$ and $\frac{d\tau}{d\sigma} = -\Lambda D''(S)\frac{dq}{d\sigma}$. Using (A9) and (A11), we can rewrite (A10) as

$$\frac{\Gamma}{1-\lambda}\frac{1}{\Lambda D''(S)}\left(\frac{\lambda - \sigma - (1-\lambda)\sigma\omega}{1-\lambda}\right)\Gamma = q\omega\Gamma$$

yielding

$$\sigma = \frac{\lambda}{1+(1-\lambda)\omega} - q\omega\frac{(1-\lambda)^2\Lambda D''(S)}{1+(1-\lambda)\omega}$$

which reduces to $\sigma = \lambda$ if $\omega = 0$, and otherwise shows that $\sigma < \lambda$ for all $\omega > 0$. To study welfare,

$$\frac{dW}{d\xi} = \left(\frac{\lambda-\sigma}{1-\lambda}\Gamma - \Lambda\tau - \sigma\omega\Gamma\right)\frac{dq}{d\xi}$$

and using $\pi_c - \pi_d = \left(\frac{1-\sigma}{1-\lambda}\right)\Gamma$ with $dq/d\xi < 0$, we obtain the statement of the proposition. ■

**Proof of Proposition 7 ($\beta$-Benevolence).** For type II, $\tau = \beta(1-\upsilon)D'(S)$. Equation (A5) is unchanged but we obtain a modified version of (A6),

$$\left[\frac{\lambda}{1-\lambda}\Gamma - \Lambda\tau\right]\frac{dq}{d\upsilon} = -\frac{\lambda}{1-\lambda}\Gamma\cdot\frac{1}{\Lambda(1-\upsilon)}\frac{S}{\eta} + \beta\frac{SD'(S)}{\eta}.$$

The total effect on environmental damage is still given by (A7), so that the last terms in each of (A6) and (A7) no longer cancel, and

$$\frac{dW}{d\upsilon} = D(S) - \frac{\lambda}{1-\lambda}\Gamma\cdot\frac{1}{\Lambda(1-\upsilon)}\frac{S}{\eta} - (1-\beta)\frac{SD'(S)}{\eta}.$$

Using (A8), and multiplying both sides by $\beta\Lambda(1-\upsilon)D'(S)/D(S)$, we obtain

$$\frac{dW}{d\upsilon} < 0 \iff \Lambda\tau - \left[\beta\lambda\frac{\pi_c - \pi_d}{\eta} + \frac{1-\beta}{\eta}\Lambda\tau\right]\frac{SD'(S)}{D(S)} < 0$$

yielding (8). ■

## A2 Robustness to Ex Ante Heterogeneity

As described in Section 2.11, suppose that firms are differentiated in terms of fixed costs of transitioning to clean technology. In particular, suppose that

$$\Gamma_i = \Gamma + \chi_i$$

with $\mathbb{E}[\chi_i] = 0$ and $H(x) \equiv \mathbb{P}(\chi_i \leq x)$. We call this "economy H." While before, equilibrium technology adoption was the jump-discontinuous function

$$\hat{q}(\tau) = \hat{q}\mathbf{1}_{\{(1-\lambda)\Psi(\tau)=\Gamma\}} + \mathbf{1}_{\{(1-\lambda)\Psi(\tau)>\Gamma\}},$$

the effect of heterogeneity is to smooth equilibrium technology,

$$\hat{q}(\tau) = H\left((1-\lambda)\Psi(\tau) - \Gamma\right).$$

The equilibrium $\hat{\tau}$ is now the fixed point of

$$\tau = D'\left((1-\delta)S_0 + k_d - \Lambda \cdot H\left((1-\lambda)\Psi(\tau) - \Gamma\right)\right), \tag{A12}$$

which will be unique if $H(\cdot)$ is increasing in the neighborhood of the solution to (A12), since $H(\cdot)$ is always nondecreasing. Define the derivative of $H$ from the left by $h$.

We can build some intuition for our geoengineering comparative statics by totally differentiating $\hat{\tau}$ with respect to $\xi$ (Type I), to obtain

$$\frac{d\hat{\tau}}{d\xi} = -\frac{1}{1/D''(S) - [\partial k_d/\partial\tau] + \Lambda^2(1-\lambda)h(0)} < 0 \tag{A13}$$

and

$$\frac{dS}{d\xi} = 1 + \frac{1}{D''(S)}\frac{d\hat{\tau}}{d\xi}. \tag{A14}$$

When either $h(0) \to +\infty$ (the case of ex ante identical firms) or $D''(S) \to 0$, the RHS of (A13) vanishes: the Pigovian tax is totally invariant to geoengineering. Otherwise, the more concentrated the distribution of firms is at $(1-\lambda)\Psi(\tau) - \Gamma$, the closer $[d\hat{\tau}/d\xi]$ gets to zero.

Types I and II geoengineering go through just as before, subject to a restriction on heterogeneity that we now calculate exactly (proofs omitted):

**Proposition A1 (Geoengineering Type I with Heterogenous Firms)** *Consider economy H. For every $\varepsilon > 0$, there exists a $\ell^{\mathrm{I}} \geq 0$ such that if $h(0) \geq \ell^{\mathrm{I}}$, then*

- $|d\hat{\tau}/d\xi| < \varepsilon.$
- $d\hat{S}/d\xi > 1 - \left(\frac{1}{D''(S)}\right)\varepsilon.$
- $d\hat{q}/d\xi < -1/\Lambda + \left(\frac{1}{\Lambda D''(S)} - \frac{1}{\Lambda}\frac{\partial E}{\partial \tau}\right)\varepsilon.$
- $dW/d\xi < 0 \iff \lambda(\pi_c - \pi_d) > \Lambda\tau + \mathcal{O}(\varepsilon).$

*In particular,*

$$\ell^{\mathrm{I}}(\varepsilon) = \frac{1}{(1-\lambda)\Lambda^2}\left[\frac{1}{\varepsilon} - \frac{1}{D''(S)} + \frac{\partial k_d}{\partial \tau}\right].$$

**Proposition A2 (Geoengineering Type II with Heterogenous Firms)** *Consider economy H. For every $\varepsilon > 0$, there exists an $\ell^{\mathrm{II}} \geq 0$ such that if $h(0) \geq \ell^{\mathrm{II}}$ then*

- $|d\hat{\tau}/d\upsilon| < \varepsilon$
- $d\hat{S}/d\upsilon > \frac{D'(S)}{(1-\upsilon)D''(S)} + \left(\frac{1}{(1-\upsilon)D''(S)}\right)\varepsilon$
- $d\hat{q}/d\upsilon < -\frac{S}{(1-\upsilon)\Lambda}\frac{1}{\eta} + \left(\frac{1}{\Lambda(1-\upsilon)D''(S)} - \frac{1}{\Lambda}\frac{\partial E}{\partial \tau}\right)\varepsilon$

- *$dW/dv < 0 \iff \eta < \eta^{\text{II}}(\lambda) + \mathcal{O}(\varepsilon)$, where $\eta^{\text{II}}(\lambda)$ is defined in Proposition 5.*

*In particular,*

$$\ell^{\text{II}}(\varepsilon) = \frac{1}{(1-\upsilon)(1-\lambda)\Lambda^2}\left[\frac{1}{\varepsilon}\frac{S}{\eta} - \frac{1}{D''(S)} + (1-\upsilon)\frac{\partial k_d}{\partial \tau}\right].$$

## A3 Stochastic Geoengineering

As described in Section 2.11, suppose that geoengineering is known to succeed only with some probability $\kappa \in (0,1)$, realized after clean investments and carbon taxation.[27]

**Proposition A3 (Stochastic Geoengineering Type I)** *Suppose an increment $d\xi$ of type I geoengineering arrives with probability $\kappa \in (0,1)$. Then expected welfare declines if and only if*

$$\lambda(\pi_c - \pi_d) > \Lambda \frac{\mathbb{E}_\kappa[D''(S)]}{D''(S-d\xi)}(\tau - (1-\kappa)b)$$

*where $\mathbb{E}_\kappa[D''(S)] = \kappa D''(S-d\xi) + (1-\kappa)D''(S)$ and $b \equiv D'(S) - D'(S-d\xi) > 0$.*

**Proof.** Totally differentiating the tax rate,

$$\hat{\tau} = \kappa D'(S-d\xi) + (1-\kappa)D'(S),$$

with respect to $\xi$, we obtain (as $\hat{\tau}$ is pinned down as before by technology IC),

$$0 = -\kappa D''(S-d\xi) - \left[\kappa\Lambda D''(S-d\xi) + (1-\kappa)\Lambda D''(S)\right]\frac{dq}{d\xi}$$

so that $\frac{dq}{d\xi} = -\kappa\frac{D''(S-d\xi)}{\mathbb{E}_\kappa[\Lambda D''(S)]}$. Consequently, the derivative of expected welfare is

$$\frac{d\mathbb{E}_\kappa[W]}{d\xi} = \left(\frac{\lambda}{1-\lambda}\Gamma - \Lambda\tau\right)\frac{dq}{d\xi} + \frac{dq}{d\xi}\Lambda\mathbb{E}_\kappa[D'(S)] + \kappa D'(S-d\xi)$$

and using $\tau = \mathbb{E}_\kappa[D'(S)]$, we obtain the following simplification:

$$\frac{d\mathbb{E}_\kappa[W]}{d\xi} = -\frac{\lambda}{1-\lambda}\Gamma \cdot \frac{\kappa D''(S-d\xi)}{\Lambda\mathbb{E}_\kappa[D''(S)]} + \kappa D'(S-d\xi)$$

which, as $\kappa > 0$, rearranges to the statement in Proposition A3. ■

**Proposition A4 (Stochastic Geoengineering Type II)** *Suppose an increment $d\upsilon$ of type II geoengineering arrives with probability $\kappa \in (0,1)$. Then expected welfare declines if and only if*

$$\eta < \frac{a}{\Lambda}\left(\frac{1-d\upsilon}{1-\kappa\cdot d\upsilon}\right)\left(\frac{\lambda(\pi_c-\pi_d)}{\tau-(1-\kappa)b'}\right)$$

*where $b' \equiv d\upsilon \cdot D'(S) > 0$.*

**Proof.** Calculate $\frac{dq}{d\upsilon} = -\kappa D'(S)/\mathbb{E}_\kappa[\Lambda(1-\upsilon)D''(S)]$, and manipulate $\frac{d\mathbb{E}[W]}{d\upsilon}$ as in the proof of Proposition 5. ■

[27] The case in which the event of geoengineering is realized before carbon taxation is essentially identical, since clean investments are still made with reference to the expected tax, though the resulting calculations involve more expectation operators, given the additional uncertainty over the realized carbon tax.

