International Climate Agreements and the Scream of Greta*

Giovanni Maggi† Robert W. Staiger‡

December 2022

Abstract

The world appears to be facing imminent peril, as countries are not doing enough to keep the Earth's temperature from rising to catastrophic levels and various attempts at international cooperation have 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? What role can international climate agreements play, if any? We address these questions through a model that features the possibility of climate catastrophe and emphasizes the role of international externalities that a country's policies exert on other countries and intertemporal externalities that current generations exert on future generations. Our analysis illuminates the role that international climate agreements can be expected to play in addressing climate change, and it points to important limitations on what such agreements can achieve, even under the best of circumstances.

*We thank for helpful comments and discussions Scott Barrett, Jonathan Bendor, Klaus Demset, Allan Hsiao, Emanuel Ornelas, Steve Redding, Jillian Stallman, Matthew Turner, participants at the 2021 NBER Future of Globalization conference, the 2021 International Trade and Environmental Policy workshop, the 2022 Annual Meetings of the ASSA, the 2022 IEFS China Annual Conference, the 2022 Political Economy Sustainability Conference, and seminar participants at Bocconi University, Boston College, Florida State University, Georegtown, the Harvard/MIT joint seminar, Penn State, Princeton University, the Nuremberg Research Seminar in Economics, Singapore Management University, Syracuse University, the University of Oslo and Yale University (Economics and the Leitner Program). Winston Chen, Wei Xiang and Yan Yan provided outstanding research assistance. Giovanni Maggi acknowledges financial support from the NSF Grant SES-1949374.

†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." (David Wallace-Wells, The Uninhabitable Earth, 2019, p.4)

1 Introduction

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

In this paper we address these questions through a formal model that features the possibility of climate catastrophe and emphasizes 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, and the extent to which international climate agreements can mitigate these problems.

Previous research has focused on the role of international climate agreements in a world without the possibility of catastrophe, and has highlighted two challenges that such agreements must meet, relating to country participation and enforcement.[3] In this paper we abstract from these well-studied challenges, and focus instead on a limitation that has not been emphasized in the formal literature, namely 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 international ("horizontal") externalities from emissions choices, it cannot address the intergenerational ("vertical") externalities exerted by those choices, nor can it address the "diagonal" externalities exerted by a country's current climate policy on future generations in other countries. A key objective of our analysis is to examine the consequences of this limitation of

[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.

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

climate agreements in a world where catastrophic outcomes are possible, and the role that climate agreements can potentially play in the presence of this limitation.

We work with a model 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 degrade the quality of the environment. This tension defines the fundamental tradeoff faced by each generation. In our core analysis we abstract from intergenerational altruism, and then later allow each generation to care about its offspring.

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

We consider two possible scenarios for climate catastrophes. In our common-brink scenario, all countries are brought to the brink of climate catastrophe at the same moment, when the global carbon stock reaches a critical level. In our heterogeneous-brink scenario, different countries have different catastrophe thresholds, thus more vulnerable countries reach the brink first. As we demonstrate, these two possibilities carry starkly different implications for outcomes along the equilibrium path and for the potential role of an ICA. We view these scenarios as two important benchmarks, with the real world arguably falling somewhere in between.

We begin with an analysis of the common-brink setting. In the absence of an ICA, 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 can avert catastrophe with an 11^{th} hour solution that has each country doing its part to halt further climate change; we show that this 11^{th} hour solution will prevail in any Pareto-undominated equilibrium. The solution involves reduced emissions levels that keep the carbon stock constant given the natural rate of atmospheric regeneration and remain at that level for all generations thereafter, and as a consequence, the brink generation and all future generations suffer a precipitous drop in utility relative to the pre-brink generations. The reason for this 11^{th} hour solution is that, while earlier generations face moderate costs of global warming as a result of their emissions choices, it is only the brink generation that faces the catastrophic implications of continuing the emissions practices of the past, and this fundamentally alters the nature of the game, with the result that the brink generation "does whatever it takes" to avoid catastrophe.

The noncooperative equilibrium in our common-brink scenario 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, because the brink generation and all future generations must accept a potentially large drop in consumption and utility relative to previous generations in order to prevent climate catastrophe. 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 argue that the unfortunate situation of the brink generation in our model may capture in a highly stylized way the essence of the plight of the climate activist Greta Thunberg and her generation, and thus our model may provide a useful lens to interpret the fraught debate between Greta's generation and the older generations currently in charge of climate policies.

We next ask: How can an ICA improve over the noncooperative equilibrium in our common-brink setting? We find that up until the world reaches the brink, the ICA plays the standard role of internalizing the international climate externalities, thus lowering emissions relative to noncooperative levels. As a by-product of this, the ICA delays the moment that the brink is reached, and may even avoid the brink altogether. But if the world reaches the brink, the role of an ICA is transformed: at that point the ICA can at most play a coordination role, if in the noncooperative equilibrium countries would fail to focus on a Pareto-undominated equilibrium and would instead go over the brink. Unlike the cooperation role played during the warming phase, for this coordination role there is no need for the ICA to enforce any commitments. And if coordination can be achieved without an ICA, then according to the common-brink scenario, at the brink of catastrophe a role for the ICA ceases to exist completely.

A remaining question is how the outcome achieved by an ICA compares with the outcome that would be implemented by a "social planner" who maximizes an intergenerational social welfare function. While the ICA negotiated by each generation in effect picks an extreme point on the Pareto frontier that places zero weight on the utility of future generations directly, we focus on the possibility that the social planner might instead place strictly positive weight on future generations directly, and hence takes into account not only the horizontal but also the vertical and diagonal climate externalities. We show that, while an ICA slows down the growth of the carbon stock relative to the noncooperative outcome, it does not do so enough relative to the social optimum. This leads to three possibilities, depending on the severity of the constraint imposed by the catastrophe threshold. If this constraint is sufficiently mild, the ICA will prevent the world from reaching the brink of catastrophe, but the carbon stock is still too large relative to the social optimum at every point in time; if the constraint is sufficiently severe, the brink will be reached both under the ICA and the social optimum, 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 social optimum. It is when the catastrophe 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 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 collapses, its citizens become climate refugees and suffer a utility cost themselves while also imposing "refugee externality costs" on the countries who receive them. In the noncooperative equilibrium 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 scenario; the possibility of a middle phase in which some countries collapse is novel to the heterogeneous-brink scenario. We show that under mild conditions the world will traverse through all three phases along the noncooperative equilibrium path – and some of the most vulnerable countries will collapse. Moreover, even small differences in the catastrophe thresholds across countries can lead to country collapse along the equilibrium path.

The heterogeneous-brink scenario provides an illuminating counterpoint to our common-brink scenario, where once the world reaches the brink countries do whatever is necessary to avoid global collapse: now each country has its *own* brink generation, who faces the existential climate crisis *alone* and up against the other countries, 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 climate refugee costs that they would incur should the country collapse. And with heterogeneous collapse points, some countries may enjoy a reasonable standard of living while others suffer 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. This scenario suggests an additional perspective from which the scream of Greta may be interpreted: the generations that will pay the highest price for the emissions choices of their ancestors now include not only those generations that must live at the brink once the catastrophe phase is brought to a halt, but also those alive during the catastrophe phase who will live in a world in which vulnerable countries are collapsing, imposing costs both on the collapsing countries themselves and on the surviving countries that will face the refugee externality costs.

We then revisit the potential role for ICAs, but now in the heterogeneous-brink setting. We find that under the ICA a (weakly) larger subset of countries survives than in the noncooperative equilibrium, but the most vulnerable countries may still collapse under the ICA. We show that this is the case even in a world where countries can make unlimited international transfers, and that when such transfers are limited by a country's resource constraints the prospects for small, vulnerable countries become even more bleak. Furthermore, 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. But we also identify a surprising possibility: the social planner may deviate in the other direction from

the ICA, allowing a country to collapse that the ICA would save. This is possible because, while the social planner redistributes welfare toward future generations relative to the ICA, the best way to do this may be to make an earlier generation face the costs of a country's collapse if that frees future generations from having to live within the constraints imposed by that country's brink.

Finally, we extend our analysis to allow for intergenerational altruism, and argue that our main qualitative insights are robust to this extension. We also identify some new strategic effects that arise in the noncooperative equilibrium, including a novel "dynamic free-rider effect." In the common-brink scenario, if the world is expected to reach the brink tomorrow, so that all countries are expected to share the burden of avoiding catastrophe, an individual country today has less incentive to keep its emissions low.[4] This effect can exacerbate the overly high level of noncooperative emissions, and may introduce an additional role for an ICA. Furthermore, moving back in time, the anticipation of possible dynamic free riding behavior can induce countries to keep their emissions below the BAU level, precisely to prevent dynamic free riding from arising in equilibrium. Interestingly, in this case the possibility of catastrophe affects equilibrium emissions even though the brink is not reached in equilibrium. We also highlight that, in the heterogeneous-brink scenario, an asymmetric version of the dynamic free-rider effect can arise, in that more resilient countries free-ride on the future efforts of less resilient countries to save themselves from collapse.

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 during the warming phase, by internalizing the horizontal climate externalities and slowing the world's march to the brink, but it has no role to play in saving the world from collapse, beyond a possible coordination role, because once the brink is reached countries have sufficient incentives to avoid catastrophe even without an ICA. And the outcome implemented by the ICA falls short of the socially optimal outcome, which requires the world to move even more slowly toward the brink and possibly avoid the brink altogether. If the catastrophe threshold varies across countries, the role of an ICA is potentially more expansive, because it may save some of the most vulnerable countries from collapse, but its limitations relative to the global social planner are potentially more devastating, because it may not save enough countries from collapse.

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

There is an emerging literature that considers optimal environmental policy in the face of climate catastrophe. Prominent examples include Barrett (2013), Lemoine and Rudik (2017) and Besley and Dixit (2019).[5] Of these papers, only Barrett (2013) considers the role of ICAs, but his model

[4] A dynamic free-rider effect arises also in Battaglini et al. (2014, 2016). In their model, higher investment by a player today induces lower investment by other players in the future. In our paper we identify a novel type of dynamic free-rider effect, which is specifically linked to the possibility of catastrophe, a possibility that is absent in the above-mentioned papers.

[5] See also Brander and Taylor (1998), who consider catastrophes in a model that links population dynamics with

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 scenario.[6]

The papers of John and Pecchenino (1997) and Kotlikoff et al. (2021a,b) are also related. Like ours, these papers consider both international and intergenerational environmental externalities, but they do not consider the possibility of catastrophes. Instead, the central message of John and Pecchenino (1997) is that cooperation between countries at a point in time may be harmful to future generations.[7] Kotlikoff et al. (2021a,b) focus on characterizing Pareto-improving carbon taxes; they are not concerned with understanding the commitments that could be negotiated in an ICA to address these environmental externalities, which is our central focus here.

Our paper is also related to the literature on the dynamics of ICAs, which includes Dutta and Radner (2004), Harstad (2012, 2021, forthcoming) and Battaglini and Harstad (2016). These papers focus on aspects of ICAs that are very different from the ones we emphasize in this paper, and they do not consider issues of intergenerational externalities or the possibility of catastrophes. In particular, Harstad (2012) and Battaglini and Harstad (2016) focus on issues of free-riding and participation in ICAs when countries can make irreversible investments in green technology that cannot be contracted upon, and Harstad (forthcoming) takes this approach one step further by considering the implications of alternative bargaining procedures.[8] Finally, Harstad (2021) focuses on the desirability of issue linkage through a trade agreement whose commitments are made contingent on forest conservation measures.

The rest of the paper proceeds as follows. Section 2 lays out our basic modeling framework. Section 3 sets out our common-brink scenario and characterizes the noncooperative equilibrium, the ICA outcome and the social optimum. Section 4 contains the parallel analysis for our heterogeneous-brink scenario. We introduce intergenerational altruism in section 5. Section 6 offers concluding comments. An Appendix provides proofs not contained in the body of the paper.

2 The Basic Modeling Framework

We begin by describing the basic elements of our modeling framework. These elements form an "umbrella" model, from which our common-brink and heterogeneous-brink scenarios then emerge

renewable resource dynamics but does not feature intergenerational or international externalities.

[6] Barrett (2013) also argues that if the catastrophic threshold is uncertain, there is a unique Nash equilibrium that can lead to catastrophe, and an ICA can achieve a Pareto improvement over that equilibrium and reduce the probability of catastrophe. We discuss how our results can incorporate uncertainty in the Conclusion.

[7] 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.

[8] 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).

as special cases.

We consider a world of M countries. Each country has an identical population of identical citizens; we normalize the (initial) population of each country to one. 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. We allow each parent to be altruistic toward its only child, and the per-capita utility of country i's generation t is given by

$$\tilde{u}_{i,t} = u_{i,t} + \beta \tilde{u}_{i,t+1}$$

where $u_{i,t}$ is material per-capita utility and $\beta \geq 0$ captures the degree of intergenerational altruism. In this setting, utility can be equivalently represented with the dynastic utility function

$$\tilde{u}_{i,t} = \sum_{s=0}^{\infty} \beta^s u_{i,t+s}. \tag{1}$$

Material utility $u_{i,t}$ 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. 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 $B(c_{i,t})$ to denote these benefits, where $c_{i,t} \geq 0$ is the level of carbon emissions of country i's generation t.[9] We therefore treat $c_{i,t}$ itself as the choice variable of country i, with the understanding that lower emissions mean lower consumption. We have in mind that each government i then implements its chosen $c_{i,t}$ with an appropriate climate policy (e.g., carbon tax). We abstract from trading relations between countries, so that we can focus on their interactions mediated through the global carbon stock.[10]

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, which we denote by C_t. This stock evolves over time according to

$$C_t = (1 - \rho) C_{t-1} + c_t^W, \text{ with } C_{-1} = 0, \tag{2}$$

where c_t^W denotes aggregate world emissions at time t, with $\rho \in [0, 1)$ 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. The

[9] Our restriction that $c_{i,t} \geq 0$ reflects the possibility of zero (net) emissions through carbon capture and other mitigation efforts. We could be more general and impose $c_{i,t} \geq c^{\min}$ 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.

[10] We briefly consider the implications of trade in our working paper (Maggi and Staiger, 2022).

relationship in (2) implies that each generation feels the impact of its own emissions.[11]

We assume that increases in the global carbon stock degrade the environment and lead to losses in material welfare. In particular, we assume that, as long as no country has collapsed, these losses rise linearly with the global carbon stock C_t according to the parameter λ and are separable from the benefits of consumption, so country i's utility level is $B(c_{i,t}) - \lambda C_t$.

Each country i is characterized by a catastrophe threshold $\tilde{C}_i$: if the carbon stock C_t exceeds this threshold, country i collapses and its citizens become climate refugees, suffering a one-time per-capita material utility cost $L > 0$. We have in mind that moderate degrees of global warming lead to moderate costs for a country, but past a certain critical level, global warming becomes catastrophic for the country, triggering its collapse. The catastrophic levels $\tilde{C}_i$ are assumed known with certainty.[12] And to focus on the main points, we assume that, aside from these critical threshold levels, countries are symmetric in all respects.

The climate refugees that escape a collapsing country impose costs on the remaining countries. In particular, we assume that a collapsing country's refugees spread uniformly across the remaining countries, with each refugee imposing a one-time material utility cost r on the country to which it immigrates. Note that a special but important case of this framework is the one where all countries have the same catastrophe threshold (the "common brink" scenario, which we analyze in the next section); of course in this case the refugee externality cost r becomes irrelevant, because all countries survive or collapse together.[13]

Our modeling framework highlights two externalities that arise in the context of climate change. One externality is international: the emissions of one 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: the emissions of a country's generation t affect the material utility of all subsequent generations $t + 1$ and beyond in that country. Moreover, these "horizontal" (international) and "vertical" (intergenerational) externalities interact to produce additional "diagonal" externalities: the emissions of one country's generation t impact the utility of future generations in all other countries.

In the sections to follow we will characterize three regimes where these externalities are addressed to varying degrees. In the noncooperative regime, neither the horizontal nor the vertical or diagonal externalities are addressed, in the sense that each country's emissions choices impose costs on other countries and on future generations that those parties did not agree to incur. The second regime we consider is one where countries can sign ICAs and international lump-sum transfers are available. As we argue below, in this case only the horizontal externalities can be addressed, not the vertical

[11] 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, who also reports an estimate of the dissipation rate ρ on the order of 0.0035 per year).

[12] We discuss the more realistic possibility that the catastrophic levels $\tilde{C}_i$ are uncertain in the Conclusion.

[13] To simplify the exposition we postpone presentation of the full notation for the utility functions net of collapse costs and refugee externality costs until the next two sections, where we lay out in more detail the common-brink and heterogenous-brink scenarios respectively.

or diagonal externalities: as future generations cannot sit at the table while the ICA is negotiated, even in the presence of intergenerational altruism the ICA's emissions choices inevitably impose costs on future generations in all countries that those parties did not agree to incur. And third, we consider the outcome that a "social planner" would implement to maximize a social welfare function. We assume that the social welfare function puts positive weight on future generations *directly*, not just indirectly through the intergenerational altruism of the initial generation; and we assume that intergenerational lump-sum transfers are unavailable, leaving emissions and international lump-sum transfers as the planner's choice variables.[14]

More specifically, we follow Fahri and Werning (2007) in postulating the following social welfare function:[15]

$$W = \sum_{t=0}^{\infty} \hat{\beta}^t U_t \tag{3}$$

where U_t is the average material utility of generation t and $\hat{\beta}$ is the social discount factor. Note that if $\hat{\beta} = \beta$ then W coincides with the initial generation's average utility, so in this case the planner places no *direct* weight on future generations. On the other hand, if $\hat{\beta} > \beta$ then the planner places positive weight directly on future generations. Thus the wedge between the social and private discount factors $(\hat{\beta} - \beta)$ captures the weight placed by the planner directly on future generations. To illustrate this point more concretely, consider for example a two-period setting: in this case, if α denotes the Pareto weight placed by the planner directly on the second generation, we would have $\hat{\beta} = \beta + \alpha$. The two-period example illustrates how the social discount factor $\hat{\beta}$ can be interpreted as a function of two fundamental parameters, the degree of intergenerational altruism (β) and the Pareto weight on the future generation (α), and importantly, it also highlights that the wedge $\hat{\beta} - \beta$ need not decrease as β rises. Notice also that in principle $\hat{\beta}$ could be greater than one, but to avoid the complications that would arise in this case we assume for simplicity that $\hat{\beta} < 1$.[16]

It is worth pausing to clarify the nature of the discrepancy between the social planner's choice and that of the ICA. When generation t chooses emissions to maximize its utility under the ICA, it takes into account the impact of these emissions on future generations only to the extent that it cares about its offspring $(\beta > 0)$. A planner who puts positive weight directly on future generations $(\hat{\beta} > \beta)$ would modify the choices of generation t and redistribute utility from generation t to subsequent generations. Notice, though, that the planner's choice does not mark a Pareto improvement over the ICA, but rather a movement along the efficiency frontier, shifting surplus from generation t to later generations.

