International Climate Agreements and the Scream of Greta*

Giovanni Maggi† Robert W. Staiger‡

Preliminary and Incomplete

November 2020

Abstract

The world appears to be in 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 11^{th}-hour solution? Will some countries, or even all, succumb on the equilibrium path? We address these questions through a formal model that features the possibility of climate catastrophe and emphasizes the role of two critical issues: the international externalities that a country's policies exert on other countries, and the intertemporal externalities that current generations exert on future generations. We examine the interaction between these two issues and explore the extent to which international agreements can mitigate the problem of climate change in their presence.

*We thank seminar participants at Bocconi University and Singapore Management University for helpful comments.

†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 in imminent peril from climate change. According to the most recent assessments of 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 11^{th}-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 the role of two critical issues: 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, relat-

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

[2] And 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.

ing to country participation and enforcement.[3] In this paper we abstract from these challenges, and focus instead on a limitation that has not been emphasized in the formal literature. 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.[4] 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.

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 unlivable in their lifetime.

We consider two possibilities for climate catastrophes. In our common-brink model, all the countries of the world 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 at a level of the global carbon stock which is lower than the level that would bring less vulnerable countries to the brink.

[3] See 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.

[4] We view this limitation as potentially even more severe than the participation and enforcement issues, because while countries can be given at least partial incentives to participate in, and comply with, climate agreements by threatening punishments in other policy areas such as trade, it is hard to think of similar ways to correct intergenerational externalities from climate policies.

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 11^{th} hour solution that has each country doing its part to halt further climate change. The solution involves reduced worldwide emissions levels that are set at the replacement rate dictated by 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 11^{th} 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 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 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 first-best outcome that would be implemented by a global social planner who 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 first best. This leads to three possible scenarios, depending on the severity of the constraint that the catastrophic global carbon stock level places on attainable steady state welfare. 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 first best; if the constraint is sufficiently severe, the brink will be reached both under the ICA and the first best, 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 first best. 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 precipitous 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-collapse-point analysis, where once the world reaches the brink of catastrophe 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 in collapse points across countries can create the possibility of 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 also explore the possibility of a "domino" effect exhibited by the collapse of countries along the equilibrium path of our heterogeneous–brink model. We show that, while there is indeed a basic domino effect at play, in that a given country can reach the brink of collapse only if the countries that rank lower in terms of resilience have all collapsed, there is also a more subtle "anti-domino" effect, in that the likelihood of a country surviving *conditional on reaching the brink* is higher if more countries have collapsed before it, for two reasons: the refugee externality imposed on each surviving country is higher when there are fewer remaining countries, and the aggregate BAU emissions are lower and hence easier to offset.

Finally, we 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.

Overall, our conclusions 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 first best outcome, 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 first best are potentially more devastating, because it may not save enough countries from collapse.

Relative to the existing literature on ICAs, our main contribution 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 how the possibility of climate catastrophe affects optimal environmental policies. Prominent examples in this literature are Barrett (2012), Lemoine and Rudik (2017) and Besley and Dixit (2018). Of these papers, only Barrett (2012) 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 of section 2.[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

[5] Barrett (2012) 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 such equilibrium and reduce the probability of catastrophe. He also emphasizes that, while in the absence of uncertainty the only possible 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 more standard models without catastrophes.

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 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), Battaglini and Harstad (2016) and Harstad (2020). 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]

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 first-best choices of a global social planners. Section 3 contains the parallel analysis for our heterogeneous-brink model. Finally, section 4 considers a number of extensions to our basic models, while section 5 concludes. Proofs not contained in the body of the paper are provided in the Appendix.

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 $t \in \{0, 1, ..., \infty\}$. 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. Each

[6] For 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).

[7] An 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.

parent is altruistic toward its only child, and the utility of a representative country's generation t is given by

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

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

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

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. 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. We therefore treat c_t itself as the choice variable of a country, with the understanding that lower emissions mean lower consumption. We have in mind that each government then implements its chosen c_t with an appropriate climate policy (e.g., carbon tax level).[9]

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, and we assume that $C_t \geq 0$ for all t. The evolution

[8] To 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.

[9] Implicit 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). Our 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.

of this stock through time depends on how long the stock persists in the atmosphere, and on the level of emissions c_t, according to

$$C_t = (1-\rho)C_{t-1} + Mc_t \text{ with } C_{-1} = 0. \tag{2.2}$$

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).[10]

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, but that past a certain critical level, a rising carbon stock would lead to a level of global warming that would trigger the collapse of civilization. We later allow the level of the carbon stock that would be catastrophic to differ by country, but for now we assume that it is common to all countries. The catastrophic level $\tilde{C}$ is assumed known with certainty.[11]

With these assumptions, we may now write the material utility of a representative country's generation t as

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

Below we will characterize various equilibria of the model, including the noncooperative equilibrium that arises when countries choose emissions levels in the absence of any agreements, the equilibrium that arises if international agreements are available, and as a benchmark the first-best outcome that a global social planner would implement beginning at time $t = 0$ to

[10] Given 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).

[11] A more realistic model would assume that $\tilde{C}$ is uncertain, but we suspect that our main qualitative insights would not be affected. If one maintains the assumption that the loss from exceeding the threshold is inifinite, then results would be unlikely to change, simply because the expected loss is infinite even if the probability of catastrophe is small. If 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 model.

maximize a social welfare function. For our characterization of the first best, we assume that the planner seeks to maximize the utility of a representative country within any generation, reflecting the symmetry across countries in our model setup; 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. In particular, we follow Fahri and Werning (2007) in postulating the following planner objective:

$$W = \sum_{t=0}^{\infty} \hat{\beta}^t u_t, \tag{2.4}$$

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 $\hat{\beta} < 1$.[12]

With this simple structure, our model highlights two externalities that arise in the context of climate change and that must both be addressed if the world is to achieve the first-best outcome. One externality is international in nature: 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 in all other countries. The other externality is intergenerational in nature: the emissions of a country's generation t affect the material utility of all subsequent generations 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.

Since the presence of intergenerational externalities is a key and novel feature of our model, it is worth pausing to clarify the nature of the deviation from the first best that is created when this externality is not internalized. To this end, suppose for a moment that there is only one

[12] In 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. To avoid this complication and stay within our infinite horizon setting, we simply assume that $\hat{\beta} < 1$.

country (so no international externality) 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 correct the choices of generation t and redistribute utility from generation t to the subsequent generations. Importantly, the same logic applies also in the presence of altruism ($\beta > 0$), because as noted above, the wedge between the social and private discount factors ($\hat{\beta} - \beta$) need not decrease as β rises. Notice also that the described move to the first best does not mark a Pareto improvement, but rather a movement along the efficiency frontier, shifting surplus from generation t to later generations.[13] As a shorthand and with some abuse of terminology we will nevertheless refer to any deviation from the first best as an "inefficiency," but it should be kept in mind that in the case of deviations that arise due to unaddressed intergenerational externalities this is not an inefficiency in the Pareto sense.

Finally, before turning to the analysis we can make a simple preliminary point: the inefficiencies 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 ($\beta = 0$). 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 first-best emissions levels chosen by the planner for each country's generation t satisfy $B'(c_t) = \frac{M}{1-\hat{\beta}}\lambda$. The overall wedge between the first-best and noncooperative emissions choices is summarized by $\frac{M}{1-\hat{\beta}} > 1$, which implies excessive emissions in the noncooperative equilibrium. 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 $\frac{1}{1-\hat{\beta}} > 1$ reflects the degree to which the intergenerational externality contributes to excessive emissions in the noncooperative equilibrium, because noncooperative choices do not account

[13] In 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.

appropriately 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 externalities that drive the inefficiencies 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 first-best and noncooperative emissions profiles are straightforward, as we have just observed. But as we establish below, the existence of a catastrophe point introduces fundamental changes to the noncooperative and first-best emissions profiles, both along the path to the catastrophe and once the brink of catastrophe is reached, as well as to the possible role of international climate agreements in addressing the inefficiencies exhibited by the noncooperative choices.

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 ($\beta = 0$) but otherwise impose no special restrictions on model parameters; in a later section 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 a later section.