## References


Acemoglu, D. (2007). Equilibrium bias of technology. *Econometrica*, **75** (5), 1371—1409.

—, Aghion, P., Bursztyn, L. and Hémous, D. (2012). The environment and directed technical change. *American Economic Review*, **102** (1), 131–166.

—, Akcigit, U., Hanley, D. and Kerr, W. (2016). Transition to clean technology. *Journal of Political Economy*, **124** (1), 52–104.

Aghion, P., Dechezleprêtre, A., Hémous, D., Martin, R. and Reenen, J. V. (2016). Carbon taxes, path dependency, and directed technical change: Evidence from the auto industry. *Journal of Political Economy*, **124** (1), 1–51.

— and Griffith, R. (2005). *Competition and Growth: Reconciling Theory and Evidence*. Zeuthen Lectures, MIT Press.

— and Howitt, P. (1992). A model of growth through creative destruction. *Econometrica*, **60** (2), 323–351.

Akcigit, U. and Kerr, W. (forthcoming). Growth through heterogenous innovation. *Journal of Political Economy*.

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

Bellman, R. (1957). *Dynamic Programming*. Princeton University Press.

Benhabib, J. and Rustichini, A. (1997). Optimal taxes without commitment. *Journal of Economic Theory*, **77**, 231–259.

Borenstein, S. (2012). The private and public economics of renewable electricity generation. *Journal of Economic Perspectives*, **26** (1), 67–92.

—, Bushnell, J., Wolak, F. and Zaragoza-Watkins, M. (2015). Expecting the unexpected: Emissions uncertainty and environmental market design. *NBER Working Paper*, (20999).

Boustan, L. P., Kahn, M. and Rhode, P. (2012). Moving to higher ground: Migration response to natural disasters in the early twentieth century. *American Economic Review: Papers & Proceedings*, **102** (3), 238–24.

Bovenberg, A. L. and Smulders, S. (1995). Environmental quality and pollution-augmenting technological change in a two-sector endogenous growth model. *Journal of Public Economics*, **57** (3), 369–391.

— and — (1996). Transitional impacts of environmental policy in an endogenous growth model. *International Economic Review*, **37** (4), 861–893.

Calvo, G. (1978). On the time consistency of optimal policy in a monetary economy. *Econometrica*, **46**, 1411–1428.

Chari, V. and Kehoe, P. (1990). Sustainable plans. *Journal of Political Economy*, **98** (4), 783–802.

Cook, J., Oreskes, N., Doran, P., Anderegg, W., Verheggen, B., Maibach, E., Carlton, J. S., Lewandowsky, S., Skuce, A., Green, S., Nuccitelli, D., Jacobs, P., Richardson, M., Winkler, B., Painting, R. and Rice, K. (2016). Consensus on consensus: A synthesis of consensus estimates on human-caused global warming. *Environmental Research Letters*, **11** (4), 1–7.

Costinot, A., Donaldson, D. and Smith, C. (2016). Evolving comparative advantage and the impact of climate change in agricultural markets: Evidence from 1.7 million fields around the world. *Journal of Political Economy*, **124** (1), 205–248.

Dell, M., Jones, B. and Olken, B. (2012). Temperature shocks and economic growth: Evidence from the last half century. *American Economic Journal: Macroeconomics*, **4** (3), 66–95.

Desmet, K. and Rossi-Hansberg, E. (2015). On the spatial economic impact of global warming. *Journal of Urban Economics*, **88**, 16–37.

Emmerling, J. and Tavoni, M. (2018). Climate engineering and abatement: A 'flat' relationship under uncertainty. *Environmental and Resource Economics*, **69** (2), 395–415.

Flannery, T. (2015). *Atmosphere of Hope: Solutions to the climate crisis*. Penguin.

Gillingham, K. and Bollinger, B. (2014). Learning-by-doing in solar photovoltaic installations. *Working Paper*.

Golosov, M., Hassler, J., Krusell, P. and Tsyvinski, A. (2014). Optimal taxes on fossil fuel in general equilibrium. *Econometrica*, **82** (1), 41–88.

Goulder, L. and Mathai, K. (2000). Optimal $CO_2$ abatement in the presence of induced technological change. *Journal of Environmental Economics and Management*, **39** (1), 1–38.

— and Schneider, S. (1999). Induced technological change and the attractiveness of $CO_2$ abatement policies. *Resource and Energy Economics*, **21** (3–4), 211–253.

Greenstone, M., Kopits, E. and Wolverton, A. (2013). Developing a social cost of carbon for US regulatory analysis: A methodology and interpretation. *Review of Environmental Economics and Policy*, **7** (1), 23–46.

Greenwood, J., Hercowitz, Z. and Huffman, G. (1988). Investment, capacity utilization, and the real business cycle. *American Economic Review*, **78** (3), 402–417.

Grimaud, A., Lafforgue, G. and Magné, B. (2011). Climate change mitigation options and directed technical change: A decentralized equilibrium analysis. *Resource and Energy Economics*, **33** (4), 938–962.

Grossman, G. and Helpman, E. (1991). Quality ladders in the theory of growth. *The Review of Economic Studies*, **58** (1), 43–61.

— and — (1994). Protection for sale. *American Economic Review*, **84** (4), 833–850.

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

— (2016). Technology and time inconsistency, Working Paper, University of Oslo.

— and Battaglini, M. (2016). Participation and duration of environmental agreements. *Journal of Political Economy*, **124** (1), 160–204.

Hartley, P., Medlock, K., Temzelides, T. and Zhang, X. (2016). Energy sector innovation and growth: An optimal energy crisis. *The Energy Journal*, **37** (1), 233–258.

Hassler, J., Krusell, P. and Olovsson, C. (2012). Energy-saving technical change. *NBER Working Paper*, (18456).

—, — and Smith, T. (2016). Environmental macroeconomics. In J. Taylor and H. Uhlig (eds.), *Handbook of Macroeconomics*, vol. 2, Elsevier, pp. 1893–2008.

Hsiang, S., Burke, M. and Miguel, E. (2013). Quantifying the influence of climate on human conflict. *Science*, **341** (6151).

IPCC (2014). *limate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom: Cambridge University Press.

Jevons, W. (1866). *The Coal Question: An Inquiry Concerning the Progress of the Nation, and the Probable Exhaustion of our Coal-mines*. London: Macmillan & Co., 2nd edn.

Jones, C. (1995). R&D-based models of economic growth. *Journal of Political Economy*, **103** (4), 759–784.

Karp, L. and Newbery, D. (1993). Intertemporal consistency issues in depletable resources. In A. Kneese and J. L. Sweeney (eds.), *Handbook of Natural Resource and Energy Economics*, vol. 3, Elsevier, pp. 881–931.

Keith, D. (2013). *A Case for Climate Engineering*. MIT Press.

—, Wagner, G. and Zabel, C. (2017). Solar geoengineering reduces atmospheric carbon burden. *Nature Climate Change*, **7**, 617–619.

Klein, P., Krusell, P. and Rull, J.-V. R. (2008). Time-consistent public policy. *The Review of Economic Studies*, **78**, 789–808.

Klette, T. J. and Kortum, S. (2004). Innovating firms and aggregate innovation. *Journal of Political Economy*, **112** (5), 986–1018.

Kopp, R., Kemp, A., Bittermann, K., Horton, B., Donnelly, J., Gehrels, W. R., Hay, C., Mitrovica, J., Morrow, E. and Rahmstorf, S. (2016). Temperature-driven global sea-level variability in the Common Era. *Proceedings of the National Academy of Sciences*, **113** (11), 1434–1441.

Kousky, C., Luttmer, E. and Zeckhauser, R. (2006). Private investment and government protection. *Journal of Risk and Uncertainty*, **33** (1), 73–100.

Kydland, F. and Prescott, E. (1977). Rules rather than discretion: The inconsistency of optimal plans. *Journal of Political Economy*, **85** (3), 473–492.

