

INTERNATIONAL CLIMATE AGREEMENTS AND THE SCREAM OF GRETA^{*}

Giovanni Maggi[†]

Robert W. Staiger[‡]

November 2021
Preliminary and Incomplete

Abstract

The world appears to be facing imminent peril, as countries are not doing enough to keep the Earth's temperature from rising to catastrophic levels and various attempts at international cooperation have failed. Why is this problem so intractable? Can we expect an 11th-hour solution? Will some countries, or even all, succumb on the equilibrium path? We address these questions through a model that features the possibility of climate catastrophe and emphasizes the role of international externalities that a country's policies exert on other countries and intertemporal externalities that current generations exert on future generations. Within this setting, we explore the extent to which international agreements can mitigate the problem of climate change. Our analysis illuminates the role that international climate agreements can be expected to play in addressing climate change, and it points to important limitations on what such agreements can achieve, even under the best of circumstances.

^{*}We thank for helpful comments and discussions Klaus Demset, Allan Hsiao, Emanuel Ornelas, Steve Redding, participants at the April 9-10 2021 NBER Future of Globalization conference and the June 9 2021 International Trade and Environmental Policy workshop, and seminar participants at Bocconi University, Florida State University, Princeton University, Singapore Management University, Syracuse University, the University of Oslo and Yale University. Winston Chen, Wei Xiang and Yan Yan provided outstanding research assistance.

[†]Department of Economics, Yale University; Graduate School of Economics, FGV-EPGE; and NBER.

[‡]Department of Economics, Dartmouth College; and NBER.

“Many perceive global warming as a sort of moral and economic debt, accumulated since the beginning of the Industrial Revolution and now come due after several centuries. In fact, ... [t]he story of the industrial world’s kamikaze mission is the story of a single lifetime – the planet brought from seeming stability to the brink of catastrophe in the years between a baptism or bar mitzvah and a funeral.”

– from David Wallace-Wells, **The Uninhabitable Earth**, 2019 page 4.

1. Introduction

The world appears to be facing imminent peril from climate change. According to the Intergovernmental Panel on Climate Change (IPCC), the costs of climate change will begin to rise to catastrophic levels if warming is allowed to surpass 1.5 degrees Celsius, and countries are not doing enough to keep the Earth’s temperature from rising beyond this level: by many accounts the world is on track to warm by almost 3 degrees Celsius by the end of the century.¹ Yet according to one estimate (Jenkins, 2014), most Americans would be unwilling to pay more than \$200 a year in support of energy-conserving policies, an amount that is “woefully short of the investment required to keep warming under catastrophic rates” (Zaki, 2019).² And various attempts at international cooperation, such as the Kyoto Protocol and the Paris Agreement on Climate Change, have also fallen short. Why is this problem so intractable? Can we expect an 11th-hour solution? Will some countries, or even all, succumb on the equilibrium path?

In this paper we address these questions through a formal model that features the possibility of climate catastrophe and emphasizes two critical issues with which efforts to address climate change must contend: the international externalities that a country’s policies exert on other countries, and the intertemporal externalities that current generations exert on future generations. We explore the problems that arise when countries act noncooperatively in this setting, and the extent to which international climate agreements can mitigate these problems.

Previous research has highlighted two challenges that a climate agreement must meet, relating to country participation and enforcement.³ In this paper we abstract from these well-studied

¹See, for example, the assessment by Climate Action Tracker at <https://climateactiontracker.org/>.

²And arguably, the policies chosen by U.S. administrations have fallen short of even this low level of the willingness of Americans to pay for such policies.

³See for example Barrett, 1994, Harstad, 2012, Nordhaus, 2015, Battaglini and Harstad, 2016 and Harstad, 2020 on the former, and Maggi, 2016 and Barrett and Dannenberg, 2018 on the latter.

challenges, and focus instead on a limitation that has not been emphasized in the formal literature on climate agreements. This limitation arises from the fact that it is not possible for a climate agreement to include *future* generations in the bargain alongside current generations. Hence, while a climate agreement can in principle address the “horizontal” externalities that arise from the international aspects of emissions choices, it cannot address the “vertical” externalities exerted by a generation’s emissions choices on future generations, nor can it address the “diagonal” externalities exerted by a country’s current climate policy on future generations in other countries. A key objective of our analysis is to examine the consequences of this limitation of climate agreements in a world where catastrophic outcomes are possible.

We work with a model world economy in which the successive generations of each country make their consumption decisions either unilaterally or within the context of an international climate agreement (ICA), and where utility is derived from consumption and from the quality of the environment. These two dimensions of utility are in tension, as consumption generates carbon emissions, which add to the global carbon stock and degrade the quality of the environment through a warming climate. This tension defines the fundamental tradeoff faced by each generation. In our core analysis we focus on a world without intergenerational altruism, and then later introduce the possibility that each generation cares about its offspring.

When born, a generation inherits the global carbon stock that was determined by the cumulative consumption decisions of the previous generations. As the carbon stock rises, the climate warms and the utility derived by the current generation from the quality of the environment falls commensurately, at least for moderate levels of warming. But if the carbon stock gets too high, the implications are catastrophic: the generation alive at the brink faces the prospect that life could go from livable to essentially unlivable in their lifetime.

We consider two possibilities for climate catastrophes. In our common-brink model, all countries are brought to the brink of climate catastrophe at the same moment, when the global carbon stock reaches a critical level. In our heterogeneous-brink model, more vulnerable countries reach the brink first. As we demonstrate, these two possibilities carry starkly different implications for outcomes along the equilibrium path and for the potential role of an ICA.

We begin with an analysis of the common-brink setting. In the absence of an ICA, we show that the equilibrium path in this setting exhibits an initial warming phase, during which each country’s emissions are constant at a “Business-As-Usual” (BAU) level. During this phase, the climate externalities imposed by the emissions choices of a given generation in a given country

on all other countries and on future generations everywhere are left unaddressed, the global stock of carbon rises suboptimally fast, and the implied degradation of the environment erodes the utility of each successive generation, until the world is brought to the brink of catastrophe. Once the brink is reached, however, the brink generation overcomes all of these externalities and averts catastrophe with an 11th hour solution that has each country doing its part to halt further climate change. The solution involves reduced worldwide emissions levels that keep the carbon stock constant given the natural rate of atmospheric regeneration and remain at that level for all generations thereafter, and it implies a discrete drop in utility for the brink generation and all future generations. The reason for this 11th hour noncooperative solution is that, while earlier generations face rising costs of global warming as their emissions contribute to a growing global stock of carbon, it is only the brink generation that faces the catastrophic implications of continuing the emissions practices of the past. And in the face of this clear and present danger, the nature of the game is fundamentally altered, with the result that the brink generation “does whatever it takes” in the noncooperative equilibrium to avoid catastrophe.

The noncooperative equilibrium of our common-brink model therefore delivers a good news/bad news message: the good news is that, while it takes a crisis to shake the world from business-as-usual behavior, when the crisis arrives the world will find a way to save itself from going over the brink; the bad news is that the world that is saved on the brink is not likely to be a nice world in which to live, both because the climate at the brink of catastrophe may be very unpleasant, and because the generation that comes of age at the brink and all future generations must accept a discrete drop in consumption and utility relative to previous generations in order prevent the climate from worsening further and resulting in global annihilation. Hence, the brink generation, once born, has an especially strong reason to regret that previous generations did not do more to address climate change.

We next ask: What is the role for an ICA to improve over the noncooperative equilibrium in our common-brink setting? One point is immediately clear: it is *not* the desire to avert a climate catastrophe that generates a role for an ICA, because once the catastrophe is at hand countries have sufficient incentives to avoid catastrophe even without an ICA. Rather, we argue that the only way that an ICA can improve over the noncooperative outcome in this setting is to internalize the international climate externalities *during the warming phase*, and thereby reduce the global carbon stock and improve the quality of the environment during this phase, and postpone – possibly forever – the world’s arrival at the brink.

A remaining question is how the outcome achieved by an ICA compares with the outcome that would be implemented by a global social planner. While the ICA negotiated by each generation in effect picks an extreme point on the Pareto frontier that places zero weight on the utility of future generations directly, we focus on the possibility that the global social planner might instead place strictly positive weight on future generations directly, and hence takes into account not only the horizontal but also the vertical and diagonal climate externalities. We show that, while an ICA slows down the growth of the carbon stock relative to the noncooperative outcome, it does not do so enough relative to the choices of the global social planner. This leads to three possible scenarios, depending on the severity of the constraint imposed by the catastrophe threshold. If this constraint is sufficiently mild, the ICA will prevent the world from reaching the brink of catastrophe, but the steady-state carbon stock is still too large relative to the social planner outcome; if the constraint is sufficiently severe, the brink will be reached both under the ICA and the social planner, but it is reached at an earlier date under the ICA; and in between these two cases, the world reaches the brink of catastrophe under the ICA but not under the choices of the social planner. It is when the carbon stock constraint lies in this third, intermediate, range that the inability of the ICA to take into account directly the interests of future generations has its most profound impact: while a global social planner would keep the world from ever arriving at the brink of climate catastrophe, an ICA will at best only postpone the arrival at the brink, and when that day arrives, the brink generation and all generations thereafter will suffer a discrete drop in welfare.

We then turn to the heterogeneous-brink setting, where countries face catastrophe at different levels of the global carbon stock. We assume that if a country were to collapse, its citizens would become climate refugees and suffer a utility cost themselves while also imposing “refugee externality costs” on the remaining countries who receive them. Along the noncooperative path the world may now pass through three possible phases: a warming phase, where warming takes place but no catastrophes occur; a catastrophe phase, where warming continues and a sequence of countries collapse; and a third phase where warming and catastrophes are brought to a halt. The first and third phases are familiar from the common-brink model; the possibility of a middle phase in which some countries collapse along the noncooperative path is novel to the heterogeneous-brink model. We show that under mild conditions the world will indeed traverse through all three phases of climate change along the equilibrium noncooperative path – and some of the most vulnerable countries will collapse.

The heterogeneous-brink model provides an illuminating counterpoint to our common-brink model, where once the world reaches the brink countries do whatever is necessary to avoid global collapse. Relative to that setting, the difference is that each country now has its *own* brink generation, who faces the existential climate crisis *alone* and up against the other countries in the world, who have no reason in the noncooperative equilibrium to internalize the impact of their emissions choices on the fate of the brink country beyond the possible climate refugee costs that they may incur should the country collapse. It is also notable that, with heterogeneous collapse points, it is entirely possible that some countries will continue to enjoy a reasonable standard of living once the global carbon stock has stabilized while others have suffered climate collapse, bringing into high relief the potential unevenness of the impacts of climate change across those countries who, due to attributes of geography and/or socioeconomic position, are more or less fortunate. And even small differences across countries can lead to country collapse along the equilibrium path: unless the brink generation of each country arrives at the same moment, the “we are all in this together” forces that enabled the world to avoid collapse in the noncooperative equilibrium of our common-brink model will be disrupted. As we demonstrate, climate refugee externalities will bring back an element of these forces, albeit only partially.

We then revisit the potential role for ICAs, but now in the setting where the catastrophe point differs across countries. We find that the ICA may or may not save a country that would collapse in the noncooperative scenario, but no country will be allowed to collapse under the ICA that would not have collapsed in the absence of the ICA. And we find that, as a result of its inability to take into account directly the interests of future generations, the ICA may allow a range of the most vulnerable countries to collapse when a global social planner would not allow this to happen. We show that these findings hold even in a world where countries can make unlimited international transfers, and that when such transfers are limited by a country’s resource constraints the prospects for small, vulnerable countries such as the Maldives becomes even more bleak.

Overall, the conclusions from our core analysis are sobering. Even abstracting from issues of free-riding in participation and compliance, our model suggests that ICAs can play only a limited role in addressing the most pressing challenges of global warming. If countries face a common threshold of catastrophe, the ICA has a potential role to play only during the warming phase, by internalizing the horizontal climate externalities and slowing the world’s march to the brink, but it falls short of achieving the outcome of a global social planner, which requires

the world to move even more slowly toward the brink and possibly avoid the brink altogether. And the ICA has no role to play in saving the world from collapse, because once the brink of catastrophe is reached countries have sufficient incentives to avoid catastrophe even without an ICA. If the catastrophe threshold varies across countries, the role of an ICA is potentially more expansive, because it may save some of the most vulnerable countries from collapse, but its limitations relative to the global social planner are potentially more devastating, because it may not save enough countries from collapse.

Relative to the existing literature on ICAs, the main contribution of our paper is to analyze the joint implications of international and intergenerational externalities in a world with the potential for catastrophic effects of climate change. We are not aware of any formal analysis that considers the interaction between these fundamental ingredients.

There is an emerging literature that considers optimal environmental policy in the face of climate catastrophe. Prominent examples include Barrett (2013), Lemoine and Rudik (2017) and Besley and Dixit (2019).⁴ Of these papers, only Barrett (2013) considers the role of ICAs, but his model is effectively static and does not consider intergenerational issues that we emphasize here. A key point in his paper is that, if the level of the carbon stock that triggers a catastrophe is known with certainty, there exists a noncooperative equilibrium in which no catastrophe occurs, and hence the only possible role for an ICA is to help countries coordinate on the “good” equilibrium – a point that is consistent with our common-brink model.⁵

The paper of John and Pecchenino (1997) is also related. Like ours, their paper considers both international and intergenerational environmental externalities, but their paper does not consider the possibility of catastrophes. Instead, the central message of their paper is that cooperation between countries at a point in time may be harmful to future generations. This is because there are two international externalities in their model: one stemming from cross-border pollution, and one related to environment-enhancing investments. Internalizing the pollution externality benefits future generations (an effect that is present also in our model), but international cooperation on the investment dimension increases the efficiency of resource

⁴See also Brander and Taylor (1998), who consider catastrophes in a model that links population dynamics with renewable resource dynamics but does not feature intergenerational or international externalities.

⁵Barrett (2013) also argues that if the catastrophic threshold is uncertain, there is a unique Nash equilibrium that can lead to catastrophe, and an ICA can achieve a Pareto improvement over that equilibrium and reduce the probability of catastrophe. He also emphasizes that, while in the absence of uncertainty the only role of an ICA is to help countries coordinate on the efficient equilibrium without catastrophe, in the setting with uncertainty the ICA has to overcome enforcement and participation issues, just as in models without catastrophes. We discuss how our results can incorporate uncertainty over the brink point in the Conclusion.

allocation and hence increases consumption, which tends to degrade the environment.

Our paper is also related to the literature on the dynamics of ICAs, which includes Dutta and Radner (2004), Harstad (2012, 2020, 2021) and Battaglini and Harstad (2016). These papers focus on aspects of ICAs that are very different from the ones we emphasize in this paper, and they do not consider issues of intergenerational externalities or the possibility of catastrophes. In particular, Harstad (2012) and Battaglini and Harstad (2016) focus on issues of free-riding and participation in ICAs when countries can make irreversible investments in green technology that cannot be contracted upon, and Harstad (2020) takes this approach one step further by considering the implications of alternative bargaining procedures.⁶ Finally, Harstad (2021) focuses on the desirability of issue linkage through a trade agreement whose commitments are made contingent on forest conservation measures.

The remainder of the paper proceeds as follows. Section 2 sets out our common-brink model and characterizes the noncooperative emissions choices, as well as those under an ICA and the choices of a global social planner. Section 3 contains the parallel analysis for our heterogeneous-brink model. Section 4 allows for the possibility of intergenerational altruism in our analysis. Finally, section 5 concludes with a brief discussion of a number of further extensions to our core models. An Appendix provides proofs not contained in the body of the paper.

2. Basic Model

2.1. Economic structure

We consider a world of M countries. Each country is identical, with a population of identical citizens that we normalize to one.⁷ Time is discrete and indexed by $t \in \{0, 1, \dots, \infty\}$. We adopt a “successive generations” setting (see Fahri and Werning, 2007), where the citizen in each country lives for one period and is replaced by a single descendant in the next period. We allow each parent to be altruistic toward its only child, and the utility of a representative country’s generation t is given by

$$\tilde{u}_t = u_t + \beta \tilde{u}_{t+1}$$

⁶For earlier analyses of ICAs that focus on issues of participation and enforcement, see for example Barrett (1994), Carraro and Siniscalco (1993) and Kolstad and Toman (2005).

⁷An identical-country assumption makes it natural to focus on symmetric equilibria of our model in which emissions choices and utility are the same across all countries of the world, allowing us to abstract in this section from the possible use of international transfers by a global social planner or in an international climate agreement. In the next section we allow countries to reach a catastrophe at different levels of the global carbon stock, and consider there the role of international transfers.

where u_t is material utility and the parameter $\beta \geq 0$ captures the degree of intergenerational altruism.⁸ In this setting, utility can be equivalently represented with the dynastic utility function

$$\tilde{u}_t = \sum_{s=0}^{\infty} \beta^s u_{t+s}. \quad (2.1)$$

Material utility u_t is derived from consumption and from the quality of the environment. But these two dimensions of utility are in tension, as consumption generates carbon emissions, which add to the global carbon stock and degrade the quality of the environment through a warming climate. This tension defines the fundamental tradeoff faced by each generation.

To highlight this tradeoff, we abstract from trading relations between countries, so that we can focus on their interactions mediated through the global carbon stock.⁹ And we adopt a reduced form approach to modeling the consumption benefits of emissions, by specifying the benefits directly as a function of emissions rather than the underlying consumption choices that generate the emissions. In particular, we use the increasing and concave function $B(c_t)$ to denote these benefits, where $c_t \geq 0$ is the level of carbon emissions of a representative country's generation t .¹⁰ We therefore treat c_t itself as the choice variable of a country, with the understanding that lower emissions mean lower consumption.¹¹ We have in mind that each government then implements its chosen c_t with an appropriate climate policy (e.g., carbon tax).

While a country's own period- t emissions generate consumption benefits for its generation t , these emissions also contribute to the global stock of carbon in the atmosphere. We denote by C_t the global carbon stock in period t .¹² The evolution of this stock through time depends

⁸To ease notation and in light of our identical-country assumption, here and throughout this section we omit country subscripts and instead present variables in terms of a representative country. In the background of course, each country makes its own choices, which turn out to be identical given that the countries are assumed to have identical attributes.

⁹We briefly consider the implications of trade in the Conclusion.

¹⁰Our restriction that $c_t \geq 0$ reflects the possibility of zero (net) emissions through carbon capture and other mitigation efforts. By this logic we could impose $c_t \geq c^{\min}$ where c^{\min} could be strictly positive or even strictly negative, but in our formal analysis it is convenient to abstract from these possibilities and equate the emissions generated by a country's best mitigation efforts with its emissions were it to collapse.

¹¹Implicit in our specification of the reduced-form benefit function $B(c_t)$ is the assumption that there is a one-to-one mapping between a country's emissions and its utility from consumption, and hence that the stock of carbon does not itself impact this mapping (e.g., by impacting a country's productivity associated with any level of emissions). In principle we could allow for such an impact with an alternative benefit function $B(c_t, C_t)$ where C_t shifts down (or up) the benefit function for a given level of emissions. If it were allowed that $B_{c_t C_t} \neq 0$ then this would complicate the paths of emissions that we derive below, but we do not believe that it would generate interesting qualitative differences in the results we emphasize, and so we opt for the tractability of our simpler modeling assumption.

¹²More accurately, C_t can be thought of as the atmospheric CO_2 concentration, an increase in which leads to

on the depreciation rate of the stock and on the level of emissions c_t , according to

$$C_t = (1 - \rho)C_{t-1} + M c_t, \text{ with } C_{-1} = 0. \quad (2.2)$$

The parameter $\rho \in [0, 1)$ reflects the natural rate of atmospheric “regeneration”: if $\rho = 1$, by the beginning of the current period the previous period’s stock of carbon is gone; if $\rho = 0$, the current period inherits the full stock of carbon from the previous period. As will become clear just below, the relationship in (2.2) implies that each generation feels the impact of its own emissions (because these emissions add to the carbon stock in the current period).¹³

We assume that increases in the global carbon stock degrade the environment and lead to losses in material welfare. We assume that these losses rise linearly with the global carbon stock C according to the parameter $\lambda > 0$, and jump to infinity if C exceeds a catastrophic level \tilde{C} . We have in mind that moderate degrees of global warming lead to moderate costs.¹⁴ But past a certain critical level, a rising carbon stock leads to a level of global warming that would trigger the collapse of civilization.¹⁵ In the next section we will allow the level of the carbon stock that would be catastrophic to differ by country, but in this section we assume that it is common to all countries, and we will sometimes refer to this model as the “common-brink model.” Finally, the catastrophic level \tilde{C} is assumed known with certainty.¹⁶ Collecting these assumptions, we may write the material utility of a representative country’s generation t as

$$u_t = \begin{cases} B(c_t) - \lambda C_t & \text{if } C_t \leq \tilde{C} \\ -\infty & \text{if } C_t > \tilde{C}. \end{cases} \quad (2.3)$$

Our model highlights two externalities that arise in the context of climate change. One externality is international: with $M > 1$ countries, the emissions of a country’s generation t contribute to the global stock of period- t carbon, which impacts the material utility of generation- t

global warming. But in the text we will refer to C_t as the carbon stock.