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

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

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

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

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

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

If the warming phase went on forever, (2.6) implies that the global carbon stock would converge to the steady state level $\frac{M}{\rho}B'^{-1}(\lambda) \equiv \tilde{C}^N$. And if the catastrophe level of the global carbon stock, $\tilde{C}$, were greater than $\tilde{C}^N$, 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 $\tilde{C}^N$. We therefore impose

$$\tilde{C} < \tilde{C}^N, \tag{Assumption 1}$$

which ensures that under BAU emissions the catastrophic level of the global climate stock would eventually be breached.[14]

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.

[14] With regard to the role of ρ in determining the critical level of the carbon stock $\tilde{C}^N$, it is relevant to note that the two major components of the carbon stock have very different rates of depreciation: CO_2 remains in the atmosphere for centuries (see, for example, Pindyck, 2020), while methane is estimated to remain in the atmosphere for around 20 years. Hence, in reality the size of ρ depends in part on the relative importance of these two components.

If the brink generation $\tilde{t}^N$ is to avoid the collapse of civilization, it must end the warming phase with an "11^{th}-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.[15] Focussing on the symmetric equilibrium where all countries adopt the same level of emissions (which is natural, and also efficient, given the assumed symmetry across countries), for generations $t \geq \tilde{t}^N$ noncooperative emissions will fall to the replacement rate dictated by the natural rate of atmospheric regeneration given by

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

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 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 $\hat{c}^N < \bar{c}^N$.[16] The utility of each generation $t \geq \tilde{t}^N$ during this second phase of the noncooperative equilibrium is then constant and given by

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

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

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

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

$$u_t^N = \begin{cases} B(\bar{c}^N) - \lambda C_t^N \text{ for } t < \tilde{t}^N \\ B(\hat{c}^N) - \lambda \tilde{C} \text{ for } t \geq \tilde{t}^N \end{cases}. \tag{2.10}$$

[15] There 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.

[16] While 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.

Note that under the noncooperative equilibrium and according to (2.10), utility must fall discretely for the brink and all subsequent generations, due to the discrete reduction in global emissions implied by

$$\hat{c}^N = \frac{\rho \tilde{C}}{M} < B'^{-1}(\lambda) = \bar{c}^N \tag{2.11}$$

that is required to prevent catastrophe once the world reaches the brink, where the inequality in (2.11) follows from Assumption 1 as we have noted. According to (2.10) and (2.11), 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:

Proposition 1. *The noncooperative equilibrium exhibits an initial warming phase, during which 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 11^{th} hour solution that halts further climate change with reduced emissions levels that are set at the replacement rate 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. The drop in utility experienced by the brink generation is larger the greater the number of countries, the smaller the regeneration capacity of the atmosphere and level of carbon stock above which climate catastrophe occurs, and the lower is the cost of moderate pre-catastrophe warming.*

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 11^{th} 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.[17]

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. And as we later demonstrate, if the catastrophe point is allowed to differ across countries then the model delivers a further piece of bad news: some of the most vulnerable countries may be pushed over the brink.

2.3. International Climate Agreements

We are now ready to consider what an international climate agreement (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 forcing 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, and that any arrangements negotiated under the ICA are perfectly enforceable provided only that under these arrangements each country is at least as well off as in the noncooperative equilibrium, which we take to be the "threat

[17] Barrett (2011) 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.

point" for the negotiations over an ICA. 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.[18] Our goal is to characterize the extent to which, as a result of this limitation alone, ICAs must inevitably fall short of the first best.

Recalling that we are focusing 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 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

$$\bar{c}^{ICA} = B'^{-1}(M\lambda). \tag{2.12}$$

According to (2.12), 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 (2.12) imply $\bar{c}^N > \bar{c}^{ICA}$, 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

$$C_t^{ICA} = (1-\rho)C_{t-1}^{ICA} + M\bar{c}^{ICA}, \text{ with } C_{-1}^{ICA} = 0, \tag{2.13}$$

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

[18] One 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.

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

$$\tilde{C} \geq \tilde{C}^{ICA}, \tag{2.14}$$

where note from their definitions that $\tilde{C}^{ICA} < \tilde{C}^{N}$ so both Assumption 1 and (2.14) will be satisfied if $\tilde{C} \in [\tilde{C}^{ICA}, \tilde{C}^{N})$. 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, $\tilde{C}^{N}$, 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 $\tilde{C}^{ICA}$ in (2.14) defines "sufficiently close" in this context.

On the other hand, if $\tilde{C}$ is below this threshold level and (2.14) is violated so that

$$\tilde{C} < \tilde{C}^{ICA}, \tag{2.15}$$

then under the ICA the brink of catastrophe will be reached in finite time, and the brink generation $\tilde{t}^{ICA}$ is determined from (2.13) as the period $\tilde{t}^{ICA}$ that satisfies $C^{ICA}_{\tilde{t}^{ICA}} = \tilde{C}$. Notice from (2.13) and (2.6) that $\tilde{t}^{ICA} > \tilde{t}^{N}$ is ensured by $\bar{c}^{N} > \bar{c}^{ICA}$, so the ICA postpones the arrival of the brink generation when (2.15) is satisfied even though it does not avoid the brink completely in this case.

If (2.15) 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 (2.15) 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 and hence have no

need for an ICA to serve this purpose.[19] As we will later demonstrate, when we allow the level of the carbon stock that would be catastrophic to differ by country, a possible role for an ICA to save some of the most vulnerable countries may arise; so the result we describe here must be qualified in that setting, but only partially.

Finally, letting c_t^{ICA} denote the path of emissions under the ICA, we can describe emissions succinctly under both (2.14) and (2.15) with

$$c_t^{ICA} = \begin{cases} \bar{c}^{ICA} \text{ for } t < \tilde{t}^{ICA} \\ \hat{c}^N \text{ for } t \geq \tilde{t}^{ICA} \end{cases}, \tag{2.16}$$

where $\tilde{t}^{ICA}$ is finite if and only if (2.15) is satisfied. Utility under the ICA is then given by

$$u_t^{ICA} = \begin{cases} B(\bar{c}^{ICA}) - \lambda C_t^{ICA} \text{ for } t < \tilde{t}^{ICA} \\ B(\hat{c}^N) - \lambda \tilde{C} \text{ for } t \geq \tilde{t}^{ICA} \end{cases}. \tag{2.17}$$

Note that under the ICA, if (2.15) is satisfied so that $\tilde{t}^{ICA}$ is finite, then (2.17) implies that utility must fall discretely for the brink and all subsequent generations, due to the discrete reduction in global emissions implied by

$$\hat{c}^N = \frac{\rho \tilde{C}}{M} < B'^{-1}(M\lambda) = \bar{c}^{ICA} \tag{2.18}$$

that is required to prevent catastrophe once the world reaches the brink, where the inequality in (2.18) follows from (2.15). Hence, according to (2.17) and (2.18) 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 $\bar{c}^N > \bar{c}^{ICA}$ 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. *The path of emissions under the ICA falls into one of two cases. (i) If $\tilde{C}$ is above a threshold level, the brink of catastrophe is never reached and the ICA emissions levels are below the noncooperative emissions levels and constant through time. (ii) If $\tilde{C}$ is below*

[19] Recall 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 16). 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.

this threshold, the brink of catastrophe will be reached under the ICA. The ICA emissions levels are below the noncooperative levels and constant through time until the brink is reached, and at that point they fall to the replacement rate dictated by the natural rate of atmospheric regeneration and remain at that level for all generations thereafter. 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. Once the brink is reached, the useful life of the ICA ends. In neither case does the ICA have a role to play in helping the world avoid climate catastrophe.