Lackner, K., Brennan, S., Matter, J., Park, A.-H., Wright, A. and van der Zwaan, B. (2012). The urgency of the development of $CO_2$ capture from ambient air. *Proceedings of the National Academy of Sciences*, **109** (33), 13156–13162.

Laffont, J.-J. and Tirole, J. (1996a). Pollution permits and compliance strategies. *Journal of Public Economics*, **62**, 85–125.

— and — (1996b). Pollution permits and environmental innovation. *Journal of Public Economics*, **62**, 127–140.

Lenton, T. and Vaughan, N. (2009). The radiative forcing potential of different climate geoengineering options. *Atmospheric Chemistry and Physics*, **9**, 5539—5561.

Mann, M., Rahmstorf, S., Kornhuber, K., Steinman, B., Miller, S. and Coumou, D. (2017). Influence of anthropogenic climate change on planetary wave resonance and extreme weather events. *Scientific Reports*.

McClellan, J., Keith, D. and Apt, J. (2012). Cost analysis of stratospheric albedo modification delivery systems. *Environmental Research Letters*, **7**, 034019.

Moreno-Cruz, J. (2015). Mitigation and the geoengineering threat. *Resource and Energy Economics*, **41** (248–263).

Morton, O. (2015). *The Planet Remade: How geoengineering could change the world*. Princeton University Press.

National Research Council (2015). *Climate Intervention: Reflecting Sunlight to Cool Earth*. Washington, DC: The National Academies Press.

Newell, R., Jaffe, A. and Stavins, R. (1999). The induced innovation hypothesis and energy-saving technological change. *Quarterly Journal of Economics*, **114** (3), 941–975.

Nordhaus, W. (1991). To slow or not to slow: The economics of the greenhouse effect. *The Economic Journal*, **101** (407), 920–937.

— (2007). A review of the 'Stern Review on the Economics of Climate Change'. *Journal of Economic Literature*, **45** (3), 686–702.

— (2008). *A Question of Balance: Weighing the Options on Global Warming Policies*. Yale University Press.

— (2014). Estimates of the social cost of carbon: Concepts and results from the DICE-2013R model and alternative approaches. *Journal of the Association of Environmental and Resource Economists*, **1** (1), 273–312.

Peltzman, S. (1975). The effects of automobile safety regulation. *Journal of Political Economy*, **83** (4), 677–726.

Phelan, C. and Stacchetti, E. (2001). Sequential equilibria in a Ramsey tax model. *Econometrica*, **69** (6).

Popp, D. (2002). Induced innovation and energy prices. *American Economic Review*, **92** (1), 160–180.

— (2004). ENTICE: Endogenous technological change in the DICE model of global warming. *Journal of Environmental Economics and Management*, **48** (1), 742–768.

Rasch, P., Tilmes, S., Turco, R., Robock, A., Oman, L., Chen, C.-C., Stenchikov, G. and Garcia, R. (2008). An overview of geoengineering of climate using stratospheric sulphate aerosols. *Philosophical Transactions of the Royal Society A*, **366** (1882), 4007–04037.

Romer, D. (1990). Endogenous technical change. *Journal of Political Economy*, **98** (5), S71–S102.

Socolow, R., Desmond, M., Aines, R., Blackstock, J., Bolland, O., Kaarsberg, T., Lewis, N., Mazzotti, M., Pfeffer, A., Sawyer, K., Siirola, J., Smit, B. and Wilcox, J. (2011). *Direct Air Capture of $CO_2$ with Chemicals*. Tech. rep., American Physical Society Panel on Public Affairs.

van der Zwaan, B. C. C., Gerlagh, R., Klaassen, G. and Schrattenholzer, L. (2002). Endogenous technological change in climate change modelling. *Energy Economics*, **24** (1), 1–19.

Weitzman, M. (2015). A voting architecture for the governance of free-driver externalities, with an application to geoengineering. *Scandinavian Journal of Economics*, **117** (4), 1049–1068.

Yan, Q., Wei, T., Korty, R., Kossin, J., Zhang, Z. and Wang, H. (2016). Enhanced intensity of global tropical cyclones during the mid-Pliocene warm period. *Proceedings of the National Academy of Sciences*, **113** (46), 12963–12967.

# Online Appendix B (Not for Publication)

In this Appendix, we first include the remainder of omitted proofs from the main text. We then extend our economy in three directions. First, we allow the stock of carbon to directly affect the production technology in the static model, as in Nordhaus (2008) and Golosov *et al.* (2014). Second, we relax the assumption in the static model that environmental damages and producer surplus are additively separable, and instead assume that society obtains welfare from the ratio of consumption to environmental damage, which turns out to be isomorphic to the latter extension up to a monotone transformation of the damages function. Third, we consider nonlinear flow utility over consumption (net of environmental damage) in our dynamic model.

## B1 Omitted proofs

**Proof of Lemma 3 (Shadow Cost of Carbon).** In equilibrium, the costate variable $\mu_t$ will satisfy the Euler-Lagrange condition $\partial H_t/\partial S = \dot{\mu}_t - \rho\mu_t$, which we can write as

$$\dot{\mu}_t = -A_t D'(S_t) + (\rho + \delta)\mu_t.$$

Dividing this equation by $A_t$, we obtain

$$\frac{\dot{\mu}_t}{A_t} = -D'(S_t) + (\rho + \delta)p_t. \tag{B1}$$

Since $p_t = \mu_t/A_t$, we have

$$\dot{p}_t = \frac{\dot{\mu}_t}{A_t} - \frac{\mu_t}{A_t^2}\frac{dA_t}{dt} = \frac{\dot{\mu}_t}{A_t} - p_t\alpha\lambda,$$

where the second equality uses the fact that $dA_t/dt = \alpha\lambda A_t$. Hence (B1) becomes

$$\dot{p}_t = -D'(S_t) + (\delta + \rho - \alpha\lambda)\, p_t,$$

which is exactly (26). Furthermore, with the transversality condition,

$$\lim_{t\to\infty} p_t S_t e^{-\rho t} = 0 \tag{B2}$$

and the initial condition $S_0$, we obtain

$$p_0 = \int_0^\infty D'(S_t)e^{-(\delta+\rho-\alpha\lambda)t}dt,$$

and the differential equation (26) admits the unique solution (27). ■

**Proof of Proposition 8 (Pigovian Best-response).** We start with a crucial lemma.

**Lemma B1 (Convergence of Clean Technology)** *Suppose that Assumptions 1′, 2′, and 3 hold. Then, $q_t \to \hat{q}$ by some finite time $T < \infty$.*

**Proof of Lemma B1.** The sequence $(q_t)_{t\geq 0}$ lives in the compact set $[0,1]$, and $(q_t)_{t\geq 0}$ is monotone since clean technology decisions are irreversible. Hence $q_t \to \tilde{q}$ .

Suppose, to obtain a contradiction, that $q_t < \tilde{q}$ for all $t$. Any tax trajectory that sustains an equilibrium in which $\dot{q}_t > 0$ for some $t \geq t'$ for every $t' < \infty$ must satisfy (22) infinitely often, which requires that for some $t \geq t'$, we have

$$\int_t^\infty \Psi(\tau_s) e^{-(\rho+\lambda)s} dt = \Gamma.$$

Let $(\tau_t)_{t\geq 0}$ be a sequence of taxes with this property.

Furthermore, the planner can attain $\hat{q}$ by always setting Pigovian taxes, which she prefers to any $\tilde{q} < \hat{q}$ (this is immediate, and also follows directly from Proposition 11, which shows that the planner prefers to induce a transition to $q_t \to \tilde{q}$). Hence we can focus on the case where $\tilde{q} > \hat{q}$. Assumption $2'$ and $\tilde{q} > \hat{q}$ together imply that for every $t' < \infty$, there is a subset of $[t', \infty)$ of positive measure for which $\tau_t > p_t$.

We next use this fact and construct a deviation from $(\tau_t)_{t\geq t'}$ to $(p_t)_{t\geq t'}$ that induces $q_{t'} \neq \tilde{q}$ for all $t' < \infty$ (with $q_t \to \tilde{q}$), completing the contradiction argument. To verify this, observe that deviating at $t'$ to $p_t$ forever (calculated with reference to the deviation path $(S'_t)_{t\geq t'}$) will fix $q_t = q_{t'}$ for all $t \geq t'$. Therefore, the time-$t'$ deviation will yield welfare

$$\mathcal{U}(t') = \int_{t'}^\infty \left[q_{t'} \left[f_c(k_c(p_t)) - k_c(p_t)\right] + (1-q_{t'}) \left[f_d(k_d(p_t)) - k_d(p_t)\right] - D(S'_t)\right] e^{-(\rho-g)t} dt,$$

while the original path with $q_t \to \tilde{q}$ will yield

$$\mathcal{Z}(t') = \int_{t'}^\infty \left[q_t \left[f_c(k_c(\tau_t)) - k_c(\tau_t)\right] + (1-q_t) \left[f_d(k_d(\tau_t)) - k_d(\tau_t)\right] - D(S_t)\right] e^{-(\rho-g)t} dt,$$

where the stock of carbon after the deviation, $S'_t$, satisfies $\dot{S}'_t = q_{t'}\gamma k_c(p_t) + (1-q_{t'})k_d(p_t) - \delta S'_t$, the original stock of carbon $S_t$ satisfies $\dot{S}_t = q_t\gamma k_c(\tau_t) + (1-q_t)k_d(\tau_t) - S_t$, and $S'_{t'} = S_{t'}$. Hence the gain from the time-$t'$ Pigovian deviation is