¹³Given that a period corresponds to a generation in our model, this feature seems broadly realistic, as existing estimates put the time it takes for current carbon emissions to translate into higher global temperatures at between 10 and 40 years (see, for example, Pindyck, 2020, who also reports an estimate of the dissipation rate ρ on the order of 0.0035 per year).

¹⁴Our main points would not be affected if we allowed some countries to gain from moderate warming.

¹⁵For Venus-like levels of atmospheric carbon dioxide this would clearly be true, but as will become clear below the relevant question here is whether BAU emissions levels would bring the world to a threshold level of carbon that would trigger collapse. While the precise meaning of “collapse” is open to interpretation, we do not require that the cost of collapse is infinite, only that it is sufficiently high, and on that basis we believe that the assumption we introduce below – that BAU emissions levels would indeed bring collapse in finite time – is warranted.

¹⁶We discuss the more realistic possibility that \tilde{C} is uncertain in the Conclusion.

in all other countries. The other externality is intergenerational: the emissions of a country's generation t affect the material utility of all subsequent generations $t + 1$ and beyond in that country. Moreover, these “horizontal” (international) and “vertical” (intergenerational) externalities interact to produce additional “diagonal” externalities: the emissions of one country's generation t impact the utility of future generations in all other countries.

Below we will characterize three scenarios where these externalities are addressed to varying degrees. In the noncooperative equilibrium that arises if countries choose emissions levels in the absence of any agreements, neither the horizontal nor the vertical or diagonal externalities are addressed, in the sense that each country's emissions choices impose costs on other countries and on future generations that those parties did not agree to incur. In the ICA equilibrium that arises if international agreements are available, we will argue below that only the horizontal externalities can be addressed, not the vertical or diagonal externalities: as future generations cannot sit at the table while the ICA is negotiated, even in the presence of intergenerational altruism the ICA's emissions choices will inevitably impose costs on future generations in all countries that those parties did not agree to incur.

And third, we will consider as a benchmark the outcome that a global social planner would implement beginning at time $t = 0$ to maximize a social welfare function. For our characterization of the social planner's choice, we assume, reflecting the symmetry across countries in our model setup, that within any generation the planner seeks to maximize the utility of a representative country; that the planner puts positive weights on future generations *directly*, not just indirectly through the intergenerational altruism of the initial generation; and that intergenerational lump-sum transfers are unavailable, leaving emissions levels (and possibly international lump-sum transfers) as the planner's only choice variable. As the planner's Pareto weights on each generation can be interpreted as the weight given to each generation in an efficient bargaining solution over emissions levels, the solution to our planner's problem can be said to address all of the externalities – horizontal, vertical and diagonal – since under this interpretation there are no costs of emissions incurred by any party that the party did not “agree” to.

In particular, to characterize the global social planner's choices we follow Fahri and Werning

(2007) in postulating the following planner objective:¹⁷

$$W = \sum_{t=0}^{\infty} \hat{\beta}^t u_t \quad (2.4)$$

where $\hat{\beta}$ is the planner's discount factor. Notice that regardless of the degree of intergenerational altruism displayed by each generation, there will be a discrete wedge between the social and private discount factor ($\hat{\beta} - \beta$) as long as the planner puts strictly positive weights on future generations directly, hence we have $\hat{\beta} > \beta$. Moreover, in general this wedge need not decrease as β rises. For example, in a two-period setting with α the Pareto weight placed by the planner on the second generation, we would have $\hat{\beta} = \beta + \alpha$. Notice also that in principle $\hat{\beta}$ could be greater than one, but to avoid the complications that would arise if this were the case we assume for simplicity that $\hat{\beta} < 1$.¹⁸

Since the presence of intergenerational externalities is one of the more novel features of our model, it is worth pausing to clarify the nature of the deviation from the social planner's choice that is created when this externality is not internalized. To this end, suppose for a moment that there is only one country (so no horizontal externality for an ICA to address) and no intergenerational altruism ($\beta = 0$). In this case, when generation t chooses emissions to maximize its utility, it ignores the impact of these emissions on future generations and simply maximizes its own material utility. A planner who puts positive weight on each generation ($\hat{\beta} > 0$) would modify the choices of generation t and redistribute utility from generation t to subsequent generations. Importantly, the same logic applies also in the presence of altruism ($\beta > 0$), because as noted above, if the social welfare function puts direct weight on future generations the wedge between the social and private discount factors ($\hat{\beta} - \beta$) need not decrease as β rises. Notice, though, that the described internalization of the intergenerational externality that the social planner orchestrates in this case does not mark a Pareto improvement, but rather a movement along the efficiency frontier, shifting surplus from generation t to later generations.¹⁹

¹⁷See also Caplin and Leahy (2004), Feng and Ke (2018), and Millner and Heal (2021).

¹⁸In the case where $\hat{\beta} \geq 1$, the infinite sum in (2.4) does not converge, so we would have to assume a finite horizon.

¹⁹In a static setting, an analogous scenario would be a world with two countries where there exists a one-way policy externality, meaning that one of the countries chooses a policy which has an externality on the other country, and where international lump-sum transfers are not possible. In such a scenario, the noncooperative policy choice would lead to the point on the Pareto frontier that maximizes the utility of the country choosing the policy, and a global planner who puts positive weight on both countries would choose a different point on the frontier, thus redistributing utility from the country choosing the policy to the other country.

For this reason, when comparing the noncooperative outcome to the ICA outcome – where only the current generation is given positive weight and only horizontal externalities are addressed – and to the social planner outcome – where strictly positive weight is also placed directly on future generations and where horizontal, vertical and diagonal externalities are all addressed – we will avoid referring to the social planner outcome as the sole benchmark for efficiency. Rather, we will instead treat these comparisons as what they are, namely, comparisons of the noncooperative outcome to two distinct points on the Pareto frontier, albeit with the ICA outcome corresponding to an extreme point on this frontier. Likewise, when we compare the ICA outcome to the social planner outcome, it should be kept in mind that we are comparing two outcomes that, while differing in their treatment of future generations, are nevertheless both efficient in the Pareto sense.

Finally, before turning to the analysis we can make a simple preliminary point: the impacts associated with horizontal and vertical externalities *reinforce* each other. This can be seen most clearly by focusing on a special and simple case of our model, in which there is no catastrophe point ($\tilde{C} = \infty$), no atmospheric regeneration ($\rho = 0$) and no intergenerational altruism ($\beta = 0$); and by comparing the noncooperative emissions choices to those chosen by the social planner. In this case it is straightforward to show and intuitive that in the noncooperative equilibrium each country's generation t would choose a level of emissions to satisfy $B'(c_t) = \lambda$ (assuming interior solutions), while the emissions levels chosen by the planner for each country's generation t satisfy $B'(c_t) = \frac{M}{1-\beta}\lambda$. The overall wedge between the planner's emissions choices and noncooperative emissions choices is summarized by $\frac{M}{1-\beta} > 1$, which implies excessive emissions in the noncooperative equilibrium relative to the planner's outcome. The wedge has two components: $M > 1$ reflects the degree to which the international externality contributes to excessive emissions in the noncooperative equilibrium, because noncooperative choices do not account appropriately for the environmental costs of a country's emissions that are imposed on other countries; and $\frac{1}{1-\beta} > 1$ reflects the degree to which the intergenerational externality contributes to excessive emissions in the noncooperative equilibrium, because noncooperative choices do not account for the environmental costs of a country's emissions that are imposed on future generations. The two externalities enter multiplicatively into this wedge, so they reinforce each other. Intuitively, this is a consequence of the above-mentioned fact that there are not only “horizontal” and “vertical” externalities, but also “diagonal” externalities.

The special case of our model described just above is useful for highlighting in simple terms

the impacts of the externalities that arise in our model. But it is also useful as a benchmark to illustrate the critical role that the catastrophe point (\tilde{C} finite) plays in our analysis of climate policy. In the absence of a catastrophe point, the social planner and noncooperative emissions profiles are straightforward, as we have just observed, as is the emissions profile under an ICA. But as we establish below, the existence of a catastrophe point introduces fundamental changes to these emissions profiles, both along the path to the catastrophe and once the brink of catastrophe is reached, and it changes the possible role of an ICA as well.

The importance of a catastrophe point for understanding the policy challenges posed by climate change is one of the central messages of our paper. To deliver this message, we henceforth focus on the case in which \tilde{C} is finite. We will proceed by focusing for now on a world without intergenerational altruism ($\beta = 0$); in section 4 we consider as well the possibility that $\beta > 0$ and show how our results extend in the presence of intergenerational altruism. Notice from (2.1) that with $\beta = 0$ there is no distinction between utility (\tilde{u}_t) and material utility (u_t), and for this reason we will simply refer to “utility” and use the notation u_t until we reintroduce intergenerational altruism in section 4.

2.2. Noncooperative Equilibrium

We begin our analysis by characterizing the noncooperative emissions choices. We will focus on Markov perfect equilibria. As we noted above, we assume $\beta = 0$ for now, so that there is no intergenerational altruism. Given $\beta = 0$, countries are effectively myopic. This implies that the noncooperative equilibrium in general has two phases.

The first phase is a “warming phase,” during which the emissions of each country’s generation t is constant at the level \bar{c}^N defined by $B'(\bar{c}^N) = \lambda$, where the marginal benefit to each country of the last unit of carbon that it emits is equal to the marginal loss of utility that it suffers as this unit of carbon is added to the global carbon stock, implying

$$\bar{c}^N = B'^{-1}(\lambda). \quad (2.5)$$

As is intuitive, (2.5) implies that \bar{c}^N is decreasing in λ , the marginal cost in terms of own utility associated with another unit of carbon emissions. We can think of \bar{c}^N as corresponding to “Business-As-Usual” (BAU) emissions levels. During the warming phase associated with these choices, the global stock of carbon grows according to

$$C_t^N = (1 - \rho)C_{t-1}^N + M\bar{c}^N, \text{ with } C_{-1}^N = 0 \quad (2.6)$$

and as the global carbon stock C_t^N grows and the cost of climate change mounts, the utility of each successive generation in every country declines according to

$$u_t^N = B(\bar{c}^N) - \lambda C_t^N. \quad (2.7)$$

If the warming phase went on forever, (2.6) implies that the global carbon stock would converge to the steady state level $\frac{M}{\rho}B'^{-1}(\lambda) \equiv C^N$. And if the catastrophe level of the global carbon stock, \tilde{C} , were greater than C^N , then BAU emissions could indeed go on forever without triggering a climate catastrophe. But the view of the majority of climate scientists is that a climate catastrophe will occur in finite time, perhaps by the end of this century, if the world stays on a BAU emissions path (see, for example, the recent reports of the IPCC). In the language of our model this view translates into a statement that \tilde{C} lies below C^N . We therefore impose

$$\tilde{C} < C^N \quad (\text{Assumption 1})$$

which ensures that under BAU emissions the catastrophic level of the global climate stock would eventually be breached.

The second phase of the noncooperative equilibrium kicks in when C_t^N reaches the brink of catastrophe \tilde{C} . This occurs for the “brink generation” $t = \tilde{t}^N$ where, ignoring integer constraints, \tilde{t}^N is defined using (2.6) by $C_{\tilde{t}^N}^N = \tilde{C}$. In effect, \tilde{t}^N represents the point in time where, in a single generation, life under BAU emissions would go from livable to unlivable.

If the brink generation \tilde{t}^N is to avoid the collapse of civilization, it must end the warming phase with an “11th-hour solution” that brings climate change to a halt. Indeed it is easy to see that if it is feasible to do so, then at any equilibrium in undominated strategies, C_t^N remains at \tilde{C} for $t = \tilde{t}^N$ and also for all subsequent generations.²⁰ Focussing on the symmetric equilibrium where all countries adopt the same level of emissions, for generations $t \geq \tilde{t}^N$ noncooperative emissions will fall to the replacement level dictated by the natural rate of atmospheric regeneration given by

$$c_t = \frac{\rho \tilde{C}}{M} \equiv \hat{c}^N \quad (2.8)$$

where $\hat{c}^N < \bar{c}^N$ is implied by Assumption 1. With $c_t = \hat{c}^N$ for generations $t \geq \tilde{t}^N$, the world remains on – but does not go over – the brink of catastrophe, so the collapse of civilization is

²⁰There are also equilibria where the world collapses, because if other countries choose very high emission levels, an individual country is indifferent over its own emission levels, so it is an equilibrium for all countries to choose very high emission levels. But it is easy to see that such equilibria are in weakly dominated strategies: starting from such an equilibrium, a country can weakly improve its payoff by lowering its emissions.

avoided. To confirm that \hat{c}^N is indeed the symmetric noncooperative equilibrium emissions level for generations $t \geq \tilde{t}^N$, we need only note that unilateral deviation to an emissions level higher than \hat{c}^N would trigger climate catastrophe and infinite loss, while deviation to a lower emissions level would not be desirable either given that $\hat{c}^N < \bar{c}^N$.²¹ The utility of each generation $t \geq \tilde{t}^N$ during this second phase of the noncooperative equilibrium is then constant and given by

$$u_t^N = B(\hat{c}^N) - \lambda\tilde{C}. \quad (2.9)$$

We may conclude that the noncooperative emissions path for each country is given by

$$c_t^N = \begin{cases} \bar{c}^N & \text{for } t < \tilde{t}^N \\ \hat{c}^N & \text{for } t \geq \tilde{t}^N. \end{cases} \quad (2.10)$$

Combining (2.9) with (2.7) we then also have the path of noncooperative utility:

$$u_t^N = \begin{cases} B(\bar{c}^N) - \lambda C_t^N & \text{for } t < \tilde{t}^N \\ B(\hat{c}^N) - \lambda\tilde{C} & \text{for } t \geq \tilde{t}^N. \end{cases} \quad (2.11)$$

Note that under the noncooperative equilibrium and according to (2.11), utility must fall discretely for the brink and all subsequent generations, due to the discrete reduction in global emissions implied by

$$\hat{c}^N = \frac{\rho\tilde{C}}{M} < B'^{-1}(\lambda) = \bar{c}^N \quad (2.12)$$

that is required to prevent catastrophe once the world reaches the brink, where the inequality in (2.12) follows from Assumption 1 as we have noted. According to (2.11) and (2.12), in order to prevent the planet from warming further, the brink generation and all future generations accept a reduced level of consumption associated with \hat{c}^N that is further below the consumption level associated with \bar{c}^N enjoyed by previous generations the greater the number of countries M , the smaller the regeneration capacity of the atmosphere ρ and level of carbon stock above which climate catastrophe occurs \tilde{C} , and the lower the cost of moderate pre-catastrophe warming λ .

Summarizing, we may now state:

²¹While we focus on the symmetric Nash equilibrium at the brink, there is also a continuum of asymmetric Nash equilibria, in which some countries cut their emissions levels below \hat{c}^N while others raise their emissions levels above \hat{c}^N and the sum of world-wide emissions remains at the level $\rho\tilde{C}$ which holds the world at the brink. It is easy to see that these asymmetric Nash equilibria are inefficient given our symmetric-country setup, and so we take the symmetric Nash equilibrium as the natural focal point. As we will discuss below, in the event that, contrary to our assumption, countries coordinate on one of the asymmetric and inefficient Nash equilibria, a coordination role for an international climate agreement would then arise in which countries agree to the symmetric and efficient Nash emissions levels \hat{c}^N and then use international lump-sum transfers to distribute according to bargaining powers the surplus gains that result from eliminating the inefficiency.

Proposition 1. *The noncooperative equilibrium of the common-brink model exhibits an initial warming phase where each country’s emissions are constant at a “Business-As-Usual” level. During this phase, the global stock of carbon rises and the world is brought to the brink of catastrophe. Once the brink is reached, a catastrophe is avoided with an 11th hour solution that halts further climate change with reduced emissions that are set at the replacement level dictated by the natural rate of atmospheric regeneration and remain at that level for all generations thereafter, and which imply a discrete drop in utility for the brink generation and all future generations.*

Notice an interesting feature of the noncooperative equilibrium described in Proposition 1: no generation up until the brink generation does anything to address the climate externalities that each generation is imposing on those of its generation residing in other countries and on future generations everywhere; and yet the brink generation overcomes all of these externalities and saves the world. The reason for this 11th hour noncooperative solution to the threat of global annihilation posed by climate change is that, while earlier generations face rising costs of global warming as their emissions contribute to a growing global stock of carbon, it is only the brink generation that faces the catastrophic implications of continuing the emissions practices of the past. And in the face of this potential catastrophe, the nature of the game is fundamentally altered, with the result that the brink generation “does whatever it takes” in the noncooperative equilibrium to avoid catastrophe.²²

Hence, Proposition 1 describes a good news/bad news feature of the noncooperative equilibrium: the good news is that, while it takes a crisis to shake the world from business-as-usual behavior, when the crisis arrives the world will find a way to save itself from going over the brink; the bad news is that the world that is saved on the brink is not likely to be a nice world in which to live, both because the climate at the brink of catastrophe may be very unpleasant, and because the brink and all future generations must accept a discrete drop in consumption and utility in order prevent the climate from worsening further and resulting in annihilation.

²²Barrett (2013) makes a related observation. He notes that the nature of the game can change if countries face a catastrophic loss function associated with climate change, but his observation is made within a static model and emphasizes the implications for the self-enforcement constraint in international climate agreements.

2.3. International Climate Agreements

We are now ready to consider what an ICA can achieve. Two important challenges that an ICA must meet relate to participation and enforcement. It is well known (see, for example, Barrett, 1994, Harstad, 2012, Nordhaus, 2015, Battaglini and Harstad, 2016 and Harstad, 2020) that ICAs create strong incentives for countries to free ride on the agreement, and that without some means of requiring participation the number of countries participating in an ICA is likely to be very small. And even among the willing participants, there is a serious question of how the commitments agreed to in the ICA can be enforced, given that the agreement must ultimately be self-enforcing and that retaliation using climate policy for this purpose is arguably ineffective (see, for example, Maggi, 2016 and Barrett and Dannenberg, 2018 on the possibility of linking trade agreements to climate agreements in this context). Together these challenges are understood to place important limitations on what an ICA can achieve.

Here we abstract from these well-studied limitations, and assume that the ICA attains full participation of all M countries in the world, that the noncooperative equilibrium is the “threat point” for the negotiations over an ICA, and that any arrangements negotiated under the ICA are perfectly enforceable by an external enforcement mechanism. Under these ideal conditions, we ask what an ICA can accomplish. Our answer highlights an additional limitation that has not been emphasized in the formal literature on climate agreements. This limitation arises from the fact that it is not possible for an ICA to include *future* generations in the bargain alongside current generations. Hence, while an ICA can in principle address the horizontal externalities that arise from the international aspects of emissions choices and that create inefficiencies in the noncooperative outcomes, it cannot address the vertical and diagonal externalities that are associated with the intergenerational aspects of the climate problem.²³ Our goal is to characterize what an ICA can achieve in the presence of this limitation, relative to the noncooperative outcome and relative to a global social planner who is not subject to this limitation.

Recalling that we are focusing for now on the case $\beta = 0$ so as to abstract from intergenerational altruism, for each generation t we characterize the ICA emissions levels as those that

²³One could imagine that in principle an implicit contract of some kind between current and future generations might be available to internalize the intergenerational externalities. But recall that altruism itself cannot address this issue. Rather, for such a contract to be implemented, future generations would have to be able to punish current generations for any deviations from the contract, and current and future generations would need to find a way to coordinate on a particular equilibrium of this kind even though communication between them is impossible. We view these challenges as essentially insurmountable.

maximize welfare of generation t in the representative country. Given our symmetric-country assumption, this is the natural ICA design to focus on, as it would emerge if countries bargain efficiently and have symmetric bargaining power.