2.4. First Best

We next consider the first-best emissions choices. These are 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 interchangeably as the first-best or the socially optimal choices.[20]

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 $B(c_t) - \lambda C_t$ and introduce the constraint $C_t \leq \tilde{C}$. To determine the first-best emissions choices, we therefore write the planner's problem as

$$\begin{aligned} & \max \sum_{t=0}^{\infty} \hat{\beta}^t [B(c_t) - \lambda C_t] \\ \text{s.t. } C_t &= (1-\rho)C_{t-1} + Mc_t \text{ for all } t \\ C_t &\leq \tilde{C} \text{ for all } t, \end{aligned} \tag{2.19}$$

where we have omitted the nonnegativity constraint $c_t \geq 0$ because it is never binding in this setting. The Lagrangian associated with the planner's problem is:

$$L = \sum_{t=0}^{\infty} \left\{ \hat{\beta}^t [B(c_t) - \lambda C_t] + \xi_t \left[C_t - (1-\rho)C_{t-1} - Mc_t \right] + \phi_t (C_t - \tilde{C}) \right\} \tag{2.20}$$

where ξ_t and ϕ_t are Lagrange multipliers. We assume that the problem is globally concave, so that we can rely on a first-order condition approach. Differentiating (2.20) with respect to c_s

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

yields the first-order condition

$$\frac{\partial L}{\partial c_s} = \hat{\beta}^s B'(c_s) - \xi_s = 0. \tag{2.21}$$

And differentiating (2.20) with respect to C_s yields the first-order condition

$$\frac{\partial L}{\partial C_s} = -\hat{\beta}^s M\lambda + \xi_s - (1-\rho)\xi_{s+1} + \phi_s = 0, \tag{2.22}$$

where we have used the fact that each C_s enters two terms of (2.20), the $t = s$ term and the $t = s+1$ term. Finally, solving (2.21) for ξ_s, substituting into (2.22) and converting s into t, yields

$$-M\lambda + B'(c_t) - (1-\rho)\hat{\beta}B'(c_{t+1}) + \hat{\beta}^{-t}\phi_t = 0. \tag{2.23}$$

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 (2.23).

Note that c_t enters equation (2.23) 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 (2.23) as

$$-M\lambda + X_t - (1-\rho)\hat{\beta}X_{t+1} = 0. \tag{2.24}$$

The solutions to (2.24) are characterized by

$$X_t = \frac{K}{\hat{\beta}^t(1-\rho)^t} + \frac{M\lambda}{1-\hat{\beta}(1-\rho)} \tag{2.25}$$

where K is an arbitrary constant. The expression in (2.25) 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 (2.25) when $K = 0$ as

$$X_t = \frac{M\lambda}{1-\hat{\beta}(1-\rho)} \equiv \bar{X}. \tag{2.26}$$

This solution has a simple interpretation. Recalling that X_t is the representative country's marginal benefit of emissions, (2.26) says that 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).

We now argue that only the constant solution described by (2.26) satisfies the first-order conditions (2.21) and (2.22). To make this argument, we consider the finite-T problem and take the limit of the solution as $T \to \infty$.

In the finite-T problem, X_T must satisfy the first-order condition $-M\lambda + X_T = 0$, which follows from (2.24). This determines the transversality condition for the finite-T problem:

$$X_T = M\lambda. \tag{2.27}$$

Note that, since $M\lambda < \bar{X}$, the curve in (2.25) that satisfies (2.27) 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 (2.26), but one of the decreasing paths.

Now consider the limit as $T \to \infty$. As T increases, the curve in (2.25) that satisfies (2.27) gets closer and closer to the constant solution described by (2.26). Indeed, as $T \to \infty$ the solution converges pointwise to (2.26).

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 $\bar{c}^{FB}$ for a representative country and for every generation is defined by $B'(\bar{c}^{FB}) = \frac{M\lambda}{1-\hat{\beta}(1-\rho)}$, implying

$$\bar{c}^{FB} = B'^{-1}\left(\frac{M\lambda}{1-\hat{\beta}(1-\rho)}\right). \tag{2.28}$$

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

$$\tilde{C} \geq \tilde{C}^{FB}, \tag{2.29}$$

where note from their definitions that $\tilde{C}^{FB} < \tilde{C}^N$ so both Assumption 1 and (2.29) will be satisfied if $\tilde{C} \in [\tilde{C}^{FB}, \tilde{C}^N)$. The intuition for this condition is analogous to that for (2.14) in the context of the ICA: if the catastrophe point $\tilde{C}$ is high and sufficiently close to the steady

state level of the BAU global carbon stock, $\tilde{C}^N$, 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; and the threshold level of the carbon stock $\tilde{C}^{FB}$ in (2.29) defines "sufficiently close" in the context of the planner's problem.

Finally, note that the first-best level of welfare achieved by a representative country's generation t, which we denote by u_t^{FB}, is given in Case 1 by

$$u_t^{FB} = B(\bar{c}^{FB}) - \lambda \bar{C}_t^{FB}, \tag{2.30}$$

where $\bar{C}_t^{FB}$ is the global stock of carbon. And $\bar{C}_t^{FB}$ evolves according to the difference equation

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

As (2.30) indicates, with the first-best emissions set at the constant level $\bar{c}^{FB}$, the utility of each generation declines through time as $\bar{C}_t^{FB}$ rises and the climate warms. It is notable that, while $\hat{\beta}$ impacts the *level* of $\bar{c}^{FB}$, it does not alter the fact that the first-best emissions level is constant through time. Evidently, in Case 1 a higher $\hat{\beta}$ induces higher welfare for later generations under the first-best 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.

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 $\tilde{C}^{FB}$ so that (2.29) is violated and instead we have

$$\tilde{C} < \tilde{C}^{FB}, \tag{2.31}$$

In this case our candidate Case-1 solution (2.28) does not work, and we need to proceed to the second guess where the state constraint $C_t \leq \tilde{C}$ binds from some $\tilde{t}^{FB}$ onward.

For $t \geq \tilde{t}^{FB}$, 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

$$c_t = \frac{\rho \tilde{C}}{M} = \hat{c}^N \text{ for } t \geq \tilde{t}^{FB}. \tag{2.32}$$

For $t < \tilde{t}^{FB}$, 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 (2.24), which yields the family

of curves (2.25). Given $\tilde{t}^{FB}$, we pick the solution (i.e., pick K) by imposing the first-order condition (2.24) at $t = \tilde{t}^{FB}$:

$$-M\lambda + X_{\tilde{t}^{FB}} - (1-\rho)\hat{\beta}\hat{X} = 0, \tag{2.33}$$

where $\hat{X} \equiv B'(\hat{c}^N)$. Again ignoring integer constraints, this requires continuity of X_t, and therefore of c_t.[21] But given (2.31) we have that $\bar{c}^{FB} > \frac{\rho\tilde{C}}{M} = \hat{c}^N$. And recalling that $\bar{c}^{FB}$ is defined by the constant solution to (2.25) with $K = 0$ so that $X_t = \bar{X}$, this implies that the first-best path of c_t for $t \leq \tilde{t}^{FB}$, which we denote by $\hat{c}_t^{FB}$, must be defined by a solution to (2.25) with $K > 0$ so that $X_t > \bar{X}$. It then follows from (2.25) together with (2.12) that $\hat{c}_t^{FB}$ begins at $t = 0$ at a level that is strictly below $\bar{c}^{ICA}$, is decreasing, and hits $\hat{c}^{FB}$ at $\tilde{t}^{FB}$.[22]

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

$$\hat{C}_t^{FB} = (1-\rho)\hat{C}_{t-1}^{FB} + M\hat{c}_t^{FB}, \text{ with } \hat{C}_{-1}^{FB} = 0. \tag{2.34}$$

Thus $\tilde{t}^{FB}$ is defined using (2.34) and $\hat{C}_{\tilde{t}^{FB}}^{FB} = \tilde{C}$. Using this condition and the analogous condition (2.13) that defines $\tilde{t}^{ICA}$ as well as the properties of $\hat{c}_t^{FB}$ described above, it is direct to confirm that $\tilde{t}^{ICA} < \tilde{t}^{FB}$.[23]

We may conclude that in Case 2, the first-best emissions for generation t in a representative country are given by

$$c_t^{FB} = \begin{cases} \hat{c}_t^{FB} \text{ for } t < \tilde{t}^{FB} \\ \hat{c}^N \text{ for } t \geq \tilde{t}^{FB} \end{cases}.$$

And the first-best level of welfare achieved by a representative country's generation t is given in Case 2 by

$$u_t^{FB} = \begin{cases} B(\hat{c}_t^{FB}) - \lambda\hat{C}_t^{FB} \text{ for } t < \tilde{t}^{FB} \\ B(\hat{c}^N) - \lambda\tilde{C} \text{ for } t \geq \tilde{t}^{FB} \end{cases}. \tag{2.35}$$

[21] If 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 (2.33).