Finally, before turning to the analysis we can make a simple preliminary point: the horizontal

[14] An alternative benchmark would be the "unconstrained first best," meaning that all generations and countries can strike a Coasian bargain where also intergenerational transfers are available. As we will argue below (see note 17), the key difference between the unconstrained first best and our social optimum amounts to the weights that are placed on the utility of future generations when choosing emissions. We choose not to focus on the unconstrained first best because intergenerational transfers are arguably not feasible in reality, as discussed below. On the other hand, our social optimum can be interpreted as the policies that would be chosen by an ICA were there to be a shift in power toward younger generations, a benchmark that seems more relevant to the current debate on climate change.

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

[16] In the case where $\hat{\beta} \geq 1$, the infinite sum in (3) does not converge, so we would have to assume a finite horizon.

and vertical externalities from emissions choices *reinforce* each other. This can be seen clearly by focusing on the case in which there is no possibility of catastrophes ($\tilde{C}_i = \infty$ for all i) and no intergenerational altruism ($\beta = 0$); and by comparing the noncooperative emissions choices to those chosen by a planner that puts positive weight on future generations ($\hat{\beta} > 0$). In this case it is easy to show and intuitive that the noncooperative equilibrium emissions satisfy $B'(c_t) = \lambda$ for all t (assuming interior solutions), while the socially optimal emissions satisfy $B'(c_t) = \frac{M}{1-\hat{\beta}(1-\rho)}\lambda$ for all t. The noncooperative emissions level is clearly above the social optimum, with the difference summarized by the wedge $\frac{M}{1-\hat{\beta}(1-\rho)} > 1$. This wedge has two components: $M > 1$ reflects the international externality, and $\frac{1}{1-\hat{\beta}(1-\rho)} > 1$ reflects the intergenerational externality. 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.[17]

This special case is useful for highlighting the impacts of the externalities that arise in our model when catastrophes are not a concern. But as we establish below, the possibility of catastrophes introduces fundamental changes to the emissions profiles under our three regimes of interest (noncooperative, ICA and social optimum), and it changes the potential role of an ICA as well.

3 The Common-Brink Scenario

We first consider a world described by the basic modeling framework laid out in section 2, but where $\tilde{C}_i = \tilde{C}$ for all i, so that all countries share a common level of the carbon stock that would bring them to the brink of catastrophe. We will refer to this as the common-brink scenario.

We focus for now on a world without intergenerational altruism ($\beta = 0$); in section 5 we consider as well the possibility that $\beta > 0$ and show how our results extend in the presence of intergenerational altruism. Notice from (1) that with $\beta = 0$ there is no distinction between utility ($\tilde{u}_{i,t}$) and material utility ($u_{i,t}$), and for this reason in what follows we will simply refer to "utility" and use the notation $u_{i,t}$ to denote the utility of country i's generation t (and similarly the notation U_t to denote the world average per-capita utility of generation t).

In the common-brink scenario countries are fully symmetric, so we can omit the country subscript i. Here, moderate degrees of global warming lead to moderate costs, but past a certain critical level a rising carbon stock leads to a level of global warming that would trigger the collapse of civilization.[18] Since in this world climate refugees have nowhere to go, we suppose L is extremely

[17] In this setting without the possibility of catastrophe, it is easy to see the difference between our social optimum and the unconstrained first best (as defined in note 14). If all countries and generations could strike a Coasian bargain and intergenerational transfers were available (and assuming transferrable utility), the emissions level would maximize the sum of utilities of all parties to the bargain, and therefore would satisfy $B'(c) = \frac{M}{\rho}\lambda$. Intuitively, the Coasian solution takes into account the externality that emissions impose on all present and future citizens in equal measure. Note also that, given our assumption $\hat{\beta} < 1$, the unconstrained first best would go further than our social optimum in cutting emissions, and so in this sense the choices of our social planner are more conservative relative to the unconstrained first best outcome.

[18] In reality, collapse on a global scale is unlikely to be the result of crossing a single climate threshold. Rather, as Wallace-Wells (2019) argues forcefully, it is the cumulative effect of the collapse of numerous ecological subsystems,

high ($L = \infty$). The utility of a representative citizen in generation t is then given by

$$u_t = \begin{cases} B(c_t) - \lambda C_t & \text{if } C_t \leq \tilde{C} \\ -\infty & \text{if } C_t > \tilde{C}. \end{cases} \tag{4}$$

3.1 Noncooperative Equilibrium

We begin our analysis of the common-brink scenario by characterizing the noncooperative emissions choices. We focus on subgame perfect equilibria of the game. Given $\beta = 0$, countries are effectively myopic, so we can simply solve for the (unique symmetric) Nash equilibrium of the static game for each level of the carbon stock C_t, and then solve for the equilibrium path of C_t. It is direct to verify that the noncooperative equilibrium in general has two phases.

The first phase is a "warming phase," during which the level of emissions is constant across generations at the level $\bar{c}^N$ defined by $B'(\bar{c}^N) = \lambda$, where each country equates the marginal benefit of emissions to its own marginal loss λ, implying

$$\bar{c}^N = B'^{-1}(\lambda). \tag{5}$$

As is intuitive, (5) implies that $\bar{c}^N$ is decreasing in λ. We can interpret $\bar{c}^N$ as the "Business-As-Usual" (BAU) emissions level. During the warming phase associated with the BAU emissions, the global stock of carbon grows according to

$$C_t = (1 - \rho)C_{t-1} + M\bar{c}^N, \text{ with } C_{-1} = 0. \tag{6}$$

Denoting C_t^N the equilibrium path of the carbon stock defined by (6), the utility of successive generations in every country declines in the warming phase according to $u_t^N = B(\bar{c}^N) - \lambda C_t^N$.

If the catastrophe threshold $\tilde{C}$ were sufficiently high, the warming phase could go on forever without crossing the threshold, and according to (6) the global carbon stock would converge to the steady state level $\frac{M}{\rho}B'^{-1}(\lambda) \equiv C^N$. But the view of the majority of climate scientists is that a climate catastrophe will occur in finite time if the world stays on a BAU emissions path (see the recent reports of the IPCC). In the context of our model this view corresponds to the statement

each cascading over their own "tipping points," that poses the most serious climate-change induced existential threat to civilization (see Lenton et al., 2008, for an attempt to identify the location of tipping points for a variety of ecological subsystems). While therefore highly stylized along this dimension, our common-brink scenario might nevertheless be viewed as approximating a world in which there are many intermediate thresholds for the carbon stock that define a step function for the cost of global warming that over an initial range is composed of many small steps (approximated in our model by a smooth, linearly increasing cost), and where the brink as we have defined it is then associated with a carbon threshold level that, if crossed, would be the tipping point for a final ecological subsystem that would prove to be the "straw that broke the camel's back." The defining feature of the common-brink scenario is that all of the countries of the world arrive at the brink together. We postpone until the next section consideration of the possibility that each country could face its own brink level of the global carbon stock, and hence that some countries may be more vulnerable to the effects of climate change than others.

that $\tilde{C}$ lies below C^N. We therefore impose the condition

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

which ensures that under BAU emissions the catastrophe threshold $\tilde{C}$ will eventually be reached.

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" $t = \tilde{t}^N$ which, ignoring integer constraints, is defined by $C_{\tilde{t}^N}^N = \tilde{C}$. In effect, $\tilde{t}^N$ represents the point in time where, in a single generation, life under BAU emissions would go from livable to unlivable.

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

$$\frac{\rho \tilde{C}}{M} \equiv \hat{c}^N \tag{7}$$

where $\hat{c}^N$ is lower than $\bar{c}^N$ by Assumption 1. Thus 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 an equilibrium emissions level for generations $t \geq \tilde{t}^N$, we need only note that unilateral deviation to a level higher than $\hat{c}^N$ would trigger climate catastrophe and infinite loss, while deviation to a lower level would not be desirable either given that $\hat{c}^N < \bar{c}^N$.

While here we emphasize the symmetric Pareto-undominated equilibrium, it should be noted that there exist two other types of equilibria. First, there are equilibria where the world collapses, because if other countries choose very high emission levels, making catastrophe inevitable, it is optimal for an individual country to also choose a high emissions level. Clearly such equilibria are Pareto-dominated by the equilibrium described above. And second, there is a continuum of asymmetric equilibria where the world survives, with some countries cutting 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 equilibria are inefficient given our symmetric-country setup, and so we take the symmetric equilibrium as the natural focal point. As we will discuss below, in the event that countries focus on one of the asymmetric and inefficient Nash equilibria, or even worse, if countries fail to coordinate at all and focus on a catastrophic equilibrium where the brink is crossed, then a coordination role for an international climate agreement would arise.

Returning to our analysis of the symmetric Pareto-undominated equilibrium, we may conclude that the noncooperative emissions path for each country is given by

$$c_t^N = \begin{cases} \bar{c}^N & \text{for } t < \tilde{t}^N \\ \hat{c}^N & \text{for } t \geq \tilde{t}^N, \end{cases} \tag{8}$$

and using our findings above, the path of noncooperative utility is given by

$$u_t^N = \begin{cases} B(\bar{c}^N) - \lambda C_t^N & \text{for } t < \tilde{t}^N \\ B(\hat{c}^N) - \lambda \tilde{C} & \text{for } t \geq \tilde{t}^N. \end{cases} \tag{9}$$

According to (9) and recalling $\hat{c}^N < \bar{c}^N$, utility must fall precipitously when the world reaches the brink;[19] in order to prevent the planet from warming further, the brink generation and all future generations have to accept a discretely lower level of emissions, and therefore of consumption, relative to the pre-brink generations. Note that the drop in the emissions level, $\bar{c}^N - \hat{c}^N = B'^{-1}(\lambda) - \frac{\rho \tilde{C}}{M}$, is larger (i) the greater the number of countries M, (ii) the smaller the regeneration capacity of the atmosphere ρ and level of carbon stock above which climate catastrophe occurs $\tilde{C}$, and (iii) the lower the cost of moderate pre-catastrophe warming λ. Summarizing, we may now state:

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

Notice an interesting feature of the noncooperative equilibrium described in Proposition 1: no generation up until the brink generation does anything to address the climate externalities that each generation is imposing on those of its generation residing in other countries and on future generations everywhere; and yet the brink generation overcomes all of these externalities and saves the world. The reason for this 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 consequences 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.[20]

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,

[19] Here and throughout we use the adjective "precipitous" to describe a decline that would remain discrete even in the limit as time in our model went from discrete to continuous.

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

because the brink and all future generations must accept a precipitous drop in consumption and utility in order to prevent the climate from worsening further and resulting in global annihilation.

3.2 International Climate Agreements

We are now ready to consider what an ICA can achieve. Two important challenges that an ICA must meet relate to participation and enforcement. It is well known (see, for example, Barrett, 1994, Harstad, 2012, Nordhaus, 2015, Battaglini and Harstad, 2016 and Harstad, forthcoming) that countries have strong incentives to free ride on ICAs, so without some means of requiring participation the number of countries participating in an ICA is likely to be 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 (but in principle, not insurmountable) limitations, and assume that the ICA attains full participation of all M countries, and that any rules negotiated under the ICA are perfectly enforceable. Given these ideal conditions, our model highlights an additional limitation that has not been emphasized in the formal literature, and that arises because 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 imposed by emissions choices, it cannot address the vertical and diagonal externalities exerted by those choices. Our goal is to characterize what an ICA can achieve in the presence of this particular limitation.[21]

Recalling that we are assuming $\beta = 0$ for now so as to abstract from intergenerational altruism, for each generation t we characterize the ICA emissions levels as those that maximize the welfare of generation t in the representative country. Given our symmetric-country setting, this is the natural ICA design to focus on, as it would emerge if countries bargain efficiently and have symmetric bargaining power.[22]

Using (4), 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

$$\bar{c}^{ICA} = B'^{-1}(M\lambda). \tag{10}$$

[21] One could imagine that an ICA might involve an implicit contract of some kind between current and future generations 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. We view this challenge as essentially insurmountable.

[22] In light of this symmetry, there is no natural role for international transfers under the ICA in our common brink scenario, but asymmetries could be introduced to generate such a role, a point that is easily seen even in a setting without the possibility of catastrophes. For example, suppose some countries are poorer than others, and the poorer countries have a higher marginal benefit of emissions (B_i') and/or attach a lower weight to the consequences of climate change (λ_i). Then an ICA may entail transfers from richer to poorer countries, as the former are willing to compensate the latter in exchange for deeper emissions cuts.

According to (10), 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 (5) and (10) imply $\bar{c}^{ICA} < \bar{c}^{N}$, because under noncooperative choices each country internalizes the costs imposed on the current generation only in its own country. Finally, with emissions levels given by $\bar{c}^{ICA}$, as long as the brink of catastrophe is not hit the carbon stock under the ICA evolves according to

$$C_t = (1-\rho)C_{t-1} + M\bar{c}^{ICA} \quad \text{with } C_{-1} = 0. \tag{11}$$

Letting C_t^{ICA} denote the path of the carbon stock defined by (11), it follows that C_t^{ICA} would eventually converge to the steady state level $\frac{M}{\rho}B'^{-1}(M\lambda) \equiv C^{ICA}$.

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

$$\tilde{C} \geq C^{ICA}. \tag{12}$$

Note that $C^{ICA} < C^{N}$, so both Assumption 1 and (12) will be satisfied if $\tilde{C} \in [C^{ICA}, C^{N})$. Intuitively, if the catastrophe threshold $\tilde{C}$ is sufficiently close to the steady state level of the carbon stock under BAU emissions, C^{N}, then only a relatively small reduction in emissions from the BAU level would be required to keep the world from reaching the brink, and by addressing the horizontal externalities the ICA will indeed deliver the required reductions. On the other hand, if

$$\tilde{C} < C^{ICA}, \tag{13}$$

then under the ICA the brink of catastrophe will be reached in finite time, with the brink generation $\tilde{t}^{ICA}$ defined by $C_{\tilde{t}^{ICA}}^{ICA} = \tilde{C}$. Notice from (11) and (6) that $\tilde{t}^{ICA} > \tilde{t}^{N}$ is ensured by $\bar{c}^{ICA} < \bar{c}^{N}$, so when (13) is satisfied the ICA postpones the arrival of the brink, even though it does not avoid it.

If (13) is satisfied, what happens under the ICA when the world reaches the brink? Clearly 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, if they can coordinate on a Pareto-undominated equilibrium. In that case, therefore, $\tilde{t}^{ICA}$ marks the end of the useful life of the ICA. On the other hand, if one admits the possibility that countries might fail to coordinate on a Pareto-undominated equilibrium, and as a result the world might collapse in the noncooperative equilibrium, then the ICA would have an important coordination role to play, namely, ensuring that countries stay away from catastrophic equilibria. Hence, the more general message of our analysis is that the role of an ICA changes dramatically once the world reaches the brink of catastrophe: at that point, the role of an ICA will at most be to address coordination failures among its member countries, and it may even become redundant and have no role to play at all.[23]

[23] There is another possible coordination role that the ICA might play when the world is at the brink of catastrophe, albeit a less dramatic one than preventing a catastrophic coordination failure. Recall that, in the noncooperative game,

Finally, letting c_t^{ICA} denote the path of emissions under the ICA, we can describe emissions succinctly under both (12) and (13) with

$$c_t^{ICA} = \begin{cases} \bar{c}^{ICA} & \text{for } t < \tilde{t}^{ICA} \\ \hat{c}^N & \text{for } t \geq \tilde{t}^{ICA} \end{cases} \tag{14}$$

where $\tilde{t}^{ICA}$ is finite if and only if (13) is satisfied. Utility under the ICA is then given by

$$u_t^{ICA} = \begin{cases} B(\bar{c}^{ICA}) - \lambda C_t^{ICA} & \text{for } t < \tilde{t}^{ICA} \\ B(\hat{c}^N) - \lambda \tilde{C} & \text{for } t \geq \tilde{t}^{ICA}. \end{cases} \tag{15}$$

Note that under the ICA, if (13) is satisfied so that $\tilde{t}^{ICA}$ is finite, then utility must fall precipitously for the brink generation, since (13) implies $\bar{c}^{ICA} > \hat{c}^N$. Hence in this case, 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, since $\bar{c}^{ICA} < \bar{c}^N$, it is also clear that the brink generation suffers a smaller decline in welfare under the ICA than in the noncooperative equilibrium. We summarize with:

Proposition 2 *In the common-brink scenario the path of emissions under the ICA falls into one of two cases. If $\tilde{C}$ is above a threshold level, then the brink of catastrophe is never reached, and the ICA emissions levels are below the noncooperative levels. Otherwise, if $\tilde{C}$ is below this threshold, then the brink of catastrophe will be reached, but at a later date than in the absence of the ICA. In this second case, the ICA emissions levels are below the noncooperative levels until the brink is reached, at which point emissions fall to the replacement rate dictated by the natural rate of atmospheric regeneration and remain at that level for all generations thereafter. As a consequence, the brink generation and all future generations experience a precipitous drop in utility relative to the pre-brink generations, but this drop is smaller than under the noncooperative equilibrium. If under the ICA the brink is reached, the ICA has a role to play in helping the world avoid climate catastrophe only insofar as it may help countries avoid a coordination failure.*

It is interesting to reflect more broadly on the evolving role of an ICA according to our common-brink scenario. Up until the world reaches the brink, the ICA plays the standard role of internalizing the international climate externalities and thereby solves a Prisoner's Dilemma. This cooperation role clearly requires the ICA to enforce commitments over participation and emissions levels. But when the world reaches the brink, the role of an ICA is transformed: at most it can solve a coordination problem for its member countries, and for this task the need to enforce commitments

we have focussed on the symmetric equilibrium in undominated strategies, which implies that for $t \geq \tilde{t}^N$ countries not only avoid catastrophe, but also adopt the efficient assignment of emissions. If instead countries coordinated on an asymmetric equilibrium (in undominated strategies) for $t \geq \tilde{t}^N$, then the ICA would have an efficiency-enhancing role to play for $t \geq \tilde{t}^{ICA}$, allowing countries to exchange emissions cuts for transfers. In particular, countries would agree to the symmetric and efficient Nash emissions levels $\hat{c}^N$ and use international lump-sum transfers to distribute according to bargaining powers the surplus gains that result from eliminating the inefficiency. In this case the ICA would have a continuing role in enhancing the efficiency properties of the emissions cuts required for survival.

disappears. In this light it is relevant to observe that while the 1997 Kyoto Protocol focused on negotiating and securing enforceable commitments from its members regarding emissions reductions, the 2016 Paris Agreement has moved to an alternative approach to emissions reductions centered on "Nationally Determined Contributions" announced voluntarily by each country; a possible interpretation of this evolution through the lens of our common-brink scenario is that the world is reaching the brink, and the role of an ICA is evolving from cooperation to coordination.[24] But it is an open question whether countries would be able to achieve the same outcome in the absence of the Paris agreement; if they could, our common brink scenario would imply that a role for the ICA ceases to exist completely once the world reaches the brink.