Using (2.3), it is direct to confirm that, for as long as the catastrophe point \tilde{C} is not hit, emissions levels under the ICA satisfy $B'(c_t) = M\lambda$ and are hence given by

$$\bar{c}^{ICA} = B'^{-1}(M\lambda). \quad (2.13)$$

According to (2.13), in any period where the catastrophe point is not hit, each country's emissions under the ICA will equate that country's marginal utility from a small increase in emissions to the marginal environmental cost, taking into account the costs imposed on the current generation in all M countries. Notice that (2.5) and (2.13) imply $\bar{c}^{ICA} < \bar{c}^N$, because under noncooperative choices each country internalizes the costs imposed on the current generation only in its own country. Finally, with emissions levels given by \bar{c}^{ICA} , as long as the brink of catastrophe is not hit the carbon stock under the ICA evolves according to

$$C_t^{ICA} = (1 - \rho)C_{t-1}^{ICA} + M\bar{c}^{ICA} \quad \text{with } C_{-1}^{ICA} = 0, \quad (2.14)$$

which defines a process of global warming in which the global carbon stock would eventually converge to the steady state level $\frac{M}{\rho}B'^{-1}(M\lambda) \equiv C^{ICA}$.

Recall that under Assumption 1 the brink of climate catastrophe will be reached under the BAU emissions of the noncooperative equilibrium. Will the ICA keep the world from ever reaching the brink? The answer is yes, if and only if

$$\tilde{C} \geq C^{ICA}, \quad (2.15)$$

where note from their definitions that $C^{ICA} < C^N$ so both Assumption 1 and (2.15) will be satisfied if $\tilde{C} \in [C^{ICA}, C^N)$. Intuitively, if the catastrophe point of the global carbon stock, \tilde{C} , is high and sufficiently close to the steady state level of the global carbon stock under BAU emissions, C^N , then only a relatively small reduction in emissions from the BAU level would be required to keep the world from reaching the brink, and the ICA will indeed deliver the required reductions; and the threshold level of the carbon stock C^{ICA} in (2.15) defines "sufficiently close" in this context.

On the other hand, if \tilde{C} is below this threshold level and (2.15) is violated so that

$$\tilde{C} < C^{ICA}, \quad (2.16)$$

then under the ICA the brink of catastrophe will be reached in finite time, and the brink generation \tilde{t}^{ICA} is determined from (2.14) as the period \tilde{t}^{ICA} that satisfies $C_{\tilde{t}^{ICA}}^{ICA} = \tilde{C}$. Notice from (2.14) and (2.6) that $\tilde{t}^{ICA} > \tilde{t}^N$ is ensured by $\bar{c}^{ICA} < \bar{c}^N$, so the ICA postpones the arrival of the brink generation when (2.16) is satisfied even though it does not avoid the brink completely in this case.

If (2.16) is satisfied, what happens under the ICA when the world reaches the brink? This might seem to be when the ICA can play its most important role, by ensuring the very survival of civilization. And clearly, given the utility function in (2.3), the ICA will not let the world go over the brink. But recall that neither would countries go over the brink in the noncooperative equilibrium. In fact, far from marking the moment when achieving international cooperation under an ICA becomes indispensable, for $t \geq \tilde{t}^{ICA}$ the ICA becomes *redundant*, because from that point forward the ICA can do no better than to replicate the noncooperative emissions choices \hat{c}^N .

Hence, \tilde{t}^{ICA} marks the end of the useful life of the ICA, and \tilde{t}^{ICA} is finite in the case when (2.16) is satisfied; and notably, in neither of the two cases we have described above does the ICA play a role in helping the world avoid climate catastrophe, for the simple reason that countries will avoid climate catastrophe in the noncooperative equilibrium in any event and hence have no need for an ICA to serve this purpose.²⁴

Finally, letting c_t^{ICA} denote the path of emissions under the ICA, we can describe emissions succinctly under both (2.15) and (2.16) with

$$c_t^{ICA} = \begin{cases} \bar{c}^{ICA} & \text{for } t < \tilde{t}^{ICA} \\ \hat{c}^N & \text{for } t \geq \tilde{t}^{ICA} \end{cases} \quad (2.17)$$

where \tilde{t}^{ICA} is finite if and only if (2.16) is satisfied. Utility under the ICA is then given by

$$u_t^{ICA} = \begin{cases} B(\bar{c}^{ICA}) - \lambda C_t^{ICA} & \text{for } t < \tilde{t}^{ICA} \\ B(\hat{c}^N) - \lambda \tilde{C} & \text{for } t \geq \tilde{t}^{ICA}. \end{cases} \quad (2.18)$$

Note that under the ICA, if (2.16) is satisfied so that \tilde{t}^{ICA} is finite, then (2.18) implies that utility must fall discretely for the brink and all subsequent generations, due to the discrete

²⁴Recall that for $t \geq \tilde{t}^N$ we have focussed on the symmetric equilibrium of the noncooperative game in which countries adopt the efficient assignment of emissions. If in the noncooperative game countries coordinated on an inefficient equilibrium for $t \geq \tilde{t}^N$ then the ICA would have a role to play for $t \geq \tilde{t}^{ICA}$, allowing countries to exchange emissions cuts for transfers and hence achieve a Pareto improvement by moving to the efficient equilibrium (see also note 21). In this case the ICA would have a continuing role in enhancing the efficiency properties of the emissions cuts required for survival, but the ICA would still play no role in helping the world avoid a climate catastrophe.

reduction in global emissions implied by

$$\hat{c}^N = \frac{\rho \tilde{C}}{M} < B'^{-1}(M\lambda) = \bar{c}^{ICA} \quad (2.19)$$

that is required to prevent catastrophe once the world reaches the brink, where the inequality in (2.19) follows from (2.16). Hence, according to (2.18) and (2.19) and similar to the noncooperative equilibrium, in order to prevent the planet from warming further, under the ICA the brink generation and all future generations accept a reduced level of consumption. However, with $\bar{c}^{ICA} < \bar{c}^N$ it is also clear that the brink generation suffers a less precipitous decline in welfare under the ICA than in the noncooperative equilibrium.

We may now summarize with:

Proposition 2. *In the common-brink model the path of emissions under the ICA falls into one of two cases. If \tilde{C} is above a threshold level, then (i) the brink of catastrophe is never reached, and (ii) the ICA emissions levels are below the noncooperative levels and constant through time. Otherwise, if \tilde{C} is below this threshold, then (i) the brink of catastrophe will be reached, (ii) the ICA emissions levels are below the noncooperative levels and constant through time until the brink is reached, at which point the useful life of the ICA ends and emissions fall to the replacement rate dictated by the natural rate of atmospheric regeneration and remain at that level for all generations thereafter, and (iii) the path of ICA emissions implies a discrete drop in utility for the brink generation relative to the previous generation, but this drop is smaller than under the path of noncooperative emissions levels. In neither case does the ICA have a role to play in helping the world avoid climate catastrophe.*

2.4. The Social Optimum

We next consider the emissions levels that a global social planner would choose in order to maximize world welfare, or equivalently, given our symmetry assumption, the welfare of the representative country as defined in (2.4). We will refer to the choices of the planner as the socially optimal choices.²⁵

Clearly, the planner will not allow the world to end in catastrophe and hence will not allow C_t to exceed \tilde{C} . Consequently, for the planner's problem we can equate u_t with $B(c_t) - \lambda C_t$

²⁵Notice that in our setting the planner problem is time-consistent, so we need only write down the planner's objective from the perspective of $t = 0$.

and introduce the constraint $C_t \leq \tilde{C}$. To determine the socially optimal emissions choices, we therefore write the planner's problem as

$$\begin{aligned} \max \quad & \sum_{t=0}^{\infty} \hat{\beta}^t [B(c_t) - \lambda C_t] \\ \text{s.t. } \quad & C_t = (1 - \rho)C_{t-1} + Mc_t \text{ for all } t \\ & C_t \leq \tilde{C} \text{ for all } t; \quad c_t \geq 0 \text{ for all } t. \end{aligned}$$

We assume that the problem is globally concave, so that we can rely on a first-order condition approach, and for simplicity we restrict attention to the case where the emissions feasibility constraint $c_t \geq 0$ is not binding.

Here we summarize the main steps of the analysis, relegating the formal proof to the Appendix. There are two cases to consider, depending on whether or not the state constraint $C_t \leq \tilde{C}$ binds for any t . We show that, if \tilde{C} is higher than a threshold level C^S , the state constraint is not binding and the brink of collapse \tilde{C} is never reached; we refer to this as Case 1. If \tilde{C} is lower than C^S , on the other hand, the state constraint is binding and the brink of collapse \tilde{C} is reached at some point in time; we refer to this as Case 2.

In Case 1, where $\tilde{C} \geq C^S$, we show that the optimal level of emissions \bar{c}^S is constant for all countries and all generations, and defined by $B'(\bar{c}^S) = \frac{M\lambda}{1-\hat{\beta}(1-\rho)}$. This solution has a simple interpretation: a country's own marginal benefit of emissions should equal the marginal environmental cost of emissions, taking into account the costs imposed on the utility of all M countries and on all future generations (discounted by the planner's discount factor $\hat{\beta}$ and accounting for the natural rate of atmospheric regeneration ρ).

In this case the carbon stock increases in a concave way and converges to the steady state level C^S . In the Appendix we show that C^S is lower than C^N , the steady state level of the BAU global carbon stock. Recalling our Assumption 1 that $\tilde{C} < C^N$, Case 1 therefore obtains if $\tilde{C} \in [C^S, C^N)$. The intuition for this case is that, if the catastrophe point \tilde{C} is high and sufficiently close to C^N , then only a relatively small reduction in emissions from the BAU level would be required to keep the world from reaching the brink, and the planner will indeed deliver the required reductions.

In Case 1, the level of welfare achieved by a representative country's generation t under the socially optimal emissions choices, which we denote by u_t^S , is given by $u_t^S = B(\bar{c}^S) - \lambda \bar{C}_t^S$, where \bar{C}_t^S is the path of the carbon stock implied by the constant emissions level \bar{c}^S . Note that, with the socially optimal emissions set at the constant level \bar{c}^S , the utility of each generation declines

through time as \bar{C}_t^S rises and the climate warms. It is notable that, while $\hat{\beta}$ impacts the *level* of \bar{c}^S , it does not alter the fact that the socially optimal emissions level is constant through time. Evidently, in Case 1 a higher $\hat{\beta}$ induces higher welfare for later generations under the socially optimal emissions choices not by tilting the emissions profile toward later generations, but by reducing the (constant) level of emissions for all generations and thereby shifting utility toward future generations in the form of a lower steady state level of atmospheric carbon and a cooler climate.

In Case 2, where $\tilde{C} < C^S$, the socially optimal carbon stock level grows over time until it reaches the brink level \tilde{C} at some point in time \tilde{t}^S . In this case, we show that the socially optimal emissions in a representative country, denoted \hat{c}_t^S , decline over time until time \tilde{t}^S , at which point they hit the level \hat{c}^N that keeps the carbon stock steady at the brink level \tilde{C} , and then remain constant at \hat{c}^N .

To gain some intuition for the result that in Case 2 the socially optimal emissions decline over time until they hit their steady state level, consider the special case where $\lambda = \rho = 0$. Then we can think of the problem as allocating a fixed amount of (nonnegative) emissions across all generations, so the optimal emissions maximize $\sum_{t=0}^{\infty} \hat{\beta}^t B(c_t)$ subject to the constraints $\sum_t c_t = \tilde{C}$ and $c_t \geq 0$. Clearly, then, whenever c_t is strictly positive it must equalize the discounted marginal benefit of emissions, $\hat{\beta}^t B'(c_t)$, across generations, so that $\hat{\beta}^t B'(c_t)$ must be constant over time, therefore the undiscounted marginal benefit $B'(c_t)$ must increase, and hence c_t must decrease over time.

In Case 2, the level of welfare achieved by a representative country's generation t under the socially optimal emissions is given by

$$u_t^S = \begin{cases} B(\hat{c}_t^S) - \lambda \hat{C}_t^S & \text{for } t < \tilde{t}^S \\ B(\hat{c}^N) - \lambda \tilde{C} & \text{for } t \geq \tilde{t}^S \end{cases} \quad (2.20)$$

where \hat{C}_t^S is the path of the carbon stock implied by the emissions path \hat{c}_t^S .

The socially optimal time path of welfare differs in interesting ways across Case 1 and Case 2. In the social optimum under Case 2, where $\tilde{C} < C^S$, the welfare achieved by each successive generation falls through time for $t < \tilde{t}^S$ – due to the warming climate implied by the rising level of atmospheric carbon as in Case 1, but in contrast to Case 1 also due to the decline in consumption implied by the falling emissions through time – until the brink generation \tilde{t}^S is reached, at which point for this generation and contrary to Case 1, both global emissions and the global carbon stock are frozen in place and the decline in welfare is halted thereafter.

Also in contrast to Case 1, in Case 2 an increase in the social discount factor $\hat{\beta}$ shifts utility to later generations both by slowing the accumulation of atmospheric carbon and keeping the planet cooler for longer *and* by tilting the emissions profile away from the earliest generations. Finally, note from (2.20) that, contrary to the noncooperative and ICA outcomes, when the brink of climate catastrophe is reached under the socially optimal level of emissions, the brink generation does not suffer a discrete drop in welfare relative to the previous generation.

We summarize the properties of the socially optimal emissions choices with:

Proposition 3. *The socially optimal path of emissions in the common-brink model falls into one of two cases. If \tilde{C} is above a threshold level, the brink of catastrophe is never reached and the socially optimal emissions levels are constant through time. Otherwise, if \tilde{C} is below this threshold the socially optimal emissions levels decline through time until the brink of catastrophe is reached, and for the brink generation and all generations thereafter the emissions remain at the replacement rate dictated by the natural rate of atmospheric regeneration. In this second case where the brink is reached, the brink generation does not suffer a discrete drop in welfare relative to the previous generation.*

2.5. Comparison of ICA and Socially Optimal Outcomes

We now compare the outcomes that are achieved under the ICA with the socially optimal outcomes that would be chosen by the planner. To this end, we begin by noting that we have $C^S < C^{ICA}$ and hence $C^S < C^{ICA} < C^N$. We can thus organize the comparison between the ICA and socially optimal outcomes into three ranges of \tilde{C} : high ($\tilde{C} \in [C^{ICA}, C^N)$), intermediate ($\tilde{C} \in [C^S, C^{ICA})$) and low ($\tilde{C} < C^S$).

Consider first the possibility that \tilde{C} falls in the high range $\tilde{C} \in [C^{ICA}, C^N)$. In this case the world will be kept below the brink of climate catastrophe by both the ICA and the planner through the implementation of constant emissions levels \bar{c}^{ICA} and \bar{c}^S respectively that are below the BAU level \bar{c}^N and that keep the global carbon stock below \tilde{C} . However, the planner dictates that the emissions choices \bar{c}^S internalize *all* the external effects of those choices, both international and intergenerational, while under the ICA emissions choices \bar{c}^{ICA} only the international climate externalities are internalized; and as a result we have $\bar{c}^S < \bar{c}^{ICA}$, with \bar{c}^S dropping further below \bar{c}^{ICA} as $\hat{\beta}$ increasing and as ρ decreases, and the steady state carbon stock delivered under the ICA is larger than the socially optimal level. The three panels of Figure 1 illustrate the time path of emissions, the global carbon stock, and the utility of a representative country

under the ICA and socially optimal emissions as well as in the noncooperative equilibrium for \tilde{C} in this range. The qualitative features of the ICA and socially optimal outcomes are similar, with the difference between the two being that the planner shifts welfare from early generations to later generations relative to the ICA by requiring lower emissions for all generations and thereby reducing the extent to which utility falls through time due to a rising global carbon stock and worsening climate.²⁶

Consider next the possibility that \tilde{C} falls in an intermediate range $\tilde{C} \in [C^S, C^{ICA})$. In this case the world would still be kept from the brink of climate catastrophe by the planner, but under the ICA the world will be brought to the brink. This is because with \tilde{C} in this intermediate range, the planner's choice of emissions \bar{c}^S is still low enough to keep the global carbon stock below \tilde{C} , but the higher level of emissions \bar{c}^{ICA} implemented during the warming phase of the ICA is no longer low enough to accomplish this. Hence, in this case the inability of the ICA to take into account directly the interests of future generations leads to a qualitative difference across the ICA and socially optimal outcomes. This is reflected in the three panels of Figure 2. As in Figure 1, here the utility of earlier generations is higher and the utility of later generations is lower under the ICA than in the social optimum, but now utility under the ICA falls precipitously for the generation alive when the brink is reached, while under the social optimum the utility of each generation evolves continuously through time. And while in this case the planner would not let utility for any generation fall to the level of utility experienced in the noncooperative equilibrium by the brink generation, under the ICA the generation alive when the brink is reached and all future generations will experience exactly that level of utility.

Finally, consider the possibility that \tilde{C} falls in the low range $\tilde{C} < C^S$. In this case the world will be brought to the brink of climate catastrophe by both the ICA and the planner, but as noted we have $\tilde{t}^N < \tilde{t}^{ICA} < \tilde{t}^S$: the ICA slows down the march to the brink relative to the noncooperative outcome, but this march is still too fast relative to the social optimum. In this case as well there are qualitative differences across the ICA and socially optimal outcomes that arise as a result of the inability of the ICA to take into account directly the interests of future generations. This is reflected in the three panels of Figure 3. In particular, here the

²⁶We have depicted the level of welfare achieved by early generations in Figure 1 as dropping under the social optimum relative to the noncooperative equilibrium, but this need not be so. If $\hat{\beta}(1 - \rho)$ is sufficiently small, the planner will raise the level of welfare achieved by the early generations as well relative to the noncooperative equilibrium, because then the planner is essentially internalizing international but not intergenerational externalities and hence mimics the ICA outcome, which provides (weakly) higher than Nash welfare for every generation.

ICA emissions remain constant at the level \bar{c}^{ICA} during the warming phase leading up to the brink and then fall precipitously to the level \hat{c}^N for the brink generation, implying an associated precipitous drop in the welfare of the brink generation relative to the previous generation. By contrast, under the socially optimal choices the emissions \hat{c}_t^S during the warming phase decline smoothly over time, and they reach the level \hat{c}^N at the brink without a discrete drop for the brink generation in either emissions or utility.

Summarizing, we may now state:

Proposition 4. *In the common-brink model, the ICA addresses the horizontal (international) externalities that are associated with emissions choices and that create inefficiencies in the noncooperative outcomes, but it cannot address the vertical and diagonal externalities that are associated with the intergenerational aspects of the climate problem. For this reason, the ICA slows down the growth of the carbon stock relative to the noncooperative outcome but not enough relative to the social optimum. More specifically: (i) If \tilde{C} is above a threshold level the ICA prevents the world from reaching the brink of catastrophe, but the steady-state carbon stock is still too large relative to the social optimum. (ii) If \tilde{C} lies in an intermediate range the world reaches the brink of catastrophe under the ICA but not under the social optimum. (iii) If \tilde{C} is below a threshold level the brink is reached both under the ICA and the social optimum, but it is reached faster under the ICA.*

It is interesting to reflect more broadly on the role of an ICA. According to the common-brink model, it is not the possibility of catastrophe that generates a role for an ICA. Rather, there is a significant role for an ICA only insofar as there are significant costs of global warming before the brink of catastrophe is reached. Indeed, if λ were zero there would be no role for an ICA according to our model. This might seem surprising, since an ICA is able to address horizontal externalities, and the possibility of catastrophe does imply extreme horizontal externalities once the world reaches the brink. But at the brink, these extreme international externalities are coupled with extreme *internalized* costs of increasing emissions, and this makes ICAs redundant as a means to avoid catastrophe *once the catastrophe is at hand*, because at that point countries have sufficient incentives to avoid catastrophe even in the noncooperative scenario.²⁷

²⁷While an ICA plays no role in the avoidance of catastrophe once the catastrophe is at hand, recall that it may improve the cross-country allocation of the costs of avoiding catastrophe, in case the countries do not focus on the efficient noncooperative equilibrium (see footnote 24).

It is also natural to wonder how the ICA affects future generations. This is not obvious *a priori*, because for each generation t the ICA is a contract that excludes future generations, and because we are focusing on a scenario without any intergenerational altruism. The answer is that in the common-brink model an ICA nevertheless benefits future generations. This is because the act of reducing emissions today under an ICA has two positive effects on future generations: first, it will leave the next generation with a lower global carbon stock, and hence reduce the environmental losses tomorrow; and second, it will at least to some extent slow down the march to the brink of climate catastrophe, and therefore put off the day of reckoning when emissions and hence consumption levels will need to fall precipitously to save the world.

Finally, returning to Figures 1-3, we may ask which of the three cases depicted in these figures most accurately reflects the true limitations faced by ICAs due to their inability to take into account directly the interests of future generations. According to the common-brink model, the answer to this question depends on the severity of the constraint that the catastrophic carbon level \tilde{C} places on attainable steady state welfare. If one takes an agnostic view regarding the relative empirical plausibility of these three scenarios, the message from the model is that the world is more likely to reach the brink of catastrophe under the ICA than under the global social planner; or more specifically, that under the ICA the world reaches the brink of catastrophe for a larger parameter region than under the planner. This is an immediate corollary of Proposition 4, which states that, fixing all other model parameters, the interval of \tilde{C} for which the world reaches the brink is wider under the ICA than under the planner.