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

[23] One might wonder whether there is another potential candidate solution: among the paths that satisfy (2.25), is there one such that the implied carbon stock C_t approaches $\tilde{C}$ as $t \to \infty$, and might this be the optimum? The answer is no. It is easy to show that there is only one solution of (2.25) 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.

As a comparison between (2.30) and (2.35) confirms, the first-best time path of welfare differs in interesting ways across Case 1 and Case 2, that is, depending on whether the catastrophic carbon stock level $\tilde{C}$ is above or below the threshold level $\tilde{C}^{FB}$. In the first best under Case 2, where $\tilde{C} < \tilde{C}^{FB}$, the welfare achieved by each successive generation falls through time for $t < \tilde{t}^{FB}$ – 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 $\tilde{t}^{FB}$ 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 (2.35) that, contrary to the noncooperative and ICA outcomes, when the brink of climate catastrophe is reached under the first best the brink generation does not suffer a discrete drop in welfare relative to the previous generation.

We summarize the first-best emissions choices with:

Proposition 3. *The first-best path of emissions falls into one of two cases. If $\tilde{C}$ is above a threshold level, the brink of catastrophe is never reached and the first-best emissions levels are constant through time. Otherwise, if $\tilde{C}$ is below this threshold the first-best 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 First Best Outcomes

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

Consider first the possibility that $\tilde{C}$ falls in the high range $\tilde{C} \in [\tilde{C}^{ICA}, \tilde{C}^{N})$. 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 $\bar{c}^{FB}$ 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 first-best emissions choices $\bar{c}^{FB}$ 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 $\bar{c}^{FB} < \bar{c}^{ICA}$, with $\bar{c}^{FB}$ 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 first-best 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 first-best emissions as well as in the noncooperative equilibrium. For $\tilde{C}$ in this range, the qualitative features of the ICA and first-best 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.[24]

Consider next the possibility that $\tilde{C}$ falls in an intermediate range $\tilde{C} \in [\tilde{C}^{FB}, \tilde{C}^{ICA})$. 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 $\bar{c}^{FB}$ 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 first-best 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 first-best, but now utility under the ICA falls precipitously for the generation alive when the brink is reached, while under the first best 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

[24] We have depicted the level of welfare achieved by early generations in Figure 1 as dropping under the first best 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.

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 $\tilde{C} < \tilde{C}^{FB}$. 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 $\tilde{t}^N < \tilde{t}^{ICA} < \tilde{t}^{FB}$: the ICA slows down the march to the brink relative to the noncooperative outcome, but this march is still too fast relative to the first best. In this case as well there are qualitative differences across the ICA and first-best 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 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 first-best choices the emissions $\hat{c}_t^{FB}$ 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. *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 and hence cannot achieve the first best. For this reason, the ICA slows down the growth of the carbon stock relative to the noncooperative outcome but not enough relative to the first best. 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 first best. (ii) If $\tilde{C}$ lies in an intermediate range the world reaches the brink of catastrophe under the ICA but not under the first best. (iii) If $\tilde{C}$ is below a threshold level the brink is reached both under the ICA and the first best, but it is reached faster under the ICA.*

It is interesting to reflect more broadly on the role of an ICA. According to our 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.[25]

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 our setting 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 our 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

[25] While 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 19).

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.[26] When this is the case it is then only the 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 profound: 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 regeneration is close enough to one, the world will reach the brink of catastrophe under the ICA but not under the first best.*

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.

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

[26] More 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.

(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. 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 first-best outcome that a global social planner would implement in this case?

To answer these questions, in this section we allow countries to reach a catastrophe at different levels of the global carbon stock. To focus sharply on the implications of heterogeneous collapse points, we assume that countries are symmetric with respect to all other parameters; we will later comment on how our results may be affected if countries are asymmetric in other dimensions as well. And to facilitate our analysis of heterogeneous collapse points, we also assume that the cost of a country's collapse that is borne by its citizens 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. The citizens of a country that collapses would suffer a utility cost from this relocation, but the level of this cost depends on the available relocation opportunities and may not be infinite.[27]

Formally, and indexing countries by i, we assume that country i reaches the brink of collapse when C rises to the level $\tilde{C}_i$. And if C exceeds $\tilde{C}_i$, country i collapses, and its citizens become climate refugees and must relocate to other countries where they incur a per-period utility cost of $\bar{L}$.[28] With these assumptions, the payoff function for country i associated with the global carbon stock C_t is now

$$u_{i,t} = \begin{cases} B(c_{i,t}) - \lambda C_t & \text{if} \quad C_t \leq \tilde{C}_i \\ -\bar{L} & \text{if} \quad C_t > \tilde{C}_i, \end{cases} \tag{3.1}$$

where $c_{i,t}$ denotes the emissions of country i in period t.

[27] As we discuss further in the Conclusion, while we focus formally on heterogeneity across countries, many of the same issues that we consider below arise *within* countries, when there is heterogeneity in climate collapse points across different regions of a country due to distinct geographical and/or socioeconomic features. This regional heterogeneity within countries 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).

[28] With this specification we are assuming that the per-period utility cost incurred by climate refugees is constant and large enough to dwarf any time-varying contributions to utility that might be associated with a worsening climate or changing emissions levels in the country to which they immigrate.

The payoff function defined in (3.1) describes how each country is impacted directly by rising levels of the global carbon stock up to and beyond the level that would lead to the country's collapse. But since in this setting countries may potentially collapse at different points in time, it is also important to consider indirect effects, and in particular how the collapse of one country affects the remaining countries. In reality the collapse of a country would generate a whole host of consequences for the surviving countries. One immediate and major consequence would arguably be the outflow of refugees from the collapsing country. Another consequence would be the destruction of international trade between the collapsed country and the surviving countries.[29]

We will capture these consequences in reduced form by assuming that the collapse of a country imposes a loss R on the rest of the world. We do not need to impose 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. We will use the climate-refugee impact as our running example, and we will refer to R as the "refugee externality" associated with a country's collapse from climate change. To preserve tractability, we assume that the externality R is shared evenly by the surviving countries, that it is independent of how many countries have collapsed in the past, that it enters utility functions in an additively separable way (so it does not impact the emissions choices of the surviving countries), and that it is borne only once, in the same period as the country collapses.[30]

Finally, since there will often 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. We order countries according to increasing $\tilde{C}_i$, so that country 1 is the country with the lowest $\tilde{C}_i$, and we assume for simplicity that the ordering is strict, i.e. no two countries have the same value of $\tilde{C}_i$. And as in the

[29] We discuss the trade consequences at greater depth in the Conclusion.

[30] Recall that we earlier normalized the identical population of each country to one, and so the emissions choices of a country that we derived earlier were also its per-capita emissions choices. If a country receives climate refugees, its population will grow, and our assumption here that its emissions choices are not impacted then implies that its per-capita emissions will fall. This feature is convenient, but as we later explain (see note 38) it does not drive any of our results. We note also that, given the linearity of the cost and given our maintained assumption in this section that $\beta = 0$, we could allow the refugee externality to generate a permanent per-period cost, and results of this section would not change.

previous section, we focus on Markov perfect equilibria.

In principle now we have two state variables, C_t and the set of countries that have survived to time t. But the set of surviving countries is itself determined by C_t as follows: if $C_t \leq \tilde{C}_1$ all M countries have survived to time t; if $\tilde{C}_1 < C_t \leq \tilde{C}_2$ the countries $\{2, 3, ..., M\}$ have survived to time t, and so on. Hence, with the set of countries that have survived to time t itself a function of C_t, 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.

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 intuitive, and we will describe it without proof.