$$\begin{aligned} \mathcal{U}(t') - \mathcal{Z}(t') \geq \; & \int_{t'}^\infty \{q_{t'} \left[f_c(k_c(p_t)) - k_c(p_t) - (f_c(k_c((\tau_t)) - k_c(\tau_t))\right] \\ & + (1-q_{t'}) \left[f_d(k_d(p_t)) - k_d(p_t) - (f_d(k_d(\tau_t)) - k_d(\tau_t))\right] \\ & - \left[D\left(S(q_{t'}, \tau_t)\right) - D(S(q_{t'}, p_t)\right]\} e^{-(\rho-g)t} dt \\ & - \|q_t - q_{t'}\| \int_{t'}^\infty \left(f_c(\tau_t) - f_d(\tau_t) + \Lambda(\tau_t) D'(S(q_{t'}, \tau_t))\right) e^{-(\rho-g)t} dt \end{aligned} \tag{B3}$$

using $\Gamma > 0$ and the bound $D(S(q_{t'}, \tau_t)) - D(S(q_t, \tau_t)) \leq \|q_t - q_{t'}\| \Lambda(\tau_t) D'(q_{t'}, \tau_t)$ that follows from convexity of $D(\cdot)$. As $p_t < \tau_t$ for a set of positive measure, and $p_t \leq \tau_t$ always,[28] the first integral

[28]The planner will always set $\tau_t \geq p_t$ for all $t \geq t'$, as $\tau_t < p_t$ reduces clean technology incentives and lowers welfare from net consumption.

in (B3) is strictly positive by definition of $p_t$, while the final term can be made arbitrarily small in finite time, since $\|q_t - q_{t'}\| \leq \|\tilde{q} - q_{t'}\|$ for all $t \geq t'$; thus, for any $\varepsilon > 0$, there exists a $t' < \infty$ such that $\|\tilde{q} - q_{t'}\| < \varepsilon$. But because $\Psi(\cdot)$ is strictly increasing and $\tau_t > p_t$ infinitely often and always on a subset of positive measure, we deduce that

$$\int_t^{\infty} \Psi(p_s) e^{-(\rho+\lambda)(t-s)} ds < \int_t^{\infty} \Psi(\tau_s) e^{-(\rho+\lambda)(t-s)} ds = \Gamma$$

for all $t \geq t'$, which contradicts (22) unless $\dot{q}_t = 0$ for all $t \geq t'$. Therefore $q_t \to \tilde{q}$ in finite time.

Finally, any limit $\tilde{q} \neq \hat{q}$ cannot be part of an equilibrium, because after $\tilde{q}$ is reached, from Proposition 8, $\tau_t = p_t$, and thus (30), (31), and (32) need to hold, and thus $\tilde{q} = \hat{q} \in (0,1)$ (where 0 and 1 are ruled out by Assumption 2$'$). ■

Let $T$ be given as in the above lemma. First, consider $t_0 \geq T$. The planner's objective is

$$\int_{t_0}^{\infty} \left[\hat{q} A_t \left[f_c(k_{ct}) - k_{ct}\right] + (1-\hat{q}) A_t \left[f_d(k_{dt}) - k_{dt}\right] - A_t D(S_t)\right] e^{-\rho(t-t_0)} dt$$

which admits the (normalized) Hamiltonian

$$\tilde{H}_t = \hat{q}\left[f_c(k_{ct}) - k_{ct}\right] + (1-\hat{q})\left[f_d(k_{dt}) - k_{dt}\right] - p_t\left[\gamma \hat{q} k_{ct} + (1-\hat{q}) k_{dt}\right]$$

where $p_t$ is given by (27). Socially optimal input levels must satisfy the necessary first-order conditions

$$\frac{\partial \tilde{H}_t}{\partial k_{ct}} = \hat{q}\left[f_c'(k_{ct}) - 1\right] - \gamma \hat{q} p_t = 0$$

and

$$\frac{\partial \tilde{H}_t}{\partial k_{dt}} = (1-\hat{q})\left[f_d'(k_{dt}) - 1\right] - (1-\hat{q}) p_t = 0$$

for all $t \geq t_0$, which are also sufficient because $\tilde{H}_t$ is strictly concave. Comparing these to the first-order conditions of firms, which are

$$f_c'(k_{ct}) - 1 = \gamma \tau_t \quad \text{and} \quad f_d'(k_{dt}) - 1 = \tau_t \quad \text{for all } t \geq t_0,$$

we conclude that $\tau_t = p_t$ for all $t \geq T$.

Second, consider $t < T$. We prove $\tau_t \leq p_t$ by backwards induction, and do this before taking the limit $\Delta \to 0$ for convenience. The planner's utility at $T - \Delta$, given by the discrete version of (15) and normalized by $1/A_t$, is

$$\sum_{s=T}^{\infty} \left[c_{T-\Delta+\Delta(s-T)} - D\left(S_{T-\Delta+\Delta(s-T)}\right)\right] e^{-\Delta(\rho-g)(s-T)},$$

which, since $q_T = \hat{q}$, we can represent recursively as

$$c_{T-\Delta} - D\left(S_{T-\Delta}\right) + e^{-(\rho-g)\Delta} \mathcal{V}_T\left(S_T, \hat{q}\right) \tag{B4}$$

where

$$\mathcal{V}_t\left(S, q\right)=\max_{\{\tau_t\}_{t=t,t+\Delta,\ldots}}\sum_{s=t}^{\infty}\left[c_{t+\Delta(s-t)}-D\left(S_{t+\Delta(s-t)}\right)\right]e^{-(\rho-g)\Delta(s-t)}$$

is the planner's continuation value conditional on stock of carbon $S$ and aggregate clean technology $q$. The first-order condition of (B4) is

$$\frac{\partial c_{T-\Delta}}{\partial\tau_{T-\Delta}}+e^{-(\rho-g)\Delta}\frac{\partial\mathcal{V}_T(S_T,\hat{q})}{\partial\tau_{T-\Delta}}=0,$$

and, since $q_t=\hat{q}$ for all $t\geq T$,

$$\begin{aligned}e^{-(\rho-g)\Delta}\frac{\partial\mathcal{V}_T(S_T,\hat{q})}{\partial\tau_{T-\Delta}}&=-\sum_{s=0}^{\infty}\frac{\partial S_{T+s\Delta}}{\partial\tau_{T-\Delta}}D'\left(S_{T+\Delta s}\right)e^{-(\rho-g)\Delta s}\\&=\left(\gamma q_{T-\Delta}\frac{\partial k_{c,T-\Delta}}{\partial\tau_{T-\Delta}}+\left(1-q_{T-\Delta}\right)\frac{\partial k_{d,T-\Delta}}{\partial\tau_{T-\Delta}}\right)p_{T-\Delta}.\end{aligned}$$

Consequently, as

$$\frac{\partial c_{T-\Delta}}{\partial\tau_{T-\Delta}}=q_{T-\Delta}\left[f_c'(k_{c,T-\Delta})-1\right]\frac{\partial k_{c,T-\Delta}}{\partial\tau_{T-\Delta}}+\left(1-q_{T-\Delta}\right)\left[f_d'(k_{d,T-\Delta})-1\right]\frac{\partial k_{d,T-\Delta}}{\partial\tau_{T-\Delta}},$$

and, by firm-level optimization,

$$f_c'(k_{c,T-\Delta})-1=\gamma\tau_{T-\Delta}\quad\text{and}\quad f_d'(k_{d,T-\Delta})-1=\tau_{T-\Delta}$$

so only $\tau_{T-\Delta}=p_{T-\Delta}$ is a best response for the planner. But then (Dynamic Technology IC) can only hold at $T-\Delta$ if $q_{T-\Delta}=0$. To see this, observe that the net present discounted value of switching to clean technology becomes

$$\begin{aligned}v_{c,T-\Delta}-v_{d,T-\Delta}&=\Psi(\tau_{T-\Delta})+\sum_{s=0}^{\infty}\Psi(\tau_{T+\Delta s})e^{-(\rho+\lambda)(s+1)\Delta}\\&=\Psi\left(p_{T-\Delta}\right)+e^{-(\rho+\lambda)\Delta}\frac{\Psi(\hat{\tau})}{\rho+\lambda}\end{aligned}\tag{B5}$$

using our result on $t\geq T$ to deduce that $\tau_{T+\Delta s}=\hat{\tau}$ for all $s\geq 0$. Therefore by (22), equation (B5) gives

$$v_{c,T-\Delta}-v_{d,T-\Delta}<\frac{\Psi(\hat{\tau})}{\rho+\lambda}=\Gamma$$

using that $\Psi'(\cdot)>0$ and $p_{T-\Delta}<p_T=\hat{\tau}$ (the latter follows from (27) since $S_T-S_{T-\Delta}>0$ and $S_{T'}=S_T=\hat{S}$ for all $T'\geq T$). Hence $q_{T-\Delta}=0$ in equilibrium. An identical argument implies that $q_{T-2\Delta}=0$. Inductively, then, $q_t=0$ for all $t<T$ and $\tau_t\leq p_t$ on $t<T$.