But something more can be said if one is willing to rule out a dystopian view of the world in which the planner would find it optimal to allow the world to arrive at the brink of catastrophe and then remain on the brink thereafter, that is, the scenario described by Figure 3. This scenario can be ruled out for any given level of \tilde{C} if the planner's discount factor accounting for the natural rate of atmospheric regeneration, $\hat{\beta}(1 - \rho)$, is sufficiently close to one. And while the most optimistic position on the severity of the constraint that \tilde{C} places on attainable steady state welfare would point to Figure 1, recall that this scenario can only apply if the cost associated with moderate degrees of global warming, λ , is above a certain threshold; so if λ is sufficiently small also Figure 1 can be ruled out.²⁸ When this is the case it is then only the

²⁸More formally, the case in Figure 3 is ruled out if $1 - \hat{\beta}(1 - \rho) < \frac{\lambda M}{B'(\frac{\rho \tilde{C}}{M})}$, and the case in Figure 1 is ruled out if $\lambda < \frac{1}{M} B'(\frac{\rho \tilde{C}}{M})$. Assuming $B'(0)$ is finite, and fixing λ below the threshold $\frac{1}{M} B'(\frac{\rho \tilde{C}}{M})$, then if $\hat{\beta}(1 - \rho)$ is close to one we are in the case of Figure 2.

middle ground associated with Figure 2 that remains. And according to Figure 2, as we have noted, the implications of the inability of ICAs to take into account directly the interests of future generations are especially dire: while a global social planner would keep the world from ever arriving at the brink of climate catastrophe, an ICA will at best only postpone the arrival at the brink, and when that day arrives, the brink generation and all generations thereafter will suffer a precipitous drop in welfare. Hence we may state:

Corollary 1. *If the cost associated with moderate degrees of global warming, λ , is sufficiently small, and if the planner’s discount factor accounting for the natural rate of atmospheric re-generation is close enough to one, the world will reach the brink of catastrophe under the ICA but not under the social optimum.*

Like the fabled boiling frog, Corollary 1 suggests that a slowly rising cost of climate change (small λ) may describe the scenario most likely to cause the world to “remain oblivious” during the warming phase, and thereby arrive at the brink of climate catastrophe under an ICA when a global social planner would not have allowed this to happen. The twist is that, unlike the frog in the fable, the world will not go off the brink under these conditions; but it will be consigned to life on the brink, a fate that the more forward-looking actions of the planner would have avoided.

This is also the case where our model may best capture in a highly stylized way the essence of the plight of the climate activist Greta Thunberg and her generation, and how their plight can be interpreted through the lens of our model as arising from the inability of climate agreements to internalize intergenerational externalities. In an address to world leaders at the United Nation’s Climate Action Summit in New York City on September 23 2019, Thunberg stated:

“This is all wrong. I shouldn’t be up here. I should be back in school on the other side of the ocean. Yet you all come to us young people for hope. How dare you! You have stolen my dreams and my childhood with your empty words. And yet I’m one of the lucky ones. People are suffering. People are dying. Entire ecosystems are collapsing. We are in the beginning of a mass extinction, and all you can talk about is money and fairy tales of eternal economic growth. How dare you! For more than 30 years, the science has been crystal clear. How dare you continue to look away and come here saying that you’re doing enough, when the politics and solutions needed are still nowhere in sight. ... The popular

idea of cutting our emissions in half in 10 years only gives us a 50% chance of staying below 1.5 degrees [Celsius], and the risk of setting off irreversible chain reactions beyond human control. Fifty percent may be acceptable to you ... [but it] is simply not acceptable to us — we who have to live with the consequences. To have a 67% chance of staying below a 1.5 degrees global temperature rise – the best odds given by the [Intergovernmental Panel on Climate Change] – the world had 420 gigatons of CO₂ left to emit back on Jan. 1st, 2018. Today that figure is already down to less than 350 gigatons. How dare you pretend that this can be solved with just ‘business as usual’ and some technical solutions? With today’s emissions levels, that remaining CO₂ budget will be entirely gone within less than 8 1/2 years. There will not be any solutions or plans presented in line with these figures here today, because these numbers are too uncomfortable. ... You are failing us. But the young people are starting to understand your betrayal. The eyes of all future generations are upon you. And if you choose to fail us, I say: We will never forgive you. We will not let you get away with this. Right here, right now is where we draw the line. The world is waking up. And change is coming, whether you like it or not.”

In terms of Figure 2, we might think of the sum total of climate agreements to date as putting the world somewhere between the noncooperative emissions path (if these agreements were completely ineffective) and the ICA path (if the agreements were maximally effective), and we might think of Greta and her generation as corresponding to the brink generation (marked in Figure 2 by \tilde{t}^N in the former case and \tilde{t}^{ICA} in the latter case). The “consequences” to which Greta refers in the quote above are then reflected in Figure 2 by the implication of the threat of a climate catastrophe experienced in her lifetime, a threat caused by the emissions of previous generations that the brink generation must now confront and that might have been avoided if previous generations had adopted socially optimal emissions policies. According to our common-brink model, the implication of this threat is that the world will indeed find an 11th hour solution which prevents the threat of climate catastrophe from materializing, much as Greta predicts. But as Figure 2 depicts, avoiding climate catastrophe at the 11th hour comes at the cost of a precipitous drop in utility for the brink generation relative to their parents, and the same low level of utility for all generations thereafter. And this solution comes about not because ICA’s will finally become more effective and deliver, but because the consequences of *not* finding a solution are so dire for the brink generation that the externalities that interfered with the attempts of previous generations to address the problem are no longer an impediment

to its solution, and at that point an ICA is not even needed.

3. The Fate of the Maldives

Thus far we have assumed that the level of the carbon stock that would be catastrophic, \tilde{C} , is the same for all countries. And under this assumption, we have argued that even in a noncooperative equilibrium catastrophe will be averted, because countries will find a way to do whatever it takes to prevent mutual collapse. But what if the global carbon stock at which a catastrophe would be triggered differs from one country to the next? And what if not all (or even any) countries have the capacity to avoid collapse on their own? For example, it is often observed that small island nations such as the Maldives are especially vulnerable to the effects of climate change and may soon face an existential threat posed by rising sea levels.²⁹ If some countries face existential threats from climate change before others, new questions arise. Under what conditions will some (or even all) of the countries collapse on the noncooperative equilibrium path? Can there be domino effects, where the collapse of one country hastens the collapse of the next? If some or all countries would collapse in the noncooperative equilibrium, can ICAs help to avoid collapse? And what is the outcome that a global social planner would implement in this case?

In this section we allow countries to reach a catastrophe at different levels of the global carbon stock, so that we may answer the questions posed above. To this end we now index countries by i and assume that country i reaches the brink of collapse when C rises to the level \tilde{C}_i ; and we order countries according to increasing \tilde{C}_i , so that country 1 is the country with the lowest \tilde{C}_i and therefore the country “least resilient to climate change,” and where we assume for simplicity that the ordering is strict, i.e. no two countries have the same value of \tilde{C}_i . To focus sharply on the implications of heterogeneous collapse points, we continue to assume that countries are symmetric with respect to all other model parameters. We will sometimes refer to this model as the “heterogeneous-brink model.”³⁰

²⁹See also Jones and King (2021), who develop a methodology for predicting a shortlist of countries that are most likely to be left standing when other countries have succumbed to climate catastrophe.

³⁰While we focus formally on heterogeneity *across* countries, it is worth noting that many of the same issues that we consider below will arise *within* countries, if there is heterogeneity in climate collapse points across different regions due to distinct geographical and/or socioeconomic features. Such regional heterogeneity raises issues for federal versus regional government emissions policy choices that are analogous to the issues we identify below for global planner/ICA versus noncooperative national emissions policy choices (see for example Lustgarten, 2020). We leave an exploration of these parallel themes to future work.

To facilitate our analysis of heterogeneous collapse points, we also assume that the cost of a country’s collapse is high but possibly finite. We have in mind that if a country were to collapse, its citizens would become climate refugees and have to relocate to other countries that have not yet collapsed.³¹ Hence, when a country collapses, the population of the world remains unchanged at M , but the population of the collapsing country must immigrate to the remaining countries.

Even absent outright collapse, the link between climate change and refugees can already be seen for example in the recent surge of migrants at the southern U.S. border, as Russonello (2021) writes in the *New York Times*:

“The main motivators of emigration from Mexico, Central America and points south are tied to climate change, violent crime and corruption – all issues that the Biden administration knows it must confront if it stands any chance of stemming the inflow of people at the border. (...) The most immediate cause of the immigration surge may be the series of deadly hurricanes that swept through Central America last year, part of a greater trend fueled by climate change. They destroyed crops and homes, especially in Honduras, leaving an estimated nine million people displaced. They’ve had six years of ongoing droughts in these areas, they have no food, no means for employment or livelihood, and they’re eating the seeds which they would normally save for planting.”

The reduced-form modeling of the costs of a country’s collapse that we adopt below is meant to capture the costs that are incurred when a climate catastrophe makes a country uninhabitable and its citizens must seek residency elsewhere.

In particular, we suppose that there are two costs associated with a country’s collapse. A first cost is borne by the citizens of the collapsing country itself: we denote by L the one-time per-capita utility cost borne by each of these citizens in the period of collapse. We will think of collapse as occurring at the end of a period, after all consumption for the period has occurred, and we will think of L as high, but finite as long as there are other surviving countries to which the citizens of the collapsing country can immigrate.

The second cost of a country’s collapse is borne by the remaining countries. We assume that the collapsing country’s citizens become climate refugees and are spread equally across the

³¹Another potential cost would be the destruction of international trade between the collapsed country and the surviving countries, a possibility we return to briefly in the Conclusion.

remaining countries of the world; and we assume that each climate refugee imposes a one-time utility cost r on the country to which it immigrates. We do not need to assume that r is positive (one can easily imagine situations where some countries may benefit from the collapse of another country), but it is arguable that in practice the consequences for other countries of a country's collapse are likely to be negative, so we will focus on this case in what follows.

Letting H_t index the most vulnerable country that has survived to time t , the number of surviving countries at t is $M - H_t + 1$, and the population of a surviving country at t is $\frac{M}{M-H_t+1}$. If country H_t collapses, since the total population of the remaining countries is $M - \frac{M}{M-H_t+1} = \frac{M(M-H_t)}{M-H_t+1}$, it follows that the one-time per-capita utility cost incurred by citizens of the remaining countries as a result of country H_t 's collapse is

$$R_{H_t} \equiv \frac{r}{M - H_t}.$$

As with L , we assume that the refugee cost R_{H_t} is incurred at the end of the period of country H_t 's collapse.³²

Notice that the refugee externality R_{H_t} is increasing in H_t . This is because in our model countries that collapse later (higher H_t) release a greater number of climate refugees ($\frac{M}{M-H_t+1}$) on a smaller rest-of-world population ($M - \frac{M}{M-H_t+1}$). But as will become clear below, our results would be unchanged if the refugee externality were independent of H_t .

To preserve tractability, we assume that both L and R_{H_t} enter utility in an additively separable way (so that they do not impact the emissions choices of countries and instead act as simple shifters of utility). With this assumption, the utility of a citizen living in country i at time t is given by:

$$u_{i,t} = B(c_{i,t}) - \lambda C_t - L \cdot E_{i,t} - R_{H_t} \cdot I_{i,t}, \quad (3.1)$$

where $E_{i,t}$ is an indicator function that equals one if country i collapses at time t , so that its population has to emigrate, and $I_{i,t}$ is an indicator function that equals one if some country other than country i collapses at time t , so that country i receives immigrants for the collapsing country. And of course, the utility function in (3.1) is defined only for the countries that have survived up to time t .

Recall that in the common-brink model of section 2, the population of each country remained constant over time and we normalized this population to one, so country-level and per-capita-

³²Since time-periods correspond to generations in our model, it seems reasonable to assume that L and R_{H_t} are incurred as one-time costs, but at the end of this section we also discuss briefly the case of permanent per-period costs.

level variables were one and the same. But with climate refugees altering the population of surviving countries when more vulnerable countries collapse, we now specify $u_{i,t}$ as the per-capita utility of a person living in country i in period t , as (3.1) reflects.³³ Note that we continue to interpret $c_{i,t}$ as the per-capita emissions of country i in period t , and $B(c_{i,t})$ as the per-capita benefit from consumption that comes from emitting at the per-capita level $c_{i,t}$.

The carbon stock now evolves according to

$$C_t = (1 - \rho)C_{t-1} + \sum_{i=H_t}^M \frac{M}{M - H_t + 1} \cdot c_{i,t} \text{ with } C_{-1} = 0. \quad (3.2)$$

Notice that according to (3.2) we now have $\frac{\partial C_t}{\partial c_{i,t}} = \frac{M}{M - H_t + 1}$: the impact of a surviving country's per-capita emissions on the carbon stock C_t grows as the number of surviving countries shrinks, because country population grows due to the absorption of climate refugees.

Finally, since there may be multiple equilibria, we assume in what follows that, if there are Pareto-rankable equilibria, countries will focus on a Pareto-undominated equilibrium.

3.1. Noncooperative Equilibrium

We first characterize the noncooperative emissions choices. As in the previous section, we focus on Markov perfect equilibria. In principle now we have two state variables, C_t and H_t , but the latter is pinned down by the former in a straightforward way, so we can continue to regard C_t as the only state variable.

Given $\beta = 0$, it is easy to see that along the noncooperative path the world may pass through three possible phases: a warming phase, where warming takes place but no catastrophes occur; a catastrophe phase, where warming continues and a sequence of countries collapse; and a third phase where warming and catastrophes are brought to a halt. The first and third phases are familiar from the analysis of the previous section where a common catastrophe point across countries was assumed; the possibility of a middle phase in which some countries collapse along the noncooperative path is novel to the current setting where catastrophe points differ across countries. Notice, too, that with $\frac{\partial C_t}{\partial c_{i,t}} = \frac{M}{M - H_t + 1}$, the BAU per-capita emissions of a surviving country now depend on H_t , so we write them as $\bar{c}_{H_t}^N$. These are implicitly defined by

$$B'(\bar{c}_{H_t}^N) = \frac{M}{M - H_t + 1} \cdot \lambda. \quad (3.3)$$

³³In writing (3.1) we have implicitly assumed that no more than one country could collapse in a given period t , but it is straightforward to accommodate the possibility of multiple-country collapses in a given period at the expense of additional notation.

From (3.3), the BAU per-capita emissions is the same across all surviving countries (so we can omit the country subscript i) but falls as the number of surviving countries shrinks (H_t rises) and the population of each surviving country rises. This simply reflects the fact that with each country collapse there are fewer remaining countries in the world; and the countries that do remain internalize a greater proportion of the global cost of their BAU emissions choices. Since the world population remains constant at M , the world BAU emissions level $M\bar{c}_{H_t}^N$ also falls with each country collapse.

To avoid uninteresting taxonomies we assume that, if a country is at its brink and there are other surviving countries, the former is not able to “save itself” by cutting its own emissions to zero if the latter choose their BAU emissions. In essence we are assuming that a country is not able to unilaterally stop the growth of the global carbon stock unless it is the lone surviving country, a feature that seems empirically plausible. Formally, since the BAU emissions $\bar{c}_{H_t}^N$ decline as H_t increases, a sufficient condition for this assumption to hold is

$$\frac{M}{2}\bar{c}_{M-1}^N > \rho\tilde{C}_{M-1}, \quad (3.4)$$

a restriction that we will maintain throughout.³⁴

To develop some intuition for how the noncooperative path is determined in this setting, it is useful first to focus on the case where $r = 0$ so that there are no refugee externalities. In this case the equilibrium path of the noncooperative game is simple, and it provides a sharp counterpoint to the equilibrium of the noncooperative game in the common-brink model of the previous section.

After an initial warming phase during which there are no catastrophes and each country selects the BAU emissions level \bar{c}_1^N , the world enters a catastrophe phase when country 1 arrives at the brink of collapse. This occurs in finite time if $M\bar{c}_1^N > \rho\tilde{C}_1$, which is implied by condition (3.4). Furthermore, condition (3.4) implies that country 1 is not able to offset the BAU emissions of the rest of the world and save itself – and with $r = 0$, the remaining countries have no reason to help country 1 survive – and hence country 1 will collapse. And by a similar logic, all countries except for the most resilient one ($i = M$) will collapse on the equilibrium path. The most resilient country is guaranteed to survive on the equilibrium path, because by

³⁴To understand condition (3.4) note that, when $H_t = M - 1$, there are only two surviving countries with population $M/2$ in each, so if country M chooses its BAU per-capita emissions \bar{c}_{M-1}^N and country $M - 1$ chooses zero emissions, total world emissions are $(M/2)\bar{c}_{M-1}^N$, and if this level exceeds $\rho\tilde{C}_{M-1}$ then country $M - 1$ will collapse once its brink is reached. And if country $M - 1$ is not able to “save itself,” neither can the more vulnerable countries ($k < M - 1$).

setting $c = \frac{\rho \tilde{C}_M}{M} \geq 0$ it can freeze the global carbon stock at the brink level \tilde{C}_M . Note that we have implicitly assumed that L is large enough to induce country M to do whatever is necessary to avoid its own collapse.³⁵

Hence, with heterogeneous brinks and $r = 0$ all countries except the least vulnerable one will collapse along the equilibrium path, provided an individual country is not able to fully offset the other countries' BAU emissions. This scenario provides an illuminating contrast to the common-brink case, where once the world reached the brink of catastrophe countries did whatever was necessary to avoid global collapse. Relative to that setting, the difference here is that each country has its *own* brink generation, who faces the existential climate crisis *alone* and up against the other countries, who do not internalize the impact of their emissions on the fate of the brink country. Notice also that even slight differences in collapse points across countries can lead to collapses along the equilibrium path: unless each country arrives at the brink in the same moment, the “we are all in this together” forces that enabled the world to avoid collapse in the noncooperative equilibrium of our common-brink model will be disrupted. As we next demonstrate, allowing for climate refugee externalities will bring back an element of these forces, albeit only partially.

To proceed, we now allow for climate refugee externalities ($r > 0$).

The game can be solved in two steps. First, for each level of C_t we characterize the equilibrium emissions choices $c_i^N(C_t)$ in each surviving country. And second, we derive the implied equilibrium path for C_t and hence for the set of countries that survive to each t . Given the absence of intergenerational altruism ($\beta = 0$), for each level of C_t we effectively have a one-shot game, of which we will study the Nash equilibria (or simply the “equilibria”).

Consider first what happens if no country is on the brink of catastrophe ($C_t \neq \tilde{C}_k$ for all k). In this case, clearly each country chooses its BAU emissions $\bar{c}_{H_t}^N$ defined by (3.3).

Next consider what happens if some country k is on the brink of catastrophe ($C_t = \tilde{C}_k$). Countries $j \geq k$ have survived up to this point, each with a population of $\frac{M}{M-k+1}$. Clearly, then, country k will survive if and only if $\frac{M}{M-k+1} \sum_{j=k}^M c_j \leq \rho \tilde{C}_k$. We now argue that there are two possible types of equilibrium.

³⁵This assumption seems natural in the context of this model, especially given that if country M collapses its citizens have nowhere left to go, so the world effectively ends. But if lags in the effects of emissions are introduced and even if there is some degree of intergenerational altruism, it might be reasonable to consider the possibility that a country would opt to enjoy a higher standard of living now even if it meant that generations further in the future would face catastrophe. We discuss the possibility of lags in the effects of emissions in the Conclusion.

The first possibility is that all countries choose the BAU emissions level \bar{c}_k^N and country k does not survive. This is always an equilibrium, because given condition (3.4), country k cannot save itself if the other countries choose \bar{c}_k^N . Also note that, if the other countries choose \bar{c}_k^N , country k 's best response is to also choose \bar{c}_k^N , because it gets to enjoy the benefit of its current-period emissions before it collapses, and therefore, given that collapse is inevitable, it can do no better than to choose its BAU emissions.