After an initial warming phase during which there are no catastrophes and each country selects the BAU emissions level $\bar{c}^N$ defined by $B'(\bar{c}^N) = \lambda$, the world enters a catastrophe phase when country 1 arrives at the brink of collapse. This occurs in finite time if $\bar{c}^N > \frac{\rho \tilde{C}_1}{M}$, a condition that we will assume is met (otherwise the possibility of catastrophes is irrelevant). Specifically, in complete analogy with Assumption 1 of the common-brink analysis of the previous section, we can define $\tilde{C}_1^N \equiv \frac{M}{\rho} B'^{-1}(\lambda)$ and impose

$$\tilde{C}_1 < \tilde{C}_1^N. \qquad \text{(Assumption 1}') \tag{}$$

Once country 1 is at its brink, there are two possibilities. If, by reducing its own emissions, country 1 can offset the rest of the world's BAU emissions and bring the global emissions down to the level $\rho \tilde{C}_1$, then it can avoid collapse. This requires $\rho \tilde{C}_1 - (M-1)\bar{c}^N \geq 0$. On the other hand, if $\rho \tilde{C}_1 - (M-1)\bar{c}^N < 0$, then country 1 cannot offset the BAU emissions of the rest of the world, and it will collapse on the equilibrium path. Thus, country 1 will survive if and only if it is able to offset the BAU emissions of the rest of the world.

More generally, once the catastrophe phase begins, the first country to survive and thus prevent further catastrophes is determined when $R = 0$ by the minimum $\tilde{C}_i$ such that $\rho\tilde{C}_i - (M - i)\bar{c}^N \geq 0$.[31] Note that country M – the most resilient country – will survive on the equilibrium path (because by setting $c = \rho\tilde{C}_M \geq 0$ it can freeze the global carbon stock at the brink level $\tilde{C}_M$).

Hence, with heterogeneous collapse points and $R = 0$, under mild conditions the world will traverse through all three phases of climate change along the equilibrium noncooperative path – and some but not all countries will collapse. These conditions describe a world where the most vulnerable country has limited ability to offset carbon emissions from the rest of the world, and where the least vulnerable country is able to keep the stock of carbon from rising once it remains the lone surviving country. This world provides an illuminating counterpoint to our earlier common-collapse-point analysis, 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 in the world, who with $R = 0$ have no reason in the noncooperative equilibrium to internalize the impact of their emissions choices on the fate of the brink country.[32] Notice also that even slight differences in collapse points across countries can create the possibility of 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-collapse-point 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 refugees ($R > 0$). Here we will offer an intuitive exposition, relegating the formal proof to the Appendix.

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

[31] We are implicitly assuming that each country prefers to set $c = 0$ rather than collapsing, that is $B(0) - \lambda\tilde{C}_i > -\bar{L}$. And while the observation we make here is obvious within our model, note that if we introduced lags in the effects of emissions, it might be possible for country M to collapse. We consider the possibility of lags in a later section, and will revisit this point then.

[32] It is also notable that, relative to the common-brink setting, 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, bringing 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.

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. We work backwards, starting from high levels of C_t. Clearly, C_t can never go beyond $\tilde{C}_M$ on the equilibrium path, so we can start with the case $C_t = \tilde{C}_M$ (ignoring as usual the discrete-time constraint). Here only country M has survived, and it will restrain its emissions just enough to avoid collapse, so $c_M^N(\tilde{C}_M) = \rho\tilde{C}_M$.

Next consider the time interval where $C_t \in (\tilde{C}_{M-1}, \tilde{C}_M)$. Here too country M is the only surviving country, and it will choose its BAU emissions, so $c_M^N(C_t) = \bar{c}^N$ for all $C_t \in (\tilde{C}_{M-1}, \tilde{C}_M)$.

We next focus on the time period t where $C_t = \tilde{C}_{M-1}$. Country $M-1$ has survived up to this point (along with country M), and will survive in period t and beyond if and only if $c_{M-1} + c_M \leq \rho\tilde{C}_{M-1}$. There are four possibilities:

(1) If $\bar{c}^N \leq \frac{\rho}{2}\tilde{C}_{M-1}$, clearly the only equilibrium is for both countries to choose their BAU emissions. In this case the carbon stock stays below $\tilde{C}_{M-1}$ and country $M-1$ survives.

(2) If $\frac{\rho}{2}\tilde{C}_{M-1} \leq \bar{c}^N \leq \rho\tilde{C}_{M-1}$, the carbon stock reaches $\tilde{C}_{M-1}$ but country $M-1$ is able to offset country M's BAU emissions and save itself by setting $c_{M-1} = \rho\tilde{C}_{M-1} - \bar{c}^N$. In this case it is clearly an equilibrium for country M to choose $\bar{c}^N$ and for country $M-1$ to offset these emissions.[33] It can also be easily shown that in this case there are no equilibria where country $M-1$ does not survive.

(3) If $\bar{c}^N > \rho\tilde{C}_{M-1}$ and R is above some threshold $\tilde{R}_{M-1}$, country $M-1$ is not able to offset country M's BAU emissions, but there are equilibria where $c_{M-1} + c_M = \rho\tilde{C}_{M-1}$ and country $M-1$ survives.[34] In these equilibria, it is in country M's own self interest to "top off" the abatement efforts of country $M-1$ so that country $M-1$ survives and a climate refugee crisis is prevented.

[33] As we show in the Appendix, there is a range of other equilibria, all of which are equivalent to the one highlighted in the text in terms of aggregate emissions and country survival outcomes. In these equilibria, $c_{M-1} + c_M = \rho\tilde{C}_{M-1}$ and country M emits less than $\bar{c}^N$ to "help" country $M-1$ avert collapse and thus avoid the refugee externality R. We have chosen to highlight in the text the "self-help" equilibrium on the grounds that it is arguably the most intuitive equilibrium, since country $M-1$ has more to lose from its own collapse than country M, but for the purposes of our main results we do not need to commit to a particular equilibrium. It is also worth noting that the self-help equilibrium does not maximize the joint payoff of the two countries, since it does not equalize their marginal benefit of emissions ($B'(\cdot)$); but also note that the many equilibria of this game 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.

[34] Unlike case (2), in this case there are also some equilibria where country $M-1$ does not survive, but these equlibria are in weakly dominated strategies and it is thus reasonable to ignore them.

(4) If $\bar{c}^N > \rho\tilde{C}_{M-1}$ and R is below the threshold $\tilde{R}_{M-1}$, then the prospect of climate refugees from country $M-1$ is not sufficient to motivate country M to help country $M-1$, and thus there is no equilibrium where country $M-1$ survives. Instead, the only equilibrium has country M choosing the BAU emissions $\bar{c}^N$ and country $M-1$ collapsing.[35]

Moving backwards, we now consider the case $C_t \in (\tilde{C}_{M-2}, \tilde{C}_{M-1})$. Over this time interval, each of countries M and $M-1$ chooses the BAU emissions $\bar{c}^N$. And more generally, when C_t is strictly in-between catastrophe points, the countries that have survived to that point choose the BAU emissions.

Finally we consider levels of C_t such that some country is on the brink (say country k) and there are at least two other surviving countries, that is $C_t = \tilde{C}_k$ with $k \leq M-2$. In this case, by the same logic as above, if $\rho\tilde{C}_k - (M-k)\bar{c}^N \geq 0$ then country k saves itself from collapse. And, as in the case analyzed above where the country at the brink is the second-most resilient country ($k = M-1$), we can say more generally that, if country k is at its brink and not able to offset the rest of the world's BAU emissions ($\rho\tilde{C}_k - (M-k)\bar{c}^N < 0$), then it will survive if and only if R is above a threshold $\tilde{R}_k$.[36] To determine this threshold, recall that we are focusing on the "self-help" equilibrium where the country at the brink does everything feasible to save itself (by reducing its carbon emissions to zero) and the remaining countries top off these efforts by each adopting the emissions level $\frac{\rho\tilde{C}_k}{M-k} < \bar{c}^N$. The no-defect condition for such an equilibrium can be written as:

$$B\left(\frac{\rho\tilde{C}_k}{M-k}\right) - \lambda\tilde{C}_k = \max_c \left[B(c) - \lambda\left(\tilde{C}_k + c - \frac{\rho\tilde{C}_k}{M-k}\right) - \frac{R}{M-k}\right]. \tag{3.2}$$

The left-hand-side of (3.2) is the payoff to a representative country other than k under the self-

[35] It is interesting to note that this equilibrium is Pareto efficient. To see this, note that the best chance to achieve a Pareto improvement over emissions levels ($c_{M-1} = 0, c_M = \bar{c}^N$) is to lower c_M from $\bar{c}^N$ to $\rho\tilde{C}_{M-1}$. But if this improves country M's payoff, then we must be in case (3), and hence this must itself be a noncooperative equilibrium.

