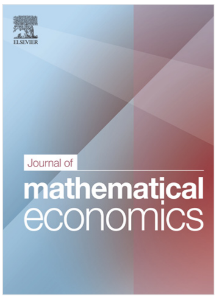

On the existence of efficient, individually rational, and fair environmental agreements

Stergios Athanasoglou

PII: S0304-4068(21)00123-3
DOI: https://doi.org/10.1016/j.jmateco.2021.102560
Reference: MATECO 102560

To appear in: *Journal of Mathematical Economics*

Received date : 25 January 2021
Revised date : 8 July 2021
Accepted date : 24 August 2021

Please cite this article as: S. Athanasoglou, On the existence of efficient, individually rational, and fair environmental agreements. *Journal of Mathematical Economics* (2021), doi: https://doi.org/10.1016/j.jmateco.2021.102560.

On the existence of efficient, individually rational, and fair environmental agreements

Stergios Athanasoglou*

January 2021; May 2021

Abstract

Suppose a group of agents are engaged in economic activity that produces emissions of pollutants. Emissions yield private benefits and impose negative externalities. The status-quo is assumed to be inefficient so that the agents are willing to negotiate an improved allocation of emissions. In this context, we are interested in allocations satisfying Pareto efficiency, individual rationality, and a principle of fairness that generalizes concepts that are encountered in practice. While the existence of such allocations is not guaranteed, we derive a necessary and sufficient condition for it. This condition is succinct and its verification is computationally tractable. Uniqueness will generally not hold, and so we describe a procedure that generates allocations with the desired properties and discuss ways of selecting from them. We apply our model to a setting of climate-change policy based on Nordhaus (2015). Our results show that it is possible to achieve a large reduction in global CO_2 emissions that enhances every region's welfare, while at the same time achieving Pareto efficiency and respecting norms of fairness.

Keywords: environmental agreement; axioms; fairness; convex optimization; climate change

JEL classifications: D62, D63, Q58

1 Introduction

The difficulty of sustaining cooperation in the management of common resources is well-established, both theoretically and empirically. International environmental agreements

*Department of Economics, University of Milan - Bicocca, Building U6, Piazza dell'Ateneo Nuovo 1, Milano 20126, Italy, stergios.athanasoglou@unimib.it. I am grateful to Stefan Ambec, Paolo Bertoletti, Laurent Drouet, Riccardo Ghidoni and Glenn Sheriff for helpful comments on earlier versions of the draft.

(IEAs) provide a salient example of this general fact (Barrett [7]). With few exceptions, it has proven very hard to negotiate effective treaties to curb global emissions of pollutants. The main reason behind the elusive nature of these agreements are the strong incentives for free-riding that countries are faced with. Free-riding occurs when it is possible to reap the benefits of cooperation without contributing to it. A large body of economic literature, too extensive to cite here, has advanced our understanding of the drivers and impacts of free-riding behavior.

In the climate context, various ways have been proposed to move the negotiations forward. Climate clubs (Nordhaus [32]), voluntary pledge-and-review frameworks such as the Paris Agreement (Sheriff [38]), agreements that focus on the supply side of fossil fuels (Harstad [24]) are just a few examples of innovative mechanisms that have been considered. Of greater relevance to the current paper are appeals to ethical norms and principles as a negotiating device. Empirical and experimental evidence suggests that equity considerations can play an important role in environmental policymaking (Lange et al. [28], Gampfer [22]). In the context of climate-change policy, equity is taken very seriously by the United Nations' International Panel on Climate Change (IPCC). A number of recent papers study the consistency of various mitigation trajectories with prominent ethical principles (Raupach et al. [34], du Pont et al. [15]).

As one might expect, there is a lively debate on the ethical principles that should underpin climate policy. This debate is far from settled [27, 15, 38]. However, even if the interested parties could agree on an appropriate standard of equity, two major problems would persist. First, the chosen ethical principle might lead to an outcome that clashes with mild requirements of efficiency such as Pareto efficiency. Second, its implementation might imply a reallocation of emissions that some agents find worse than the status-quo, and thus violate individual rationality. Both issues present major challenges to an agreement's acceptability. As regards efficiency, it would be difficult to justify choosing an allocation over another that would Pareto-improve upon it. Analogously, it would be hard to convince the relevant parties to participate in a collective effort if doing so would leave some of them worse off.