The second possibility is that country k survives, and the countries' emissions levels satisfy $\frac{M}{M-k+1} \sum_{j=k}^M c_j = \rho \tilde{C}_k$. Intuitively, this type of equilibrium can exist only if the refugee externality is large enough, so that countries $j > k$ have an incentive to "top off" the mitigation efforts of country k , ensure country k 's survival, and avoid a climate refugee crisis. It is easy to show that equilibria with survival of country k exist if and only if it is an equilibrium for country k to do everything feasible to save itself by setting $c_k = 0$ and for the remaining countries ($j > k$) to top off this effort in a symmetric way.³⁶ Note that, given $c_k = 0$, the maximum symmetric emissions level for each country $j > k$ that ensures survival of country k is $\frac{\rho \tilde{C}_k}{M-k}$, and since the population of each country at this stage is $\frac{M}{M-k+1}$, the corresponding per capita emissions level is $\frac{M-k+1}{M} \cdot \frac{\rho \tilde{C}_k}{M-k} \equiv c_k^{save} < \bar{c}_k^N$. We will refer to this as the "self-help" equilibrium. This is arguably the most intuitive equilibrium among those in which country k survives, because it minimizes the less vulnerable countries' gain from deviating. In what follows, if there are equilibria where country k survives, we will restrict our focus to the self-help equilibrium, but our main results do not depend on this equilibrium selection assumption.³⁷

We now write down the conditions under which the self-help equilibrium described above arises. The no-defect condition for a country $j > k$ can be written as follows:

$$G_k \equiv \left[B(\bar{c}_k^N) - \lambda \left(\tilde{C}_k + \frac{M}{M-k+1} \cdot (\bar{c}_k^N - c_k^{save}) \right) \right] - \left[B(c_k^{save}) - \lambda \tilde{C}_k \right] \leq R_k \quad (3.5)$$

The left-hand side of (3.5) is the gross per-capita gain to a country $j > k$ were it to defect from the self-help equilibrium and cause country k to collapse. The term in the first square brackets

³⁶To see this, note that (i) raising c_k above zero would require reducing emissions for some country $j > k$, and this would increase the latter country's temptation to defect; and (ii) assigning asymmetric emissions to countries $j > k$ would make it harder to sustain such an equilibrium, because it would increase the temptation to defect for the countries that have a bigger burden.

³⁷It is worth noting that the self-help equilibrium does not maximize the joint payoff of the surviving countries, since it does not equalize their marginal benefit of emissions ($B'(\cdot)$); but also note that the equilibria with survival of country k (if they exist) are not Pareto-rankable, since international transfers are not used in a noncooperative equilibrium. This suggests that, if countries indeed focus on the self-help equilibrium, one of the potential roles of an ICA will be to allow countries to move to the efficient allocation of emissions through the use of transfers. We will come back to this point in the next subsection.

is the per-capita payoff to a country $j > k$ were it to deviate to the BAU emissions level \bar{c}_k^N and cause country k to collapse, and the term in the second square brackets is the per-capita payoff to a country $j > k$ under the self-help equilibrium emissions levels in which country k does not collapse. Note that, when defecting, country j 's emissions increase by the amount $\frac{M}{M-k+1} (\bar{c}_k^N - c_k^{save})$, and this causes the carbon stock to exceed \tilde{C}_k by the same amount. The right-hand side is the per-capita cost that country $j > k$ incurs as a result of k 's collapse. Thus, as suggested by intuition, an equilibrium with survival of country k exists only if the refugee externality R_k that the collapse of country k would impose on the remaining countries is large enough.

The no-defect condition for country k can be written as:

$$G_k^0 \equiv \left[B(\bar{c}_k^N) - \lambda \left(\tilde{C}_k + \frac{M}{M-k+1} \cdot \bar{c}_k^N \right) \right] - \left[B(0) - \lambda \tilde{C}_k \right] \leq L \quad (3.6)$$

The left-hand side of (3.6) is country k 's per-capita gain in defecting from the self-help equilibrium, in terms of the additional welfare it enjoys for the period up until the moment of its collapse. The term in the first square brackets is the per-capita welfare that country k would enjoy were it to deviate to the BAU emissions level \bar{c}_k^N , and the term in the second square brackets is its per-capita welfare under the self-help equilibrium emissions levels. The term on the right-hand side is the per-capita cost that the collapse of country k , precipitated by its own defection from the self-help equilibrium, would impose on its citizens at the end of the period.

An equilibrium with survival of country k exists if and only if both (3.6) and (3.5) are satisfied. Notice that, if this is the case, the equilibrium where country k collapses (described as the first possibility above) is Pareto-dominated.³⁸ It can also be shown that there are no other possible equilibria in addition to the ones we just described. Thus, given our assumption that countries focus on Pareto-undominated equilibria, we can conclude that if country k is at the brink of catastrophe ($C_t = \tilde{C}_k$), this country survives in equilibrium if and only if (3.6) and

³⁸To see this, consider first country k . Since (3.6) is satisfied, country k prefers to emit zero and survive rather than deviating from the self-help equilibrium, emitting \bar{c}_k^N , and collapsing at the end of the period. But k 's payoff would be smaller still if all countries emitted at the level \bar{c}_k^N , as would be true in the equilibrium where country k collapses, because the carbon stock would be larger under the latter scenario. So country k is better off in the self-help equilibrium than in the equilibrium where country k collapses. What about a country $j > k$? The argument is similar. If the self-help equilibrium exists, then we must have (3.5), and hence country j prefers the self-help equilibrium to deviating from c_k^{save} , emitting \bar{c}_k^N , and having country k collapse at the end of the period. But j 's payoff would be smaller still if all countries emitted at the level \bar{c}_k^N as would be true in the equilibrium where country k collapses, because the carbon stock would be larger under the latter scenario. So country $j > k$ is better off in the self-help equilibrium than in the equilibrium where country k collapses. The claim then follows.

(3.5) are satisfied.

Finally note that, if $C_t = \tilde{C}_M$, so that the only surviving country $j = M$ is at the brink of catastrophe, this country will restrain its emissions just enough to avoid collapse. Since at this stage the entire world population M is located in this country, its per capita emissions are given by $c_M^N(\tilde{C}_M) = \frac{\rho \tilde{C}_M}{M}$. If the world reaches this stage, the carbon stock stops growing and will stay at the level \tilde{C}_M forever thereafter.

Our next observation is that the gains from defection from the self-help equilibrium, G_k and G_k^0 , decrease with the number of countries that have collapsed in the past, and hence with k . Consider first G_k . Intuitively, as more and more countries collapse, the gain from defection G_k becomes smaller, for three reasons. First, there are fewer surviving countries and hence more people in any country $j > k$ that considers defection, implying that a defection has greater (negative) consequences for the global climate. Second, there are fewer surviving countries and hence more people in the “brink country” (k) emitting zero carbon, and this increases c_k^{save} ; and third, \tilde{C}_k increases, and this also increases c_k^{save} . The fact that c_k^{save} increases through these last two channels means that countries $j > k$ can afford to emit more while still saving country k , thus the sacrifice needed to save country k becomes smaller. A similar argument applies to G_k^0 , except that only two of the three effects described above are at work, since c_k^{save} is replaced by 0.³⁹

Since G_k and G_k^0 are decreasing in k and R_k is increasing in k , we can say that, conditional on a given country k being at the brink, this country is more likely to survive if more countries have collapsed in the past. The following lemma summarizes the key features of the equilibrium outcome conditional on country k being on the brink:

Lemma 1. *Conditional on country k being on the brink ($C_t = \tilde{C}_k$): (i) If (3.5) and (3.6) are satisfied, country k survives with the help of emissions reductions below BAU levels by*

³⁹Formally, rewrite G_k as $G_k = \left[B(\bar{c}_k^N) - \lambda \frac{M}{M-k+1} \bar{c}_k^N \right] - \left[B(c_k^{save}) - \lambda \frac{M}{M-k+1} c_k^{save} \right]$. Note that the term in the first square brackets is maximized by \bar{c}_k^N , so we can apply the envelope theorem and write:

$$\frac{dG_k}{dk} = -\lambda \frac{M}{(M-k+1)^2} (\bar{c}_k^N - c_k^{save}) - \frac{d}{dc_k^{save}} \left(B(c_k^{save}) - \lambda \frac{M}{M-k+1} c_k^{save} \right) \cdot \frac{dc_k^{save}}{dk}. \quad (3.7)$$

The first term on the right-hand side of (3.7) is negative, since $\bar{c}_k^N > c_k^{save}$. Turning to the second term and recalling that $c_k^{save} = \frac{M-k+1}{M} \cdot \frac{\rho \tilde{C}_k}{M-k}$, it is easy to check that $\frac{dc_k^{save}}{dk} > 0$. Next note that $\frac{d}{dc_k^{save}} (\cdot) > 0$ because c_k^{save} is lower than \bar{c}_k^N , the emissions level that maximizes $B(c) - \lambda \frac{M}{M-k+1} c$. This ensures that the second term on the right-hand side of (3.7) is negative as well. We can conclude that the derivative in (3.7) is negative, and hence $\frac{dG_k}{dk} < 0$ as claimed. A similar argument can be applied to show that also G_k^0 is decreasing in k .

other countries; (ii) If (3.5) or (3.6) is violated, country k collapses and the surviving countries continue to choose their BAU emissions. (iii) Country k is more likely to survive if more countries have collapsed in the past.

The above analysis raises an interesting question: Is there a “domino effect” in our model when countries collapse on the equilibrium path? At one level the answer is yes, for the simple reason that if a country at the brink is able to avert its own collapse (with the help of other countries), no more dominos will fall. In other words, a given country i can reach the brink only if all the countries that are more vulnerable than country i (that is countries $1, 2, \dots, i - 1$) have already collapsed. In this sense our model exhibits a domino effect. However, as Lemma 1 and the preceding discussion highlights, *conditional on a country reaching the brink* the likelihood of collapse is lower if more countries have collapsed in the past, so in this sense there is also an “anti-domino effect” in our model.⁴⁰

Having characterized the equilibrium emissions conditional on the global carbon stock C_t , it is straightforward to back out the implied equilibrium path for C_t and hence for the set of countries that survive to each t . In the initial phase, all countries are present and the growth of C_t is dictated by the BAU emissions level \bar{c}_1^N . Once C_t reaches the level that endangers country 1 (\tilde{C}_1), if the refugee externality that the collapse of country 1 would exert on a representative citizen of the rest of the world is not severe enough ($R_1 < G_1$), then country 1 collapses at the end of the period and the rest of the world carries on with their BAU emissions level \bar{c}_2^N . In a similar fashion, the growth of the carbon stock will cause the sequential collapse of further countries.

Note that our model allows for the internal cost of collapse L to be low enough that country 1 itself prefers to collapse rather than cutting emissions to zero, which would *a fortiori* imply that country 1 collapses in equilibrium, but empirically it seems plausible that L is high enough that the no-defect condition for the country at the brink (3.6) is not binding. For this reason we emphasize this case in our discussion, but all of our results are valid more generally.

The condition $R_1 < G_1$, under which a subset of countries collapses on the noncooperative equilibrium path, is arguably quite weak. In reality, the negative externality felt by other countries if a country like the Maldives suffers an early collapse will be limited, due both to the

⁴⁰There are also forces outside our model that would push in the direction of an anti-domino effect, namely: (a) if part of the collapsing country’s population perishes, the total world population will fall, and this will push in the direction of lower aggregate BAU emissions; and (b) to the extent that other resources, such as land and capital, are lost when a country collapses, this will push further in the same direction.

relatively small population of climate refugees that would be released by these countries and to the fact that the associated refugee externality triggered by this early collapse will be shared across many countries.

When does the string of catastrophes end? Recall that the refugee externality R_k increases as more and more countries collapse, while the gain from defecting G_k decreases, thus the process will stop when either (i) R_k rises above G_k , at which point the surviving countries become proactive and help the country at the brink avoid collapse, or (ii) country M becomes the lone surviving country, in which case this country will take care of itself and stop the growth of the carbon stock. In either case, we can state the following:

Proposition 5. *Suppose the catastrophe point \tilde{C}_i differs across countries: (i) If the refugee externality imposed by the collapse of the most vulnerable country on the remaining countries is not severe enough, then a non-empty subset of countries will collapse on the noncooperative equilibrium path. This is true even if the differences between catastrophe points (\tilde{C}_i) across countries are small. (ii) A given country i can reach the brink only if the countries that are more vulnerable (countries $1, 2, \dots, i - 1$) all have collapsed (a basic “domino effect”). But the likelihood of country i surviving conditional on having reached the brink ($C_t = \tilde{C}_i$) is higher if more countries have collapsed before it (an “anti-domino effect”).*

Note the contrast with the common-brink model, where catastrophe never happens on the equilibrium noncooperative path. When asymmetries in the collapse points are introduced, the result changes dramatically, and equilibrium catastrophes become likely; moreover, the conditions under which a given country collapses on the equilibrium path are not affected by the distance between the catastrophe point of this country and those of other countries, as long as the catastrophe points are different, so the asymmetries need not be large.

It is also notable that, when collapse points are heterogeneous, it is entirely possible that some countries could continue to enjoy a reasonable level of utility once the carbon stock has stabilized while others have suffered climate collapse. Hence, the heterogeneous-brink model brings into high relief the possibly uneven impacts of climate change across those countries who, due to attributes of geography and/or socioeconomic position, are more or less fortunate.

3.2. International Climate Agreements

We next revisit the potential role for ICAs, but now in a setting where the catastrophe point differs across countries. In the case of symmetric countries analyzed in the common-brink model of the previous section there was no role for international transfers, so we abstracted from them. But in the present setting where collapse points are heterogeneous across countries, international transfers become relevant. Moreover, such transfers play a prominent role in real world discussions of approaches to address climate change (see, for example, Mattoo and Subramanian, 2013), and allowing them therefore seems important. So in the context of ICAs (and later, the social optimum) we will introduce international transfers explicitly into the model. In our formal analysis we will assume that there is no limit on the potential size of these transfers, so that we can continue to focus on the inability of the ICA to take the interests of future generations into account as the source of potential shortcomings of ICA outcomes relative to the social optimum. We will also comment, however, on how our results would be affected if the size of international transfers were limited by resource constraints.

While we allow for international transfers, we rule out intergenerational transfers. Assuming away transfers across generations seems reasonable for two reasons. First, such transfers would be relevant if different generations could strike a Coasian bargain to correct the intergenerational environmental externalities, but as we have already noted such a Coasian bargain is problematic if not impossible, since different generations may not even be present at the same time. And second, unlike international transfers, intergenerational transfers do not figure prominently in the climate debate.

Formally, we model international transfers as lump-sum transfers of an outside good that enters additively into utility and that does not contribute to emissions. We can think of each country as endowed with a fixed amount of this outside good which it can either consume itself or transfer to other countries, but we assume that the endowment is large enough that it never imposes a binding constraint on transfers for any country, so we can keep this endowment in the background. We denote by $Z_{i,t}$ the (positive or negative) per-capita transfer made by country i at time t in terms of the outside good. The utility of a citizen living in a (surviving) country i at time t can thus be written, building on (3.1), as $\hat{u}_{i,t} \equiv u_{it} - Z_{it}$ (where we omit the fixed endowment of the outside good from the utility function to simplify notation). Note that the absence of intergenerational transfers implies $\sum_{i=1}^M Z_{i,t} = 0$ for all t .

We are now ready to analyze the equilibrium path of carbon emissions and of the carbon

stock under an ICA, and the implications for the collapse and survival of countries.

For a given generation t , the ICA specifies emissions levels for each of the countries that have survived to date in order to maximize world welfare and will then use international transfers to divide up the surplus across countries according to their bargaining powers. We keep international transfers mostly in the background, and focus below on determining the emissions levels that maximize world welfare.

As all countries are symmetric except for the threshold \tilde{C}_i , efficiency again dictates that the ICA must choose the same per-capita emissions in period t for each country that has survived to that point, so as before we can omit the country subscript i on ICA emissions.

Under the assumption that $\beta = 0$ and recalling that the population of country H_t , the most vulnerable country that has survived to time t , is $\frac{M}{M-H_t+1}$ while the population of the remaining countries is $M - \frac{M}{M-H_t+1} = \frac{M(M-H_t)}{M-H_t+1}$ and using $R_{H_t} = \frac{r}{M-H_t}$, we can then write the average per-capita world welfare at time t as

$$U_t = B(c_t) - \lambda C_t - \left(\frac{L+r}{M-H_t+1} \right) \cdot I_t \quad (3.8)$$

where I_t is an indicator function that equals one if country H_t collapses at time t and zero otherwise (and recalling our assumption that parameters are such that at most one country collapses at a given time t). The ICA for generation t will choose c_t to maximize U_t .

Consider first the initial warming phase, in which the carbon stock is below the catastrophe point for country 1 ($C_t < \tilde{C}_1$). In this phase the ICA selects the symmetric level of emissions that maximizes the common per-period payoff, which is given by $\bar{c}^{ICA} \equiv B'^{-1}(M\lambda)$, just as in the warming phase of the common-brink setting analyzed in the previous section: as before, the ICA internalizes the international climate externalities that travel through λ , and hence lowers emissions below the BAU level $\bar{c}_1^N = B'^{-1}(\lambda)$.

Recalling from our analysis of the common-brink model that $C^{ICA} \equiv \frac{M}{\rho} B'^{-1}(M\lambda)$ is the level to which the carbon stock would eventually converge given the emissions level \bar{c}^{ICA} , it follows that if $\tilde{C}_1 \geq C^{ICA}$, the carbon stock never reaches \tilde{C}_1 under the ICA and so country 1 is never brought to the brink of catastrophe and ICA emissions remain at the level \bar{c}^{ICA} forever. Moreover, recall that under condition (3.4) country 1 does reach the brink of catastrophe under the noncooperative scenario (and may or may not collapse), thus in the case $\tilde{C}_1 \geq C^{ICA}$ the ICA prevents any country from ever reaching the brink. On the other hand, if $\tilde{C}_1 < C^{ICA}$, country 1 is brought to the brink under the ICA, and we need to consider what happens next.

Suppose, then, that $\tilde{C}_1 < C^{ICA}$, and country 1 has reached the brink of catastrophe under the ICA. To avoid a taxonomy of uninteresting cases, we focus on the case in which, absent an ICA, a non-empty subset of the most vulnerable countries would collapse. That is, if country \bar{k}^N denotes the marginal surviving country in the noncooperative equilibrium, we are focusing on the case $\bar{k}^N > 1$, which recall is guaranteed under the rather weak condition $R_1 < G_1$.

In this case, when country 1 reaches the brink of catastrophe it would collapse in the absence of an ICA, and the generation alive in the world at this moment now faces a very different international cooperation problem than the problem faced by previous generations. In particular, the world is now confronted with a stark choice: it can cooperate to save country 1 from collapse, or it can let country 1 collapse at the end of period t and carry on without it.

The availability of international lump-sum transfers ensures that the ICA will make this choice so as to maximize the average per-capita world welfare from the point of view of generation t as defined in (3.8). This implies that country 1 will be saved under the ICA if and only if the global loss from the collapse of country 1 exceeds the (minimum) cost to the world of cutting emissions by the sufficient amount to stop the growth of the carbon stock; or put differently, country 1 will be saved if and only if it would be willing to compensate the rest of the world for contributing to stop the growth of C_t .

The global average per-capita loss from the collapse of country 1 is comprised of country 1's own loss L (the per-capita utility cost borne by the citizens of country 1 times its normalized initial population of one) and the refugee externality r that its collapse (and, in light of its normalized initial population, its release of one refugee) would impose on others, averaged over the world population M . And recalling that under our assumptions collapse occurs at the end of a period after all consumption for the period has occurred, it follows that if country 1 is allowed to collapse the ICA will nevertheless implement the country-level emissions \bar{c}^{ICA} for that period. On the other hand, the efficient way to save country 1 is for all countries (including country 1) to reduce per capita emissions to the level $\rho\tilde{C}_1/M$, since efficiency requires the marginal benefit from emissions to be equalized across countries.

More generally and with the above logic in mind, if a given country k is at the brink of catastrophe under the ICA, it will be saved if and only if

$$\Gamma_k \equiv \left[B(\bar{c}^{ICA}) - \lambda \left((1 - \rho)\tilde{C}_k + M\bar{c}^{ICA} \right) \right] - \left[B\left(\frac{\rho\tilde{C}_k}{M} \right) - \lambda\tilde{C}_k \right] \leq \frac{L + r}{M - k + 1} \equiv \psi_k \quad (3.9)$$

The left-hand side of (3.9), Γ_k , is the difference in gross per-capita welfare between having all

countries continue to emit at the level \bar{c}^{ICA} and letting country k collapse at the end of the period, and having all countries emit at the per capita level $\rho\tilde{C}_k/M$ and saving country k from collapse. The right hand side of (3.9), ψ_k , is the global average per-capita refugee cost imposed on the world by the collapse of country k .