[36] As we detail in the Appendix, when $\rho\tilde{C}_k - (M-k)\bar{c}^N < 0$ and R lies between the lower threshold $\tilde{R}_k$ and an upper threshold $\hat{R}_k$, there are both equilibria with and without survival of country k, but the equilibrium where country k does not survive is Pareto-dominated by some other equilibrium where country k survives, and hence we can ignore it given our assumption that countries do not play a Pareto-dominated equilibrium. Intuitively, it may be in the interest of the more resilient countries to save country k, but there may be a coordination problem among them: it may be an equilibrium for all of them to choose their BAU emissions, because a single country may not find it worthwhile to cut emissions enough to save country k all by itself; and it may also be an equilibrium for all of them to pitch in and do their part to save country k, because given that country k is on the brink, a unilateral deviation to a higher level of emissions would cause country k to collapse, imposing the refugee cost R on the surviving countries. It is not hard to see that the former equilibrium is Pareto dominated by the latter.

help equilibrium emissions levels in which country k does not collapse, and the right-hand-side is the payoff to the representative country were it to defect from this emissions level and cause country k to collapse. The threshold $\tilde{R}_k$ is the value of R that satisfies the no-defect condition (3.2) with equality.

It is intuitive and easy to show using (3.2) that the threshold $\tilde{R}_k$ decreases with the number of countries that have collapsed in the past, and hence with k.[37] This is because as more countries collapse the burden of their climate refugees is shared by fewer and fewer surviving countries, who therefore have a stronger incentive to help the country on the brink avoid a collapse. Thus, 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, for two reasons: the condition $\rho\tilde{C}_k - (M-k)\bar{c}^N \geq 0$ is more likely to be satisfied, so the country is more likely to be able to save itself; and R is more likely to be above the threshold $\tilde{R}_k$, so the other countries are more likely to have the incentive to help prevent its collapse.

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 ($C_t = \tilde{C}_k$): (1) if $\rho\tilde{C}_k - (M-k)\bar{c}^N \geq 0$ then country k saves itself from collapse regardless of R; (2) if $\rho\tilde{C}_k - (M-k)\bar{c}^N < 0$ and R is above a threshold $\tilde{R}_k$, then country k survives with the help of emissions reductions below BAU levels by other countries; (3) if $\rho\tilde{C}_k - (M-k)\bar{c}^N < 0$ and $R < \tilde{R}_k$, then country k collapses and the surviving countries continue to choose their BAU emissions. The threshold $\tilde{R}_k$ decreases with k.*

Proof: See Appendix.

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 (possibly 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

[37] In particular, ignoring country integer constraints and totally differentiating (3.2) with respect to $\tilde{R}_k$ and k yields

$$\frac{d\tilde{R}_k}{dk} = -\left[(M-k)\cdot(B'(\frac{\rho\tilde{C}_k}{M-k}) - \lambda)\cdot[\frac{\rho}{M-k}\frac{\partial\tilde{C}_k}{\partial k} + \frac{\rho\tilde{C}_k}{(M-k)^2}]\right] < 0$$

where we have used the envelope theorem and the fact that $\frac{\rho\tilde{C}_k}{M-k} < \bar{c}^N$.

$1, 2, ..., i-1$) have all 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, because then the country at the brink shares the world with fewer countries, and hence (a) the refugee externality imposed on each surviving country is higher, and (b) the aggregate BAU emissions are lower and hence it is more likely that the country at the brink is able to offset them.[38] So in this sense there is also an "anti-domino effect" in our model.[39]

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. Once C_t reaches the level that endangers country 1 ($\tilde{C}_1$), under the mild conditions highlighted above this country collapses and the rest of the world marches on with their BAU emissions. In a similar fashion, the growth of the carbon stock will cause the sequential collapse of further countries, and at some point the carbon stock stops growing and the string of catastrophes ends, either because the increasing potential refugee externality persuades the surviving countries to be proactive and help the country at the brink avoid collapse, or because the country at the brink is able to offset the remaining countries' BAU emissions, or both. The following proposition summarizes the qualitative predictions of the model regarding the survival and collapse of countries on the equilibrium path:

Proposition 5. *Suppose the catastrophe point $\tilde{C}_i$ differs across countries: (i) If the most vulnerable country is not able to offset the rest of the world's BAU emissions and the refugee externality exerted by its collapse is not severe enough, then a non-empty subset of countries*

[38] The reason that aggregate BAU emissions fall when a country collapses in our model is mechanical, given our assumption that a country receiving climate refugees does not alter its BAU emissions level. But we note that our model abstracts from three other forces that would go in the same direction: (a) if each surviving country re-optimizes its BAU emissions level after receiving climate refugees, aggregate BAU emissions will fall, because with each surviving country now comprising a greater fraction of the world population, the BAU emissions levels of each surviving country would be lower on a per-capita basis. (b) If part of the collapsing country's population dies, the total world population will fall, and this will push in the same direction of lower aggregate BAU emissions; (c) 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. For this reason we feel justified in emphasizing this channel, even though our simple model delivers it in a mechanical way.

[39] Our model is suggestive of a further force that goes against domino effects. While we have assumed that the refugee externality on the rest of the world (R) does not depend on the cumulative number of collapsed countries, it would be reasonable to suppose that surviving countries' populations increase as they absorb the climate refugees from the countries that collapse before them, and so they generate more climate refugees if they collapse. This would further reinforce the anti-domino effect highlighted in the text.

collapses 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 $1, 2, ..., i-1$) 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 ("anti-domino effect"), because the refugee externality imposed on each surviving country is then higher, and because the aggregate BAU emissions are lower and hence it is easier to offset them.

Proof: See Appendix.

The conditions summarized in Proposition 5(i), under which a subset of countries collapses on the noncooperative equilibrium path, are arguably quite weak. First, in reality a highly vulnerable country like the Maldives has no ability to offset the rest of the world's BAU emissions; and second, the negative externality felt by other countries if one of these vulnerable countries suffers an early collapse will be limited, due both to the 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.

Note also the contrast with the earlier case of a common catastrophe point across countries, 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.

Finally, notice that the cost that a collapsing country incurs itself ($\bar{L}$) does not affect the set of countries that survive along the equilibrium noncooperative path: only the cost that the collapsing country would impose on other countries (R) is relevant for its survival. The former cost will become relevant when we consider the ICA equilibrium and the social planner optimum, and it accounts for a key difference in survival outcomes between these settings and the noncooperative equilibrium.

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 first-best choices of the planner) we will introduce international transfers explicitly into the model. In our formal analysis we will also assume that there is effectively 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 first best. We will also comment, however, on how our results would be effected 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.[40]

Formally, we model international transfers as lump-sum transfers of an outside good that enters additively into utility. 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) transfer made by country i at time t in terms of the outside good. The utility of generation t in country i is then given by

$$u_{i,t} = \begin{cases} B(c_{i,t}) - \lambda C_t - Z_{i,t} & \text{if} \quad C_t \leq \tilde{C}_i \\ -\bar{L} & \text{if} \quad C_t > \tilde{C}_i, \end{cases}$$

where we have omitted 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.

[40] Moreover, in the microfounded model presented in the Appendix, intergenerational transfers are not even feasible under the assumption that goods are nonstorable.

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 and possibly transfers between them.

Consider first the initial warming phase, in which the carbon stock is below the catastrophe point for country 1 ($C_t < \tilde{C}_1$).[41] Recall that we are abstracting from intergenerational altruism by setting $\beta = 0$. Given that in this phase there are no differences in payoff functions across countries, and assuming symmetric bargaining powers, it follows that the ICA selects the symmetric level of emissions that maximizes the common per-period payoff, just as in the warming phase of the common-brink setting analyzed in the previous section. Thus for all $C_t < \tilde{C}_1$ the ICA emissions level for each of the M countries is given by $\bar{c}^{ICA}_{\{M\}} \equiv B'^{-1}(M\lambda)$. As before, the ICA internalizes the international climate externalities that travel through λ, and hence lowers emissions below the BAU level $\bar{c}^N = B'^{-1}(\lambda)$. Note that no international transfers are needed in this phase.