It might seem surprising that, once the world reaches the brink, there is no longer need for an agreement that addresses international externalities by enforcing commitments, since the possibility of catastrophe does imply extreme international externalities. But at the brink, these extreme international externalities are coupled with extreme *internalized* costs of increasing emissions, and this makes any enforcement role of ICAs unnecessary.

3.3 The Social Optimum

We next consider the path of emissions that a social planner would choose in order to maximize the intergenerational social welfare function (3). Given our symmetric setting and recalling (4), the planner's objective can be written as $W = \sum_{t=0}^{\infty} \hat{\beta}^t u_t$. We focus on the case in which the planner places positive weight directly on future generations, and since we are abstracting for now from intergenerational altruism ($\beta = 0$), we allow for any $\hat{\beta} \in (0, 1)$.[25]

Clearly, the planner will not allow the world to end in catastrophe and hence will not allow C_t to exceed $\tilde{C}$. Consequently, we can equate u_t with $B(c_t) - \lambda C_t$ and introduce the constraint $C_t \leq \tilde{C}$, which we will henceforth refer to as the "brink constraint." We therefore write the planner's problem as

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

For simplicity we restrict attention to the case where the emissions feasibility constraint $c_t \geq 0$ is not binding.

Here we summarize the main steps of the analysis, relegating the formal proof to the Appendix. There are two cases to consider, depending on whether or not the brink constraint $C_t \leq \tilde{C}$ binds for any t. We show that, if $\tilde{C}$ is higher than a threshold level C^S, the brink constraint is not binding

[24] See Barstad (forthcoming) for an alternative interpretation of the evolution from Kyoto to Paris that focuses on changes in the numbers of major carbon polluting countries over this period.

[25] 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 $t = 0$.

and the brink of collapse $\tilde{C}$ is never reached; we refer to this as Case 1. If $\tilde{C}$ is lower than C^S, on the other hand, the brink constraint is binding and the brink of collapse $\tilde{C}$ is reached at some point in time; we refer to this as Case 2.

In Case 1, where $\tilde{C} \geq C^S$, the optimal level of emissions $\bar{c}^S$ is constant for all countries and all generations, and defined by $B'(\bar{c}^S) = \frac{M\lambda}{1-\hat{\beta}(1-\rho)}$. This solution has a simple interpretation: a country's marginal benefit from its own emissions should equal the marginal environmental cost of emissions, taking into account the costs imposed on all M countries and on all future generations (discounted by the social discount factor $\hat{\beta}$ and accounting for the natural rate of atmospheric regeneration ρ). It is easy to show and intuitive that $\bar{c}^S < \bar{c}^{ICA} < \bar{c}^N$: the planner goes further than the ICA in cutting emissions relative to the BAU level. In this case the carbon stock increases in a concave way and converges to the steady state level $C^S = \frac{M\bar{c}^S}{\rho}$.

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

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

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

For Case 2 it is easy to see that the level of welfare falls through time for $t < \tilde{t}^S$ – due to the warming climate as in Case 1, but in contrast to Case 1 also due to the decline in consumption implied by the falling emissions – until the brink generation $\tilde{t}^S$ is reached, at which point 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, it can be shown that 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 that, contrary to the noncooperative and ICA outcomes, when the brink of catastrophe is reached under the socially optimal level of emissions, the brink generation does not suffer a precipitous drop in welfare relative to the previous generation.

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

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

3.4 Comparison of ICA and Socially Optimal Outcomes

We now compare the socially optimal outcome characterized in the previous section with the ICA outcome examined in Section 3.2. To this end, we begin by noting that $C^S < C^{ICA} < C^N$ and recalling that we assume $\tilde{C} < C^N$ so that the brink is reached in finite time under the noncooperative equilibrium. We can thus organize the comparison between the ICA and socially optimal outcomes into three ranges of $\tilde{C}$: high ($\tilde{C} \in [C^{ICA}, C^N)$), intermediate ($\tilde{C} \in [C^S, C^{ICA})$) and low ($\tilde{C} < C^S$).

Consider first the possibility that $\tilde{C}$ falls in the high range $\tilde{C} \in [C^{ICA}, C^N)$. In this case the world will be kept below the brink of catastrophe by both the ICA and the planner through constant emissions levels ($\bar{c}^S$ and $\bar{c}^{ICA}$ respectively) that are below the BAU level $\bar{c}^N$ and that keep the carbon stock below $\tilde{C}$. However, the socially optimal emissions $\bar{c}^S$ take into account both international *and* intergenerational externalities, while the ICA emissions $\bar{c}^{ICA}$ take into account only the international externalities; and as a result we have $\bar{c}^S < \bar{c}^{ICA}$, with $\bar{c}^S$ dropping further below $\bar{c}^{ICA}$ as $\hat{\beta}$ increased and as ρ decreases, and the steady state carbon stock delivered under the ICA is larger than the socially optimal level. The three panels of Figure 1 illustrate the time paths of the emissions and of the carbon stock, and the utility of a representative country under the ICA, the social planner and the noncooperative equilibrium for $\tilde{C}$ in this range. The qualitative features of the ICA and socially optimal outcomes are similar, with the difference between the two being that the planner shifts welfare from early generations to later generations relative to the ICA by requiring lower emissions for all generations and thereby reducing the extent to which utility falls through time due to a rising global carbon stock and worsening climate.[26]

Consider next the possibility that $\tilde{C}$ falls in the intermediate range $\tilde{C} \in [C^S, C^{ICA})$. In this case the world would still be kept away from the brink of catastrophe by the planner, but under the ICA the world will be brought to the brink. This is because the socially optimal emissions level $\bar{c}^S$ is still low enough to keep the carbon stock below $\tilde{C}$, but the emissions level $\bar{c}^{ICA}$ implemented by the ICA during the warming phase is no longer low enough to accomplish this. Hence, in this case

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

the inability of the ICA to take into account directly the interests of future generations leads to a qualitative difference across the ICA and socially optimal outcomes. This is reflected in the three panels of Figure 2. As in Figure 1, here the utility of earlier generations is higher and the utility of later generations is lower under the ICA than in the social optimum, but now utility under the ICA falls precipitously for the brink generation, while utility under the social optimum evolves smoothly through time. And while in this case the planner would not let utility for any generation fall to the level experienced in the noncooperative equilibrium by the brink generation, under the ICA the brink generation and all subsequent generations will experience exactly that level of utility.

Finally, consider the possibility that $\tilde{C}$ falls in the low range $\tilde{C} < C^S$. In this case the world will be brought to the brink of climate catastrophe by both the ICA and the planner, but as noted we have $\tilde{t}^N < \tilde{t}^{ICA} < \tilde{t}^S$: the ICA slows down the march to the brink relative to the noncooperative outcome, but this march is still too fast relative to the social optimum. In this case as well there are qualitative differences across the ICA and socially optimal outcomes that arise as a result of the inability of the ICA to take into account directly the interests of future generations. This is reflected in the three panels of Figure 3. Here the ICA emissions remain constant at the level $\bar{c}^{ICA}$ during the warming phase 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 social optimum the emissions $\hat{c}_t^S$ during the warming phase decline smoothly over time, and they reach the level $\hat{c}^N$ at the brink without a precipitous drop for the brink generation in either emissions or utility.

Summarizing, we may now state:

Proposition 4 *In the common-brink scenario, the ICA addresses the horizontal (international) externalities from carbon emissions that create inefficiencies in the noncooperative outcome, but it does not take into account the vertical and diagonal externalities that are associated with the intergenerational aspects of the climate problem. For this reason, the ICA slows down the growth of the carbon stock relative to the noncooperative outcome but not enough relative to the social optimum. More specifically: (i) If $\tilde{C}$ is above a threshold level the ICA prevents the world from reaching the brink of catastrophe, but the carbon stock at every point in time is still too large relative to the social optimum. (ii) If $\tilde{C}$ lies in an intermediate range the world reaches the brink of catastrophe under the ICA but not under the social optimum. (iii) If $\tilde{C}$ is below a threshold level the brink is reached both under the ICA and the social optimum, but under the social optimum it is reached later and the brink generation does not suffer a precipitous drop in welfare.*

It is also natural to wonder how the ICA signed by a given generation t affects future generations. This is not obvious *a priori*, because for each generation the ICA is a contract that excludes future generations. Nevertheless, it is direct to verify that in our model an ICA must benefit 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 slow down the

march to the brink of 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 is most relevant for interpreting the current debate on the challenges posed by climate change, and in particular the 'scream of Greta' referenced in the title of our paper. In an address to world leaders at the United Nation's Climate Action Summit in New York City on September 23, 2019, Greta Thunberg stated:

> "You have stolen my dreams and my childhood with your empty words. ... Entire ecosystems are collapsing. We are in the beginning of a mass extinction ... How dare you! For more than 30 years, the science has been crystal clear... The popular idea of cutting our emissions in half in 10 years only gives us a 50% chance of staying below 1.5 degrees [Celsius], and the risk of setting off irreversible chain reactions beyond human control. Fifty percent may be acceptable to you ... [but it] is simply not acceptable to us — we who have to live with the consequences. [...] How dare you pretend that this can be solved with just 'business as usual' and some technical solutions? [...] You are failing us. But the young people are starting to understand your betrayal. The eyes of all future generations are upon you. And if you choose to fail us, I say: We will never forgive you. We will not let you get away with this. Right here, right now is where we draw the line. The world is waking up. And change is coming, whether you like it or not."

At a broad level, the plight of Greta and her generation can be interpreted through the lens of our model as arising from the inability of ICAs to take into account intergenerational externalities. But more specifically, it is arguably the intermediate case depicted by Figure 2 that best captures the essence of this plight. In the case of Figure 2, the implications of the inability of ICAs to take into account directly the interests of future generations are especially dire: while a global social planner would keep the world from ever arriving at the brink of climate catastrophe, an ICA will 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 relative to the pre-brink generations. In terms of the fundamental parameters of our model, the case of Figure 2 obtains when the marginal cost of moderate degrees of global warming (λ) is small enough and the social discount factor ($\hat{\beta}$) is sufficiently close to one. Intuitively, if λ is small enough the brink will be reached under the ICA; lowering λ also makes it more likely that the brink will be reached under the social optimum, but if $\hat{\beta}$ rises sufficiently this will more than offset the impact of lowering λ and so the brink will not be reached under the social optimum. We record this observation in:

Corollary 1 *If the cost associated with moderate degrees of global warming, λ, is not too large and the social discount factor ($\hat{\beta}$) is close enough to one, the world will reach the brink of catastrophe under the ICA but not under the social optimum.*

In terms of Figure 2, we might think of the impact of climate agreements to date as putting the world somewhere between the noncooperative emissions path (if these agreements were completely

ineffective) and the ICA path (if the agreements were maximally effective), and we might think of Greta and her generation as corresponding to the brink generation (marked in Figure 2 by $\tilde{t}^N$ in the former case and $\tilde{t}^{ICA}$ in the latter case). The "consequences" to which Greta refers in the quote above are then reflected in Figure 2 by the implication of the threat of a climate catastrophe experienced in her lifetime, a threat caused by the emissions of previous generations that the brink generation must now confront and that might have been avoided if previous generations had adopted socially optimal emissions policies. According to our common-brink scenario, the implication of this threat is that the world will indeed find an 11^{th} hour solution which prevents the threat of climate catastrophe from materializing, much as Greta predicts. But as Figure 2 depicts, avoiding climate catastrophe at the 11^{th} hour comes at the cost of a precipitous drop in utility for the brink generation relative to their parents, and the same low level of utility for all generations thereafter.

The perspective provided by the common-brink scenario that we describe here remains relevant also for the heterogeneous-brink scenario that we consider next. But as we will show, the heterogeneous-brink scenario provides an additional perspective from which the scream of Greta may be interpreted, by extending this interpretation to a setting in which a sequence of countries may suffer climate collapse along the equilibrium path.

4 The Heterogeneous-Brink Scenario

We now focus on a scenario where countries reach the brink of catastrophe at different levels of the global carbon stock. In terms of our umbrella framework of Section 2, we now assume that each country i has its own catastrophe threshold $\tilde{C}_i$.

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.[27] If some countries face existential threats from climate change before others, new questions arise. Under what conditions will some (or even all) of the countries collapse on the noncooperative equilibrium path? Can there be domino effects, where the collapse of one country hastens the collapse of the next? If some or all countries would collapse in the noncooperative equilibrium, can ICAs help to avoid collapse? And what is the outcome that a global social planner would implement in this case?

We order countries according to increasing $\tilde{C}_i$, so that country 1 is the country with the lowest $\tilde{C}_i$ and therefore the country "least resilient" (or "most vulnerable") to climate change, while country M has the highest $\tilde{C}_i$ and is hence the most resilient country. We also assume that this ordering is strict, i.e. no two countries have the same value of $\tilde{C}_i$.

Recall from section 2 that there are two costs associated with a country's collapse. The first is a one-time per-capita utility cost L suffered by the citizens of the collapsing country. We assume

[27] And it is not just through rising sea levels that global-warming induced climate changes can have differential effects on countries. For example, Lenton et al. (2008) describe a tipping point exhibited by global-warming-induced changes in the amplitude of the El Nino - Southern Oscillation that would trigger drought conditions in Southeast Asia. See also Jones and King (2021), who develop a methodology for predicting a shortlist of countries that are most likely to be left standing when other countries have succumbed to climate catastrophe.

collapse occurs at the end of a period, after consumption has occurred. We allow L to be finite as long as there are other surviving countries to which the citizens of the collapsing country can immigrate, while we assume that this cost is infinite (or prohibitively high) for the citizens of the *last* surviving country who, as in the common-brink scenario, facing collapse would have nowhere to go.[28] The second cost of a country's collapse is borne by the remaining countries: the collapsing country's citizens become climate refugees (spreading equally across the remaining countries), and each climate refugee imposes a one-time utility cost r on the country to which it immigrates.

Letting $H_t \in \{1, 2, ..., M\}$ index the most vulnerable country that has survived to time t, the number of surviving countries at t is $M - H_t + 1$, and recalling that the total world population has measure M, the population of a surviving country at t is $\frac{M}{M-H_t+1}$. If country H_t collapses, since the total population of the remaining countries is $M - \frac{M}{M-H_t+1} = \frac{M(M-H_t)}{M-H_t+1}$, it follows that the one-time per-capita utility cost incurred by citizens of the remaining countries as a result of country H_t's collapse is

$$R(H_t) \equiv \frac{r}{M - H_t}.$$

As with L, we assume that the refugee cost $R(H_t)$ is incurred at the end of the period of country H_t's collapse. Notice also that the refugee externality $R(H_t)$ is increasing in H_t. This is because in our model countries that collapse later (higher H_t) release a greater number of climate refugees ($\frac{M}{M-H_t+1}$) on a smaller rest-of-world population ($M - \frac{M}{M-H_t+1}$).

Recall that in the common-brink scenario the population of each country remained constant over time and we normalized this population to one, so country-level and per-capita-level variables were one and the same. But with climate refugees altering the population of surviving countries when more vulnerable countries collapse, country-level and per-capita-level variables will diverge. So we now specify $u_{i,t}$ at the per-capita level, as the utility of a person living in country i in period t. We continue to interpret $c_{i,t}$ as the per-capita emissions of country i in period t, and $B(c_{i,t})$ as the associated per-capita benefit.

To preserve tractability, we assume that both L and $R(H_t)$ enter utility in an additive way, so the utility of a citizen living in country i at time t is given by:

$$u_{i,t} = B(c_{i,t}) - \lambda C_t - L \cdot E_{i,t} - R(H_t) \cdot I_{i,t}, \tag{16}$$

where $E_{i,t}$ is a dummy that equals one if country i collapses at time t, so that its population has to emigrate, and $I_{i,t}$ is a dummy that equals one if some country other than country i collapses at time t, so that country i receives immigrants from the collapsing country.[29] And of course, the utility function in (16) is defined only for countries that survive to time t.

[28] With a slight abuse of notation we use the same symbol L to denote the internal cost of collapse for any country, with the understanding that the value of L is infinite for the last surviving country (M).

[29] In writing (16) we are implicitly focussing on the case where at most one country collapses in a given period t, but it is straightforward to accommodate the possibility of multiple-country collapses in a given period.

Finally, the carbon stock in our heterogeneous-brink scenario evolves according to

$$C_t = (1-\rho)C_{t-1} + \sum_{i=H_t}^{M} \frac{M}{M - H_t + 1} \cdot c_{i,t} \text{ with } C_{-1} = 0. \tag{17}$$

Notice that according to (17) we now have $\frac{\partial C_t}{\partial c_{i,t}} = \frac{M}{M-H_t+1}$: the impact of a surviving country's per-capita emissions on the carbon stock C_t grows as the number of surviving countries shrinks, because country population grows due to the absorption of climate refugees.

4.1 Noncooperative Equilibrium

We first characterize the noncooperative emissions choices. As in the previous section, given $\beta = 0$ we effectively have a one-shot game for each generation, so we can solve the game by first characterizing the equilibrium emissions for each level of the carbon stock (for the countries that are alive at that level of the carbon stock), and then backing out the equilibrium path for the carbon stock and hence for the set of countries that survive to each point in time.

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 phase is familiar from the analysis of the common-brink scenario; the possibility of a middle phase in which some countries collapse, as well as the possibility of a third phase in which a single country lives at the brink forever, are novel to the heterogeneous-brink setting. Notice, too, that with $\frac{\partial C_t}{\partial c_{i,t}} = \frac{M}{M-H_t+1}$, the BAU per-capita emissions level of a surviving country now depends on H_t, so we write it as $\bar{c}^N(H_t)$. Clearly this is given by

$$\bar{c}^N(H_t) = B'^{-1}\left(\frac{M\lambda}{M - H_t + 1}\right). \tag{18}$$

From (18), the BAU per-capita emissions are the same across all surviving countries (so we can omit the country subscript i) but fall as the number of surviving countries shrinks (H_t rises) and the population of each surviving country rises. This reflects the fact that with each country collapse there are fewer remaining countries; and the countries that do remain internalize a greater proportion of the global cost of their BAU emissions choices. Since the world population remains constant at M, the world BAU emissions level $M\bar{c}^N(H_t)$ also falls with each country collapse.