The marginal surviving country under the ICA, which we denote \bar{k}^{ICA} , is the lowest k such that condition (3.9) is satisfied. Notice that ψ_k is increasing in k , for essentially the same reasons that the refugee externality R_k is increasing: each climate refugee bears a one-time utility cost L and imposes a one-time utility cost r on the country to which it migrates, and countries that collapse later (higher k) release a greater number of climate refugees ($\frac{M}{M-k+1}$) on a smaller rest-of-world population ($M - \frac{M}{M-k+1}$). Thus the global cost imposed by the collapse of a country becomes higher as more and more countries collapse. Furthermore, it is direct to verify that Γ_k decreases with k .⁴¹ Thus, the ICA will let the most vulnerable country collapse if and only if $\Gamma_1 > \psi_1$, and in this case, the sequence of country collapses under the ICA will stop when ψ_k rises above Γ_k .

We can now turn to a key question: will the ICA save some countries that would collapse in its absence? Or will more countries collapse under the ICA than in its absence? To answer this question we need to compare the marginal surviving countries under the ICA (\bar{k}^{ICA}) and under the noncooperative scenario (\bar{k}^N).

Recall that \bar{k}^{ICA} is the lowest k such that condition (3.9) is satisfied, while \bar{k}^N is the lowest k such that the no-defect conditions (3.5) and (3.6) are both satisfied. First note that the right hand side of (3.9), ψ_k , which is the global average per-capita refugee cost imposed by the collapse of country k , is the population-weighted average of the right hand sides of (3.5) and (3.6), which are respectively the per-capita refugee costs imposed by the collapse of country k on the remaining countries (R_k) and on country k itself (L). On the other hand, as we show in the Appendix, the left hand side of (3.9), Γ_k , is lower than the population-weighted average of the right hand sides of (3.5) and (3.6), G_k and G_k^0 . Intuitively, under the ICA the emissions level is symmetric across countries, while under the self-help equilibrium they are asymmetric,

⁴¹To see this, note that

$$\frac{d\Gamma_k}{dk} = \lambda\rho\frac{d\tilde{C}_k}{dk} - B'\left(\frac{\rho\tilde{C}_k}{M}\right)\frac{\rho}{M}\frac{d\tilde{C}_k}{dk} = \frac{\rho}{M}\frac{d\tilde{C}_k}{dk}\left(M\lambda - B'\left(\frac{\rho\tilde{C}_k}{M}\right)\right).$$

Next note that $\frac{d\tilde{C}_k}{dk} > 0$ and $B'\left(\frac{\rho\tilde{C}_k}{M}\right) > M\lambda$ because $\frac{\rho\tilde{C}_k}{M} < \bar{c}^{ICA}$ and \bar{c}^{ICA} is defined by $B'(\bar{c}^{ICA}) = M\lambda$. This implies $\frac{d\Gamma_k}{dk} < 0$.

with the brink country k choosing zero emissions, and since the gross benefit function $B(c)$ is concave, the average gain from raising emissions is higher in the latter case. This in turn implies that \bar{k}^N must be weakly higher than \bar{k}^{ICA} , so the set of countries that survive under the ICA is weakly larger than the set of countries that survive under the noncooperative scenario.

It is important to note that the ICA has a potential role to play in saving countries from collapse for two distinct reasons. One was highlighted just above and has to do with the “gains” side of the tradeoff (Γ_k is lower than the average of G_k and G_k^0), and the other has to do with the refugee-cost side of the tradeoff. To see this, suppose as a thought experiment that Γ_k were *equal* to population-weighted average of G_k and G_k^0 , in which case (3.9) would be a side-by-side weighted average of (3.5) and (3.6). Then it is easy to see that the set of surviving countries would *still* be weakly larger under the ICA than under the noncooperative scenario, and strictly larger for a whole parameter region.⁴² This is intuitive, since in the noncooperative scenario, if a given country causes the collapse of country k by increasing its own emissions, it exerts a negative refugee externality on all other countries (and this is true also if country k causes its own collapse).

Finally it is arguable that the empirically relevant case would have the internal cost of collapse L large enough so that a country at the brink would always do everything feasible to avoid collapse (i.e. reduce its emissions to zero). In this case, condition (3.6) would not be binding, and hence only the refugee externality r would be relevant for determining \bar{k}^N , while both L and r would still be relevant for determining \bar{k}^{ICA} .

The following proposition summarizes the comparison between the set of countries that survive in the noncooperative equilibrium and under the ICA.

Proposition 6. *When the catastrophe points \tilde{C}_i differ across countries: (i) A (weakly) larger subset of countries survive under the ICA than in the noncooperative scenario, but the ICA may still let some countries collapse; (ii) If the internal cost of collapse L is large enough that a country at the brink is always willing to cut its emissions to zero in order to survive, only the external cost r matters for the set of countries that survive in the noncooperative scenario,*

⁴²We can illustrate this point with an example. Suppose there are only two countries ($M = 2$), so the population of each country is $1/2$, and country 1 is at the brink ($C_t = \tilde{C}_1$). Further suppose, as mentioned in the text, that $\Gamma_1 = (G_1 + G_1^0)/2$, and recall that in this case $\psi_1 = (r + L)/2$. Then (3.5) and (3.6) reduce respectively to $G_1 \leq r$ and $G_1^0 \leq L$, while (3.9) reduces to $G_1 + G_1^0 \leq r + L$. These conditions define parameter regions in (L, r) space, and it is easy to see that the region where (3.5) and (3.6) are both satisfied is a proper subset of the region where (3.9) is satisfied, as claimed in the text.

whereas the combined internal and external cost of collapse ($r + L$) matters for the set of countries that survive under the ICA.

Recall that we have assumed that countries do not face binding constraints in their ability to make transfers. This assumption, together with our assumption that ICAs do not face issues of free-riding in participation and compliance, will allow us to focus sharply on the shortcomings that ICAs face relative to the social optimum because of the inability of future generations to sit at the bargaining table, when we compare the ICA outcome with the social planner's choices in the next subsection. But it is important to highlight that, if international transfers are limited because of resource constraints, the ICA will have a more limited ability to save countries from collapse than we have characterized here.

Specifically, it is intuitive and can be shown that if international transfers are constrained the ICA lets (weakly) more countries collapse than if international transfers are unlimited; and furthermore, that the ICA is less likely to save a given country if the country faces a more severe constraint on international transfers (because even if it would be efficient to save the country in the presence of unlimited international transfers, the ICA can orchestrate this outcome only if the country has enough resources to compensate the remaining countries and ensure that they receive at least their threat-point payoff when they cut their emissions). This would suggest that smaller countries (like the Maldives) are less likely to be able to look to an ICA to save them from climate catastrophe, because they face more severe resource constraints and hence have less ability to make the substantial transfers to the rest of the world that would be needed under an ICA to achieve this feat.

3.3. The Social Optimum

We next turn to the social optimum in the setting where the catastrophe point differs across countries. As international lump-sum transfers are available, the planner maximizes average per-capita world welfare and uses international transfers to redistribute the surplus in each period across countries according to their Pareto weights (which we can leave in the background).

Given the social discount factor $\hat{\beta} \geq 0$, the objective of the social planner can be written as

$$W = \sum_{t=0}^{\infty} \hat{\beta}^t U_t,$$

where U_t is given by (3.8). The planner's problem can then be written as:

$$\begin{aligned} \max \quad & \sum_{t=0}^{\infty} \hat{\beta}^t \left[[B(c_t) - \lambda C_t] - \left(\frac{L+r}{M-H_t+1} \right) \cdot I_t \right] \\ \text{s.t. } \quad & C_t = (1-\rho)C_{t-1} + \sum_{i=1}^M c_{i,t} \text{ for } t \geq 1 \\ & c_{i,t} \geq 0 \text{ for all } i, t. \end{aligned}$$

Again for simplicity we restrict attention to the case where the emissions feasibility constraints $c_{i,t} \geq 0$ are not binding. Given the discontinuities in the payoff functions when the catastrophe point differs across countries, the planner's problem is no longer amenable to a first-order approach as it was in our common-brink model, and there is no simple set of optimality conditions that we can write down. But we can establish some qualitative properties of the socially optimal solution with direct arguments.

We focus on a novel feature of the planner's decision in the heterogeneous-brink model: How many countries does the planner save from climate catastrophe? We establish that a planner's concern for the utility of future generations – as embodied in the social discount factor $\hat{\beta} > 0$ – does not necessarily translate into saving more countries from collapse. In fact, as we next establish, it is possible that a social planner with a higher discount factor will choose a path for emissions that eventually implies a *higher* carbon stock and causes *more* countries to succumb to climate catastrophe than would the same planner with a lower discount factor. And we then consider what this implies for the comparison between the planner's choices and the choices of an ICA regarding the number of countries that are allowed to collapse along the optimal path.

How could it be that a planner with a higher discount factor might choose to allow the carbon stock to rise further and cause more countries to succumb to climate catastrophe than would the same planner with a lower discount factor? The intuition for this surprising result turns out to be as simple as it is illuminating. A planner with a higher discount factor will certainly wish to shift utility toward future generations. The question, though, is how best to do this. Cutting emissions today and lowering the carbon stock that will be inherited by future generations has a direct effect that increases their utility, and this suggests that a planner with a higher discount factor will make emissions choices that lead to a lower carbon stock at any moment in time, and therefore emissions choices that imply (weakly) fewer countries succumbing to climate catastrophe along the optimal path, than would the same planner if it had a lower discount factor. But raising emissions today and crossing the brink of catastrophe

generates a climate refugee cost that is borne primarily (and under our assumption that this is a one-time cost, solely) by the generation alive today; and if the brink constraint is binding on the emissions choices for future generations that would have been made by the planner in the absence of the brink, then raising emissions today – and incurring today the refugee costs of crossing the brink in order to relax the constraint faced by those alive tomorrow – works indirectly to shift utility toward future generations. As we demonstrate, this indirect effect can dominate the direct effect of the higher carbon stock that would go in the other direction.

More specifically, we show in the Appendix that if for a given social discount factor $\hat{\beta}_1 > 0$ the socially optimal plan does not bring the marginal surviving country to the brink of collapse, then a small increase in the social discount factor, say to $\hat{\beta}_2 > \hat{\beta}_1$, can only reduce (or maintain) the number of countries collapsing along the optimal path. We then show that if instead at the initial discount factor $\hat{\beta}_1$ the marginal surviving country under the optimal plan reaches the brink of catastrophe, then it is possible that the higher discount factor $\hat{\beta}_2$ will lead the country to collapse along the optimal path. In the Appendix we prove:

Proposition 7. *Suppose the catastrophe point \tilde{C}_i differs across countries. If, under the social planner's discount factor $\hat{\beta}_1$, the most vulnerable surviving country along the optimal path stays strictly below the brink, then a higher discount factor $\hat{\beta}_2 > \hat{\beta}_1$ can only lead to the same or fewer countries collapsing along the optimal path. But if under $\hat{\beta}_1$ the most vulnerable surviving country along the optimal path survives at the brink, then it is possible that the higher discount factor $\hat{\beta}_2$ will lead to additional countries collapsing along the equilibrium path.*

And owing to the fact that a social planner with discount factor $\hat{\beta}_1 = 0$ would implement the ICA outcome, we can also state the following:

Corollary 2. *As judged by a social planner with $\hat{\beta} > 0$, the number of countries that collapse under an ICA – and the long-term extent of global warming – can be either too high or too low.*

Together, Proposition 7 and its Corollary confirm that, when countries face heterogeneous climate collapse points, greater concern for the welfare of future generations does not necessarily translate neatly into a reduction in the socially optimal carbon stock at every point in time and an increase in the number of countries that survive along the optimal path.

Finally, it is worth emphasizing the central role played in these findings by our assumption that the refugee costs L and r associated with country collapse are not permanent, but rather are borne primarily by the generation alive at the time of the country's collapse, a feature that also played a central role in the intuition we gave for these findings at the outset of this subsection. In our formal model, we have adopted an extreme version of this assumption, namely, that these are one-time costs, but our results would survive as long as these costs are concentrated sufficiently on the current generation. On the other hand, if L and r were instead assumed to reflect constant refugee costs that are incurred in the period of country collapse and every period thereafter, then increasing $\hat{\beta}$ would unambiguously make it (weakly) less likely that a given country k collapses under the social planner's optimal plan. The reason is that (i) the direct effect of increasing the carbon stock inherited by future generations on their utility would be the same as above; and (ii) the indirect effect would be absent, because conditional on being at the brink, if refugee costs were constant and incurred in every period into the infinite future, crossing the brink would *not* shift utility toward future generations. Hence, it is important for Proposition 7 and its Corollary that the refugee costs to the world that would be associated with a country's collapse are concentrated sufficiently on the current generation.

4. Intergenerational Altruism

TBA

5. Conclusion

The world appears to be facing imminent peril from climate change, as countries are not doing enough to keep the Earth's temperature from rising to catastrophic levels and various attempts at international cooperation have failed. Why is this problem so intractable? Can we expect an 11th-hour solution? Will some countries, or even all, succumb to climate catastrophe on the equilibrium path? In this paper we have addressed these questions through a formal model that features the possibility of climate catastrophe and emphasizes the role of the international externalities that a country's policies exert on other countries and the intertemporal externalities that current generations exert on future generations. We have examined the interaction between these features and have explored the extent to which international agreements can mitigate the problem of climate change in their presence. Our analysis delivers novel insights

on the role that international climate agreements can be expected to play in addressing climate change, and it points to important limitations on what such agreements can achieve, even under the best of circumstances.

We have adopted many strong assumptions to carry out our analysis. For example, throughout we have ignored potentially important “domestic” problems which might also frustrate attempts to address climate change, such as the political power of the fossil fuel industry. Putting such issues aside has the advantage of producing a modeling framework that is capable of generating novel insights with maximum clarity. But it also carries the risk of abstraction from important features that should not be ignored in a more complete analysis of the issues. In this light, it seems appropriate to conclude with a discussion of a number of further extensions of our model and analysis which seem especially salient. Our purpose here is not to provide a full analytical treatment of these extensions. Rather, we keep our discussion brief, and focus only on a few key points.

Uncertainty and/or heterogeneous beliefs We have abstracted from uncertainty, but of course uncertainty is an important feature of climate change and the challenge that it poses for the world. Moreover, and relatedly, beliefs about climate change are heterogeneous, with some countries and some groups within countries firmly believing that a climate catastrophe will occur if the world continues along its BAU path, while others are skeptical of such claims or even deny outright that the threat of climate catastrophes are real. How would these features affect our analysis?

To explore this question, we focus on uncertainty over the level of the catastrophic global carbon stock. In our common-brink model, a very stylized way to introduce such uncertainty is to assume that with probability κ the critical level of the carbon stock which, if exceeded, would cause catastrophic loss, is \tilde{C} , and with probability $(1 - \kappa)$ this critical level is infinite. With \tilde{C} satisfying Assumption 1, this would imply that with probability κ a climate catastrophe will occur if the world continues along its BAU path and with probability $(1 - \kappa)$ there is no climate catastrophe to worry about. And similarly in our heterogeneous-brink model, we can assume that with probability κ_i the critical level of the carbon stock for country i is \tilde{C}_i and with probability $(1 - \kappa_i)$ it is infinite.

With these assumptions and assuming also that all agents in the model are risk-neutral, it is straightforward to show that introducing the implied uncertainty about the location of the

critical carbon thresholds into our analysis does not change our basic findings. In the common-brink model, as long as the cost of exceeding \tilde{C} continues to be infinite, any strictly positive probability κ that the critical carbon stock is \tilde{C} rather than infinite will ensure that countries find a way to avoid the expected infinite cost ($\kappa \times \infty + (1 - \kappa) \times \lambda = \infty$) of exceeding \tilde{C} even in the noncooperative equilibrium, just as in our common-brink analysis of section 2.⁴³ And similarly, in the heterogeneous-brink model, country i will avoid collapse with certainty in the noncooperative equilibrium if $\kappa_i R_i$ and $\kappa_i L$ are above some thresholds and will collapse with probability κ_i otherwise, while under the ICA and socially optimal choices country i will avoid collapse with certainty if $\kappa_i \psi_i$ is above some threshold and will collapse with probability κ_i otherwise. In this setting, our main findings of section 3 continue to hold, and in particular, the set of surviving countries under the ICA is weakly larger than under the noncooperative scenario, but may be larger or smaller than under the social optimum.

We can also consider the possibility that beliefs are heterogeneous across countries with a simple reinterpretation of the parameter κ in our common-brink model, by assuming that $X \in \{\tilde{C}, \infty\}$ is the true level of the critical carbon stock and that κ now represents the fraction of countries that believe the critical carbon stock is \tilde{C} , with the remaining fraction $(1 - \kappa)$ of countries believing that the critical carbon stock is infinite and hence that there is no climate catastrophe to worry about. And similarly for the heterogeneous-brink model, we can capture the possibility of heterogeneous beliefs by assuming that $X_k \in \{\tilde{C}_k, \infty\}$ is the true level of the critical carbon stock for country k and that κ_k now represents the fraction of countries that believe the critical carbon stock for country k is \tilde{C}_k , with the remaining fraction $(1 - \kappa_k)$ of countries believing that the critical carbon stock for country k is infinite.

Interestingly, this reinterpretation suggests that heterogeneous beliefs about the risks posed by climate change could potentially be more devastating to the world than uncertainty about the position of the critical thresholds. For example, with heterogeneous beliefs it is easy to see that there is an important new possibility in the common-brink model: if $X = \tilde{C}$ so that the true level of the critical carbon stock is \tilde{C} and if κ is sufficiently small so that the fraction of climate skeptics and deniers in the world is sufficiently high, then it is possible that in the

⁴³If one assumes instead that the loss from exceeding the threshold \tilde{C} is very high but finite (say \bar{L}), and \tilde{C} is a random variable with a bounded support, then the expected loss will be continuous but rising very steeply for C in the support of \tilde{C} . In this case, fixing the distribution of \tilde{C} , as \bar{L} goes to infinity the expected loss function converges to the one we assumed, and we conjecture that the results would then be approximately the same as those of our common-brink model.

noncooperative equilibrium and contrary to our common-brink model of section 2 the world will trigger a climate catastrophe, because the climate skeptics and deniers are sufficiently prevalent in the world to preclude the possibility that the climate believers of the world could do enough on their own to avoid the catastrophe, much as can happen for the most vulnerable countries who face a climate crisis alone in our heterogeneous-brink model of section 3. A further implication is then that in the presence of heterogeneous beliefs a potential new role for ICAs could also arise in the common-brink model, namely, that through their reductions in emissions, ICAs might help keep the global carbon stock below \tilde{C} and thereby help the world avoid a climate catastrophe that would occur on the noncooperative equilibrium path. Similar possibilities can arise in the heterogeneous-brink model when beliefs are heterogeneous.⁴⁴

Lags We have assumed that each generation incurs the warming generated by its own emissions, and that crossing the critical carbon-stock threshold will subject the generation that crosses the threshold to a climate catastrophe. If we think of a generation as representing a span of 20 to 30 years, then our analysis can accommodate lags on the order of a decade or two in the process by which a rising carbon stock leads to warming temperatures, and our findings apply without modification in the presence of such lags.

But the presence of lags in the effects of emissions may lead to a “tipping point” phenomenon, whereby if the global carbon stock crosses a certain threshold, an inevitable and irreversible process is set in motion that leads to climate catastrophe several generations into the future.⁴⁵ Clearly, in the absence of intergenerational altruism such a generation-spanning lag could cause a climate catastrophe to occur in the noncooperative equilibrium of the common-brink model, contrary to our findings in section 2. Still, if there is any intergenerational altruism and provided the cost of a climate catastrophe in the common-brink model is infinite, then there will be no climate catastrophe in the noncooperative equilibrium and our earlier analysis will apply in the presence of intergenerational lags of this nature, much as is true with uncertainty about the position of the critical climate stock as discussed above. And with finite costs of a climate

⁴⁴For example, if a sufficient fraction of the rest of the world does not believe that country k has reached the brink of catastrophe at \tilde{C}_k when in fact country k really is on the brink of collapse, then country k may not be able to acquire the help from the world that it needs to avoid collapse, because too few countries believe that they would face refugee externalities from country k if they don’t step up their efforts to reduce emissions and keep the global carbon stock from exceeding \tilde{C}_k .

⁴⁵On the possibility of climate tipping points and the likely intergenerational lags associated with them, see, for example, Lenton et al (2019).

catastrophe, the same statement applies provided that the degree of intergenerational altruism is sufficiently high.

Introducing such lags into our heterogeneous-brink model of section 3 would have similar effects with regard to individual countries, increasing the likelihood of country-level climate catastrophes under the noncooperative scenario, the ICA and the social optimum. Still, it is not clear that there would be an impact on the relative outcomes across these three scenarios, and hence it is not clear that there would be new implications for the comparison between noncooperative, ICA and socially optimal outcomes that are our focus.