We also note that we cannot show $\tau_t\not<p_t$ for all $t<T$, but the discussion of Remark 2 below gives precise conditions for which indeed $\tau_t\geq p_t$, and therefore $\tau_t=p_t$, for all $t\geq 0$. ■

**Discussion of Remark 2 (Counterexample to Pigovian Taxes).** Consider the Pigovian equilibrium candidate in which $\tau_t=p_t$ for all $t$, and

$$q_t=\begin{cases}0 & \text{for } t<T\\ \hat{q} & \text{for } t\geq T\end{cases}$$

where $T$ is defined by $p_t = \hat{\tau}$, for $\hat{\tau}$ from the unique BGP. Let $(S_t)_{t\geq 0}$ be the pollution stock associated with this Pigovian trajectory.

The net benefit to a deviation $(\tau'_t)_{t\leq T}$ inducing a faster carbon trajectory $S'_t \geq S_t$ on $[0,T]$ and an earlier transition at $t' < T$, equals

$$\begin{aligned}\mathcal{G}(t',\lambda) = &\int_0^T \left[f(k_d(\tau'_t)) - k_d(\tau'_t) - [f(k_d(p_t)) - k_d(p_t)] - D(S'_t) + D(S_t)\right] e^{-(\rho-g)t}dt \\ &+ e^{-(\rho-g)T}\left(e^{(\rho-g)(T-t')} - 1\right)\left(\frac{\lambda(\hat{\pi}_c - \hat{\pi}_d)}{\rho+\lambda} - \Lambda\hat{\tau}\right)\end{aligned} \tag{B6}$$

since the earlier transition creates additional flow profits on $[T',T]$ of

$$\hat{q}\left(f_c(\hat{k}_c) - \hat{k}_c - \left[f_d(\hat{k}_d) - \hat{k}_d\right]\right) + f_d(\hat{k}_d) - \hat{k}_d - f_d(k_d(p_t)) + k_d(p_t)$$

incurs the fixed cost $\Gamma$ at $T'$, and by (22),

$$f_c(\hat{k}_c) - \hat{k}_c - \left[f_d(\hat{k}_d) - \hat{k}_d\right] - \Gamma = \frac{\lambda(\hat{\pi}_c - \hat{\pi}_d)}{\rho+\lambda} - \Lambda\hat{\tau}.$$

Equation (B6) implies that this strategy is preferred if and only if

$$\lambda \geq \tilde{\lambda} = \inf\{\lambda > 0 : \mathcal{G}(t'_0,\lambda) > 0\} \tag{B7}$$

where $t'_0 = \arg\max_{t\leq T}\mathcal{G}(t,\lambda)$. Note that if $\mathcal{G}(t'_0,\lambda) \leq 0$ for all $\lambda$, then $\tilde{\lambda} = +\infty$. In particular, since the integral in (B6) is negative by the definition of $p_t$, it must be that

$$\frac{\tilde{\lambda}(\hat{\pi}_c - \hat{\pi}_d)}{\rho+\tilde{\lambda}} > \Lambda\hat{\tau}$$

verifying the remark. ■

**Proof of Proposition 9 (Uniqueness of BGP).** Using (31), the (farsighted) Pigovian tax given by (32) becomes

$$\hat{\tau} = \frac{1}{\delta+\rho-\alpha\lambda}D'\left(\frac{\hat{q}\gamma k_c(\hat{\tau}) + (1-\hat{q})k_d(\hat{\tau})}{\delta}\right), \tag{B8}$$

which is a decreasing function of $\hat{q}$. Hence there exists at most one $\hat{q} \in (0,1)$ that solves (32) when $\hat{\tau}$ is given by (30). ■

**Proof of Proposition 10 (Dynamic Equilibrium).** We know that $q_t \to \hat{q}$ in finite time by Lemma B1. In addition,

**Lemma B2 (Monotone Pollution)** *In equilibrium, $(S_t)_{t\geq 0}$ is everywhere nondecreasing.*

**Proof of Lemma B2.** Assumption $2'$ implies that $S_0 \leq \hat{S}$. Suppose also that $S_t \leq \hat{S}$ for all $t$ (we prove this below). Now suppose, in order to obtain a contradiction, that $\dot{S}_t < 0$ for some $t$. But

$$\dot{S}_t < 0 \iff \tau_t > \hat{\tau} \geq p_t \tag{B9}$$

because

$$
\begin{aligned}
0 &= \hat{q}k_c(\hat{\tau}) + (1-\hat{q})k_d(\hat{\tau}), \hat{q}) - \delta\hat{S} \\
&\le q_t k_c(\tau_t) + (1-q_t)k_d(\tau_t) - \delta S_t \quad \text{for all } t \le T,
\end{aligned}
\tag{B10}
$$

where the first line follows from the definition of $\hat{\tau}$ and $\hat{q}$, and the second follows from $q_t \le \hat{q}$ (by monotone convergence of $q_t \uparrow \hat{q}$, proven in Lemma B1), $\tau_t \le \hat{\tau}$, and $S_t \le \hat{S}$ by assumption. But (B9) can never be optimal; deviating downwards to $\hat{\tau}$ will strictly improve welfare since $\hat{\tau} \ge p_t$ and $\hat{\tau}$ cannot affect $(q_t)_{t\ge0}$.

Finally, $S_t \uparrow \hat{S}$, and until $T$, at which point $\dot{S}_T = 0$ and the economy is on the BGP. Hence $S_0 \le \hat{S}$ implies $S_t \le \hat{S}$ for all $t$, completing the proof of the lemma. ■

To conclude the proof of Proposition 10, note also that $S_t$ is in fact increasing when either $q_t < \hat{\tau}$ and $\tau_t < \hat{\tau}$ from (B10), and thus we can conclude that there exists $T < \infty$ such that $S_T = \hat{S}$, $p_T = \hat{\tau}$ and $q_T = \hat{q}$, which completes the proof of the proposition. ■

**Proof of Proposition 11 (Second-best Dynamic Policy with Commitment).** Let $T^\star$ denote the first-best switching time (when the planner controls both input decisions and technology choices) and

$$T \ge T^\star$$

denote the equilibrium switching time without commitment specified in Proposition 10.

Observe that $T = T^\star$ (and therefore $\hat{\tau}_t = \tau_t^{SB} = \tau_t^\star = p_t$ for all $t \ge 0$ and $\hat{q} = q^{SB} = q^\star$) if and only if $\lambda = 0$. This follows by comparing the first-order condition for the planner to those of firms. In particular, consider the full Hamiltonian that incorporates the constraint on the evolution of the stock of clean technology, which is

$$\tilde{H}_t = q_t\left[f_c(k_{ct}) - k_{ct}\right] + (1-q_t)\left[f_d(k_{dt}) - k_{dt}\right] - \dot{q}_t\Gamma + p_t\left[q_t\gamma k_{ct} + (1-q_t)k_{dt}\right] + \mathcal{Q}_t\dot{q}_t$$

and note that the first-order conditions $\tilde{H}_{\dot{q}_t} = \mathcal{Q}_t - \Gamma = 0$ and $\dot{\mathcal{Q}}_t = -\tilde{H}_{q_t} + \rho\mathcal{Q}_t$ imply

$$f_c(k_c(\tau_t)) - k_c(\tau_t) - \left[f_d(k_d(\tau_t)) - k_d(\tau_t)\right] + p_t\Lambda(\tau_t) - \rho\Gamma = 0, \tag{B11}$$

with complementary slackness. Now we can see that this coincides with firms' first-order conditions,

$$\int_t^\infty \left[\pi_c(\tau_s) - \pi_d(\tau_s) + \tau_s\Lambda(\tau_s)\right] e^{-(\rho+\lambda)(s-t)} ds \le \Gamma$$

(with equality if $q_t > 0$) when $\tau_t = p_t$ if and only if $\lambda = 0$. This proves the second part of the proposition.

Otherwise, when $\lambda > 0$, (B11) illustrates that the planner prefers a tax policy that induces convergence to $q^{SB} > \hat{q}$, and in particular, she will obtain strictly higher welfare by committing to the strategy $\tau_t^{SB}$ that maximizes welfare subject to (22) at $T^{SB}$, i.e., subject to

$$\int_{T^{SB}}^\infty \Psi(\tau_t^{SB}) e^{-(\rho+\lambda)(t-T^{SB})} dt = \Gamma, \tag{B12}$$

in order to induce a transition $q^{SB} > \hat{q}$ on $[T^{SB}, \infty)$, where $T^{SB} \leq T$. Note that $\tau_t^{SB} \geq p_t^{SB}$ for all $t \geq 0$, since any $\tau_t' < p_t$ will reduce clean technology incentives by $\Psi' > 0$ (Assumption 1′) and reduce welfare by definition of $p_t^{SB}$. Finally, Assumption 2′ implies that $\tau_t^{SB} > p_t^{SB}$ for some $t \in [T^{SB}, \infty)$ if (B12) holds, completing the proof of the proposition. ∎

**Proof of Proposition 12 (Dynamic Geoengineering Type I).** (*Taxes, clean technology*). Recall that Type I geoengineering corresponds to an increase in $\xi$, so that the damages function now equals $D(S - \xi)$. Differentiating the BGP equilibrium tax

$$\hat{\tau} = \frac{D'(S(\hat{\tau}, \hat{q}) - \xi)}{\rho + \delta - \alpha\lambda}.$$

with respect to $\xi$, we obtain

$$\frac{d\hat{\tau}}{d\xi} = \frac{D''(\cdot)}{\rho - g + \delta}\left[\frac{\partial S}{\partial \tau}\frac{d\hat{\tau}}{d\xi} - \frac{\Lambda}{\delta}\frac{dq}{d\xi} - 1\right] + \frac{\alpha D'(\cdot)}{(\rho - g + \delta)^2}\frac{d\lambda}{d\xi} \tag{B13}$$