It is therefore important to know whether allocations satisfying a given standard of fairness are consistent with Pareto efficiency and individual rationality. If they are not, then this would suggest that it may be necessary to rethink the fairness requirement. Conversely, if suitable allocations do exist, it would be useful to have a practical way of computing them.

Contribution. This paper attempts to address the above points. First, we propose a principle of fairness that generalizes concepts that are commonly invoked in climate negotiations. The fairness principle's primary motivation is positive. In the climate context, countries propose criteria such as GDP, GDP per capita, historical emissions and others, to determine and rank country-level mitigation targets (Sheriff [38]). Different criteria generally lead to different rankings and it is no surprise that a country's advocacy for a particular criterion can be self-serving (see e.g., Lange et al. [29]). It thus becomes important to offer a clear-eyed account of a given fairness principle's implications: in particular, whether it is consistent with Pareto efficiency and individual rationality. The present paper develops a simple way of carrying out this task.

To this end, we demonstrate that an allocation satisfying Pareto efficiency, individual rationality and fairness may not always exist. In our analysis, this impossibility result is driven by agent asymmetry in pollution damages. Subsequently, using insights from convex optimization, we provide a simple necessary and sufficient condition for the existence of allocations satisfying the desired properties. The verification of this condition is analytically and computationally tractable. Our approach is constructive so that, when suitable allocations exist, we actually exhibit one. The individually rational and fair allocation that minimizes aggregate emissions, a constrained environmental optimum of sorts, plays a pivotal role in these results.

In general, assuming existence holds, there will be a multiplicity of solutions that satisfy Pareto efficiency (PE), individual rationality (IR) and fairness (F). This raises the question of how to select from the set of PE-IR-F allocations. A natural choice involves optimizing an appropriate objective function over this set, for instance maximizing welfare or minimizing emissions. Unfortunately, our setting does not permit a systematic adoption of this approach as the set of PE-IR-F allocations is nonconvex. We thus use the insights of our theoretical analysis to propose a simple procedure for generating suitable allocations. Having simulated a set of PE-IR-F allocations, one may then choose from among its elements on the basis of whichever criterion one sees fit, including the two mentioned above.

We provide a proof of concept of the theoretical analysis that is relevant to global climate policy. Using the model, data, and calibrated parameters of Nordhaus [32], we compute a PE-IR-F allocation of emissions to 15 macro regions. The notion of fairness that we employ, a special case of the criterion introduced in the paper's theoretical section, requires that countries with greater GDP per capita contribute proportionally more to emissions mitigation. With the aid of the simulation-based procedure mentioned earlier, we generate 1600 PE-IR-F allocations. We then use two metrics to gauge a generated allocation's performance: welfare gains and emissions reductions, both with respect to a 2011 base year.[1]

The main finding of our numerical exercise is that, along the simulated PE-IR-F

[1] The analysis focuses on year 2011 because that is the year examined by Nordhaus [32].

frontier, aggregate welfare gains are very small but emissions reductions substantial. In addition, within the set of generated PE-IR-F allocations we observe very small variation in welfare gains and relatively greater (though still small) variation in emissions reductions. Given these features of the empirical results, we propose to set aside the welfare metric and select the PE-IR-F allocation that minimizes aggregate emissions. This allocation yields a 17.5% decrease in emissions compared to base year 2011. This means that, within Nordhaus's framework, it is possible to achieve an immediate and sizable reduction in emissions that enhances every region's welfare, while at the same time achieving Pareto efficiency and respecting norms of fairness.

Related Work. To the best of our knowledge, this is the first paper that attempts to rigorously address the compatibility of Pareto efficiency, individual rationality and fairness in the management of common resources. The concept of fairness that it introduces is a novel generalization of criteria commonly encountered in policy debates.