The possibility of a future technological fix Our analysis abstracts from technological change, but it is important to consider the possibility that a technological fix to the climate problem might arrive at some point in the future. As in Besley and Dixit (2019), we consider here the possibility that a discontinuous technological “breakthrough” might occur, producing a “silver-bullet” technology that solves the climate problem once and for all.

To fix ideas, we imagine that each country takes an iid technology draw in each period, and if a draw is successful a silver-bullet technology is discovered, in which case the catastrophe threshold \tilde{C} jumps to infinity forever thereafter. It is easy to see that the possibility of a silver-bullet discovery would have no impact on our findings in the common-brink model of section 2, beyond the obvious impact that if the silver-bullet technology is found then all countries would revert to BAU emissions levels thereafter. But it would have a number of interesting implications within the heterogeneous-brink model of section 3.

First, there would now be a novel benefit of slowing down the march to the brink for any country, in the sense that with more time before a country reaches the brink there is now a greater chance that the silver-bullet technology will appear and the country will be saved from the brink by a technological fix. Moreover, in the presence of intergenerational altruism, there is now a new (and negative) international externality associated with a given country’s collapse: the probability of the arrival of a technological fix at some point in the future is diminished with each country that collapses and gives up its ability to take a technology draw, making all surviving countries worse off in expectation.⁴⁶ This would make collapse along the equilibrium

⁴⁶The assumption that a collapsing country gives up its technological draw might seem in tension with the assumption that its citizens/scientists are assumed to migrate to other countries when a country collapses, but the assumption could be rationalized on the grounds that R&D also involves substantial sunk capital costs that could not be transferred to other countries and would therefore be lost when a country succumbs to a climate

path less likely for the noncooperative choices and also for outcomes under the ICA and the social optimum in the heterogeneous-brink model, though again it is not clear that there would be an impact on the relative outcomes across these three scenarios.

Finally, if it were allowed that investment in technology is a choice that countries make, with increases in investment leading to a greater chance that the technology draw will result in the silver bullet, then these investments themselves would be a natural policy choice for ICAs to cover, in addition to the emissions choices that we have focused on. We leave an exploration of this possibility to future research.

Mitigation versus adaptation We have focused on mitigation (i.e., reduced emissions) as a response to the climate problem, but an alternative and possibly complementary response is adaptation (e.g., building sea walls). It is interesting to consider two possible interpretations of investments in adaptation within our modeling framework. One possibility is that adaptation lowers the cost that a country experiences from moderate levels of warming, related to our parameter λ ; this could be captured by allowing each country i to have its own λ_i which it could lower at a cost. A second possibility is that adaptation raises the critical global carbon stock \tilde{C}_i at which country i faces a catastrophe, and again country i could raise \tilde{C}_i at a cost. Under the first interpretation, the possibility of adaptation could be added to either our common-brink model or our heterogeneous-brink model; to consider the second interpretation it is only the heterogeneous-brink model that is relevant.

While mitigation generates a positive externality across countries and across generations and is hence to be generally encouraged under the ICA and in the social optimum relative to the noncooperative outcome, adaptation may generate a negative externality. Under the first interpretation where adaptation reduces λ_i , the externality is unambiguously negative, because country i 's noncooperative choice of emissions in both the common-brink and heterogeneous-brink models will be higher when it invests in adaptation and lowers λ_i . Clearly, then, the ICA and (to a greater extent) the social planner will require that countries increase mitigation (cut back on emissions) and *decrease* adaptation relative to noncooperative choices. Under the second interpretation, where country i 's investments in adaptation raise \tilde{C}_i , the international externality may be positive or negative. In particular, if \tilde{C}_i rises to $\tilde{C}_i + \Delta_i$ but stays below \tilde{C}_{i+1} , countries $i + 1$ and higher may well benefit, especially if the refugee externality r is large:

catastrophe.

intuitively, the larger is r the more aligned are countries' interests in raising the catastrophe threshold for any one country. But if \tilde{C}_i rises above \tilde{C}_{i+1} , country $i + 1$ may be worse off, especially if r is low (and more generally country $j > i$ may be worse off if \tilde{C}_i rises above \tilde{C}_{i+j}). Thus the implications for the treatment of adaptation in ICAs and the social optimum relative to noncooperative levels may be different across the two interpretations.

Adaptation under the second interpretation also suggests a reason that the brink levels in the heterogeneous-brink model might be positively correlated with country wealth and income: even abstracting from natural variation across countries in the susceptibility to climate change due to geography, if richer countries are more able to invest in adaptation than poorer countries, then in the noncooperative equilibrium it is likely to be the poorer countries who are most vulnerable to the greatest costs of climate change, and the ones most likely to suffer collapse from a climate catastrophe. How ICAs and the social optimum would alter these distributional impacts of climate change from what would transpire in the noncooperative equilibrium is an important question that we leave for future research.

Trade We have abstracted from international trade in our formal analysis, but the links between climate issues and trade policy are central to the debate over the appropriate response to global warming (see Esty, 1994, for an early expression of these links, and see Nordhaus 2015, more recently). In our heterogeneous-brink model, we have emphasized climate refugees as an important example of the international externality that is imposed on surviving countries when a country succumbs to a climate catastrophe. And as we have shown, the externalities that a collapsing country would impose on the remaining countries plays an important role in determining – under both the noncooperative emissions choices and also the ICA and socially optimal choices – the path of emissions and whether or not countries collapse along the equilibrium path under these three scenarios.

But another potentially important externality associated with a country's collapse is the loss of the gains from trade with the collapsing country that the remaining countries may suffer. Viewed from this lens, it is direct to see that our model implies a link between trade policy and the climate issues we study: the bigger the gains from trade, the larger will be the negative externalities imposed on surviving countries when a country collapses due to climate change, and by the logic of our model the fewer countries are likely to collapse along the equilibrium path. This suggests in turn that effective cooperation on trade issues, to the extent that such

cooperation enhances the size of the gains from trade, can by itself help to reduce the most extreme costs of climate change.

6. Appendix

Proof of Proposition 3

The Lagrangian associated with the planner's problem is:

$$L = \sum_{t=0}^{\infty} \left\{ \hat{\beta}^t [B(c_t) - \lambda C_t] + \xi_t [C_t - (1 - \rho)C_{t-1} - M c_t] + \phi_t (C_t - \tilde{C}) \right\} \quad (6.1)$$

where ξ_t and ϕ_t are Lagrange multipliers. Differentiating (6.1) with respect to c_s yields the first-order condition

$$\frac{\partial L}{\partial c_s} = \hat{\beta}^s B'(c_s) - \xi_s = 0. \quad (6.2)$$

And differentiating (6.1) with respect to C_s yields the first-order condition

$$\frac{\partial L}{\partial C_s} = -\hat{\beta}^s M \lambda + \xi_s - (1 - \rho)\xi_{s+1} + \phi_s = 0 \quad (6.3)$$

where we use the fact that each C_s enters two terms of (6.1), the $t = s$ term and the $t = s + 1$ term. Finally, solving (6.2) for ξ_s , substituting into (6.3) and converting s to t , yields

$$-M \lambda + B'(c_t) - (1 - \rho)\hat{\beta} B'(c_{t+1}) + \hat{\beta}^{-t} \phi_t = 0. \quad (6.4)$$

The transversality condition is non-standard and requires some care, so we address it below.

To proceed, we will follow a guess-and-verify approach. There are two cases to consider, depending on whether or not the state constraint $C_t \leq \tilde{C}$ binds for any t .

Case 1: the brink is never reached.

We first suppose that the state constraint never binds, so we set $\phi_t = 0$ for all t in (6.4).

Note that c_t enters equation (6.4) only through $B'(c_t)$, so we can let $X_t \equiv B'(c_t)$ and treat X_t as the unknown rather than c_t , keeping in mind that X_t is decreasing in c_t . We can thus rewrite (6.4) as the first-order linear difference equation

$$-M \lambda + X_t - (1 - \rho)\hat{\beta} X_{t+1} = 0. \quad (6.5)$$

The solutions to (6.5) are characterized by

$$X_t = \frac{K}{\hat{\beta}^t (1 - \rho)^t} + \frac{M \lambda}{1 - \hat{\beta}(1 - \rho)} \quad (6.6)$$

where K is an arbitrary constant. The expression in (6.6) defines a family of curves, one of which is constant (for $K = 0$), while others are increasing and convex (for $K > 0$) and still others are decreasing and concave (for $K < 0$). For future reference, we write the constant solution to (6.6) when $K = 0$ as

$$X_t = \frac{M\lambda}{1 - \hat{\beta}(1 - \rho)} \equiv X. \quad (6.7)$$

We now argue that only the constant solution described by (6.7) satisfies the first-order conditions (6.2) and (6.3). To make this argument, we consider the finite- T problem and take the limit of the solution as $T \rightarrow \infty$.

In the finite- T problem, X_T must satisfy the first-order condition $-M\lambda + X_T = 0$, which follows from (6.5). This determines the transversality condition for the finite- T problem:

$$X_T = M\lambda. \quad (6.8)$$

Note that, since $M\lambda < X$, the curve in (6.6) that satisfies (6.8) must have $\frac{K}{\hat{\beta}^T(1-\rho)^T} < 0$ and hence $K < 0$. This establishes that in the finite- T problem, the optimum path for X_t is not the constant solution described by (6.7), but one of the decreasing paths.

Now consider the limit as $T \rightarrow \infty$. As T increases, the curve in (6.6) that satisfies (6.8) gets closer and closer to the constant solution described by (6.7). Indeed, as $T \rightarrow \infty$ the solution converges pointwise to (6.7).

Thus our candidate solution for Case 1 is the constant solution $X_t = \bar{X}$, and using $X_t \equiv B'(c_t)$, the associated level of emissions for a representative country and for every generation, which we denote by \bar{c}^S , is defined by $B'(\bar{c}^S) = \frac{M\lambda}{1 - \hat{\beta}(1 - \rho)}$, implying

$$\bar{c}^S = B'^{-1} \left(\frac{M\lambda}{1 - \hat{\beta}(1 - \rho)} \right) \quad (6.9)$$

This is the optimum if the implied carbon stock never reaches \tilde{C} . It is easy to see that, if the emissions level is \bar{c}^S per country, the carbon stock increases in a concave way and converges to the steady state level $\frac{M}{\rho}\bar{c}^S \equiv C^S$, hence the condition for \bar{c}^S to be the solution is

$$\tilde{C} \geq C^S \quad (6.10)$$

where note from their definitions that $C^S < C^N$ so both Assumption 1 and (6.10) will be satisfied if $\tilde{C} \in [C^S, C^N)$.

Finally note that the global stock of carbon, denoted \bar{C}_t^S , evolves in Case 1 according to the difference equation

$$\bar{C}_t^S = (1 - \rho)\bar{C}_{t-1}^S + M\bar{c}^S \quad \text{with } \bar{C}_{-1}^S = 0.$$

Case 2: the brink is reached in finite time

Now suppose that the critical level of the carbon stock \tilde{C} is below the threshold level C^S so that (6.10) is violated and instead we have

$$\tilde{C} < C^S. \quad (6.11)$$

In this case our candidate Case-1 solution (6.9) does not work, and we need to proceed to the second guess where the state constraint $C_t \leq \tilde{C}$ binds from some \tilde{t}^S onward.

For $t \geq \tilde{t}^S$, under this guess C_t stays constant at the threshold level \tilde{C} , hence c_t must be set at the replacement rate dictated by the natural rate of atmospheric regeneration given by

$$c_t = \frac{\rho\tilde{C}}{M} = \hat{c}^N \quad \text{for } t \geq \tilde{t}^S. \quad (6.12)$$

For $t < \tilde{t}^S$, the guess is that the state constraint does not bind, so $\phi_t = 0$, and hence we arrive at the same system of first-order difference equations as (6.5), which yields the family of curves (6.6). Given \tilde{t}^S , we pick the solution (i.e., pick K) by imposing the first-order condition (6.5) at $t = \tilde{t}^S$:

$$-M\lambda + X_{\tilde{t}^S} - (1 - \rho)\hat{\beta}\hat{X} = 0 \quad (6.13)$$

where $\hat{X} \equiv B'(\hat{c}^N)$. Again ignoring integer constraints, this requires continuity of X_t , and therefore of c_t .⁴⁷ But given (6.11) we have that $\bar{c}^S > \frac{\rho\tilde{C}}{M} = \hat{c}^N$. And recalling that \bar{c}^S is defined by the constant solution to (6.6) with $K = 0$ so that $X_t = \bar{X}$, this implies that the socially optimal path of c_t for $t \leq \tilde{t}^S$, which we denote by \hat{c}_t^S , must be defined by a solution to (6.6) with $K > 0$ so that $X_t > \bar{X}$. It then follows from (6.6) together with (2.13) that \hat{c}_t^S begins at $t = 0$ at a level that is strictly below \bar{c}^{ICA} , is decreasing, and hits \bar{c}^S at \tilde{t}^S .⁴⁸

Finally, to determine \tilde{t}^S , we use the condition that the path of C_t implied by the path of emissions \hat{c}_t^S , which we denote \hat{C}_t^S , reaches \tilde{C} at \tilde{t}^S . The path \hat{C}_t^S is the solution to the difference

⁴⁷If we take the integer constraint into account, there will (generically) be a period (say $\tilde{t}^{FB} - 1$) where X_t is between \hat{X} and the level defined by (6.13).

⁴⁸Depending on the functional form of B (and in particular on its third derivative), the implied path of \hat{c}_t^S for $t \leq \tilde{t}^S$ may be concave or convex. For example if B is quadratic, the path is concave, but if B is logarithmic the path is convex.

equation

$$\hat{C}_t^S = (1 - \rho)\hat{C}_{t-1}^S + M\hat{c}_t^S \quad \text{with } \hat{C}_{-1}^S = 0. \quad (6.14)$$

Thus \tilde{t}^S is defined using (6.14) and $\hat{C}_{\tilde{t}^S}^S = \tilde{C}$. Using this condition and the analogous condition (2.14) that defines \tilde{t}^{ICA} as well as the properties of \hat{c}_t^S described above, it is direct to confirm that $\tilde{t}^{ICA} < \tilde{t}^S$.⁴⁹

We may conclude that in Case 2, the socially optimal emissions for generation t in a representative country are given by $c_t^S = \hat{c}_t^S$ for $t < \tilde{t}^S$ and $c_t^S = \hat{c}^N$ for $t \geq \tilde{t}^S$.

Proof of Proposition 6

Recall that the marginal surviving country under the noncooperative equilibrium is the lowest k that satisfies

$$\begin{aligned} G_k &\equiv \left[B(\bar{c}_k^N) - \lambda \left(\tilde{C}_k + \frac{M}{M-k+1} \cdot \left(\bar{c}_k^N - \frac{\rho \tilde{C}_k}{M-k} \right) \right) \right] - \left[B \left(\frac{\rho \tilde{C}_k}{M-k} \right) - \lambda \tilde{C}_k \right] \leq \frac{r}{M-k} \\ G_k^0 &\equiv \left[B(\bar{c}_k^N) - \lambda \left(\tilde{C}_k + \frac{M}{M-k+1} \cdot \bar{c}_k^N \right) \right] - \left[B(0) - \lambda \tilde{C}_k \right] \leq L \end{aligned}$$

We now argue that if the above conditions are satisfied, also the following condition is satisfied, so country k will survive under the ICA:

$$\Gamma_k \equiv \left[B(\bar{c}^{ICA}) - \lambda \left((1 - \rho)\tilde{C}_k + M\bar{c}^{ICA} \right) \right] - \left[B \left(\frac{\rho \tilde{C}_k}{M} \right) - \lambda \tilde{C}_k \right] \leq \frac{L + r}{M - k + 1} \equiv \psi_k$$

Noting that ψ_k is the population-weighted average of $\frac{r}{M-k}$ and L (where the weights are respectively $\frac{M-k}{M-k+1}$ and $\frac{1}{M-k+1}$), all we need to show is that Γ_k is no higher than the population-weighted average of G_k and G_k^0 , that is:

$$\Gamma_k \leq \frac{M-k}{M-k+1} G_k + \frac{1}{M-k+1} G_k^0$$

⁴⁹One might wonder whether there is another potential candidate solution: among the paths that satisfy (6.6), is there one such that the implied carbon stock C_t approaches \tilde{C} as $t \rightarrow \infty$, and might this be the optimum? The answer is no. It is easy to show that there is only one solution of (6.6) such that the associated path of C_t converges to a strictly positive level, and that is the $K = 0$ solution, with the associated carbon stock converging to $\bar{C} = \frac{M\bar{c}}{\rho} > \tilde{C}$. For all solutions with $K > 0$, the path of X_t diverges to infinity, thus the path of c_t goes to zero, and hence also C_t converges to zero.

or equivalently

$$\begin{aligned}
& B(\bar{c}^{ICA}) - B\left(\frac{\rho\tilde{C}_k}{M}\right) - \lambda\left(M\bar{c}^{ICA} - \rho\tilde{C}_k\right) \\
& \leq \frac{M-k}{M-k+1} \left[B(\bar{c}_k^N) - B\left(\frac{\rho\tilde{C}_k}{M-k}\right) - \frac{\lambda M}{M-k+1} \left(\bar{c}_k^N - \frac{\rho\tilde{C}_k}{M-k} \right) \right] \\
& \quad + \frac{1}{M-k+1} \left[B(\bar{c}_k^N) - B(0) - \frac{\lambda M}{M-k+1} \bar{c}_k^N \right]
\end{aligned}$$

Letting $\alpha \equiv \frac{M-k}{M-k+1}$ and $v(c) \equiv B(c) - \lambda \frac{M}{M-k+1} c$, and rearranging, we need to show

$$v(\bar{c}_k^N) - v(\bar{c}^{ICA}) + \lambda \frac{M-k}{M-k+1} (M\bar{c}^{ICA} - \rho\tilde{C}_k) + \left[\alpha v\left(\frac{\rho\tilde{C}_k}{M-k}\right) + (1-\alpha)v(0) - v\left(\frac{\rho\tilde{C}_k}{M}\right) \right] \geq 0$$

This condition is satisfied because: (i) $v(c)$ is maximized by \bar{c}_k^N (recalling that the population of each country at this stage is $\frac{M}{M-k+1}$), so $v(\bar{c}_k^N) > v(\bar{c}^{ICA})$; (ii) concavity of $v(c)$ implies that the square parenthesis is positive; and (iii) for country k to reach the brink under the ICA, it must be $M\bar{c}^{ICA} > \rho\tilde{C}_k$. The claim follows.

Proof of Proposition 7

Let \bar{k}^S denote the marginal surviving country under the optimal plan given discount factor $\hat{\beta}_1$. Suppose first that the planner's discount factor is given by $\hat{\beta}_1 > 0$ and that country \bar{k}^S is not brought to the brink of collapse. This implies that under the optimal plan, the steady state level of the carbon stock remains strictly below $\tilde{C}_{\bar{k}^S}$. If $\hat{\beta}$ is then increased from $\hat{\beta}_1$ to $\hat{\beta}_2$, the optimal carbon path will be adjusted to shift utility to future generations. But if $\tilde{C}_{\bar{k}^S}$ did not impose a binding constraint for any generation under the optimal plan with discount factor $\hat{\beta}_1$, then increasing the carbon stock to exceed the brink will not relax any constraint for future generations, and will instead only leave future generations with a higher inherited carbon stock, which by itself reduces their utility. So in this case, the only way to shift utility to the future when $\hat{\beta}$ increases from $\hat{\beta}_1$ to $\hat{\beta}_2$ is to reduce the carbon stock from the initial level, and that ensures that a higher discount factor $\hat{\beta}_2 > \hat{\beta}_1$ can only lead to the same or fewer countries collapsing along the optimal path.

Now suppose that the optimal plan under the discount factor $\hat{\beta}_1$ brings the marginal surviving country \bar{k}^S to the brink of collapse, implying that under the optimal plan the steady state level of the carbon stock reaches $\tilde{C}_{\bar{k}^S}$ and remains there forever. We wish to show that it

is possible in this case that the higher discount factor $\hat{\beta}_2$ will lead country \bar{k}^S to collapse along the optimal path.