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

increases, a sufficient condition for this assumption to hold is

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

a restriction that we will maintain throughout.[30]

To develop some intuition for how the noncooperative equilibrium 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 is simple and provides a sharp counterpoint to the equilibrium path for the common-brink scenario.

After an initial warming phase during which there are no catastrophes and each country selects the BAU emissions level $\bar{c}^N(1)$, the world enters a catastrophe phase when country 1 arrives at the brink of collapse. This occurs in finite time if $M\bar{c}^N(1) > \rho\tilde{C}_1$, which is implied by Assumption 2. Furthermore, Assumption 2 implies that country 1 is not able to offset the BAU emissions of the rest of the world – and with $r = 0$, the remaining countries have no reason to help country 1 survive – thus country 1 will collapse. And by a similar logic, all countries except for the most resilient one ($i = M$) will collapse on the equilibrium path. The most resilient country is guaranteed to survive on the equilibrium path, because by setting $c = \frac{\rho\tilde{C}_M}{M} \geq 0$ it can freeze the global carbon stock at the brink level $\tilde{C}_M$.[31]

Hence, with heterogeneous brinks and $r = 0$, all countries except the most resilient one will collapse, provided only that no individual country is able to fully offset the other countries' BAU emissions. This scenario provides an illuminating contrast to the common-brink setting. Relative to that setting, where once the world reaches the brink of catastrophe the "we are all in this together" forces are strong and countries do whatever is necessary to avoid global collapse, now each country has its *own* brink generation, who faces the existential climate crisis *alone* and up against the other countries, who do not internalize the impact of their emissions on the fate of the brink country. As we next demonstrate, allowing for climate refugee externalities will bring back an element of these forces, albeit only partially.

To proceed, we now suppose that $r > 0$. To simplify the exposition, as with our analysis of the common-brink scenario, we ignore integer constraints.

If in period t no country is on the brink of catastrophe ($C_{t-1} \neq \tilde{C}_{H_t}$), clearly each country chooses its BAU emissions $\bar{c}^N(H_t)$ defined by (18). Things are more interesting if in period t country H_t is on the brink of catastrophe ($C_{t-1} = \tilde{C}_{H_t}$). Countries $j \geq H_t$ have survived up to this point, each with a population of $\frac{M}{M-H_t+1}$, and it follows that country H_t will survive if and only if $\frac{M}{M-H_t+1}\sum_{j=H_t}^{M} c_j \leq \rho\tilde{C}_{H_t}$. We now argue that there are two possible types of equilibrium.

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

[31] Recall that we have assumed that L is infinite if country M is the last surviving country, so this country will do whatever is necessary to avoid its own collapse.

The first possibility is that all countries choose the BAU emissions level $\bar{c}^N(H_t)$ and country H_t does not survive. This is always an equilibrium, because given Assumption 2 no country (including country H_t itself) can unilaterally save country H_t if the other countries choose $\bar{c}^N(H_t)$. Also note that, if the other countries choose $\bar{c}^N(H_t)$, country H_t's best response is to also choose $\bar{c}^N(H_t)$, because it gets to enjoy the benefit of its current-period emissions before it collapses, and therefore, given that collapse is inevitable, it can do no better than to choose its BAU emissions.

The second possibility is that country H_t survives, and the countries' emissions levels satisfy $\frac{M}{M-H_t+1}\sum_{j=H_t}^{M} c_j = \rho\tilde{C}_{H_t}$. Intuitively, this type of equilibrium can exist only if the refugee externality is large enough, so that countries $j > H_t$ have an incentive to "top off" the mitigation efforts of country H_t, ensure country H_t's survival, and avoid a climate refugee crisis. We will now examine the conditions under which such an equilibrium exists.

First note that equilibria with survival of country H_t exist if and only if there is an equilibrium where aggregate emissions equal $\rho\tilde{C}_{H_t}$ and countries $j > H_t$ choose symmetric emissions.[32] Thus, in such an equilibrium, country H_t emits c^b (where b is for "brink") and each country $j > H_t$ emits c^{nb} (where nb is for "not brink") with

$$\frac{M}{M-H_t+1}\cdot c^b + (M-H_t)\cdot\frac{M}{M-H_t+1}\cdot c^{nb} = \rho\tilde{C}_{H_t} \tag{19}$$

A country $j > H_t$ has no incentive to defect from such equilibrium if and only if

$$G^{nb}(c^{nb},H_t) \equiv \left[B(\bar{c}^N(H_t)) - \lambda\left(\tilde{C}_{H_t} + \frac{M}{M-H_t+1}\cdot\left(\bar{c}^N(H_t) - c^{nb}\right)\right)\right] - \left[B(c^{nb}) - \lambda\tilde{C}_{H_t}\right] \leq R(H_t). \tag{20}$$

The left-hand side of (20) is the gross per-capita gain to a country $j > H_t$ from defecting from the equilibrium described above: the term in the first square brackets is the per-capita payoff to country $j > H_t$ from deviating to the BAU emissions level $\bar{c}^N(H_t)$ and causing country H_t to collapse; the term in the second square brackets is the per-capita payoff to country $j > H_t$ under the proposed equilibrium. The right-hand side is the per-capita cost that country $j > H_t$ incurs as a result of H_t's collapse.

Turning to the no-defect condition for country H_t, this condition can be written as:

$$G^{b}(c^{b},H_t) \equiv \left[B(\bar{c}^N(H_t)) - \lambda\left(\tilde{C}_{H_t} + \frac{M}{M-H_t+1}\cdot\left(\bar{c}^N(H_t) - c^{b}\right)\right)\right] - \left[B\left(c^{b}\right) - \lambda\tilde{C}_{H_t}\right] \leq L. \tag{21}$$

The left-hand side of (21) is country H_t's gross per-capita gain from defecting from the proposed equilibrium: the term in the first square brackets is the per-capita welfare that country H_t would enjoy were it to deviate to the BAU emissions level $\bar{c}^N(H_t)$, and the term in the second square brackets is its per-capita welfare under the proposed equilibrium. The term on the right-hand side

[32] To see this, note that assigning asymmetric emissions to countries $j > H_t$ would make it harder to sustain such an equilibrium, because it would increase the temptation to defect for the countries that have a bigger burden. And similarly, if aggregate emissions were less than $\rho\tilde{C}_{H_t}$ some countries would have a bigger mitigation burden and hence a stronger temptation to defect.

is the per-capita cost that the collapse of country H_t, precipitated by its own defection from the self-help equilibrium, would impose on its citizens at the end of the period.

An equilibrium with survival of country H_t exists if and only if there is a pair (c^b, c^{nb}) that satisfies (19), (20) and (21). Notice that, if such an equilibrium exists, it Pareto-dominates the equilibrium where country H_t collapses.[33] Thus, maintaining our emphasis on Pareto-undominated equilibria as we did in the common-brink setting (we will again comment in a later section on the role of ICAs in addressing possible coordination failures), we can conclude that if country H_t is at the brink of catastrophe it survives if and only if the conditions stated above are satisfied.

Finally note that, if $C_{t-1} = \tilde{C}_M$, so that the only surviving country $j = M$ is at the brink of catastrophe in period t, this country will restrain its emissions just enough to avoid collapse. Since at this stage the entire world population M is located in this country, its per capita emissions are given by $\frac{\rho \tilde{C}_M}{M}$. If the world reaches this stage, the carbon stock stops growing and will stay at the level $\tilde{C}_M$ forever thereafter.

Our next observation is that a country at the brink is more likely to survive if more countries have collapsed in the past, that is when H_t is higher. Formally, we show in the Appendix that there must be a cutoff value $H' \in \{1, ..., M-1\}$ such that the conditions for survival ((19), (20) and (21)) are satisfied for $H_t > H'$ but not for $H_t \leq H'$.

To gain some intuition for this result, first recall that the refugee externality cost $R(H_t)$ increases with H_t, so the cost that the collapse of country H_t would impose on the countries that are not at the brink increases with H_t. Second, when there are fewer surviving countries and hence more people in any country, the BAU emissions level $\bar{c}^N(H_t)$ is lower, as each country internalizes a larger fraction of the global cost of its emissions, so starting from a given emissions level, a defecting country has less to gain from increasing emissions to $\bar{c}^N(H_t)$. And third, the catastrophe threshold $\tilde{C}_{H_t}$ increases with H_t, so other things equal smaller mitigation efforts are required to save country H_t from collapse.

The following lemma summarizes the key features of the equilibrium outcome conditional on a country being on the brink:

Lemma 1 *If country H_t is on the brink in period t ($C_{t-1} = \tilde{C}_{H_t}$) then: (i) If there exist (c^b, c^{nb}) that satisfy (19), (20) and (21), country H_t survives with the help of emissions reductions below BAU levels by other countries; (ii) Otherwise, country H_t collapses and the surviving countries continue to choose their BAU emissions. (iii) Country H_t is more likely to survive if more countries have collapsed in the past (i.e. if H_t is higher).*

[33] To see this, consider first country H_t. Since (21) is satisfied, country H_t prefers the equilibrium with survival to emitting $\bar{c}^N(H_t)$ and collapsing at the end of the period. But H_t's payoff would be smaller still in the equilibrium where all countries emit $\bar{c}^N(H_t)$ and country H_t collapses, because the carbon stock would be larger under the latter scenario. So country H_t is better off in the equilibrium with survival than in the equilibrium where country H_t collapses. The argument is similar for countries $j > H_t$. Country j prefers the equilibrium with survival to emitting $\bar{c}^N(H_t)$ and having country H_t collapse. But, using the same logic as above, j's payoff would be smaller still in the equilibrium where all countries emit $\bar{c}^N(H_t)$ and country H_t collapses. So also countries $j > H_t$ are better off in the equilibrium with survival than in the equilibrium where country H_t collapses. The claim then follows.

The above analysis raises an interesting question: Does the heterogeneous-brink scenario exhibit a "domino effect" when countries collapse on the equilibrium path? At one level the answer is yes, for the simple reason that a given country i can reach the brink only if the more vulnerable countries countries $1, 2, ..., i-1$ have already collapsed. In this sense the heterogeneous-brink scenario exhibits a domino effect. However, as Lemma 1 highlights, a country who has reached the brink is less likely to collapse if more countries have collapsed in the past, so in this sense there is also an "anti-domino effect" in the heterogeneous-brink scenario.

Having characterized the equilibrium emissions conditional on the carbon stock, it is easy to back out the implied equilibrium path for C_t and hence for the set of countries that survive to each t. In the initial phase, all countries are present and the growth of C_t is dictated by the BAU emissions level $\bar{c}^N(1)$. Once country 1 reaches the brink, if conditions (19), (20) and (21) cannot be jointly satisfied for $H_t = 1$ then country 1 collapses at the end of period t and the rest of the world carries on with their BAU emissions level $\bar{c}^N(2)$. In a similar fashion, the growth of the carbon stock will cause the sequential collapse of further countries.

It is clear from the discussion above that a sufficient condition for country 1 – and hence a non-empty subset of countries – to collapse on the equilibrium path is that the refugee externality that the collapse of country 1 would exert on a representative citizen of the rest of the world is not severe enough. More specifically, it is easy to show that country 1 will collapse if $R(1) < G^{nb}(c^{save}, 1)$, where c^{save} denotes the maximum emissions level for countries $j > H_t$ that keeps country H_t alive if the latter reduces its emissions to zero. This sufficient condition is arguably quite weak. In reality, the negative externality felt by other countries if a country like the Maldives suffers an early collapse will be limited, due both to the relatively small population of climate refugees that would be released and to the fact that the associated refugee externality triggered by this early collapse would be shared across many countries.

When does the string of catastrophes end? Recalling that the internal cost of collapse L is assumed to be infinite for country M, the process will stop either when conditions (19), (20) and (21) can be jointly satisfied, or country M becomes the lone surviving country, in which case this country will take care of itself and stop the growth of the carbon stock.

The following proposition summarizes our main findings for the noncooperative equilibrium in the heterogeneous-brink setting:

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

Note the contrast with the common-brink scenario, where catastrophe never happens on the equilibrium noncooperative path. When heterogeneous 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, so even slight differences in collapse points across countries can lead to collapses along the equilibrium path: unless each country arrives at the brink at the same time, the "we are all in this together" forces that enabled the world to avoid collapse in the noncooperative equilibrium of our common-brink scenario will be disrupted, and these forces can only be resurrected if the collapse of the country on the brink would impose large enough refugee externalities on the remaining countries to incentivize them to do what it takes to help the country avoid collapse. It is also notable that, when collapse points are heterogeneous, some countries may enjoy a reasonable level of utility while others suffer climate collapse, and once the carbon stock has stabilized, a country that survives at the brink may experience a lower standard of living than the other surviving countries. Hence, the heterogeneous-brink scenario highlights the possibly uneven impacts of climate change across those countries who, due to their geographic and/or socioeconomic attributes, are more or less fortunate.

Our findings for the heterogeneous-brink setting also suggest an additional perspective from which Greta's protest can be interpreted: the lax climate policies of previous generations will impose severe costs not only on those generations that will live at the brink once the catastrophe phase has ended, but also on those who will have to live through the catastrophe phase in which vulnerable countries collapse and the surviving countries suffer disruptions due to climate refugees.

4.2 International Climate Agreements

We next revisit the potential role for ICAs, but now in the context of the heterogeneous-brink scenario. In the common-brink scenario where countries were symmetric there was no role for international transfers, so we abstracted from them. But when collapse points are heterogeneous across countries, international transfers become relevant. So in the context of ICAs we now allow for international transfers.

Given the availability of international transfers, for a given generation t the ICA specifies emissions for each country that has survived to date in order to maximize world welfare, and then uses international transfers to divide up the surplus across countries. We keep the determination of international transfers in the background, and focus below on determining the emissions levels that maximize world welfare.[34] As all countries are symmetric except for the threshold $\tilde{C}_i$, efficiency again dictates that the ICA choose the same per-capita emissions in period t for each country that has survived to that point, so as before we omit the country subscript i on ICA emissions.

Under the assumption that $\beta = 0$, recalling that the population of country H_t is $\frac{M}{M-H_t+1}$ while the population of the remaining countries is $M - \frac{M}{M-H_t+1} = \frac{M(M-H_t)}{M-H_t+1}$ and using $R(H_t) = \frac{r}{M-H_t}$,

[34] A simple way to "microfound" this approach would be to assume that countries can make lump-sum transfers of an outside good that enters additively into utility, and each country is endowed with a large enough amount of the outside good that it never faces a binding constraint on transfers.

we can write the average per-capita world welfare at time t as

$$U_t = B(c_t) - \lambda C_t - \left(\frac{L+r}{M - H_t + 1}\right) \cdot I_t \tag{22}$$

where I_t is a dummy that equals one if country H_t collapses at time t. The ICA for generation t will choose c_t to maximize U_t.

Consider first the initial warming phase, in which the carbon stock is below the catastrophe point for country 1 ($C_{t-1} < \tilde{C}_1$). In this phase the ICA selects the symmetric level of emissions $\bar{c}^{ICA} \equiv B'^{-1}(M\lambda)$ that maximizes the common per-period payoff, just as in the warming phase of the ICA in the common-brink scenario: as before, the ICA internalizes the horizontal climate externalities, and hence lowers emissions below the BAU level $\bar{c}^N(1) = B'^{-1}(\lambda)$.

Can the warming phase go on forever under the ICA? Recall that under our Assumption 2, country 1 will reach the brink of catastrophe under the noncooperative scenario (and may or may not collapse), so the warming phase reaches an end in the absence of an ICA. And just as in the common-brink scenario, $C^{ICA} \equiv \frac{M}{\rho}\bar{c}^{ICA}$ is the level to which the carbon stock would eventually converge given the emissions level $\bar{c}^{ICA}$. Thus, if $\tilde{C}_1 \geq C^{ICA}$ the ICA prevents any country from ever reaching the brink and the warming phase can go on forever. On the other hand, if $\tilde{C}_1 < C^{ICA}$, country 1 is brought to the brink under the ICA, just as it would be in the ICA's absence, and we need to consider what happens next. To avoid a taxonomy of uninteresting cases, we focus on the case in which, absent an ICA, a non-empty subset of countries would collapse. Recall that this is guaranteed under the rather weak condition $R(1) < G(1)$.

Suppose, then, that $\tilde{C}_1 < C^{ICA}$, and that country 1 has reached the brink. The generation alive at this moment now faces a very different international cooperation problem than the problem faced by previous generations. In particular, the world is now confronted with a stark choice: it can cooperate to save country 1 from collapse, or it can let country 1 collapse at the end of period t and carry on without it. The availability of international lump-sum transfers ensures that the ICA will make this choice so as to maximize the average per-capita world welfare of generation t as defined in (22). This implies that country 1 will be saved under the ICA if and only if the global loss from the collapse of country 1 exceeds the (minimum) cost to the world of cutting emissions by an amount sufficient to stop the growth of the carbon stock.

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

More generally and with the above logic in mind, if a given country H_t is at the brink of

catastrophe under the ICA, it will be saved if and only if

$$\Gamma(H_t) \equiv \left[B(\bar{c}^{ICA}) - \lambda\left((1-\rho)\tilde{C}_{H_t} + M\bar{c}^{ICA}\right)\right] - \left[B\left(\frac{\rho\tilde{C}_{H_t}}{M}\right) - \lambda\tilde{C}_{H_t}\right] \leq \frac{L+r}{M-H_t+1} \equiv \psi(H_t) \tag{23}$$

The left-hand side of (23) is the difference in gross per-capita welfare between (i) setting emissions $\bar{c}^{ICA}$ for all countries and letting country H_t collapse at the end of the period, and (ii) setting emissions $\rho\tilde{C}_{H_t}/M$ for all countries and saving country H_t from collapse. The right hand side of (23) is the global average per-capita loss from the collapse of country H_t, which is given by $L+r$ times country H_t's population $\frac{M}{M-H_t+1}$ averaged over the world population M.