The study of fairness and its various declinations has been a central concern of economic theory (Young [42], Roemer [36], Fleurbaey [19], Fleurbaey and Maniquet [20], World Bank [41]). Of particular relevance to the current paper is the concept of equality of opportunity (Roemer [35], Fleurbaey [19], Fleurbaey and Maniquet [20], Fleyrbaey and Peragine [21], Roemer and Trannoy [37]). Inspired by earlier work in political philosophy, this framework distinguishes between two kinds of characteristics that determine an outcome: *circumstances* (or, irrelevant characteristics) which are exogenous to the agent, and *effort* (or, relevant characteristics) for which an agent can be held accountable. Clearly, categorizing characteristics as relevant or irrelevant can be a controversial exercise which does not admit a clean resolution. Nonetheless, given such a categorization, the equality of opportunity paradigm defines fairness as the minimization of differences in outcomes due to circumstances. In contrast, the framework does not consider differences due to effort ethically questionable. Extending this reasoning even further, justice requires that agents be compensated for disadvantages in their circumstances, while they be held responsible (that is, rewarded or penalized) for their effort. While there is broad agreement on its basic principles, the exact way the equality-of-opportunity ideal is pursued differs significantly across its various formalizations (Roemer and Trannoy [37])

The fairness criterion that we propose and analyze in this work is not directly related to the above literature. Tailored to the environmental setting, it is of a more practical bent than the philosophically sophisticated work of [35, 19, 20]. Working with mitigation targets as a primary unit of analysis, this criterion operationalizes different notions of fairness that are commonly discussed in international environmental agreements. As such, it channels existing equity principles instead of proposing new ones or refining existing ones.

An important building block of the fairness criterion is an indicator that is deemed to be relevant in ranking agents' obligations to mitigate emissions (e.g., income, historical emissions, life expectancy, etc). Taking this as input, an allocation is said to be fair if it results in mitigation targets that are ordered in accordance with the given indicator data. In particular, the more accountable an agent is for reducing emissions according to the given indicator, the more demanding is his/her proportional mitigation target. In the case of climate change policy, the paper's fairness criterion nests the following views on who should contribute proportionally more to emissions mitigation: (i) nations with higher GDP per capita; (ii) nations with higher historical/cumulative emissions (iii) nations with higher current per capita emissions; (iv) nations that are more exposed to the negative effects of climate change; and (v) all nations should have the same targets. Evidently, the choice of indicator implies a particular ethical worldview and we highlight some prominent ones in the following section, where we also draw some parallels to the equality-of-opportunity paradigm.

There is a rich economics literature on international environmental agreements (IEAs), originating with the seminal papers of Barrett [6],Carraro and Siniscalco [11] and Chander and Tulkens [12, 13], but its focus is quite different from ours. Using the tools of nooncoperative [11, 6, 7, 19] as well as cooperative [12, 13] game theory, the main objective of this line of research is to study the effect of strategic behavior on the stability of environmental agreements. Special attention is given to the incentives for free-riding and to the serious constraints they impose on the design of environmental treaties. The axiomatic foundations of the IEA model are not studied in a systematic way.

Problems inspired by water allocation provide a small but meaningful counterpoint to the above body of work. Ambec and Sprumont [1], Ambec and Ehlers [2], Ansink and Weikard [4, 5], Van den Brink et al [39] and Ozturk [33] all employ axiomatic approaches to study the properties of different water-sharing schemes. Broadly speaking, this strand of the literature blends axiomatic analysis with cooperative game theory to study the normative properties of transboundary water agreements. Departing from the setting of water management, Ambec and Ehlers [3] study the axiomatic properties of the polluter-pays (PP) principle. Often invoked in policy-making circles, the PP principle requires that polluting agents bear the cost of the damages their emissions cause. Ambec and Ehlers characterize the equilibrium welfare distribution that the PP induces with axioms that echo the concepts of individual rationality and personal responsibility. Motivated by global climate policy, de Villemeur and Leroux [14] investigate the properties of various cost-sharing mechanisms. These mechanisms take the form of transfer schedules that redistribute the costs incurred by the stock of greenhouse gases. Adapting the framework of Bossert and Fleurbaey [8] to a setting with externalities, De Villemeur and Leroux explore the tension between the responsibility countries hold for their emissions and the compensation they are entitled to receive. The mechanisms they study strike a necessary balance between holding countries accountable for their contribution to the problem and compensating them for damages that are beyond their control.