To see that this is possible, note first that for a planner with discount factor $\hat{\beta}_1$, the payoff (expressed in per-capita terms) from remaining at the brink $\tilde{C}_{\bar{k}^S}$ forever is

$$\frac{1}{1 - \hat{\beta}_1} \left[B \left(\frac{\rho \tilde{C}_{\bar{k}^S}}{M - \bar{k}^S + 1} \right) - \lambda \tilde{C}_{\bar{k}^S} \right], \quad (6.15)$$

while the payoff (expressed in per-capita terms) from going over the brink today is

$$B(\bar{c}_{\bar{k}^S}(\hat{\beta}_1)) - \lambda \tilde{C}_{\bar{k}^S}^+(\hat{\beta}_1) - \frac{L + r}{M - \bar{k}^S + 1} + \hat{\beta}_1 V_{\bar{k}^S}(\hat{\beta}_1, \tilde{C}_{\bar{k}^S}^+(\hat{\beta}_1)), \quad (6.16)$$

where $\bar{c}_{\bar{k}^S}(\hat{\beta}_1)$ is the optimal emissions level conditional on going over the brink $\tilde{C}_{\bar{k}^S}$ for the period where the brink is exceeded (i.e., today), $\tilde{C}_{\bar{k}^S}^+(\hat{\beta}_1) \equiv (1 - \rho)\tilde{C}_{\bar{k}^S} + M\bar{c}_{\bar{k}^S}(\hat{\beta}_1)$ is the level of the carbon stock at the beginning of the next period implied by $\bar{c}_{\bar{k}^S}(\hat{\beta}_1)$, and $V_{\bar{k}^S}(\hat{\beta}_1, \tilde{C}_{\bar{k}^S}^+(\hat{\beta}_1))$ is the value function (expressed in per-capita terms) beginning in the period after the the brink $\tilde{C}_{\bar{k}^S}$ is exceeded (i.e., beginning tomorrow) when the initial carbon stock is $\tilde{C}_{\bar{k}^S}^+(\hat{\beta}_1)$.

To demonstrate this possibility result, it is convenient to suppose that the values of L and r are such that under the optimal plan for $\hat{\beta}_1$ the planner is *indifferent* between remaining at the brink $\tilde{C}_{\bar{k}^S}$ and exceeding the brink. Using (6.15) and (6.16) we then have with this choice of L and r the initial indifference condition

$$B(\bar{c}_{\bar{k}^S}(\hat{\beta}_1)) - \lambda \tilde{C}_{\bar{k}^S}^+(\hat{\beta}_1) + \hat{\beta}_1 V_{\bar{k}^S}(\hat{\beta}_1, \tilde{C}_{\bar{k}^S}^+(\hat{\beta}_1)) - \frac{1}{1 - \hat{\beta}_1} \left[B \left(\frac{\rho \tilde{C}_{\bar{k}^S}}{M - \bar{k}^S + 1} \right) - \lambda \tilde{C}_{\bar{k}^S} \right] = \frac{L + r}{M - \bar{k}^S + 1}. \quad (6.17)$$

With L and r fixed at their initial levels, the right-hand side of (6.17) is independent of $\hat{\beta}$, so we want to show that the left-hand side of (6.17) can be increasing in $\hat{\beta}$. The derivative of the left-hand side of (6.17) with respect to $\hat{\beta}$ can be written as

$$\frac{1}{1 - \hat{\beta}} \left(\frac{L + r}{M - \bar{k}^S + 1} + \frac{d}{d\hat{\beta}} \left((1 - \hat{\beta}) \left[B(\bar{c}_{\bar{k}^S}(\hat{\beta}_1)) - \lambda \tilde{C}_{\bar{k}^S}^+(\hat{\beta}_1) + \hat{\beta}_1 V_{\bar{k}^S}(\hat{\beta}_1, \tilde{C}_{\bar{k}^S}^+(\hat{\beta}_1)) \right] \right) \right), \quad (6.18)$$

where we have used the initial indifference condition (6.17). If we can show that the expression in (6.18) can be strictly positive, then we will have shown that a small increase in $\hat{\beta}$ starting from $\hat{\beta}_1$ can lead to a strict preference for exceeding the brink $\tilde{C}_{\bar{k}^S}$, causing country \bar{k}^S to collapse and increasing the number of countries that succumb to a catastrophe along the optimal path.

The expression in (6.18) is strictly positive if and only if

$$\frac{L+r}{M-\bar{k}^S+1} > -\frac{d}{d\hat{\beta}} \left((1-\hat{\beta}) \left[B(\bar{c}_{\bar{k}^S}(\hat{\beta}_1)) - \lambda \tilde{C}_{\bar{k}^S}^+(\hat{\beta}_1) + \hat{\beta}_1 V_{\bar{k}^S}(\hat{\beta}_1, \tilde{C}_{\bar{k}^S}^+(\hat{\beta}_1)) \right] \right). \quad (6.19)$$

The left-hand side of (6.19) is strictly positive. Hence, if we can find model parameters such that the right-hand side is equal to zero, we will have established the possibility result. But this is straightforward. Suppose, for example, that $\lambda = 0$ and that $\tilde{C}_{\bar{k}^S+1}$ is sufficiently high so that $\tilde{C}_{\bar{k}^S+1}$ will never bind along the optimal path, even if $\tilde{C}_{\bar{k}^S}$ were exceeded. Then conditional on exceeding $\tilde{C}_{\bar{k}^S}$, the utility of each subsequent generation would be a constant \bar{V} (because consumption would be a constant, and the rising carbon stock would cause no disutility when $\lambda = 0$ and $\tilde{C}_{\bar{k}^S+1}$ is sufficiently high so that $\tilde{C}_{\bar{k}^S+1}$ never binds), and hence

$$B(\bar{c}_{\bar{k}^S}(\hat{\beta}_1)) - \lambda \tilde{C}_{\bar{k}^S}^+(\hat{\beta}_1) + \hat{\beta}_1 V_{\bar{k}^S}(\hat{\beta}_1, \tilde{C}_{\bar{k}^S}^+(\hat{\beta}_1)) = \frac{1}{1-\hat{\beta}_1} \bar{V},$$

implying

$$\frac{d}{d\hat{\beta}} \left((1-\hat{\beta}) \left[B(\bar{c}_{\bar{k}^S}(\hat{\beta}_1)) - \lambda \tilde{C}_{\bar{k}^S}^+(\hat{\beta}_1) + \hat{\beta}_1 V_{\bar{k}^S}(\hat{\beta}_1, \tilde{C}_{\bar{k}^S}^+(\hat{\beta}_1)) \right] \right) = \frac{d\bar{V}}{d\hat{\beta}} = 0.$$

Hence, if $\lambda = 0$ and $\tilde{C}_{\bar{k}^S+1}$ is sufficiently high so that $\tilde{C}_{\bar{k}^S+1}$ will never bind, and if under the discount factor $\hat{\beta}_1$ the most vulnerable surviving country along the optimal path survives (under conditions of indifference ensured by the choice of L and r) at the brink, then a slightly higher discount factor $\hat{\beta}_2$ will lead to additional countries collapsing along the optimal path.

Finally we note that there are other parameterizations of the model that could also generate this possibility. Suppose, for example, that λ is strictly positive but that the brink point for country $\bar{k}^S + 1$ is close to $\tilde{C}_{\bar{k}^S}$, and in particular suppose that model parameters are such that $\tilde{C}_{\bar{k}^S+1} = \tilde{C}_{\bar{k}^S}^+(\hat{\beta}_1)$ and that the brink at $\tilde{C}_{\bar{k}^S+1}$ would not be exceeded along the optimal path. Then, with $\tilde{C}_{\bar{k}^S+1}$ reached at the end of the period in which $\tilde{C}_{\bar{k}^S}$ is exceeded, the utility of each subsequent generation would be a constant \tilde{V} (now this will be so because emissions will be held constant at the level that holds the carbon stock at the level $\tilde{C}_{\bar{k}^S+1}$, and with the carbon stock fixed, λ need not be equal to zero for this result); and if $\tilde{C}_{\bar{k}^S+1}$ is itself close enough to $\tilde{C}_{\bar{k}^S}$ then $\left[B(\bar{c}_{\bar{k}^S}(\hat{\beta}_1)) - \lambda \tilde{C}_{\bar{k}^S}^+(\hat{\beta}_1) + \hat{\beta}_1 V_{\bar{k}^S}(\hat{\beta}_1, \tilde{C}_{\bar{k}^S}^+(\hat{\beta}_1)) \right]$ can be brought arbitrarily close to $\frac{1}{1-\hat{\beta}_1} \tilde{V}$, ensuring that (6.19) can be made to hold.

7. References

- Barrett, Scott (1994), “Self-enforcing international environmental agreements,” *Oxford Economic Papers* 46: 878–894.
- Barrett, Scott (2003), **Environment and Statecraft: The Strategy of Environmental Treaty-Making**, Oxford University Press, Oxford.
- Barrett, Scott (2013), “Climate treaties and approaching catastrophes,” *Journal of Environmental Economics and Management* 66(2): 235–250.
- Barrett, Scott and Astrid Dannenberg (2018), “Coercive Trade Agreements for Supplying Global Public Goods,” mimeo.
- Battaglini, Marco and Bård Harstad (2016), “Participation and Duration of Environmental Agreements,” *Journal of Political Economy* 124(1): 160–204.
- Besley, Timothy and Avinash Dixit (2019), “Environmental catastrophes and mitigation policies in a multiregion world,” *Proceedings of the National Academy of Sciences* 166 (12): 5270–5276.
- Brander, James and Taylor, M. Scott (1998), “The Simple Economics of Easter Island: A Ricardo-Malthus Model of Renewable Resource Use,” *American Economic Review* 88(1): 119–38.
- Caplin, Andrew and John Leahy (2004), “The Social Discount Rate,” *Journal of Political Economy* 112(6): 1257–1268.
- Carraro, C., and D. Siniscalco (1993), “Strategies for the International Protection of the Environment,” *Journal of Public Economics* 52(3): 309–28.
- Dutta, Prajit K., and Roy Radner (2004), “Self-Enforcing Climate-Change Treaties,” *Proceedings of the National Academy of Science* 101: 4746–51.
- Esty, Daniel C. (1994), **Greening the GATT: Trade, Environment, and the Future**, Institute for International Economics, Washington DC.
- Farhi, Emmanuel and Ivan Werning (2007), “Inequality and Social Discounting,” *Journal of Political Economy* 115(3): 365–402.

- Feng, Tangren and Shaowei Ke (2018), “Social Discounting and Intergenerational Pareto,” *Econometrica* 86(5): 1537-1567.
- Harstad, Bård (2012), “Climate Contracts: A Game of Emissions, Investments, Negotiations, and Renegotiations,” *The Review of Economic Studies* 79(4): 1527–1557.
- Harstad, Bård (2020), “Pledge-and-Review Bargaining: From Kyoto to Paris,” *Journal of Political Economy* 124(1): 160-204.
- Harstad, Bård (2021), “Trade and Trees,” mimeo, University of Oslo, March 28.
- Jenkins, Jesse D. (2014), “Political economy constraints on carbon pricing policies: What are the implications for economic efficiency, environmental efficacy, and climate policy design?,” *Energy Policy* 69: 467-477.
- John, Andrew and Rowena A. Pecchenino (1997), “International and Intergenerational Environmental Externalities,” *Scandinavian Journal of Economics* 99(3): 371–387.
- Jones, Aled and Nick King (2021), “An Analysis of the Potential for the Formation of ‘Nodes of Persisting Complexity’,” *Sustainability* 13, 8161: <https://doi.org/10.3390/su13158161>.
- Kolstad, C. D., and M. Toman (2005), “The Economics of Climate Policy,” **Handbook of Environmental Economics** 3: 1562-93.
- Lemoine, Derek and Ivan Rudik (2017), “Steering the Climate System: Using Inertia to Lower the Cost of Policy,” *American Economic Review*, 107(10): 2947–2957.
- Lenton, Timothy M., Rockstrom, Johan, Gaffney, Owen, Rahmstorf, Stefan, Richardson, Katherine, Steffan, Will and Hans Joachim Schellnhuber (2019), “Climate tipping points – too risky to bet against,” *Nature* (Comment) 575, November 28: 592-595.
- Lustgarten, Abrahm (2020), “How Climate Migration Will Reshape America: Millions will be displaced. Where will they go?,” *The New York Times Magazine*, September 15.
- Maggi, Giovanni (2016), “Issue Linkage,” in K. Bagwell and R.W. Staiger (eds.), **The Handbook of Commercial Policy**, vol. 1B, Elsevier.
- Mattoo, Aaditya and Arvind Subramanian (2013), **Greenprint: A New Approach to Cooperation on Climate Change**, Center for Global Development, Washington, D.C.

- Millner, Antony and Geoffrey Heal (2021), “Choosing the Future: Markets, Ethics, and Rapprochement in Social Discounting,” NBER Working Paper No 28653, April.
- Nordhaus, William D. (2015), “Climate Clubs: Overcoming Free-riding in International Climate Policy,” *American Economic Review* 105(4): 1339-70.
- Pindyck, Robert S. (2020), “What We Know and Don’t Know about Climate Change, and Implications for Policy,” NBER Working Paper No 27304.
- Russonello, Giovanni (2021), “What’s Driving the Surge at the Southern Border?,” *The New York Times*, April 5.
- Wallace-Wells, David (2019), **The Uninhabitable Earth: Life After Warming**, Tim Duggan Books, New York.
- Zaki, Jamil (2019), “Caring about tomorrow: Why haven’t we stopped climate change? We’re not wired to empathize with our descendants,” *The Washington Post* (Outlook), August 22.

Figure 1: ICA, Planner and Noncooperative Outcomes (High \tilde{C})

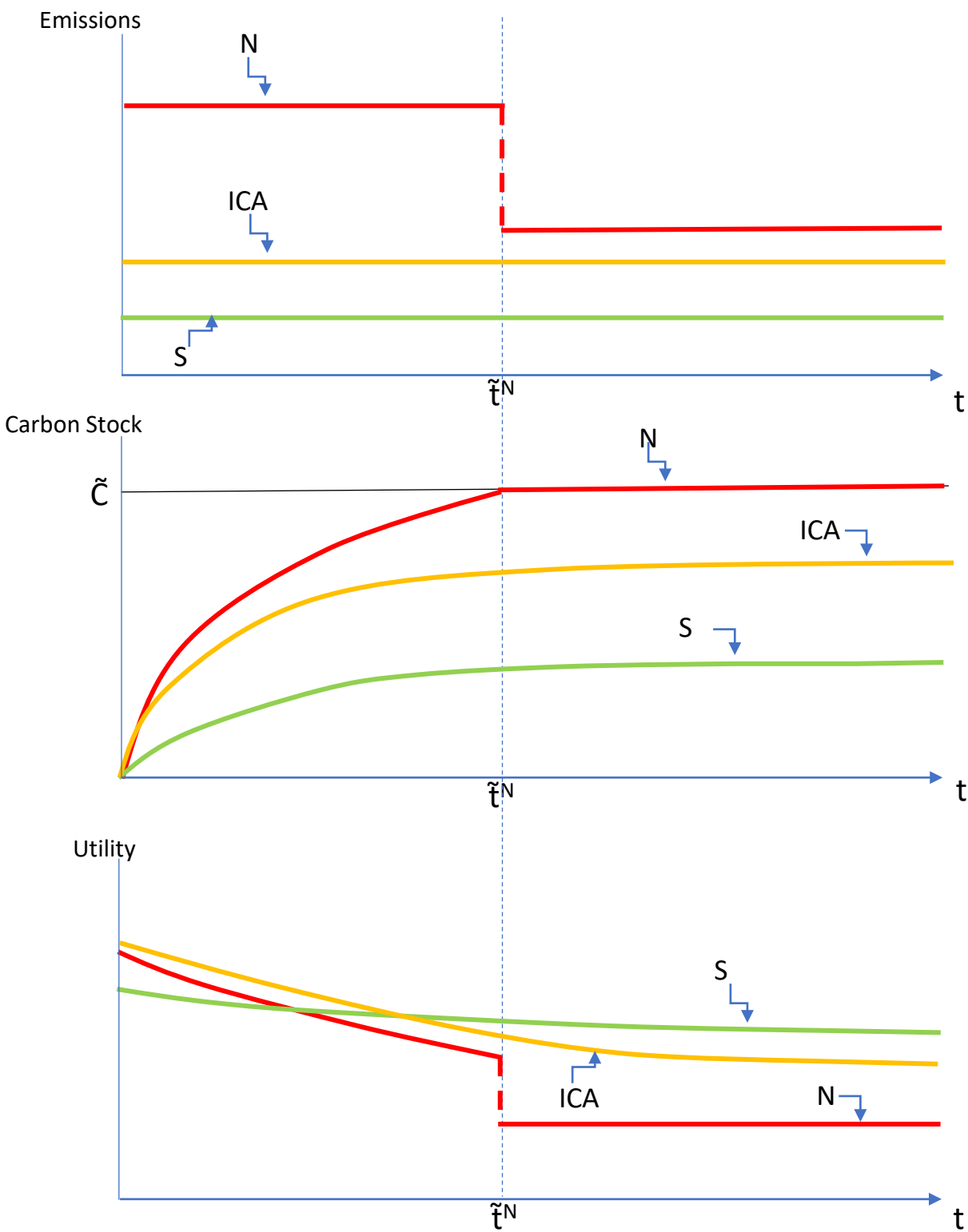


Figure 2: ICA, Planner and Noncooperative Outcomes (Intermediate \tilde{C})

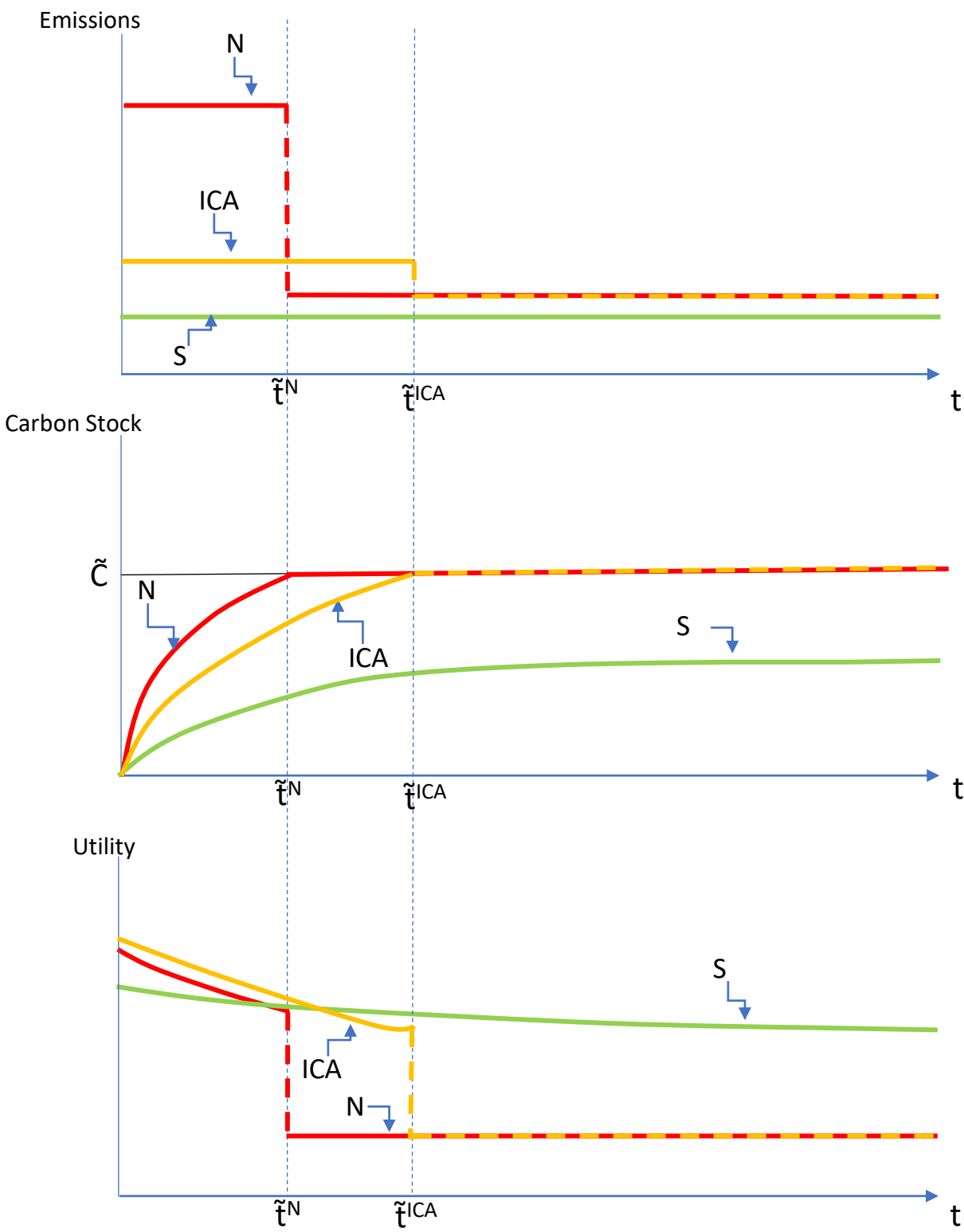


Figure 3: ICA, Planner and Noncooperative Outcomes (Low \tilde{C})