The marginal surviving country under the ICA is the lowest H_t such that condition (23) is satisfied, which we denote H^{ICA}. Notice that $\psi(H_t)$ is increasing in H_t, for essentially the same reasons that the refugee externality $R(H_t)$ is increasing, as we discussed above: countries that collapse later release a greater number of climate refugees on a smaller rest-of-world population. Furthermore, it is direct to verify that $\Gamma(H_t)$ decreases with H_t.[35] Thus, the ICA will let the most vulnerable country collapse if and only if $\Gamma(1) > \psi(1)$, and in this case, the sequence of country collapses under the ICA will stop when $\psi(H_t)$ rises above $\Gamma(H_t)$.

We can now turn to a key question: Will the ICA save some countries that would have collapsed in its absence? Or is it possible that more countries could collapse under the ICA than in its absence?

Recall that the marginal surviving country in the noncooperative equilibrium is the lowest value of H_t such that (19), (20) and (21) can be jointly satisfied. Denoting this value by H^N, the question is how H^N compares with H^{ICA}. As we now argue, the answer is that $H^{ICA} \leq H^N$, so the set of countries that survive under the ICA must be weakly larger than under the noncooperative equilibrium.

First note that the right hand side of (23), $\psi(H_t)$, is the population-weighted average of the right hand sides of (20) and (21), which are respectively $R(H_t)$ and L. On the other hand, as we show in the Appendix, the left hand side of (23), $\Gamma(H_t)$, is weakly lower than the population-weighted average of the right hand sides of (20) and (21), $G^b(c^b, H_t)$ and $G^{nb}(c^{nb}, H_t)$, for any distribution of emissions (c^b, c^{nb}) that satisfies (19) and therefore keeps the carbon stock at $\tilde{C}_{H_t}$. The reason is that under the ICA the emissions levels are symmetric across countries, while the emissions levels (c^b, c^{nb}) that keep the carbon stock at $\tilde{C}_{H_t}$ in the noncooperative equilibrium are in general asymmetric; and since the gross benefit function $B(c)$ is concave, the average gain from raising emissions is higher in the latter case. This in turn implies that $H^{ICA} \leq H^N$.

The set of surviving countries is weakly larger under the ICA than under the noncooperative equilibrium for two distinct reasons. The first reason was highlighted just above and relates to the "gains-from-collapse" side of the tradeoff (Γ is lower than the average of G^b and G^{nb}); the second

[35] To see this, note that $\Gamma'(H_t) = \lambda\rho\frac{d\tilde{C}_{H_t}}{dH_t} - B'\left(\frac{\rho\tilde{C}_{H_t}}{M}\right)\frac{\rho}{M}\frac{d\tilde{C}_{H_t}}{dH_t} = \frac{\rho}{M}\frac{d\tilde{C}_{H_t}}{dH_t}\left(M\lambda - B'\left(\frac{\rho\tilde{C}_{H_t}}{M}\right)\right)$. Next note $\frac{d\tilde{C}_{H_t}}{dH_t} > 0$ and $B'\left(\frac{\rho\tilde{C}_{H_t}}{M}\right) > M\lambda$ since $\frac{\rho\tilde{C}_{H_t}}{M} < \bar{c}^{ICA}$ and $\bar{c}^{ICA}$ is defined by $B'(\bar{c}^{ICA}) = M\lambda$. This implies $\Gamma'(H_t) < 0$.

reason relates to the "refugee-cost-from-collapse" side of the tradeoff. To see this second reason, suppose that Γ were *equal* to the population-weighted average of G^b and G^{nb}, in which case (23) would be a side-by-side weighted average of (20) and (21). Then it is easy to see that the set of surviving countries would *still* be weakly larger under the ICA than under the noncooperative scenario, and strictly larger for a whole parameter region.[36] This is intuitive, since in the noncooperative scenario, if a given country causes the collapse of country H_t by increasing its own emissions, it exerts a negative refugee externality on all other countries (and this is true also if country H_t causes its own collapse); this is a negative externality that the ICA would internalize.

We have established above that $H^{ICA} \leq H^N$, and if $H^{ICA} < H^N$ the ICA will save some countries from collapse. A remaining question is whether there is any role for an ICA if $H^{ICA} = H^N$. In this case there are two possible roles that the ICA might play. First, it is possible that the most vulnerable surviving country under the ICA never reaches the brink, in which case the steady-state level of the carbon stock is lower under the ICA than in the noncooperative scenario, and hence the ICA plays a role in keeping the climate cooler. Second, in the discussion above we supposed that in the noncooperative scenario countries coordinate on Pareto-undominated equilibria, but if such coordination fails, then the number of countries that collapse in the noncooperative scenario may be higher than under the ICA, and hence just as in our common-brink scenario the ICA would have a coordination role to play.

We can now summarize our main result on the comparison between the set of countries that survive in the noncooperative equilibrium and under the ICA in the heterogeneous-brink scenario:

Proposition 6 *When the catastrophe points $\tilde{C}_i$ differ across countries, a (weakly) larger subset of countries survive under the ICA than in the noncooperative scenario, but the ICA may still let some countries collapse.*

It is interesting to note that, even if a country is "saved" by the ICA, so that it avoids the cost of collapse L, it may have to pay a price for this in the form of a transfer to compensate more resilient countries for cutting their emissions relative to the BAU level. Thus under the ICA more vulnerable countries may still have to accept a lower standard of living than more resilient countries, even though the ICA allows them to avoid collapse.[37]

[36] An example can illustrate the point. Suppose there are only two countries ($M = 2$), so the population of each country is 1/2, and country 1 is at the brink in period t ($C_{t-1} = \tilde{C}_1$). Further suppose, as mentioned in the text, that $\Gamma = (G^b + G^{nb})/2$, and recall that in this case $\psi = (r + L)/2$. Then (20) and (21) reduce respectively to $G^{nb} \leq r$ and $G^b \leq L$, while (23) reduces to $G^b + G^{nb} \leq r + L$. These conditions define parameter regions in (L, r) space, and it is clear that the region where (20) and (21) are both satisfied is a proper subset of the region where (23) is satisfied.

[37] We have left the determination of transfers in the background of the model, but it is easy to see that a country saved by the ICA will have to compensate the remaining countries if, for example, the threat point for the ICA negotiations is given by the noncooperative equilibrium and the country at the brink does not have more bargaining power than the more resilient countries. It is also important to emphasize that, since countries differ in our model only in the catastrophe threshold $\tilde{C}_i$, the only possible role for international transfers is for more vulnerable countries to compensate more resilient countries, but as we have observed (see note 22), in a richer model where countries differ in other dimensions as well, international transfers could run in different directions, and this could soften the particular conclusion we highlight just above.

Finally, we have assumed that countries do not face binding constraints in their ability to make transfers, but if international transfers are limited because of resource constraints the ICA will have a more limited ability to save countries from collapse than we have characterized here. Intuitively, the ICA will be less likely to save a given country if the country faces a more severe constraint on international transfers, because even if it would be efficient to save the country in the presence of unlimited international transfers, the ICA can orchestrate this outcome only if the country has enough resources to compensate the remaining countries for cutting their emissions. This suggests that smaller countries (like the Maldives) with more severe resource constraints are less likely to be able to look to an ICA to save them from climate catastrophe.

4.3 The Social Optimum

We next consider the social optimum in the context of the heterogeneous-brink scenario. As international lump-sum transfers are available, within any generation the planner maximizes average per-capita world welfare and uses transfers to redistribute the surplus in each period across countries according to their Pareto weights (which we leave in the background).

As we noted in section 2, with the social discount factor $\hat{\beta} \geq 0$ the objective of the social planner can be written as $W = \sum_{t=0}^{\infty} \hat{\beta}^t U_t$, where U_t is now given by (22), yielding the planner's problem

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

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

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

How could it be that a planner with a higher discount factor might choose to allow the carbon stock to rise further and cause more countries to succumb to climate catastrophe than would the same planner with a lower discount factor? After all, a planner with a higher discount factor will certainly wish to shift utility toward future generations. The question, though, is how best to do this. Cutting emissions today and lowering the carbon stock that will be inherited by future generations has a direct effect that increases the utility of future generations, and this suggests that a planner with a higher discount factor will make emissions choices that lead to a lower carbon stock at any moment in time – and therefore emissions choices that imply (weakly) fewer countries succumbing to climate catastrophe along the optimal path – than would the same planner if it had a lower discount factor. But raising emissions today and crossing the brink of catastrophe for some country would generate a climate refugee cost that is borne primarily (and under our assumption that this is a one-time cost, solely) by the generation alive today; and if this brink constraint is binding on the emissions choices that the planner would have made for future generations in the absence of the brink, then raising emissions today – and incurring today the refugee costs of crossing the brink in order to relax the constraint faced by those alive tomorrow – works indirectly to shift utility toward future generations. As we demonstrate, this indirect effect can dominate.

More specifically, we show in the Appendix that, if for a given discount factor $\hat{\beta}$ the marginal surviving country reaches the brink of catastrophe under the optimal plan, then it is possible that increasing the discount factor, say to $\hat{\beta}+\varepsilon$, will lead the country to collapse along the optimal path, so that the set of surviving countries under the optimal plan shrinks. And owing to the fact that the social optimum for $\hat{\beta} = 0$ corresponds to the ICA outcome, we obtain the following result:

Proposition 7 *Suppose the catastrophe point $\tilde{C}_i$ differs across countries. As judged by a social planner with $\hat{\beta} > 0$, the number of countries that collapse under an ICA – and the long-term extent of global warming – can be either too high or too low.*

Finally, it is worth emphasizing the central role played in this finding by our assumption that the refugee costs L and r associated with country collapse are not permanent, but rather are borne primarily by the generation alive at the time of the country's collapse, as it is only due to this assumption that crossing the brink today can shift utility toward future generations. Our model adopts an extreme version of this assumption, namely, that these are one-time costs, but our results would survive as long as these costs are concentrated sufficiently on the current generation.

5 Intergenerational Altruism

We now extend our analysis to allow citizens to care about their offspring ($\beta > 0$). We revisit both the common-brink scenario of section 3 and the heterogeneous-brink scenario of section 4.

Common brink Consider first the noncooperative outcome in the common-brink scenario. Recall that if agents place positive weight β on their offspring they will maximize the present value of

the stream of future utilities with a discount factor β, as defined by the dynastic utility function given in (1), even though each agent has only a single offspring. In our infinite-horizon setting, this would give rise to equilibria where players can sustain cooperative behavior by punishing past actions (e.g. trigger-strategy equilibria), or in other words, self-enforcing agreements. But when characterizing the ICA outcome we have abstracted from issues related to self-enforcing agreements and have assumed that externally enforced agreements are available. In this light it seems reasonable when considering noncooperative equilibria in a setting where β is positive to focus on equilibria where current behavior is not conditioned on past actions, and then compare those equilibria with externally-enforced agreements. One way to do this is to focus on a finite-horizon game, because in this case the logic of backward induction unravels self-enforcing agreements. Thus in this section we will consider a version of our game with T periods.[38] Ideally we would like to consider a game with large T, but the characterization of noncooperative outcomes when β is positive turns out to be analytically quite complex, due to the possibility of catastrophe. So our strategy in this section is to solve analytically a two-period version of the game, which will bring about some new dynamic insights, and then turn to numerical methods to obtain results for a larger number of periods.

We therefore consider now the two-period version of our common-brink scenario with $\beta > 0$, focusing on subgame perfect equilibria when characterizing the noncooperative outcome. To streamline the exposition we suppose that there are only two countries, A and B. The formal setup is identical in all other respects to the common-brink scenario we analyzed in section 3.

A first observation concerns the BAU path of emissions, which we define as the equilibrium path of per-country emissions if the catastrophe threshold $\tilde{C}$ never binds (i.e. it has no impact on equilibrium emissions), and which we denote by (c_1^{BAU}, c_2^{BAU}). It is easy to see that $c_1^{BAU} < c_2^{BAU}$, since the first generation cares about the next generation, whereas the second generation has no offspring. In what follows we will focus on the more interesting case where the catastrophe threshold $\tilde{C}$ is binding so that the noncooperative equilibrium differs from the BAU emission level.

We proceed by backward induction. The analysis of the subgame ($t = 2$) is straightforward. If the carbon stock inherited from period 1, C_1, is lower than a critical level $\bar{C}_1$, the catastrophe threshold is not binding in period 2 and countries choose their BAU emissions c_2^{BAU}; but if the carbon stock is between $\bar{C}_1$ and $\tilde{C}$, the catastrophe threshold is binding, and the two countries will

[38] Two comments are in order. First, in the literature on dynamic games there is another approach to capturing "non-cooperative" behavior, namely focusing on Markov-perfect equilibria of the infinite-horizon game, where current behavior is conditioned only on the current value of the state variable(s) and not on the players' past actions. This approach is effective in infinite-horizon games where it is not possible to credibly punish past actions by using Markov strategies. But in games with global public resources such as ours, where the stock of the resource is the state variable, it has been shown (see, e.g., Dutta and Sundaram, 1993, and Battaglini et al., 2014) that there is a vast multiplicity of Markov-perfect equilibria, including some where players punish past actions indirectly by conditioning current behavior on the current value of the state variable. Thus, in our framework, characterizing non-cooperative behavior by focusing on Markov-perfect equilibria in an infinite-horizon game is not viable. The second comment is that, in some finite-horizon games where the stage game has multiple equilibria, some cooperation can be sustained by specifying that bad actions will be punished by switching to a worse equilibrium of the stage game (see for example Dixit, 1987). In our model, the stage game can have multiple equilibria, for example if players are at the brink of catastrophe, but we will assume that the equilibrium played by the government is not conditional on past actions. This seems natural given that we want to abstract from self-enforcing agreements.

cut their emissions below c_2^{BAU} in a symmetric way to avoid catastrophe (we can ignore levels of the carbon stock higher than $\tilde{C}$ because they cannot arise in equilibrium).

Next focus on period 1. Here the equilibrium emissions depend crucially on the tightness of the catastrophe threshold $\tilde{C}$. The key step for characterizing the equilibrium emissions is to understand the shape of a country's reaction function. With symmetric countries, it suffices to describe country A's reaction function, which is depicted in Figure 4(a). We draw two loci in Figure 4(a): the combination of emissions such that $c_1^A + c_1^B = \bar{C}_1$ and the combination of emissions such that $c_1^A + c_1^B = \tilde{C}$ (which we refer to as the Brink1 locus). Both of these loci are lines with slope -1. In the region left of the $c_1^A + c_1^B = \bar{C}_1$ locus (which we refer to as the No-Brink region) the brink is never reached; in the region between the two loci (which we refer to as the Brink2 region) the brink is reached at $t = 2$; along the Brink1 locus the brink is reached in the current period ($t = 1$); and in the region to the right of the $c_1^A + c_1^B = \tilde{C}$ locus, catastrophe would occur.

We are now ready to describe the shape of country A's period-1 reaction function. To this end, let us fix $\tilde{C} > 0$ and consider how country A's optimal emissions c_1^A vary as we increase c_1^B from 0 to $\tilde{C}$. If c_1^B is between 0 and some critical level $\bar{c}_1 \geq 0$, the brink constraint does not bind and country A chooses the BAU emission level c_1^{BAU}. The interval $(0, \bar{c}_1)$ will be empty if $\tilde{C}$ is small enough, but Figure 4(a) focuses on the case where this interval is nonempty. Note that this part of the reaction function lies entirely in the No-Brink region. At $c_1^B = \bar{c}_1$, country A's best response jumps up from the BAU level to a level between c_1^{BAU} and c_2^{BAU}, bringing countries into the Brink2 region; and as c_1^B increases beyond $\bar{c}_1$, country A's best response decreases with slope flatter than -1 until it reaches the Brink1 locus, and then it runs along that locus (thus it decreases with slope -1) until $c_1^B = \tilde{C}$, at which point country A's best reply is to emit zero.[39]

Why does country A's reaction function take on this peculiar shape? While it is intuitive that the reaction function is flat at the BAU level if c_1^B is low and is decreasing if c_1^B is higher, the feature that at some point it jumps above the BAU level and stays above that level for a whole interval of c_1^B is due to an interesting effect that deserves emphasis. We label this the *dynamic free-rider (DFR) effect*: anticipating that, if the world reaches the brink in the next period, the burden of cutting emissions to save the world will be shared by both countries, in the current period an individual country has a stronger incentive to raise its emissions, and for this reason the best-response emissions may be above the BAU level.[40] Another interesting feature of the reaction

[39] The part of the reaction function that runs along the Brink1 locus may be empty. For example, if $\rho = 0$ and the utility function satisfies the Inada condition ($u(c) \to -\infty$ as $c \to 0$), then reaching the brink at $t = 1$ cannot be an equilibrium for any $\tilde{C} > 0$: emitting a positive amount at $t = 1$ that causes the world to arrive at the brink at $t = 1$ cannot be an equilibrium, because a country would deviate and move some of those emissions to $t = 2$.

[40] The upward jump in the reaction function is due to the fact that, if c_1^B is relatively close to $\bar{c}_1$, country A's payoff function has two local maxima. The first one is c_1^{BAU}, which is the optimal emission subject to (c_1^A, c_1^B) lying in the No-Brink region; this is defined by the first-order condition $u'(c_1^A) - \lambda = \beta\lambda$ and depicted in Figure 4(a) as the dashed horizontal line within the No-Brink region. The second local maximum is the optimal emission subject to (c_1^A, c_1^B) lying in the Brink2 region; this is defined by the first-order condition $u'(c_1^A) - \lambda = \beta u'\left(\frac{\tilde{C} - c_1^A - c_1^B}{2}\right)$ and is depicted as the dashed curve within the Brink2 region. Note that the difference between the two first-order condition is that in the former one, the future marginal cost of emissions is the cost of an increase in the future carbon stock at $t = 2$, while in the latter one it is the cost of a (shared) reduction in consumption at $t = 2$. At $c_1^B = \bar{c}_1$ the two local maxima attain the same value, and so at this point the optimal c_1^A jumps from one to the other.

function is that it has slope flatter than -1 in the Brink2 region. This is because, conditional on the brink being reached at $t = 2$, if country B increases its emissions at $t = 1$, thus forcing country A to reduce its own emissions in order to avoid catastrophe, the optimal way for country A to achieve this reduction is to spread it over both periods for consumption-smoothing reasons.

With the help of Figure 4(a) it is now easy to characterize how the (symmetric) noncooperative equilibrium emissions vary with $\tilde{C}$, given that the equilibrium is determined by the intersection between the reaction function and the diagonal. The key observation is that, as $\tilde{C}$ falls, the reaction function shifts horizontally to the left. In what follows we assume that if there are two equilibria the countries focus on the more efficient one, but the qualitative conclusion would not change in the opposite case. The resulting equilibrium first-period emissions are labeled c_1^e in Figure 4(b).