The above axiomatic papers differ from our work in two important ways. First, their primary objective is to explore the foundations of specific cost-sharing mechanisms that are invoked in policy-making. The existence per se of allocations satisfying a desired set of axioms is not of central concern. The papers of Ambec and Sprumont [1] and Ambec and Ehlers [2, 3] are good examples of this general tendency. A second difference between the prior literature and the present paper is that we do not allow for transfers. Instead, we assume an environment in which utility is non-transferable, akin to the frameworks of Barrett [6] and Nordhaus [32]. This may be viewed as a weakness of our approach, since transfers between countries are often incoroprated in the formulation of environmental agreements. At the same time, the introduction of transfers is not guaranteed to be helpful as it may result in more complicated and unstable agreements that are hard to enforce (Nordhaus [32], Weikard et al [40]).[2] A further challenge that transfers pose is that countries are often reluctant to make them (e.g., in the initial stages of the Kyoto protocol) and, even if they do go through with them, issues regarding credibility persist.

On the applied side, there are a number of papers that investigate equitable emissions allocations in the context of global climate-change policy (for a review see Hohne et al. [27]). As these papers are very different in style and scope than ours, we will not describe them in detail. The general approach of this line of research is to start with a given budget of emissions and use integrated assessment modeling to study the implications of various burden-sharing schemes for individual countries and regions. For instance, Raupach et al. [34] and Du Pont et al. [15] assume a 2 degree Celsius target on global warming and investigate the economic effects of allocation mechanisms con-

[2] It is also worth noting that, in contrast to some of the literature that models IEAs with the use of repeated or dynamic games with stocks (e.g., Dutta and Radner [16], Harstad [24, 25]), our framework is static. While this is consistent with much previous work in both the axiomatic and non-axiomatic strands of the literature (e.g., [1, 2, 4, 3, 11, 6, 12, 13, 40, 39, 32, 30, 18, 33]), the dynamic nature of most stock externalities means that it is an assumption that merits attention. For example, Harstad [25] shows that under certain conditions short-term environmental agreements can be considerably less effective than longer-term ones. That being said, a static framework can be an appropriate approximation for a dynamic model with long periods. It can also capture the fact that certain countries are loath to commit to long-term emissions trajectories, instead focusing on short-term targets. The Paris Agreement itself, with its focus on periodic five-year emissions targets and stock-taking exercises, partly adheres to this structure [38]. Finally, static dynamics allow for a more straightforward axiomatic analysis.

sidered by the IPCC (inspired by different ethical principles). Meanwhile. in a recent contribution Sheriff [38] takes the opposite approach. Fixing emissions to the allocation proposed in the 2015 Paris Agreement, he determines the cost-sharing agreements and ethical principles that are consistent with it.

Paper outline. Section 2 describes the model and introduces the formal axioms. Section 3 establishes a necessary and sufficient condition for Pareto efficiency. Section 4 begins by demonstrating the generic impossibility of Pareto efficiency, individual rationality and fairness. It then establishes a necessary and sufficient condition for existence to hold and discusses ways of generating and selecting suitable allocations. Section 5 applies the theoretical results to the climate-change setting of Nordhaus [32]. Section 6 concludes.

2 Model Description

2.1 Preliminaries

There are $I \geq 2$ agents indexed by $i = 1, ..., I$. Each agent i's emissions are denoted by $e_i \geq 0$, and the group's emissions *allocation* is given by vector $\boldsymbol{e} = (e_1, ..., e_I)$. Aggregate (or total) emissions are given by

$$e = \sum_{i=1}^{I} e_i.$$

For all $i = 1, ..., I$, agent i's utility is given by

$$u_i(\boldsymbol{e}) = b_i(e_i) - c_i(e),$$

where $b_i : \Re_+ \mapsto \Re$ is a strictly concave, twice continuously differentiable benefit function and $c_i : \Re_+ \mapsto \Re_+$ is a non-negative, increasing, twice continuously differentiable and convex damage function satisfying $c_i(0) = 0$.

Agent i's initial emissions are given by $\tilde{e}_i$. The allocation $\tilde{\boldsymbol{e}} = (\tilde{e}_1, ..., \tilde{e}_I)$ can be thought of as a *status-quo* outcome that the agents wish to improve upon in a coordinated fashion. A reasonable modeling assumption would be to assume that $\tilde{\boldsymbol{e}}$ is an inefficient Nash equilibrium.

Finally, for all $i = 1, ..., I$, let D_i be an *equity-relevant indicator* of agent i. The D_i's can be thought of as an indicator or a statistic that is deemed to be relevant in determining and ranking countries' obligations to mitigate emissions. For example, in a setting where agents are individual countries, $\{D_i : \quad i = 1, 2, .., I\}$ could equal countries' per capita GDP levels, average life expectancy, historical emissions, status-quo emissions per capita, or its Human Development Index score, among others.