Noting that $d\lambda/d\xi = 0$ because the supply of scientists is fixed, and $d\hat{\tau}/d\xi = 0$, equation (B13) implies that

$$\frac{d\hat{q}}{d\xi} = -\frac{\Lambda}{\delta} < 0 \quad \text{and} \quad \frac{d\hat{\tau}}{d\xi} = 0.$$

(*Welfare*). The derivative of flow utility from production with respect to $\hat{q}$ is

$$f_c(\hat{k}_c) - \hat{k}_c - (f_d(\hat{k}_d) - \hat{k}_d) - \rho\Gamma$$

which we can write as $\lambda\Gamma - \Lambda\hat{\tau}$, or equivalently

$$\frac{\lambda}{\lambda + \rho}(\pi_c - \pi_d) - \Lambda\hat{\tau}$$

since $\frac{\lambda}{\lambda+\rho}(\pi_c - \pi_d) = \lambda\Gamma$ via rearranging the condition that $\pi_c - \pi_d = (\rho + \lambda)\Gamma$. Total flow environmental damages respond as

$$\frac{d}{d\xi}D(\hat{S} - \xi) = -D'(\cdot) + \frac{d\hat{S}}{d\xi}D'(\cdot) = 0$$

using $d\hat{S}/d\xi = 1$. ∎

**Proof of Proposition 13 (Dynamic Geoengineering Type II).** Now the total derivative of the stationary farsighted Pigovian tax, $\hat{\tau} = (1 - \upsilon)(\rho - g + \delta)^{-1}D'(\hat{S})$, equals

$$\frac{d\hat{\tau}}{d\upsilon} = -\frac{D'(\hat{S})}{\rho - g + \delta} + (1 - \upsilon)\frac{D''(\hat{S})}{\rho - g + \delta}\left[\frac{\partial S}{\partial \tau}\frac{d\hat{\tau}}{d\upsilon} - \frac{\Lambda}{\delta}\frac{d\hat{q}}{d\upsilon}\right] + \frac{\alpha D'(\hat{S})}{(\rho - g + \delta)^2}\frac{d\lambda}{d\upsilon}$$

which, with $d\hat{\tau}/d\upsilon = 0$ and $d\lambda/d\upsilon = 0$, implies that

$$\frac{d\hat{q}}{d\upsilon} = -\frac{1}{1 - \upsilon}\frac{\delta\hat{S}}{\Lambda}\frac{1}{\eta} \tag{B14}$$

The total effect on the stock of carbon equals

$$(1-\upsilon)\frac{d\hat{S}}{d\upsilon} = \frac{D'(\hat{S})}{D''(\hat{S})} = \frac{\hat{S}}{\eta} \tag{B15}$$

so that the welfare-relevant flow term equals

$$-(1-\upsilon)\frac{d\hat{S}}{d\upsilon}D'(\hat{S}) + D(\hat{S}) = D(S) - \frac{\hat{S}D'(\hat{S})}{\eta}$$

Using $d\hat{\tau}/d\upsilon = 0$ and $d\hat{q}/d\upsilon$ from (B14), and using the fact that flow output shifts as

$$\left(\frac{\lambda}{\rho+\lambda}(\hat{\pi}_c - \hat{\pi}_d) - \Lambda\hat{\tau}\right)\frac{d\hat{q}}{d\upsilon},$$

then

$$-\Lambda\tau\frac{d\hat{q}}{d\upsilon} = \frac{\delta}{\rho-g+\delta}\frac{SD'(S)}{\eta}$$

which we can combine with the previous expression, dividing by $\hat{S}D'(\hat{S})$,

$$\frac{dW}{d\upsilon} < 0 \iff -\frac{\delta}{\eta}\left((\rho-g+\delta)\frac{\lambda}{\rho+\lambda}\frac{\hat{\pi}_c-\hat{\pi}_d}{\Lambda\hat{\tau}}\right) - \frac{\delta}{\rho-g+\delta}\frac{1}{\eta} + \frac{1}{a} < 0$$

with $a \equiv \hat{S}D'(\hat{S})/D(\hat{S}) > 1$ as in the static case. This reduces to $dW/d\upsilon < 0$ if and only if

$$\eta < \eta^{\mathrm{II}}(\lambda) \equiv a\left(\delta(\rho-g+\delta)\frac{\lambda}{r+\lambda}\frac{\hat{\pi}_c-\hat{\pi}_d}{\Lambda\hat{\tau}} + \frac{\delta}{\rho-g+\delta}\right)$$

which yields the proposition. ■

## B2 Alternative Specification of Environmental Damages

Abbreviating output (net of the fixed costs of clean technology) by

$$Y(k,q) = q(f_c(k_c) - k_c - \Gamma) - (1-q)(f_d(k_d) - k_d),$$

we let welfare equal

$$[1-D(S)]Y(k,q) \tag{B16}$$

where $D(\cdot)$ is the same increasing, convex, and twice continuously differentiable damage function as before. Firms of type $j \in \{c,d\}$ access production technologies

$$[1-D(S)](f_j(k)-k)$$

and the fixed cost of switching is $[1-D(S)]\Gamma$.

**Lemma B3 (Pigovian Taxation)** *In economy B.2, optimal taxation satisfies*

$$\hat{\tau} = D'(S)Y(k,q). \tag{B17}$$

**Proof.** Differentiate (B16) with respect to $k$ to obtain

$$[1 - D(S)][q(f_c'(k_c) - 1) + (1 - q)(f_d'(k_d) - 1)] - Y(k, q)D'(S)(1 - q + q\gamma) = 0.$$

At a per-unit carbon tax $\tau$, each firm maximizes its profits $[1 - D(S)][f_j(k) - k] - \tau\gamma_j k$ by setting

$$[1 - D(S)][f_c'(k_c) - 1] = \gamma\tau \quad \text{and} \quad [1 - D(S)][f_d'(k_d) - 1] = \tau$$

so (B17) will implement the Pigovian allocation. ■

**Lemma B4 (Equilibrium Technology)** *If*

$$\frac{\lambda}{1 - \lambda}\Gamma < \Lambda\left(\frac{D''(S)}{D'(S)} + D'(S)\right), \tag{B18}$$

*then the interior equilibrium $(\hat{\tau}, \hat{q})$ in economy B.2 is unique.*

**Proof.** The proof of Proposition 2 (the interior equilibrium) carries through if $\hat{\tau}$ and $q$ are still strict strategic substitutes. Indeed, $\Psi'(\tau) > 0$ as before, so the equilibrium IC is unique. However, there may exist multiple $q$'s that satisfy this, since the second term in

$$\frac{\partial\hat{\tau}}{\partial q} = -\Lambda D''(S) + \frac{\partial Y}{\partial q}D'(S)$$

can make $\tau(q)$ not everywhere decreasing in $q$. However, as

$$\frac{\partial Y}{\partial q} = \frac{\lambda}{1 - \lambda}\Gamma - \Lambda\tau$$

in the interior, assuming (B18) will guarantee that $\tau'(q) < 0$. ■

What happens after geoengineering? When environmental damages affect production directly, geoengineering's effect on the stock of carbon is still weak, but is no longer exactly zero:

**Proposition B1 (Geoengineering Type I)** *Suppose that Assumptions 1 and 2 hold. In economy B.2, consider a geoengineering technology improvement of type I that increases $\xi$ by a small amount $d\xi$. Then we have*

- $d\hat{\tau}/d\xi = 0$ *(there is no effect on the equilibrium carbon tax).*
- $d\hat{q}/d\xi = -d\xi/\Lambda(\hat{\tau})$ *(investment in clean technology declines).*
- $dE/d\xi = 1 > 0$ *(emissions increase, through lower $\hat{q}$).*
- $dD/d\xi = 0$ *(environmental damages remain constant).*
- $dW/d\xi < 0$ *if and only if $\lambda(\pi_c - \pi_d) > \Lambda\tau$ (welfare may decline).*

**Proof.** The incentive-compatible tax satisfies

$$\hat{\tau} = YD'(S - \xi),$$

so totally differentiating with respect to $\xi$ and using that $d\hat{\tau}/d\xi = 0$ in the interior yields

$$0 = \frac{dY}{d\xi}D'(\cdot) + \frac{dS}{d\xi}YD''(\cdot) - YD''(\cdot)$$

or an increase in the stock of carbon (through $\hat{q}$) by

$$\frac{dS}{d\xi} = 1 - \frac{D'(\cdot)}{YD''(\cdot)}\frac{dY}{d\xi}, \tag{B19}$$

which differs from our benchmark case by the presence of the second term that depends on $dY/d\xi$.