If $\tilde{C}$ is large enough, the catastrophe constraint has no impact on the equilibrium emissions, so $c_1^e = c_1^{BAU}$. If $\tilde{C}$ lies in the intermediate interval $(\tilde{C}_1, \tilde{C}_2)$, the equilibrium emissions are above c_1^{BAU}: this is the reflection of the DFR effect highlighted above. And if $\tilde{C}$ is below $\tilde{C}_1$, the equilibrium emissions are below c_1^{BAU}: here the DFR effect is still at work (because it's still true that the burden of saving the world tomorrow will be shared), but since the catastrophe threshold is tight the DFR effect is outweighed by the pressure to keep emissions low in both periods, coupled with the need to smooth consumption over time.

The analysis above suggests a number of further questions. How does the DFR effect manifest itself in the noncooperative outcomes, if at all, with a larger number of periods ($T > 2$)? And are there new strategic effects that arise with $T > 2$? Our next step is to employ a numerical approach to examine the game for a larger number of periods so that we can shed light on the answers to these questions. Due to computational constraints, we extend the number of periods to $T = 4$, striking a compromise between achieving a high level of precision with the available computational resources and having a sufficient number of periods to allow for the main interesting effects to arise.

The key results of our numerical analysis are illustrated in Figures 5(a)-5(d). In each figure, the top panel plots the evolution of the global carbon stock C_t over time while the bottom panel plots the evolution of emissions c_t. Noncooperative equilibrium magnitudes are shown as solid red lines, and the BAU magnitudes are shown as dashed blue lines.

For the model parameters described in Figure 5(a), the brink of catastrophe is reached at the end of period 2, and the DFR effect causes noncooperative emissions to rise above their BAU level in period 1. Here, the first generation free rides on the efforts that the second generation in the other country will make to avoid going over the brink in period 2. For the model parameters described in Figure 5(b), the DFR effect causes noncooperative emissions to rise above their BAU level in period 1 but the brink of catastrophe is not reached until the end of period 3. Here, the first generation free rides on the efforts of the next *two* generations in the other country, as both of those generations will reduce emissions to deal with the approaching brink at the end of period 3 (with generation 2 spreading emission cuts over two periods for consumption-smoothing reasons).

For the model parameters described in Figure 5(c), the brink is avoided altogether. But this is accomplished with the help of the first two generations, who keep their emissions below the BAU

level so as to prevent the carbon stock from rising to the point where the DFR effect would kick in for generation 3 and the associated inefficiencies would be incurred. In other words, here the earlier generations reduce emissions as a commitment device to avoid the costs of the DFR effect for later generations. Finally, for the model parameters described in Figure 5(d), the brink is again avoided altogether, but now this is due solely to the efforts of generation 2, whose emissions are kept below the BAU level so as to prevent the carbon stock from rising to the point where the DFR effect would kick in for generation 3. And as Figure 5(d) depicts, here there is a "second-order" DFR effect that drives the emissions level of generation 1 *above* its BAU level, as generation 1 free-rides on the future efforts of the other country's generation 2 to avoid the DFR effect for generation 3.

Having considered the noncooperative outcome in the common-brink scenario with intergenerational altruism, we next consider the potential role that an ICA can play for improving over the noncooperative equilibrium. Clearly, a key insight from our section-3 analysis continues to hold with $\beta > 0$: the ICA has a role to play by slowing down the pace of warming while the world is not yet at the brink, but once the world reaches the brink, the ICA no longer has a role to play (beyond possibly helping to solve a coordination failure). For example, in the model with $T = 4$ analyzed numerically above, if the world reaches the brink under the ICA at the end of period $t = 2$ or $t = 3$, at that point the ICA can be abandoned without any loss. At the same time, our analysis here suggests that the ICA may have an additional role to play relative to the case of no intergenerational altruism: if a DFR effect is at play in the noncooperative scenario, so that emissions are above the BAU level in the run-up to the brink, the ICA can address this dynamic free riding, in addition to addressing the more standard (static) free riding behavior.

Finally, we revisit the comparison between the ICA solution and the social optimum with intergenerational altruism. In our section-3 analysis we assumed $\beta = 0$ and allowed the social optimum to place some weight directly on future generations, so that $\hat{\beta} > \beta = 0$. Now we focus on the case $\beta > 0$, with the social discount factor again allowed to be higher than β, so that $\hat{\beta} > \beta > 0$. Clearly, comparing the ICA with the social optimum is equivalent to comparing the social optimum for two different values of $\hat{\beta}$, a lower level $\hat{\beta}_0 = \beta$ (corresponding to the ICA outcome) and a higher level $\hat{\beta}_1 > \hat{\beta}$ (corresponding to the social optimum). Viewed in this light, the only difference relative to our analysis of section 3 is that there we focused on the case $\hat{\beta}_0 = 0$, while now we focus on the case $\hat{\beta}_0 > 0$. For this comparison we are not bound by the same complexities that arise in the noncooperative setting analyzed above, so we can consider an arbitrary number of periods T.

Consider first the impact of raising $\hat{\beta}$ on the optimal emission path c_t. This impact hinges on whether or not the catastrophe constraint is binding. It is not hard to show that: (a) if the catastrophe constraint is not binding for $\hat{\beta} = \hat{\beta}_0$ (and hence is not binding for the higher level $\hat{\beta}_1$ either), raising $\hat{\beta}$ lowers the optimal level of emissions c_t for all generations (with the exception of generation T, who has no offspring); (b) if the catastrophe constraint is binding both for $\hat{\beta}_0$ and $\hat{\beta}_1$, so that for both levels of $\hat{\beta}$ the level of emissions c_t is declining until the brink is reached and then stays constant, raising $\hat{\beta}$ reduces the level of emissions c_t for earlier generations and leads to a more gradual decline in c_t, with the brink being reached later; and (c) if the catastrophe constraint

is binding for $\hat{\beta} = \hat{\beta}_0$ but not for $\hat{\beta} = \hat{\beta}_1$, raising $\hat{\beta}$ again lowers emissions for all generations (again with the exception of generation T). The impact of raising $\hat{\beta}$ on the optimal utility path u_t is more straightforward. It is intuitive and easy to show that raising $\hat{\beta}$ dictates a redistribution of utility from earlier generations to later generations, with the only caveat that the utility of generations who live at the brink under both levels $\hat{\beta}_0$ and $\hat{\beta}_1$ is not affected at all.

The comparison between the ICA solution and the social optimum is an immediate corollary of the observations above. The key qualitative change relative to the case $\beta = 0$ analyzed in section 3 is that in the case $\beta > 0$, the paths of emissions and utility under the ICA are smoother when the catastrophe constraint is binding on the ICA, with the "brink generation" suffering a less precipitous drop in the level of utility. But we note that, if β is positive but sufficiently small, under the ICA the brink generation will still suffer a larger drop in utility relative to the previous generations. And it remains true that the social optimum will smooth out these paths relative to the ICA case. Thus, at a broad level, the main qualitative insights for the common-brink scenario that we highlighted in the case $\beta = 0$ case continue to hold in the case $\beta > 0$.

Heterogenous brinks We now turn to the heterogeneous-brink scenario. Mirroring our analysis just above, we consider the two-period version of this scenario with $\beta > 0$, focusing on subgame perfect equilibria when characterizing the noncooperative outcome, and we suppose that there are only two countries, A (who we take to be more resilient) and B (who we take to be more vulnerable). The formal setup is identical in all other respects to the heterogenous-brink scenario we analyzed in section 4. As above, our strategy here is to solve analytically a two-period version of the game while turning to numerical methods to obtain results for a larger number of periods.

A first observation is that, if the internal and external refugee costs (L and R) are infinite, the equilibrium outcome in this scenario is the same as in the common-brink scenario considered just above, since the more resilient countries will view the collapse of a more vulnerable country as a catastrophe also for themselves. And the same statement applies if L and R are sufficiently large.

Next we turn to the more realistic case where the external refugee cost R is lower than the internal cost L. But before proceeding, we make a second observation. Given $\beta > 0$, if R is sufficiently close to zero the externality caused by the collapse of a country on the surviving countries may be positive. The reason is that, if a country collapses and its citizens migrate, in the following period the world population will be concentrated in a smaller number of countries, who therefore will better internalize the horizontal externalities from emissions, thus leading to a more efficient level of emissions. If R is sufficiently small, this positive "internalization effect" will outweigh the direct negative externality R, thus the *net* externality is positive. And if this is the case, a pure-strategy equilibrium may fail to exist.[41] In light of this observation, and since empirically the negative impact of climate refugees on the receiving countries is arguably of first-order importance,

[41] To see this intuitively in our two-country setup, note that if the net refugee externality is positive, A has an incentive to increase its emissions just enough to push B over the brink. But given A's emissions, B's best response is to reduce its emissions just enough to keep the carbon stock exactly at the brink level. A could increase its emissions enough to make it infeasible for B to save itself, but this discrete increase in emissions may be too far from its BAU level to be worth it, and as a consequence there may be no pure strategy equilibrium.

we assume in what follows that R is not too small relative to β, so that the net refugee externality is non-positive and a pure-strategy equilibrium exists. We can now turn to the question of whether our main qualitative results of section 4 continue to hold if β is positive.

Focus first on Proposition 5(i), which states that if R is not too large, some countries will collapse on the noncooperative equilibrium path. While we do not allow R to be close to zero when $\beta > 0$ for the reasons discussed above, we can confirm that with $\beta > 0$ there continues to exist a region of parameters (where R is neither too small nor too large) such that some countries collapse on the equilibrium path.[42] Also the insight of Proposition 5(ii) continues to hold with $\beta > 0$: by the very nature of our model the basic "domino effect" is still present when $\beta > 0$; and the "anti-domino effect" is also still present, in the sense that the refugee externality is more severe if more countries have collapsed in the past.

Focus next on the results of Propositions 6 and 7. Proposition 6 says that the ICA leads to a weakly smaller number of countries collapsing than in the noncooperative scenario. To probe the robustness of this result when $\beta > 0$, recall that the proof of Proposition 6 focuses on the tradeoff that determines the marginal surviving country, and in particular on a comparison across the ICA and the noncooperative equilibrium of the "gains-from-collapse" side of this tradeoff with the "refugee-cost-from-collapse" side of the tradeoff. But the refugee cost from collapse is incurred in the period when collapse occurs, so this side of the tradeoff is independent of β. And while the gains-from-collapse side of the tradeoff will be impacted by the level of β, the reason that we describe in the run up to Proposition 6 as to why the gains from collapse are smaller in the context of the ICA than in the context of the noncooperative equilibrium is still valid when $\beta > 0$. With these arguments, it can be shown that the results of Proposition 6 extend to the two-period version of our heterogeneous-brink scenario with $\beta > 0$. To get some sense of the robustness of this result for a larger number of periods, we turn to our numerical analysis, again for the case of two countries and $T = 4$. We considered a wide range of parameter values (subject to the condition that a pure-strategy equilibrium exists), and in all cases we found that a weakly larger number of countries survive under the ICA than in the noncooperative equilibrium. Regarding Proposition 7, we note that this proposition continues to hold with $\beta \geq 0$ for the simple reason that it is a possibility result, and the case $\beta = 0$ is a special case of the more general case $\beta \geq 0$.

Our final observation concerns the DFR effect. While this effect is sharpest in the common-brink scenario, it is present also in the heterogeneous-brink scenario, but now with a twist, namely that it is asymmetric and tilted toward the more resilient countries. For example, suppose that the world is relatively close to the brink of the weaker country in our two-country setup: anticipating that the weaker country will do whatever it takes to save itself, in the previous periods the stronger country may free-ride on that anticipated effort and increase its emissions above the BAU level, while the weaker country emits below the BAU level. We can confirm this asymmetric DFR effect analytically in the two-period, two-country version of the game: it is not hard to show that there

[42] This can be confirmed numerically: for the case of two countries and $T = 4$, we checked that there exist parameter values with $\beta > 0$ such that the more vulnerable country collapses in a (pure-strategy) equilibrium.

is a region of parameters such that at $t = 1$ the strong country emits above the BAU level and the weak country emits below the BAU level, and at $t = 2$ the brink of the weak country is reached but not crossed, with the weak country emitting less than the strong country.

6 Conclusion

In this paper we have presented a novel framework for studying the role of international climate agreements. Our framework features the possibility of climate catastrophe, and it emphasizes the role of both the international externalities that a country's policies exert on other countries and the intertemporal externalities that current generations exert on future generations. We have used this framework to examine the interaction between these features, and we have explored the extent to which international agreements can mitigate the problem of climate change in their presence. Our analysis delivers novel insights on the role that international climate agreements can be expected to play in addressing climate change, and it points to important limitations on what such agreements can achieve, even under the best of circumstances.

We have adopted many strong assumptions to carry out our analysis. In our working paper Maggi and Staiger (2022), we discuss a number of further extensions of our framework and analysis which seem especially salient, including the introduction of uncertainty and/or heterogeneous beliefs, the existence of long lags in the impact of emissions on climate, the possibility of a future technological fix to the climate problem, investments in adaptation in addition to mitigation, and the role of trade. Here we keep our discussion brief, and focus only on the first of these extensions. Our purpose is not to provide a full analytical treatment, but rather to identify a few key points.

In our formal model we have abstracted from uncertainty, but of course uncertainty is an important feature of climate change and the challenge that it poses for the world.[43] Moreover, and relatedly, beliefs about climate change are heterogeneous, with some countries and some groups within countries firmly believing that a climate catastrophe will occur if the world continues along its BAU path, while others are skeptical of such claims or even deny outright that the threat of climate catastrophes are real. How would these features affect our analysis?

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

With these assumptions, it is straightforward to show that introducing the implied uncertainty about the location of the critical carbon thresholds into our analysis does not change our basic

[43] Uncertainty regarding the precise location of thresholds and tipping points is widely emphasized in the climate literature (see, for example, Lenton et al., 2008, 2019, and Rockstrom et al., 2009).

findings. In the common-brink scenario, as long as the cost of exceeding $\tilde{C}$ continues to be infinite, any strictly positive probability κ that the critical carbon stock is $\tilde{C}$ rather than infinite will ensure that countries will not exceed the level $\tilde{C}$ in the noncooperative equilibrium, just as in our common-brink analysis of section 3.[44] And similarly, in the heterogeneous-brink scenario, country i will avoid collapse with certainty in the noncooperative equilibrium if $\kappa_i R_i$ and $\kappa_i L$ are above some thresholds and will collapse with probability κ_i otherwise, while under the ICA and socially optimal choices country i will avoid collapse with certainty if $\kappa_i \psi_i$ is above some threshold and will collapse with probability κ_i otherwise. In this setting, our main findings of section 4 continue to hold, and in particular, the set of surviving countries under the ICA is weakly larger than under the noncooperative scenario, but may be larger or smaller than under the social optimum.

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

Interestingly, this reinterpretation suggests that heterogeneous beliefs about the risks posed by climate change could potentially be more devastating to the world than uncertainty about the position of the critical thresholds. For example, with heterogeneous beliefs it is easy to see that there is an important new possibility in the common-brink scenario: if $X = \tilde{C}$ so that the true level of the critical carbon stock is $\tilde{C}$ and if κ is sufficiently small so that the fraction of climate skeptics and deniers in the world is sufficiently high, then it is possible that in the noncooperative equilibrium and contrary to our common-brink scenario of section 3 the world will trigger a climate catastrophe, because the climate skeptics and deniers are sufficiently prevalent in the world to preclude the possibility that the climate believers of the world could do enough on their own to avoid the catastrophe, much as can happen for the most vulnerable countries who face a climate crisis alone in our heterogeneous-brink scenario of section 4. A further implication is then that in the presence of heterogeneous beliefs a potential new role for ICAs could also arise in the common-brink scenario, namely, that through their reductions in emissions, ICAs might help keep the global carbon stock below $\tilde{C}$ and thereby help the world avoid a climate catastrophe that would occur on the noncooperative equilibrium path. And similar possibilities can arise in the heterogeneous-brink

[44] The expected cost of exceeding $\tilde{C}$ is infinite ($\kappa \times \infty + (1-\kappa) \times \lambda = \infty$), so if agents are risk-neutral their expected utility for $C > \tilde{C}$ is minus infinity, and of course if agents are risk-averse this conclusion is strengthened. Also note that, if the loss from exceeding the threshold $\tilde{C}$ is very high but finite (say $\bar{L}$), and $\tilde{C}$ is a random variable with a bounded support, then the expected loss will be continuous but rising very steeply for C in the support of $\tilde{C}$. In this case, fixing the distribution of $\tilde{C}$, as $\bar{L}$ goes to infinity the expected loss function converges to the one we assumed, and we conjecture that the results would then be approximately the same as those of our common-brink scenario.

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

7 Appendix

Proof of Proposition 3

The Lagrangian associated with the planner's problem is:

$$L = \sum_{t=0}^{\infty} \left\{ \hat{\beta}^t [B(c_t) - \lambda C_t] + \xi_t \left[C_t - (1-\rho) C_{t-1} - M c_t \right] + \phi_t (C_t - \tilde{C}) \right\} \tag{24}$$

where ξ_t and ϕ_t are Lagrange multipliers. Differentiating (24) with respect to c_s yields the first-order condition

$$\frac{\partial L}{\partial c_s} = \hat{\beta}^s B'(c_s) - \xi_s = 0. \tag{25}$$

And differentiating (24) with respect to C_s yields the first-order condition

$$\frac{\partial L}{\partial C_s} = -\hat{\beta}^s M\lambda + \xi_s - (1-\rho)\xi_{s+1} + \phi_s = 0 \tag{26}$$

where we use the fact that each C_s enters two terms of (24), the $t = s$ term and the $t = s+1$ term. Finally, solving (25) for ξ_s, substituting into (26) and converting s to t, yields

$$-M\lambda + B'(c_t) - (1-\rho)\hat{\beta} B'(c_{t+1}) + \hat{\beta}^{-t}\phi_t = 0. \tag{27}$$

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 brink constraint $C_t \leq \tilde{C}$ binds for any t.

Case 1: the brink is never reached.

We first suppose that the brink constraint never binds, so we set $\phi_t = 0$ for all t in (27).