In analogy with our common-brink analysis of ICAs, we can define $\tilde{C}^{ICA}_1 \equiv \frac{M}{\rho} B'^{-1}(M\lambda)$ as the level to which the carbon stock would eventually converge under the emissions level $\bar{c}^{ICA}_{\{M\}}$, and note that the carbon stock never reaches $\tilde{C}_1$ under the ICA – and so country 1 is never brought to the brink of catastrophe – if $\tilde{C}_1 \geq \tilde{C}^{ICA}_1$, and in this case ICA emissions remain at the level $\bar{c}^{ICA}_{\{M\}}$ forever. By their definitions we have $\tilde{C}^{ICA}_1 < \tilde{C}^N_1$, and so it follows that both Assumption $1^{/}$ and this condition will be met if $\tilde{C}_1 \in [\tilde{C}^{ICA}_1, \tilde{C}^N_1)$. If $\tilde{C}_1 < \tilde{C}^{ICA}_1$, on the other hand, country 1 is brought to the brink under the ICA, and we need to consider what happens next.

Suppose, then, that $\tilde{C}_1 < \tilde{C}^{ICA}_1$, 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 a non-empty subset of the most vulnerable countries, say countries $\{1, ..., \tilde{k}^N - 1\}$, would collapse in the absence of an ICA, that is, in the noncooperative equilibrium (recall from our analysis of the noncooperative equilibrium path that this subset is non-empty under rather weak conditions). Formally, ignoring country integer constraints we can use the no-defect condition in (3.2) with

[41] Notice that here we analyze the equilibrium path moving forward, while in the noncooperative scenario we proceeded backwards. The reason is expositional simplicity: in the noncooperative case we chose to work backwards because the analysis is simpler when there are only two surviving countries, while in the ICA the description of the equilibrium path is simpler if we proceed forward.

fixed R to solve for the marginal surviving country $\tilde{k}^N$ in the noncooperative equilibrium,

$$B\left(\frac{\rho\tilde{C}_{\tilde{k}^N}}{M-\tilde{k}^N}\right)-\lambda\tilde{C}_{\tilde{k}^N}=\max_c\left[B(c)-\lambda\left(\tilde{C}_{\tilde{k}^N}+c-\frac{\rho\tilde{C}_{\tilde{k}^N}}{M-\tilde{k}^N}\right)-\frac{R}{M-\tilde{k}^N}\right], \tag{3.3}$$

where our assumption is then that $\tilde{k}^N > 1$. Hence, 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 now faces a stark choice: it can cooperate to save country 1 from collapse, or it can let country 1 collapse, and march on.

With the availability of international lump-sum transfers, the ICA will choose emissions levels to maximize global welfare from the point of view of generation t (and use the transfers to distributed the surplus according to bargaining powers), and this implies that country 1 will be saved if and only if the global loss from the collapse of country 1, which is comprised of country 1's own loss $\bar{L}$ and the refugee externality R that its collapse would impose on others, 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 is willing to compensate the rest of the world for contributing to stop the growth of C_t. Note that the efficient way to save country 1 is for all countries to reduce emissions to the level $\rho\tilde{C}_1/M$, since efficiency requires the marginal benefit from emissions to be equalized across countries. And if country 1 is allowed to collapse, the remaining $M-1$ countries would under the ICA choose the optimal emissions level.

More generally and with the above logic in mind, if the brink country $\tilde{k}^{ICA}$ is at the margin of survival and collapse under the ICA, then we have that $\tilde{k}^{ICA}$ is defined by

$$B\left(\frac{\rho\tilde{C}_{\tilde{k}^{ICA}}}{M-(\tilde{k}^{ICA}-1)}\right)-\lambda\tilde{C}_{\tilde{k}^{ICA}}=\left(\frac{M-\tilde{k}^{ICA}}{M-(\tilde{k}^{ICA}-1)}\right)\cdot\max_c\left[B(c)-\lambda\left((1-\rho)\tilde{C}_{\tilde{k}^{ICA}}+(M-\tilde{k}^{ICA})c\right)\right]-\frac{R+\bar{L}}{M-(\tilde{k}^{ICA}-1)}, \tag{3.4}$$

where the left hand side is global welfare when all countries emit at the level $\rho\tilde{C}_{\tilde{k}^{ICA}}/[M-(\tilde{k}^{ICA}-1)]$ and country $\tilde{k}^{ICA}$ survives, and the right hand side is global welfare when country $\tilde{k}^{ICA}$ collapses and the other countries re-optimize their emissions levels under the ICA.

Using (3.4) it is easily confirmed and intuitive that $\tilde{k}^{ICA}$ rises as R and/or $\bar{L}$ fall, indicating that more countries will collapse under the ICA when the refugee externality from collapse is

small, or when the utility cost of collapse for the citizens of the collapsing country are small. Notice, too, that using (3.3), the same conclusion for $\tilde{k}^N$ can be drawn with regard to R, namely, more countries will collapse in the absence of an ICA when the refugee externality from collapse is small; but the utility cost of collapse for the citizens of the collapsing country $\bar{L}$ plays no role in the determination of $\tilde{k}^N$. Finally, using (3.4) and (3.3) and ignoring country integer constraints, it is direct to establish that $\tilde{k}^{ICA} < \tilde{k}^N$: given our focus on the case in which a non-empty subset of the most vulnerable countries would collapse in the absence of an ICA, the ICA will save some, though not necessarily all, of these countries.

Summarizing, we may now state:

Proposition 6. *When the catastrophe points $\tilde{C}_i$ differ across countries, the ICA may or may not save a country that would collapse in the noncooperative scenario. More specifically: (i) No country will be allowed to collapse under the ICA that would not have collapsed in the absence of the ICA; (ii) If a range of the most vulnerable countries would collapse in the absence of an ICA, the ICA will save some of the least vulnerable in this range from collapse; (iii) Conditional on a country reaching the brink of catastrophe under the ICA, that country is less likely to be saved by the ICA if the internal and external costs of the country's collapse (R and $\bar{L}$) are lower.*

Proof: See Appendix.

Finally, we have assumed that countries do not face binding constraints in their ability to make transfers. This assumption, together with the assumption that ICAs do not face participation or enforcement issues, will allow us to focus sharply on the shortcomings relative to the first best that ICAs face because of the inability of future generations to sit at the bargaining table, when we compare the ICA outcome with the first best outcome in the next section. But it is important to highlight that, if international transfers are limited because of resource constraints, the ICA will have even more limited ability to save countries from collapse. 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 is efficient to save the country the ICA can orchestrate this outcome only if the country has enough resources to compensate the remaining countries for cutting 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. First Best

We next turn to the first-best emissions levels that would be chosen by a global social planner in the setting where the catastrophe point differs across countries. Given that international lump-sum transfers are available, the planner maximizes global welfare and uses international transfers to redistribute the "pie" across countries according to their Pareto weights (which we can leave in the background), so the problem can be written as:

$$\begin{aligned} & \max \sum_{t=0}^{\infty} \sum_{i=1}^{M} \hat{\beta}^t u_{i,t} \\ \text{s.t. } C_t &= (1-\rho)C_{t-1} + \sum_{i=1}^{M} c_{i,t} \text{ for } t \geq 1 \\ c_{i,t} &\geq 0 \text{ for all } i,t \end{aligned}$$

where the utility $u_{i,t}$ is given by (3.1). Again we focus for simplicity on the case where the emissions feasibility constraints $c_{i,t} \geq 0$ are not binding.

Given the discontinuities in the payoff functions $u_{i,t}$, 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 first-best solution with direct arguments.

Recall that under Assumption $1^{/}$ we are focusing on the case in which country 1 would collapse in the noncooperative scenario, and recall also from our discussion leading up to Proposition 6 that under the ICA, if country 1 is brought to the brink of collapse ($C_t = \tilde{C}_1$) then it will be saved if and only if $\bar{L} + R$ is above some critical level, which we now denote by $\left(\bar{L} + R\right)_1^{ICA}$.