Welfare, which equals $(1 - D(\cdot))Y$, will shift as

$$\frac{dW}{d\xi} = (1 - D(\cdot))\frac{dY}{d\xi} - Y\frac{dD(\cdot)}{d\xi} + YD'(\cdot).$$

Noting that $dD/d\xi = [dS/d\xi]D'(\cdot)$, and using (B19), the direct geoengineering gain $YD'(\cdot)$ cancels with the first term of $dS/d\xi$ and the total effect on welfare reduces to

$$\frac{dW}{d\xi} = \left(1 - D(\cdot) + \frac{(D'(\cdot))^2}{D''(\cdot)}\right)\frac{dY}{d\xi},$$

and it is apparent that

$$\frac{dW}{d\xi} > 0 \iff \frac{dY}{d\xi} = (\lambda(\pi_c - \pi_d) - \Lambda\tau) > 0,$$

which is the same condition in our benchmark economy. ■

The intuition is that exactly as before, the (IC) constraint still pins down $\hat{\tau}$, so $d\hat{\tau} = 0$. To sustain the incentive-compatible $\hat{\tau}$ after an exogenous removal of $d\upsilon > 0$ carbon from the initial stock, emissions $E$ need to increase to sustain the previous level of $S$, through smaller $q$. However, with fewer clean firms, the economy incurs lower fixed costs, increasing output. On the other hand, the optimal tax formula (B17) increases in output, which helps to increase the post-geoengineering Pigovian price back to $\hat{\tau}$. Hence emissions need not respond by "as much," i.e., $dE \in (0, d\xi)$.

A type II modification will require a lower $q$ to increase emissions and sustain the incentive-compatible $\hat{\tau}$, as exactly as in proposition 4. However, lowering $q$ will affect total producer surplus, which as in the type I comparative static will affect the tax level through the income effect. Our analogue to Proposition 5 is therefore:

**Proposition B2 (Geoengineering Type II in Economy B2)** *Suppose that Assumptions 1 and 2 hold. In economy B2, consider a geoengineering technology improvement of type II with* $\upsilon \in [0, 1)$*, and let* $\eta = SD''(S)/D'(S)$ *be the elasticity of the marginal damage function. Then we have*

- *$d\hat{\tau}/d\upsilon = 0$ (there is no effect on the equilibrium carbon tax).*
- *$dS/d\upsilon > 0$ (the total stock of carbon increases).*
- *$d\hat{q}/d\upsilon < 0$ (investment in clean technology declines).*
- *$dE/d\upsilon > 0$ (emissions increase, through lower $\hat{q}$).*
- *$dD/d\upsilon > 0$ if and only if $\eta \leq \eta^*$ (environmental damage increases if the damage function is not too convex), where $\eta^* \geq 1$.*
- *$dW/d\upsilon < 0$ if and only if $\eta \leq \eta_B^{\mathrm{II}}(\lambda)$ defined below (welfare may decline).*

**Proof.** As before, $d\hat{\tau}/d\upsilon = 0$. Differentiating

$$\hat{\tau} = (1-\upsilon)YD'(\cdot),$$

we obtain

$$0 = \frac{dY}{d\upsilon}(1-\upsilon)D'(S) - YD'(S) + (1-\upsilon)Y\frac{dS}{d\upsilon}D''(S),$$

from which we conclude that

$$\frac{dS}{d\upsilon} = \frac{D'(S)/D''(S)}{1-\upsilon} - \frac{D'(S)}{YD''(S)}\frac{dY}{d\upsilon}. \tag{B20}$$

Welfare satisfies $W = [1-(1-\upsilon)D(S)]Y$, so

$$\frac{dW}{d\upsilon} = \frac{dY}{d\upsilon}(1-\upsilon)(1-D(S)) + YD(S) - (1-\upsilon)Y\frac{dD(S)}{d\upsilon},$$

and using that

$$\frac{dD}{d\upsilon} = \frac{dS}{d\upsilon}D'(S)$$

and our formula for $dS/d\upsilon$ in (B20), we obtain

$$\frac{dW}{d\upsilon} = \left(1-(1-\upsilon)D + (1-\upsilon)\frac{(D'(S))^2}{D''(S)}\right)\frac{dY}{d\upsilon} + YD(S) - Y\frac{1}{\eta}SD'(S)$$

so that the condition for environmental damages is the same as in the benchmark economy, and the threshold for welfare to decline becomes

$$\eta < \eta_B^{\mathrm{II}}(\lambda) \equiv \left(a + \frac{1-(1-\upsilon)D(S)}{SD'(S)}\frac{dY}{d\upsilon}\right)^{-1}\left(1+(1-\upsilon)\frac{dY}{d\upsilon}\frac{1}{Y}\right)$$

where $dY/d\upsilon = [\lambda(\pi_c - \pi_d) - \Lambda\tau][d\hat{q}/d\upsilon]$, and $a \equiv SD'(S)/D(S) > 1$ as before. ■

## B3 Another Alternative Environmental Damage Specification

We next briefly discuss the robustness of our results to a specification in which environmental damages directly affect consumption preferences. In particular, suppose now there exists a representative household who consumes all of the output in each period,

$$C = q(f_c(k_c) - k_c - \Gamma) + (1-q)(f_d(k_d) - k_d) \tag{B21}$$

and garners increasing, concave, and differentiable utility

$$U\left(\frac{C}{D(S)}\right). \tag{B22}$$

Differentiating (B22) with respect to $k_j$, we obtain the first-order conditions

$$U'(\cdot)\frac{f_j'(k_j) - 1}{D(S)} - U'(\cdot)\gamma_j D'(S)\frac{C}{[D(S)]^2} = 0$$

for $j \in \{c, d\}$, which can be implemented as a decentralized equilibrium with a per-unit-carbon tax of

$$\hat{\tau} = \frac{D'(S)}{D(S)}C. \tag{B23}$$

Observe that if we transform damages into $D(S) = \exp \tilde{D}(S)$, so that $D'(S)/D(S) = \tilde{D}'(S)$, then expression (B23) is exactly the tax as in economy B2. Moreover, since $U'$ will not alter the sign of $dW$, all of our results from economy B2 go through, except with the modified damage function. Noting that in particular, the elasticity of $D(S) = \exp \tilde{D}(S)$ satisfies

$$\eta = SD'(S) + \frac{D''(S)}{D'(S)}S,$$

the exponential transformation increases the curvature of our damages function.

## B4 Concave Preferences

In the text, we focused on linear preferences, even in the dynamic model, which greatly simplified the analysis. We now show that our results generalize when consumers have a concave utility function, so that their dynamic preferences are given at each $t$ by

$$\int_t^\infty U\left(C_s - A_s D(S_s; \xi, \upsilon)\right) e^{-\rho(s-t)} ds,$$

where the utility function $U(\cdot)$ is increasing, twice continuously differentiable, and concave. The specification where damages and consumption are additive arguments of $U(\cdot)$ is similar to Greenwood *et al.* (1988). Our results can also be extended to different formulations, but those introduce additional income effects, further complicating the relevant conditions. In addition, to sustain a BGP, we also assume a constant elasticity of marginal consumption (relative risk aversion), $\theta = -xU''(x)/U'(x)$, so that $U$ can be represented by

$$U(x) = \frac{x^{1-\theta}}{1-\theta}$$

for $\theta \neq 1$ and the limit $U(x) = \log(x)$ if $\theta = 1$ (and $\theta = 0$ recovers the linear specification of the main text).

The household thus maximizes (B4)

$$\frac{1}{1-\theta}\int_t^\infty \left(\tilde{C}_s\right)^{1-\theta} e^{-\rho(s-t)} ds,$$

at each time $t \geq 0$, where we define net consumption as $\tilde{C}_s = C_s - A_s D(S_s; \xi, \upsilon)$.

### B4.1 Household Optimization

The Ramsey equation (derived from household intertemporal optimization with endogenous savings) yields a market interest rate of

$$r_t = \rho + \theta g_t \tag{B24}$$

where $g_t \equiv \dot{\tilde{C}}_t / \tilde{C}_t$ denotes the growth rate of average net consumption, defined as $\tilde{C}_t = C_t - A_t D(S_t; \xi, \upsilon)$. We derive (B24) in Appendix B4.7.

### B4.2 Firms

Now that the interest rate differs in general from $\rho$, we must define firms' value functions slightly differently. Equation (20) for the value of a clean incumbent with quality $n$ at time $t$ becomes

$$\begin{aligned} V_{ct}(n) &= \mathbb{E}_t\left[A^n \int_t^\infty \pi_c(\tau_s) e^{-r_s(s-t)} \mathbf{1}_{\{\text{extant at } s\}} ds\right] \\ &= A^n \int_t^\infty \pi_c(\tau_s) e^{-(r_s+\lambda)(s-t)} ds, \end{aligned} \tag{B25}$$

Likewise, the expected net present discounted value at $t$ for a $d$-type firm with quality $n$ becomes

$$V_{dt}(n) = \sup_{\nu \geq 0}\left[\left[V_{c,t+\nu}(n) - A^n\Gamma\right] e^{-\int_t^{t+\nu}(r_s+\lambda)ds} + A^n \int_t^{t+\nu} \pi_d(\tau_s) e^{-(r_s+\lambda)(s-t)} ds\right] \tag{B26}$$

since they can switch at any time $t + \nu \geq t$. With these new value functions, and their normalized counterparts $v_{ct}$ and $v_{dt}$, R&D, input and clean technology switching decisions all go through just as in the benchmark case.