Note that c_t enters equation (27) 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 (27) as the first-order linear difference equation

$$-M\lambda + X_t - (1-\rho)\hat{\beta} X_{t+1} = 0. \tag{28}$$

The solutions to (28) are characterized by

$$X_t = \frac{K}{\hat{\beta}^t (1-\rho)^t} + \frac{M\lambda}{1 - \hat{\beta}(1-\rho)} \tag{29}$$

where K is an arbitrary constant. The expression in (29) 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 (29)

when $K = 0$ as

$$X_t = \frac{M\lambda}{1 - \hat{\beta}(1-\rho)} \equiv X. \tag{30}$$

We now argue that only the constant solution described by (30) satisfies the first-order conditions (25) and (26). 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 (28). This determines the transversality condition for the finite-T problem:

$$X_T = M\lambda. \tag{31}$$

Note that, since $M\lambda < X$, the curve in (29) that satisfies (31) 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 (30), but one of the decreasing paths.

Now consider the limit as $T \to \infty$. As T increases, the curve in (29) that satisfies (31) gets closer and closer to the constant solution described by (30). Indeed, as $T \to \infty$ the solution converges pointwise to (30).

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

$$\bar{c}^S = B'^{-1}\left(\frac{M\lambda}{1 - \hat{\beta}(1-\rho)}\right) \tag{32}$$

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

$$\tilde{C} \geq C^S \tag{33}$$

where note from their definitions that $C^S < C^N$ so both Assumption 1 and (33) will be satisfied if $\tilde{C} \in [C^S, C^N)$.

Finally note that the global stock of carbon, denoted $\bar{C}_t^S$, evolves in Case 1 according to the difference equation $\bar{C}_t^S = (1-\rho)\bar{C}_{t-1}^S + M\bar{c}^S$ with $\bar{C}_{-1}^S = 0$.

Case 2: the brink is reached in finite time

Now suppose that the critical level of the carbon stock $\tilde{C}$ is below the threshold level C^S so that (33) is violated and instead we have

$$\tilde{C} < C^S. \tag{34}$$

In this case our candidate Case-1 solution (32) does not work, and we need to proceed to the second

guess where the brink constraint $C_t \leq \tilde{C}$ binds from some $\tilde{t}^S$ onward.

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

$$c_t = \frac{\rho \tilde{C}}{M} = \hat{c}^N \quad \text{for } t \geq \tilde{t}^S. \tag{35}$$

For $t < \tilde{t}^S$, the guess is that the brink constraint does not bind, so $\phi_t = 0$, and hence we arrive at the same system of first-order difference equations as (28), which yields the family of curves (29). Given $\tilde{t}^S$, we pick the solution (i.e., pick K) by imposing the first-order condition (28) at $t = \tilde{t}^S$:

$$-M\lambda + X_{\tilde{t}^S} - (1-\rho)\hat{\beta}\hat{X} = 0 \tag{36}$$

where $\hat{X} \equiv B'(\hat{c}^N)$. Again ignoring integer constraints, this requires continuity of X_t, and therefore of c_t.[45] But given (34) we have that $\bar{c}^S > \frac{\rho \tilde{C}}{M} = \hat{c}^N$. And recalling that $\bar{c}^S$ is defined by the constant solution to (29) with $K = 0$ so that $X_t = \bar{X}$, this implies that the socially optimal path of c_t for $t \leq \tilde{t}^S$, which we denote by $\hat{c}_t^S$, must be defined by a solution to (29) with $K > 0$ so that $X_t > \bar{X}$. It then follows from (29) together with (10) that $\hat{c}_t^S$ begins at $t = 0$ at a level that is strictly below $\bar{c}^{ICA}$, is decreasing, and hits $\hat{c}^S$ at $\tilde{t}^S$.[46]

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

$$\hat{C}_t^S = (1-\rho)\hat{C}_{t-1}^S + M\hat{c}_t^S \quad \text{with } \hat{C}_{-1}^S = 0. \tag{37}$$

Thus $\tilde{t}^S$ is defined using (37) and $\hat{C}_{\tilde{t}^S}^S = \tilde{C}$. Using this condition and the analogous condition (11) that defines $\tilde{t}^{ICA}$ as well as the properties of $\hat{c}_t^S$ described above, it is direct to confirm that $\tilde{t}^{ICA} < \tilde{t}^S$.[47] We may conclude that in Case 2, the socially optimal emissions for generation t in a representative country are given by $c_t^S = \hat{c}_t^S$ for $t < \tilde{t}^S$ and $c_t^S = \hat{c}^N$ for $t \geq \tilde{t}^S$. QED

Proof of Lemma 1

Parts (i) and (ii) of the Lemma were proved in the text, so we only need to prove part (iii).

We need to show that there is a cutoff value $H' \in \{1, ..., M-1\}$ such that the conditions for survival ((19), (20) and (21)) are satisfied for $H_t > H'$ but not for $H_t \leq H'$. This is equivalent to

[45] 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 (36).

[46] Depending on the third derivative of the B function, the implied path of $\hat{c}_t^S$ for $t \leq \tilde{t}^S$ may be concave or convex. For example if B is quadratic, the path is concave, but if B is logarithmic the path is convex.

[47] One might wonder whether there is another potential candidate solution: among the paths that satisfy (29), 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 (29) 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.

showing that, if there exist (c^b, c^{nb}) that satisfies (19), (20) and (21) for $H_t = H_t'$, this is true also for any $H_t > H_t'$. This is what we prove next.

Let $c^{nb}(c^b, H_t)$ be the value of c^{nb} that satisfies (19). Thus country H_t survives if and only if there exists c^b such that

$$\begin{aligned} G^{nb} &= \left[B(\bar{c}^N(H_t)) - \frac{M\lambda}{M-H_t+1}\bar{c}^N(H_t)\right] - \left[B(c^{nb}(c^b,H_t)) - \frac{M\lambda}{M-H_t+1}\cdot c^{nb}(c^b,H_t)\right] \leq R(H_t) \\ G^{b} &= \left[B(\bar{c}^N(H_t)) - \frac{M\lambda}{M-H_t+1}\bar{c}^N(H_t)\right] - \left[B(c^{b}) - \frac{M\lambda}{M-H_t+1}\cdot c^{b}\right] \leq L. \end{aligned}$$

Since R increases with H_t, it suffices to show that $\frac{dG^{nb}}{dH_t} < 0$ and $\frac{dG^b}{dH_t} < 0$ for any c^b in the relevant range. Note that the term in the first square brackets of the expressions above is maximized by $\bar{c}^N(H_t)$, so we can apply the envelope theorem and write:

$$\frac{dG^{nb}}{dH_t} = -\frac{M\lambda}{(M-H_t+1)^2}\left(\bar{c}^N - c^{nb}\right) - \left(B'(c^{nb}) - \frac{M\lambda}{M-H_t+1}\right)\cdot\frac{dc^{nb}}{dH_t} \tag{38}$$

$$\frac{dG^{b}}{dH_t} = -\frac{M\lambda}{(M-H_t+1)^2}\left(\bar{c}^N - c^{b}\right) \tag{39}$$

The first term on the right-hand side of (38) and the first term on the right-hand side of (39) are both negative, since both c^b and c^{nb} are lower than $\bar{c}^N$ in the relevant range. Turning to the second term of (38), it is easy to check that $\frac{dc^{nb}}{dH_t} > 0$. Next note that $B'(c^{nb}) > \frac{M\lambda}{M-H_t+1}$ because c^{nb} is lower than $\bar{c}^N$, the emissions level that maximizes $B(c) - \frac{M\lambda}{M-H_t+1}c$. We can conclude that $\frac{dG^{nb}}{dH_t} < 0$ and $\frac{dG^b}{dH_t} < 0$ as claimed. QED

Proof of Proposition 6

Recall that a country H_t at the brink survives under the noncooperative equilibrium if (19), (20) and (21) are satisfied for some (c^b, c^{nb}). We now argue that if these conditions are satisfied, also the following condition is satisfied, so country H_t will survive also under the ICA:

$$\Gamma(H_t) \equiv \left[B(\bar{c}^{ICA}) - \lambda\left((1-\rho)\tilde{C}_{H_t} + M\bar{c}^{ICA}\right)\right] - \left[B\left(\frac{\rho\tilde{C}_{H_t}}{M}\right) - \lambda\tilde{C}_{H_t}\right] \leq \frac{L+r}{M-H_t+1} \equiv \psi(H_t)$$

Noting that $\psi(H_t)$ is the population-weighted average of $\frac{r}{M-H_t}$ and L (where the weights are respectively $\frac{M-H_t}{M-H_t+1}$ and $\frac{1}{M-H_t+1}$), all we need to show is that $\Gamma(H_t)$ is no higher than the population-weighted average of $G^{nb}(c^{nb}, H_t)$ and $G^b(c^b, H_t)$ for any (c^b, c^{nb}) that satisfies (19), that is:

$$\Gamma(H_t) \leq \frac{M-H_t}{M-H_t+1}\cdot G^{nb}(c^{nb},H_t) + \frac{1}{M-H_t+1}\cdot G^b(c^b,H_t)$$

or equivalently

$$\begin{aligned} & B(\bar{c}^{ICA}) - B\left(\frac{\rho\tilde{C}_{H_t}}{M}\right) - \lambda\left(M\bar{c}^{ICA} - \rho\tilde{C}_{H_t}\right) \\ \leq\ & \frac{M-H_t}{M-H_t+1}\left[B(\bar{c}^N(H_t)) - B\left(c^{nb}\right) - \frac{\lambda M}{M-H_t+1}\left(\bar{c}^N(H_t) - c^{nb}\right)\right] \\ & + \frac{1}{M-H_t+1}\left[B(\bar{c}^N(H_t)) - B\left(c^{b}\right) - \frac{\lambda M}{M-H_t+1}\left(\bar{c}^N(H_t) - c^{b}\right)\right] \end{aligned}$$

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

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

This condition holds for any (c^b, c^{nb}) that satisfies (19) because: (i) $v(c)$ is maximized by $\bar{c}^N(H_t)$ (recalling that the population of each country at this stage is $\frac{M}{M-H_t+1}$), so $v(\bar{c}^N(H_t)) > v(\bar{c}^{ICA})$; (ii) concavity of $v(c)$ implies that the square parenthesis is nonnegative; and (iii) for country H_t to reach the brink under the ICA, it must be $M\bar{c}^{ICA} > \rho\tilde{C}_{H_t}$. QED

Proof of Proposition 7

It is straightforward to construct examples where an increase in $\hat{\beta}$ leads to a larger set of countries surviving in steady state, so we focus on the opposite possibility: we will construct an example where an increase in $\hat{\beta}$ leads to fewer countries surviving in steady state.

Suppose that the only country with a finite brink point is country 1, so $\tilde{C}_1$ is finite while $\tilde{C}_i$ is infinite for $i = 2, ..., M$. Further suppose that $\lambda = \rho = 0$.

In this case, it is clear that under the optimal plan the carbon stock C_t will reach $\tilde{C}_1$ and possibly go beyond it, depending on the global cost of country 1's collapse $L + r$.

It is also clear that, if $L + r$ is prohibitively high (or infinite), this planner problem is equivalent to the common-brink planner problem where all countries face the same brink $\tilde{C}_1$. Thus if $L + r$ is sufficiently high it is optimal for the carbon stock to grow until the brink level $\tilde{C}_1$ and stop there, so country 1 (and all other countries) survive under the optimal plan. Now, fixing $\hat{\beta}$, consider decreasing $L + r$ to the point where the planner is indifferent between stopping at the brink level $\tilde{C}_1$ and crossing the brink. Let $(L + r)_0$ denote this level of $L + r$.

We now show that, given the parameter configuration described above, if we increase the discount factor, say from $\hat{\beta}$ to $\hat{\beta}+\varepsilon$, the planner's indifference will be broken in favor of letting country 1 collapse.

Note first that for discount factor $\hat{\beta}$ the value (expressed in per-capita terms) of remaining at the brink $\tilde{C}_1$ forever is

$$\frac{B(0)}{1-\hat{\beta}}, \tag{40}$$

where we used the fact that emissions must be zero for the carbon stock to be frozen at $\tilde{C}_1$. On the other hand, the value of going over the brink today (expressed in per-capita terms) is

$$\frac{B(c^0)}{1-\hat{\beta}} - \frac{L+r}{M}, \tag{41}$$

where $c^0 \equiv B'^{-1}(0)$ is the optimal emissions level conditional on crossing the brink $\tilde{C}_1$ (or equivalently, the emissions level that would be optimal if crossing the brink were costless). Let us now set $L + r = (L + r)_0$, so that the planner is indifferent between staying at the brink $\tilde{C}_1$ and crossing it. The indifference condition can be written as

$$\frac{B(c^0) - B(0)}{1-\hat{\beta}} = \frac{L+r}{M} \tag{42}$$

Note that the left hand side of (42) is increasing in $\hat{\beta}$. Hence, given the parameter values we considered, a slightly higher discount factor $\hat{\beta} + \varepsilon$ will break the planner's indifference in favor of crossing the brink and letting country 1 collapse. QED

8 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, Salvatore Nunnari and Thomas Palfrey (2014), "Dynamic Free Riding with Irreversible Investments," *The American Economic Review* 104(9): 2858-2871.

Battaglini, Marco and Bård Harstad (2016), "Participation and Duration of Environmental Agreements," *Journal of Political Economy* 124(1): 160-204.

Besley, Timothy and Avinash Dixit (2019), "Environmental catastrophes and mitigation policies in a multiregion world," *Proceedings of the National Academy of Sciences* 166 (12): 5270-5276.

Brander, James and Taylor, M. Scott (1998), "The Simple Economics of Easter Island: A Ricardo-Malthus Model of Renewable Resource Use," *American Economic Review* 88(1): 119-38.

Caplin, Andrew and John Leahy (2004), "The Social Discount Rate," *Journal of Political Economy* 112(6): 1257-1268.

Carraro, C., and D. Siniscalco (1993), "Strategies for the International Protection of the Environment," *Journal of Public Economics* 52(3): 309-28.

Dixit, Avinash (1987), "Strategic Aspects of Trade Policy," in Truman Bewley (ed.), **Advances in Economic Theory: Fifth World Congress**, Cambridge University Press.

Dutta, Prajit K., and Roy Radner (2004), "Self-Enforcing Climate-Change Treaties," *Proceedings of the National Academy of Science* 101: 4746–51.

Dutta, Prajit K., and Rangarajan K. Sundaram (1993), "The Tragedy of the Commons?," *Economic Theory* 3(3): 413-426.

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

Feng, Tangren and Shaowei Ke (2018), "Social Discounting and Intergenerational Pareto," *Econometrica* 86(5): 1537-1567.

Harstad, Bård (2012), "Climate Contracts: A Game of Emissions, Investments, Negotiations, and Renegotiations," *The Review of Economic Studies* 79(4): 1527–1557.

Harstad, Bård (2021), "Trade and Trees," mimeo, University of Oslo.

Harstad, Bård (forthcoming), "Pledge-and-Review Bargaining: From Kyoto to Paris," *Economic Journal*.

Jenkins, Jesse D. (2014), "Political economy constraints on carbon pricing policies: What are the implications for economic efficiency, environmental efficacy,and climate policy design?," *Energy Policy* 69: 467-477.

John, Andrew and Rowena A. Pecchenino (1997), "International and Intergenerational Environmental Externalities," *Scandinavian Journal of Economics* 99(3): 371–387.

Jones, Aled and Nick King (2021), "An Analysis of the Potential for the Formation of 'Nodes of Persisting Complexity'," *Sustainability* 13, 8161: https://doi.org/10.3390/su13158161.

Kolstad, C. D., and M. Toman (2005), "The Economics of Climate Policy," **Handbook of Environmental Economics** 3: 1562-93.

Kotlikoff, Laurence, Felix Kubler, Andrey Polbin, Jeffrey Sachs and Simon Scheidegger (2021a), "Making Carbon Taxation a Generational Win Win," *International Economic Review* 62(1): 3-46.

Kotlikoff, Laurence, Felix Kubler, Andrey Polbin and Simon Scheidegger (2021b), "Can Today's and Tomorrow's World Uniformly Gain from Carbon Taxation?," mimeo, October.

Lemoine, Derek and Ivan Rudik (2017), "Steering the Climate System: Using Inertia to Lower the Cost of Policy," *American Economic Review*, 107(10): 2947–2957.

Lenton„ Timothy M., Hermann Held, Elmar Kriegler, Jim W. Hall, Wolfgang Lucht, Stefan Rahmstorf and Hans Joachim Schellnhuber (2008), "Tipping elements in the Earth's climate system," *PNAS* 105(6): 1786-1793.

Lenton, Timothy M., Rockstrom, Johan, Gaffney, Owen, Rahmstorf, Stefan, Richardson, Katherine, Steffan, Will and Hans Joachim Schellnhuber (2019), "Climate tipping points – too risky to bet against," *Nature* (Comment) 575, November 28: 592-595.

Maggi, Giovanni (2016), "Issue Linkage," in K. Bagwell and R.W. Staiger (eds.), **The Handbook of Commercial Policy**, vol. 1B, Elsevier.

Maggi, Giovanni and Robert W. Staiger (2022), "International Climate Agreements and the Scream of Greta," NBER Working Paper No. 30681. November.

Millner, Antony and Geoffrey Heal (2021), "Choosing the Future: Markets, Ethics, and Rapprochement in Social Discounting," NBER Working Paper No 28653, April.

Nordhaus, William D. (2015), "Climate Clubs: Overcoming Free-riding in International Climate Policy," *American Economic Review* 105(4): 1339-70.

Pindyck, Robert S. (2020), "What We Know and Don't Know about Climate Change, and Implications for Policy," NBER Working Paper No 27304.

Rockstrom, Johan et al. (2009), "A safe operating space for humanity," *Nature* 461(24): 472-475.

Wallace-Wells, David (2019), **The Uninhabitable Earth: Life After Warming**, Tim Duggan Books, New York.

Zaki, Jamil (2019), "Caring about tomorrow: Why haven't we stopped climate change? We're not wired to empathize with our descendants," *The Washington Post* (Outlook), August 22.