2.2 Properties

We begin by defining the relevant standards of efficiency and participation that we will work with.

Property 1 *An allocation* $\boldsymbol{e}$ *is* **Pareto efficient** *if there does not exist an allocation* $\boldsymbol{e}'$ *such that* $u_i(\boldsymbol{e}') \geq u_i(\boldsymbol{e})$ *for all* $i = 1, ..., I$ *and* $u_i(\boldsymbol{e}') > u_i(e)$ *for some agent* i*.*

Property 2 *An allocation* $\boldsymbol{e}$ *is* **individually rational** *if for all* $i = 1, ..., I$, $u_i(\boldsymbol{e}) \geq u_i(\tilde{\boldsymbol{e}})$*.*

Pareto efficiency provides an uncontroversial, minimal standard of efficiency. Individual rationality ensures that all agents wish to participate in the reallocation scheme. The justification behind it is that, if negotiations fail, the agents will fall back on the status-quo allocation $\tilde{\boldsymbol{e}}$, which is typically a Nash equilibrium.[3] Therefore, for a proposed allocation to be viable, it must weakly outperform the status-quo for all agents.

In contrast to the previous two properties, the concept of fairness is harder to pinpoint. For this reason we adopt a flexible approach that allows for significant generality. Recall the equity-relevant indicators $\{D_i : \ i = 1, ..., I\}$. These data are important building blocks of the equity principle we propose.

Property 3 *Suppose* $D_{i_1} \leq ... \leq D_{i_I}$*. An allocation* $\boldsymbol{e}$ *is* **fair** *if the ratios* $\{\frac{e_i}{\tilde{e}_i} : i = 1, ..., I\}$ *are weakly decreasing in equity-relevant indicators* $\{D_i : i = 1, ...I\}$*. That is,* $\boldsymbol{e}$ *is fair if*

$$\frac{e_{i_1}}{\tilde{e}_{i_1}} \geq ... \geq \frac{e_{i_I}}{\tilde{e}_{i_I}}.$$

Our fairness criterion takes as input an equity-relevant indicator D_i that is deemed relevant to determining and ranking an agent i's obligation to mitigate emissions. Subsequently, it requires that the ratios $\frac{e_i}{\tilde{e}_i}$ be weakly decreasing in D_i. This means that the more accountable an agent is for reducing emissions according to indicator D_i, the more demanding is his mitigation target (note that a smaller value of $\frac{e_i}{\tilde{e}_i}$ corresponds to a smaller value for e_i, and thus a more stringent mitigation target).

A clarifying example might be useful: suppose the indicators $\{D_i, \ i = 1, ..., I\}$ denote GDP per capita levels for countries $1, 2, ..., I$. Consider now the United States and China. Property 3 stipulates that, since the US has higher GDP per capita than China, the US mitigation target, in relative terms, should be more ambitious than the

[3] A similar standard of participation is used by Martimort and Sand-Zantman [30].

Chinese one. That is, the US should mitigate its emissions by a proportionally greater amount compared to China.

This modeling choice might seem restrictive since it does not directly take into account the agents' utilities and overall welfare distribution. So additional comments are in order.

The focus on emissions ratios $\frac{e_i}{\bar{e}_i}$ can be justified in two ways, one substantive and the other technical. On the substantive side, IEAs are frequently stipulated in terms of individual emissions reductions with respect to a given base year.[4] Presumably this is because relative emissions reductions deliver a simple and easily comparable metric that can form the basis of negotiations. Since fairness considerations are essential to the formulation of IEAs, it would seem sensible to define a fairness criterion in terms of the defining quantities of the IEA itself, i.e. emissions ratios. Doing so would allow for a direct and unambiguous application of the fairness principle to the IEA.

On the technical side, expressing a fairness criterion in terms of emissions ratios leads to clear and tractable optimization problems. By contrast, making direct reference to utility functions introduces a host of challenges. For instance, suppose the fairness axiom required that the ratios $\frac{u_i(\boldsymbol{e})}{u_i(\bar{\boldsymbol{e}})}$ to be nonincreasing in equity-relevant indicators D_i. This would be a perfectly reasonable, and in some ways quite compelling, criterion to adopt. The problem is that it implies constraints on allocations that are nonconvex and thus not readily amenable to systematic analysis. Were we to use such a formulation, it would be difficult to study the compatibility of fairness with other properties in ways that go beyond ad-hoc heuristics.