### B4.3 Optimal Taxation

The planner maximizes household utility, (B4). The Hamiltonian which generalizes (24) of the main text is

$$H_t(k_t, S_t) = U\left(\tilde{C}_t\right) - \mu_t\left[E_t - \delta S_t\right].$$

Optimal input decisions $k_{ct}, k_{dt}$ satisfy $\partial H_t/\partial k_{ct} = \partial H_t/\partial k_{dt} = 0$, or

$$A_t U'(\tilde{C}_t)\left[f_d'(k_{dt}) - 1\right] = \mu_t \quad \text{and} \quad A_t U'(\tilde{C}_t)\left[f_c'(k_{ct}) - 1\right] = \gamma\mu_t,$$

which coincide with private firm input decisions when the planner levies a per-unit emissions tax of $\mu_t/\left(A_t U'(\tilde{C}_t)\right)$. Consequently, we use the normalization

$$p_t = \frac{\mu_t}{A_t U'(\tilde{C}_t)},$$

which differs from the one in the main text only by the presence of the non-constant marginal utilities. Along the equilibrium path, the shadow price of the stock of carbon $\mu_t$ will satisfy the Euler-Lagrange condition $\partial H_t/\partial S = \dot{\mu}_t - \rho\mu_t$. The Pigovian shadow cost of carbon emissions is identical to that of the main text, except with an endogenous interest rate $(r_t)_{t\geq 0}$. However, we can weaken Assumption 3 of the main text to the following:

**Assumption 3′ (Growth with Concave Preferences)** $(1-\theta)\alpha\varphi Z \leq \rho + \delta$.

Note that Assumption 3′ always holds if $\theta \geq 1$, regardless of the other parameter values. Consequently,

**Lemma B5 (Shadow Cost of Carbon with Concave Utility)** *Suppose Assumption 3′ holds. Then, along any optimal path,*

$$p_t = \int_t^\infty D'(S_s)e^{-(\delta + r_s - \alpha\lambda)(s-t)}ds, \tag{B27}$$

*for all $t \geq 0$, where $(r_s)_{s\geq 0}$ denotes the Ramsey interest rate given by (B24).*

**Proof.** See Appendix B4.7. ■

### B4.4 Dynamic Equilibrium

We define an equilibrium in the concave economy exactly as in the text in section 3.5, except appended with the natural condition that

- Given each of the other equilibrium objects, the interest rate $r_t^*$ satisfies (B24).

Moreover, to be precise, we need to rewrite our assumption for interiority such that it takes into account the dependence of farsighted Pigovian taxes on the interest rate.

**Assumption 2″ (Conditions for Dynamic Interior Equilibrium)** *For all $t \geq 0$,*

$$\Gamma \in \left( \int_t^\infty \Psi(\underline{\tau}_s) e^{-(r_s+\lambda)(s-t)} ds, \; \int_t^\infty \Psi(\overline{\tau}_s) e^{-(r_s+\lambda)(s-t)} ds \right)$$

*where*

$$\overline{\tau}_t = \int_t^\infty D' \left( S_0 e^{-\delta s} + \int_0^s k_d(\overline{\tau}_\nu) e^{-\delta(s-\nu)} d\nu \right) e^{-(\delta + r_s - \alpha\lambda)(s-t)} ds$$

*and*

$$\underline{\tau}_t = \int_t^\infty D' \left( S_0 e^{-\delta s} + \int_0^t k_d(\underline{\tau}_\nu) e^{-\delta(t-\nu)} d\nu + \gamma \int_t^s k_c(\underline{\tau}_\nu) e^{-\delta(s-\nu)} d\nu \right) e^{-(\delta + r_s - \alpha\lambda)(s-t)} ds,$$

*and $(r_s)_{s\geq 0}$ satisfies (B24).*

### B4.5 BGP

Just as before, if the stock of carbon converges to some $\hat{S}$, (31) will hold just as before for all subsequent $t$ hence. Hence growth reduces to as before

$$g = (A-1)\lambda$$

and the interest rate (B24) simplifies to $\hat{r} = \rho + \theta\alpha\lambda$. The BGP farsighted Pigovian condition (32) becomes

$$\hat{\tau} = \frac{D'(\hat{S})}{\delta + \rho - (1-\theta)\alpha\lambda}, \tag{B28}$$

equation (22) becomes

$$\frac{\Psi(\hat{\tau})}{\rho + (1+\theta\alpha)\lambda} = \Gamma, \tag{B29}$$

and we obtain an analogue to Proposition 9.

**Proposition B3 (Existence, Uniqueness of the BGP)** *Suppose Assumptions 1′, 2″, and 3′ hold. Then there exists a unique interior solution $(\hat{S}, \hat{q}, \hat{\tau})$ to (31), (B28), and (B29).*

**Proof.** Omitted. ■

### B4.6 Geoengineering with Concave Preferences

We conclude by noting that the versions of the dynamic geoengineering propositions in the main text, augmented to account for the endogenous interest rate dynamics, are substantively identical.

### B4.7 Proofs omitted above

**Proof of the Ramsey equation (B24).** Suppose that our household can save at a rate $r_t$, with savings $a_t$, and costate $s(t)$. Then

$$\dot{a}_t = r_t a_t - A_t c_t + \int_0^1 A^{n_{it}}(f(k_{it}) - k_{it})di,$$

where recall $c_t = C_t/A_t$. Denoting normalized net consumption by $\tilde{c}_t = c_t - D(S_t)$, the household's Hamiltonian may be written as

$$H = \frac{1}{1-\theta}A_t^{1-\theta}(\tilde{c}_t)^{1-\theta} + s(t)\left[r_t a_t - A^{n_t}c_t + \int_0^1 A^{n_{it}}(f(k_{it}) - k_{it})di\right]$$

with discount rate $\rho$, or rather, since

$$A_t^{1-\theta} = \exp\{(1-\theta)(A-1)t\lambda\},$$

we can write the transformed

$$\tilde{H} = \frac{\tilde{c}_t^{1-\theta}}{1-\theta} + s(t)\left[r_t a_t - A_t c_t + \int_0^1 A^{n_{it}}(f(k_{it}) - k_{it})di\right]$$

with a transformed discount rate $\tilde{\rho} \equiv \rho + (1-A)(1-\theta)\lambda$. The first-order savings conditions of $\tilde{H}$ are

$$H_{c_t} = \tilde{c}_t^{-\theta} - A_t s(t) = 0 \tag{B30}$$

$$\dot{s}(t) = -H_{a_t} + \tilde{\rho}s(t). \tag{B31}$$

Differentiating (B30) with respect to time, we obtain

$$-\theta(\tilde{c}_t)^{-1-\theta}\dot{\tilde{c}}_t = A_t\dot{s}(t) + (A-1)\lambda A_t s(t)$$

which we divide by $A_t s(t)$ to obtain via (B30)

$$\frac{\dot{s}(t)}{s(t)} = -\theta\frac{\dot{\tilde{c}}_t}{\tilde{c}_t} - (A-1)\lambda$$

so that, using $H_{a_t} = r_t$, the definition of $\tilde{\rho}$, and the differential equation (B31), we obtain

$$-\theta\frac{\dot{\tilde{c}}_t}{\tilde{c}_t} - (A-1)\lambda = -r_t + \rho + (1-\theta)(1-A)\lambda.$$

Hence

$$r_t = \rho + \theta\frac{\dot{\tilde{c}}_t}{\tilde{c}_t} + \theta\alpha\lambda \tag{B32}$$

is the equilibrium interest rate. Noticing that

$$\frac{\dot{\tilde{c}}_t}{\tilde{c}_t} = \frac{\dot{\tilde{C}}_t}{\tilde{C}_t} - (A-1)\lambda$$

yields equation (B24). ■

**Proof of Lemma B5 (Shadow Cost of Carbon in GHH).** In equilibrium, we can write the necessary Euler-Lagrange condition $\partial H_t/\partial S = \dot{\mu}_t - \rho\mu_t$ as

$$\dot{\mu}_t = -A_t U'(\tilde{C}_t) D'(S_t) + (\rho + \delta)\mu_t.$$

Dividing this equation as before by $A_t U'(\tilde{C}_t)$, we obtain

$$\frac{\dot{\mu}_t}{A_t U'(\tilde{C}_t)} = -D'(S_t) + (\rho + \delta) p_t. \tag{B33}$$

Recall our change-of-variables

$$p_t = \frac{\mu_t}{A_t U'(\tilde{C}_t)}$$

and note in particular that it satisfies

$$\begin{aligned}\dot{p}_t &= \frac{\dot{\mu}_t}{A_t U'(\tilde{C}_t)} - \frac{\mu_t}{\left(A_t U'(\tilde{C}_t)\right)^2} \frac{d}{dt}\left[A_t U'(\tilde{C}_t)\right] \\ &= \frac{\dot{\mu}_t}{A_t U'(\tilde{C}_t)} - p_t \left(\theta \frac{\dot{\tilde{c}}_t}{\tilde{c}_t} + (1 - A)(1 - \theta)\lambda\right)\end{aligned}$$

using

$$\frac{\frac{d}{dt}\left[A_t U'(\tilde{C}_t)\right]}{A_t U'(\tilde{C}_t)} = \frac{\frac{d}{dt} A_t^{1-\theta}}{A_t^{1-\theta}} + \frac{\frac{d}{dt}\tilde{c}_t^{-\theta}}{\tilde{c}_t^{-\theta}},$$

and $A_t = \exp\{t\lambda(A - 1)\}$. With (B32), equation (B33) becomes

$$\dot{p}_t = -D'(S_t) + (\delta + r_t - \alpha\lambda) p_t,$$

which with $p_0$ determined by transversality gives (B27). ■