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

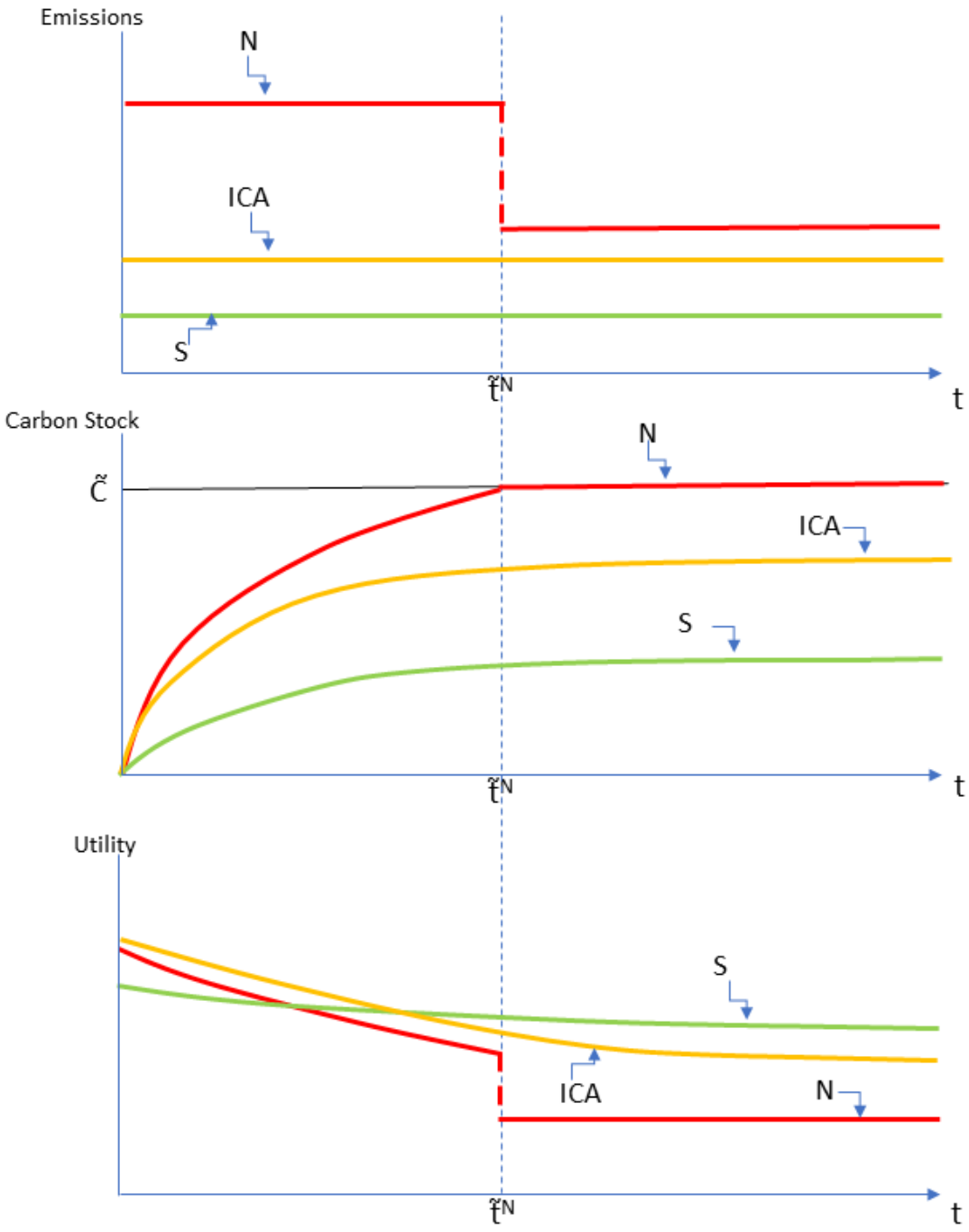

Figure 2: ICA, Planner and Noncooperative Outcomes (Intermediate $\tilde{C}$)

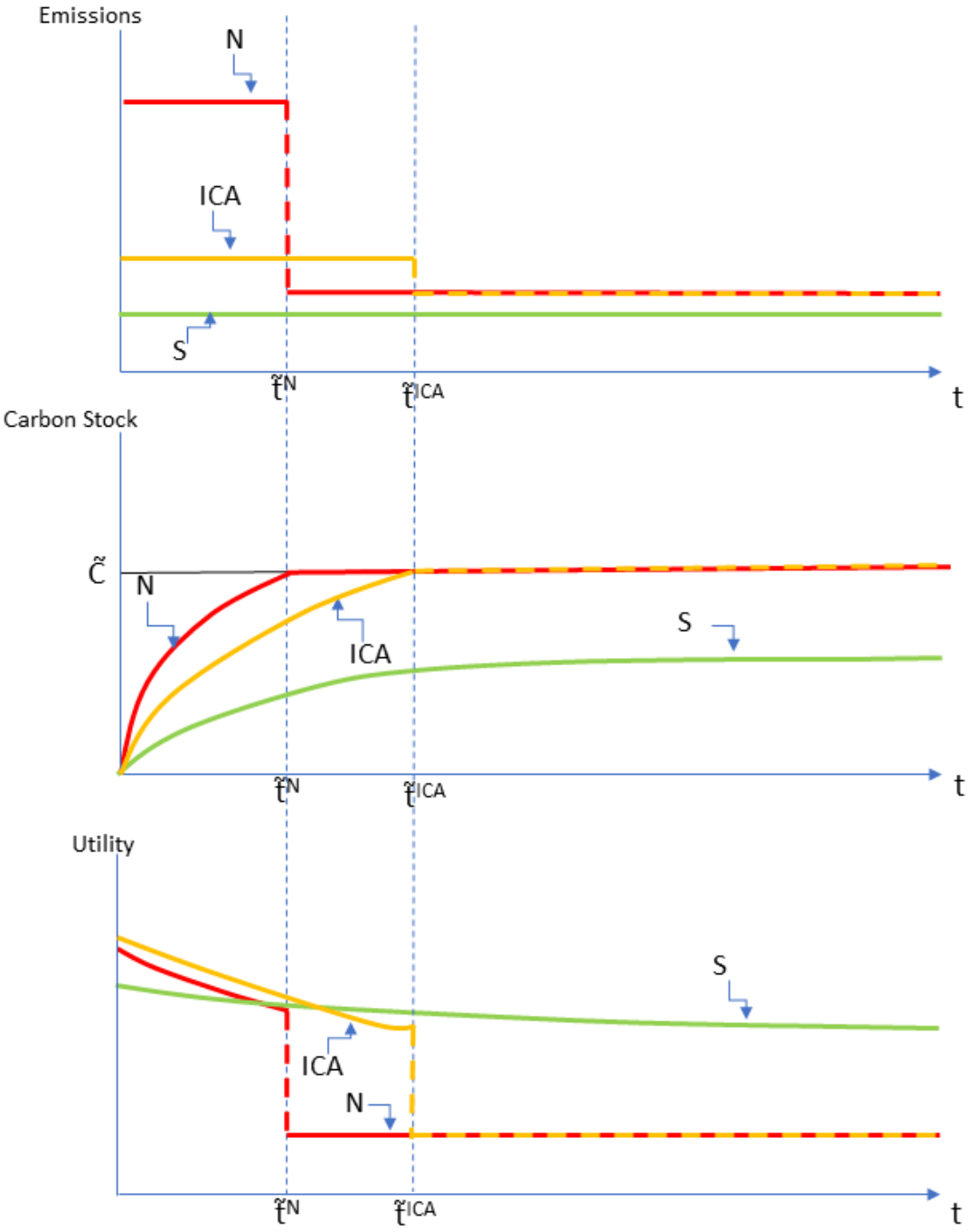

Figure 3: ICA, Planner and Noncooperative Outcomes (Low $\tilde{C}$)

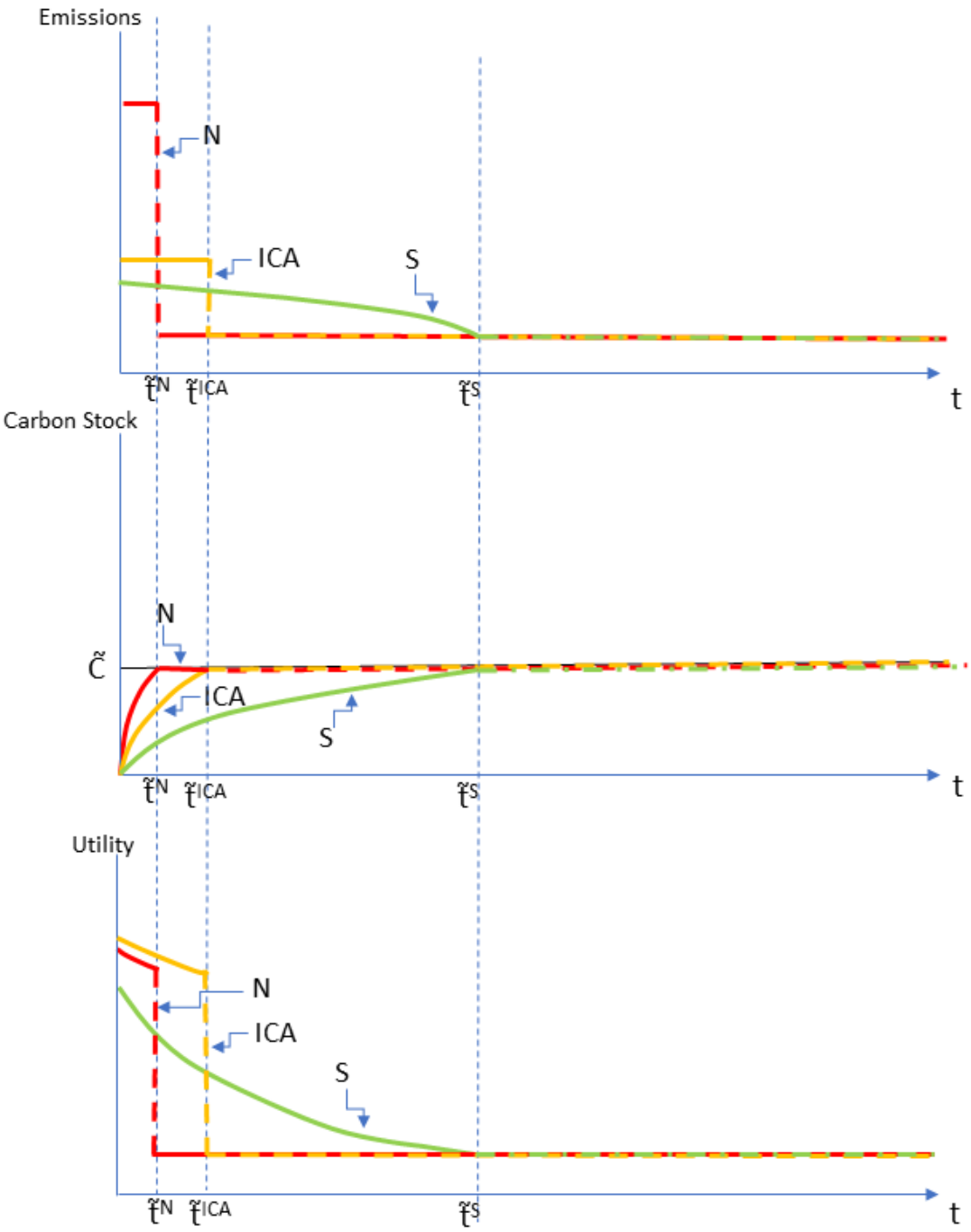

Figure 4(a): Country A's period-1 reaction function for $\beta > 0$ and a fixed $\tilde{C}$ in the two-period Common-Brink Scenario

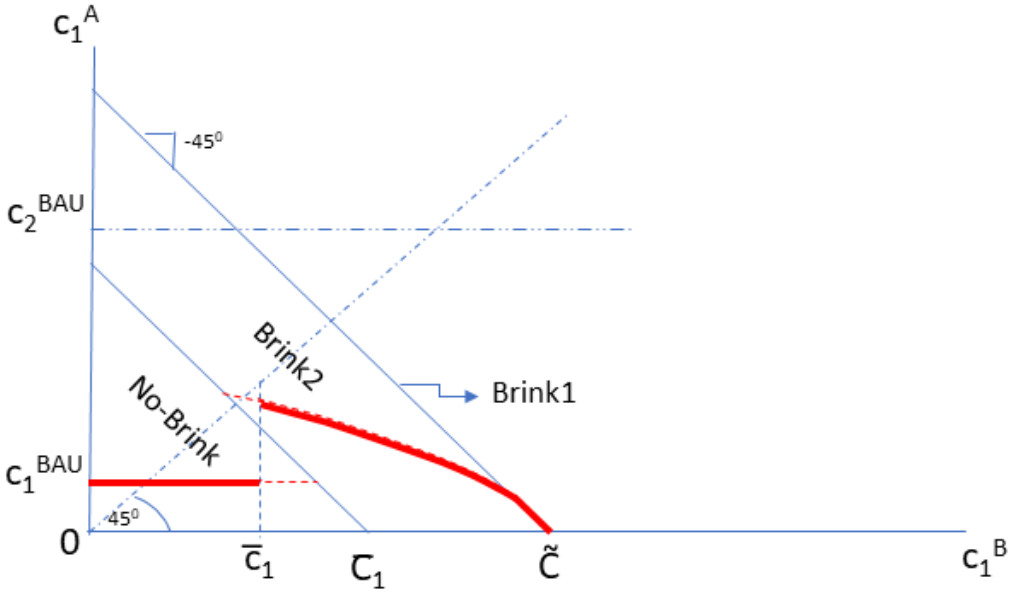

Figure 4(b): Equilibrium period-1 emissions for $\beta > 0$ as a function of $\tilde{C}$ in the two-period Common-Brink Scenario

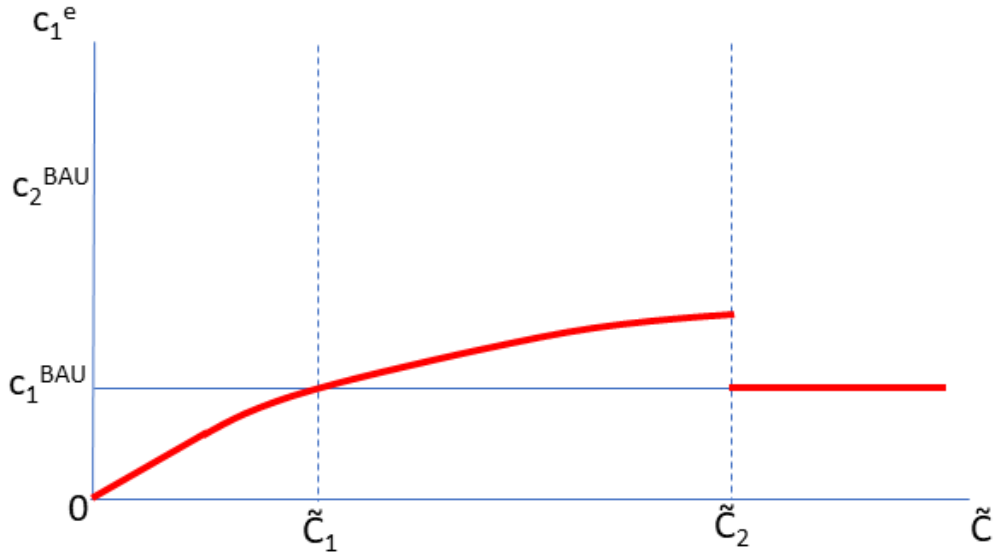

Figure 5(a): Common Brink, Noncooperative Equilibrium

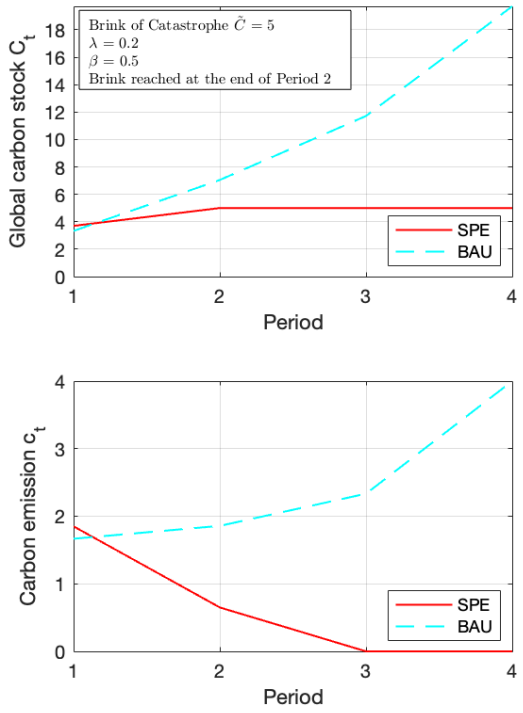

Figure 5(b): Common Brink, Noncooperative Equilibrium

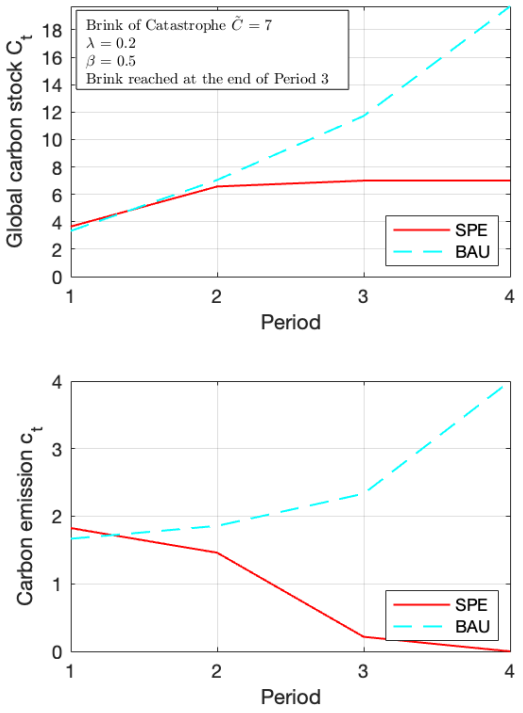

Figure 5(c): Common Brink, Noncooperative Equilibrium

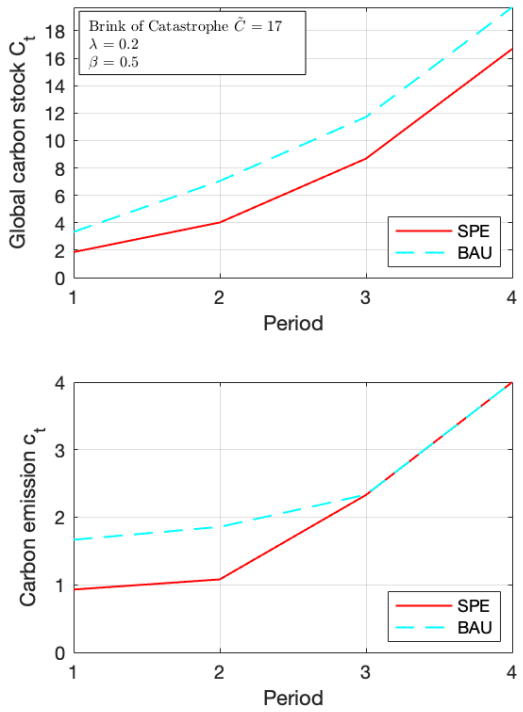

Figure 5(d): Common Brink, Noncooperative Equilibrium