It is worth highlighting that Property 3 generalizes many equity principles that are invoked in practice. We list a few that are commonly encountered in global climate negotiations:[5]

(i) The *Brazilian proposal* dictates that a country's mitigation targets be commensurate with its contribution to the problem, consistent to the polluter-pays principle. It thus holds that mitigation efforts should be increasing to countries' stock of cumulative emissions. In our framework it would correspond to Property 3 with D_i being equal to country i's cumulative emissions.

(ii) The *equality* approach is predicated on the notion that each person has an equal claim to the production of emissions and the welfare that derives from it. It

[4] The Paris Agreement with its focus on "nationally determined contribution" (NDCs) exemplifies this fact. For example, the initial NDC of the European Union is a "binding target of an at least 40% domestic reduction in greenhouse gas emissions by 2030 compared to 1990". While there is variation in the exact formalization of NDCs across countries they can be transformed into equivalent measures (see Sheriff [38]).

[5] The following discussion borrows significantly from Sheriff [38].

requires that mitigation targets be decreasing in current per capita emissions. In the framework of Property 3, $D_i = \tilde{e}_i/P_i$, where P_i denotes the population of country i.

(iii) The *capability* approach states that mitigation targets should reflect countries' ability to incur the cost of mitigation. Thus, they should be decreasing in countries' GDP levels or other indices of socioeconomic development such as the Human Development Index (HDI). In our context, this would imply D_i is equal to a country i's GDP, GDP per capita, or HDI score.

(iv) The *benefits* approach states that the intensity of mitigation targets should be greater for countries which have more to gain from mitigation. That is, if a country has more to gain from the containment of climate change, then it should contribute more to mitigation efforts. In our framework, $D_i = -R_i$, where R_i is an index that measures a country's risk of exposure to the negative effects of climate change (e.g., extreme weather, sea level rise, etc).

(iv) The *sovereignty* approach holds that mitigation targets should be equal across countries. In our framework, this can be captured by a modification of Property 3 whereby $\frac{e_1}{\tilde{e}_1} = ... = \frac{e_I}{\tilde{e}_I}$ irrespective of any equity-relevant indicator D_i.

We note that it may be, and in fact often is, hard for countries to agree on an appropriate equity-relevant indicator D_i on which to base the ordering of mitigation targets. In the case of climate negotiations, the debate often revolves around issues of equity with consensus proving elusive. To complicate matters further, there is evidence that countries advocate for the adoption of criteria because of self-serving motivations (see e.g., Lange et al. [29]). In the presence of such disagreement, the value-added of our approach can be to provide a litmus test for a proposed criterion by determining whether it is compatible with PE and IR. Should this test fail, this result could serve as an indication that the proposed criterion is misguided and should be set aside. Of course, unless this procedure eliminates all but one recommendation, ambiguity will persist as regards the choice of D_i. One possible (though by no means the only) way of overcoming this impasse would be to have the interested parties vote on their preferred D_i and select the choice that receives most votes.

Multidimensional extension. Property 3 can be extended to accommodate multiple dimensions of equity. Suppose we have two sets of equity-relevant indicators[6], $\{D_i^1 : i = 1, ..., I\}$ and $\{D_i^2 : i = 1, ..., I\}$, implying different orderings of the agents: $D_{i_1}^1 \leq ... \leq D_{i_I}^1$ and $D_{j_1}^2 \leq ... \leq D_{j_I}^2$. To take an example from the climate context,

[6] The same reasoning can be applied to the case of more than two sets.