A first point is intuitive, and follows a similar logic to that in the common-brink setting: each given country is less likely to reach the brink of collapse under the first best than under the ICA. The reason is that the first-best global carbon stock can never exceed the global carbon stock under the ICA, because in addition to the international externalities the planner also

internalizes the intergenerational externalities, so if a country never reaches the brink under the emissions choices of the ICA then it won't under the planner's choices either.

The next point is that, conditional on the carbon stock reaching $\tilde{C}_1$, the planner will save country 1 if and only if $\bar{L}+R$ is above a threshold $\left(\bar{L}+R\right)_1^{FB}$ that is lower than $\left(\bar{L}+R\right)_1^{ICA}$, and so country 1 is more likely to be saved under the first best than under the ICA. The intuition follows from the observation that the planner takes into account the consequences for current and future generations of whether or not to save country 1 while the ICA only considers the consequences for the current generation; and while the value of saving country 1 for the planner is $\frac{1}{1-\hat{\beta}}$ times the corresponding value for the ICA, since the carbon stock will then remain at $\tilde{C}_1$ forever and hence the future will be stationary, the value of letting country 1 collapse for the planner is less than $\frac{1}{1-\hat{\beta}}$ times the corresponding value for the ICA, because if country 1 is allowed to collapse the carbon stock will grow in the future, and this will impose costs on future generations which the planner takes into account (and which are ignored by the ICA).[42] Moreover, this result applies to any country k that reaches the brink of catastrophe both under the ICA and under the first best: if country k is saved by the ICA it will also be saved by the planner, and as a consequence, the planner allows weakly fewer countries to collapse than under the ICA (and strictly fewer for some parameter configurations).

More formally, suppose that under the ICA country k is brought to the brink of catastrophe (that is, C_t reaches $\tilde{C}_k$), but it survives. If the planner is myopic ($\hat{\beta}=0$) it will make exactly the same choices as the ICA, and hence it too will save country k. Is it possible that a more patient planner (with $\hat{\beta}>0$) would allow country k to collapse? We now establish that the answer is no.

To this end, recall from the discussion leading up to (3.4) that the condition for country k to be saved under the ICA can be written as:

$$(M-k+1)\left[B(\frac{\rho\tilde{C}_k}{M-k+1})-\lambda\tilde{C}_k\right]\geq(M-k)\max_c\left[B(c)-\lambda\left(\tilde{C}_k+(M-k)c-\rho\tilde{C}_k\right)\right]-R-\bar{L}. \tag{3.5}$$

For future reference, we let $v_{M-k}^{ICA}(\tilde{C}_k)$ denote the right hand side of (3.5), that is, the maximum joint payoff for the $M-k$ most-resilient countries given that the carbon stock is $\tilde{C}_k$ and country

[42] Notice that if country 1 collapses under the ICA, the planner's decision to save country 1 under these circumstances does not mark a Pareto improvement over the ICA outcome: as we have observed above (see also note 35), in this case the ICA is Pareto efficient along the dimension of that country's survival. Instead, the planner's optimum simply reflects a different point on the Pareto frontier where positive weight has been placed on the welfare of the *future* generations of country 1 (whereas they receive zero weight under the ICA).

k collapses (so only $M-k$ countries remain). The condition for country k to be saved by the planner can then be written as:

$$\frac{1}{1-\hat{\beta}}(M-k+1)\left[B(\frac{\rho\tilde{C}_k}{M-k+1})-\lambda\tilde{C}_k\right]\geq v^{FB}_{M-k}(\tilde{C}_k) \tag{3.6}$$

where $v^{FB}_{M-k}(\tilde{C}_k)$ is the value function for the planner given that the carbon stock is $\tilde{C}_k$ and country k collapses. The left hand side of (3.6) is the discounted value of stopping the growth of carbon (in an efficient way, by allocating emissions evenly across countries).[43]

As we observed above, if $\hat{\beta}=0$ the first best coincides with the ICA solution, thus $v^{FB}_{M-k}(C_t)=v^{ICA}_{M-k}(C_t)$ for all C_t if $\hat{\beta}=0$. We now argue that $v^{FB}_{M-k}(\tilde{C}_k)<\frac{1}{1-\hat{\beta}}v^{ICA}_{M-k}(\tilde{C}_k)$ if $\hat{\beta}>0$, from which it then follows immediately that, if (3.5) is satisfied, then also (3.6) must be, and hence fewer countries collapse under the first best than under the ICA. The argument is straightforward. Suppose for a moment that $\rho=1$. Then the environment is effectively static, and the first best given the initial condition $C_t=\tilde{C}_k$ coincides with the ICA solution, and the corresponding value for the social planner is $v^{FB}_{M-k}(\tilde{C}_k)=\frac{1}{1-\hat{\beta}}v^{ICA}_{M-k}(\tilde{C}_k)$. Now decrease ρ from one: an envelope argument establishes that this must reduce the maximum attainable value for the planner, and hence $v^{FB}_{M-k}(\tilde{C}_k)<\frac{1}{1-\hat{\beta}}v^{ICA}_{M-k}(\tilde{C}_k)$ if $\hat{\beta}>0$ and $\rho<1$.

The discussion above suggests a simple but important insight. In the best of circumstances, when ICAs face no country-participation or enforcement issues and unlimited international transfers are available, ICAs may have a role to play in avoiding catastrophic collapses, but can only be an imperfect substitute for the planner in this regard, owing to the inability of future generations to have a seat at the ICA negotiating table. We summarize with:

Proposition 7. *When the global carbon stock at which a catastrophe would be triggered differs from one country to the next, the first-best outcome that would be implemented by a global social planner may or may not let some of the most vulnerable countries collapse; but the planner will allow weakly fewer countries to suffer collapse than would the ICA.*

We have thus far assumed that international lump-sum transfers are available without limit. We now briefly consider the polar opposite case, in which such transfers are not available at all. A heuristic discussion will suffice here.

[43] Note that, in writing the left hand side of (3.6), we are using the fact that, given the Markovian nature of the problem which implies that the optimum can be represented as a mapping from the state variable C_t to the emissions vector, if it is optimal to set per-country emissions at $\frac{\rho\tilde{C}_k}{M-k+1}$ when the carbon stock is $\tilde{C}_k$, so that the carbon stock will be $\tilde{C}_k$ again in the next period, then it is optimal to do the same in all future periods.

Suppose initially that the planner assigns symmetric Pareto weights to all countries. This provides a natural starting point, given that countries are symmetric in all respects except for the collapse points. Clearly, with symmetric Pareto weights transfers do not matter in the planner's problem, because the planner does not face participation constraints (that is, there is no requirement that the planner must make all countries better off than they would be in the noncooperative equilibrium, in the same way that there is no requirement that all generations must be made better off than in the noncooperative equilibrium). So in this case the first best is the same as in the case where transfers are available. And since, as we have observed, in the absence of transfers the ICA lets more countries collapse than in the presence of transfers, we can conclude that the number of countries that the planner "saves" relative to the ICA is larger in the absence of transfers.

Now allow Pareto weights to be asymmetric. Since countries differ only in the dimension of their vulnerability to climate change ($\tilde{C}_i$), it is natural to consider two salient cases: a scenario where the planner attaches larger weights to more vulnerable countries, and the opposite scenario where the more resilient countries have larger weights. Since more vulnerable countries fare worse than more resilient countries in the noncooperative equilibrium and under the ICA, the first scenario can be interpreted as one where the planner is averse to cross-country inequality, and therefore might be more natural than the opposite scenario. If the planner is inequality-averse in the sense just described, intuitively this will further widen the distance between the ICA outcome and the first best outcome in terms of the number of countries that are allowed to collapse. On the other hand, if the more resilient countries are those that carry higher Pareto weights, this will push in the opposite direction and reduce the number of countries that collapse under the ICA but would be saved by the planner.

4. Extensions

TBA

5. Conclusion

TBA

6. Appendix

TBA

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 (2018), "Environmental catastrophes and mitigation policies in a multiregion world," *Proceedings of the National Academy of Sciences* 16: 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.

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.

Farhi, Emmanuel and Ivan Werning (2007), "Inequality and Social Discounting," *Journal of Political Economy* 115(3): 365-402.

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.

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.

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.

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.

Nordhaus, W. 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.

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.