the D_i^1's could be GDP per capita and the D_i^2's could be current emissions. Suppose, further, that we would like the fairness criterion to take both equity-relevant indicators into account when ordering emissions ratios. For this purpose it is sufficient to extend the definition of Property 3 in the following way: For every pair of agents $i, j \in \{1, ..., I\}$, if (i) $D_i^1 \leq D_j^1$ and $D_i^2 \leq D_j^2$, set $\frac{e_i}{\bar{e}_i} \geq \frac{e_j}{\bar{e}_j}$; otherwise if (ii) $D_i^1 \leq D_j^1$ and $D_j^2 \geq D_i^2$, set $\frac{e_i}{\bar{e}_i} = \frac{e_j}{\bar{e}_j}$. After we have completed this operation for all pairs of agents we will be left with a weak ordering of the emissions ratios that is consistent with both $D_{i_1}^1 \leq ... \leq D_{i_I}^1$ and $D_{j_1}^2 \leq ... \leq D_{j_I}^2$. That is, for all pairs of agents $i, j \in \{1, ..., I\}$, we will have $\frac{e_i}{\bar{e}_i} \geq \frac{e_j}{\bar{e}_j}$ whenever either $D_i^1 \leq D_j^1$ or $D_i^2 \leq D_j^2$, with equality holding if $D_i^1 \leq D_j^1$ and $D_j^2 \geq D_i^2$.

A potential shortcoming of the above multidimensional extension of Property 3 is that it may often result in pairs of countries being assigned equal mitigation targets. It could be argued, not without reason, that such ties limit the usefulness of the proposed criterion. There are various ways to engage with this critique. The first is simply to accept the possibility of equal targets as a natural implication of simultaneously entertaining different conceptions of fairness. Different ethical visions, encapsulated by different choices for D_i, generally imply different orderings for country mitigation targets. Without an assessment of the relative standing of these ethical principles, a reasonable way forward is to "declare a tie" and set equal targets for the affected countries. Conversely, another way of dealing with the conflicting recommendations of different D_i's is by imposing $\frac{e_i}{\bar{e}_i} \geq \frac{e_j}{\bar{e}_j}$ only when a majority of indicators favor country i over j. An example here might be useful: suppose D^1, D^2, D^3 denote GDP per capita, historical emissions, and current emissions, respectively. An extension of Property 3 to this multidimensional setting might be the following: set $\frac{e_i}{\bar{e}_i} \geq \frac{e_j}{\bar{e}_j}$ if $D_i^k \geq D_j^k$ for at least two $k \in \{1, 2, 3\}$. Another solution still might be to consider an aggregate index of equity-relevant indicators (an example here might be the HDI index) and order country mitigation targets on that basis.

To be sure, all of the above operations, though reasonable, are completely ad-hoc. It is thus unwise to assert that they resolve the underlying issue. At the same time, coming up with a rigorous, theoretically grounded way of adjudicating the relative importance of different indicators (and the ethical stances they codify via Property 3) is far from obvious, which means that such rules-of-thumb might be, at least for now, the best one can hope for.

Links to the equality-of-opportunity framework. The relation between Property 3 and the more sophisticated equality-of-opportunity frameworks of Roemer [35], Fleurbaey [19] and others, is not immediate. Unlike their work, the proposed model does not explicitly distinguish between between relevant and irrelevant characteristics, nor does it clearly delineate how such characteristics affect payoffs and outcomes. Fragments of these ideas are present, but only in the background. In what follows we attempt to illustrate how the proposed criterion might connect to the responsibility/compensation framework.

In most cases, irrelevant characteristics would include agents' benefit and damage functions, as they are (at least in the short term) out of their control. For example, if a country has high climate-related damages due to its geography, it (arguably) should not be held responsible for its bad luck and some form of compensation is called for. Conversely, relevant characteristics would likely include status-quo (i.e., $\tilde{\boldsymbol{e}}$), and historical (i.e. cumulative) emissions, as well as other equity-relevant D_i indicators that are used to determine the relative rank of mitigation targets.[7] Such indicators might include GDP per capita or other indicators of economic development, since the pursuit of economic growth often involves the emission of greenhouse gases and countries tend to be aware of the environmental effect of their economic activity. Within the context of the compensation/responsibility framework, we could point to all of the above indicators as relevant characteristics that countries should be held responsible for. Accordingly, a greater degree of responsibility is codified in the form of a lower emissions ratio $\frac{e_i}{\tilde{e}_i}$, i.e., a more ambitious mitigation target.

The above being said, the relation between equity-relevant indicators and relevant characteristics is far from exact. Some indicators might lend themselves to ambiguous interpretations while others might be more clearly problematic.[8] To sum up, the connection of Property 3 to the responsibility/compensation framework, though present, is not particularly sharp or well-articulated.

3 Characterizing the set of Pareto efficient allocations

In this section we introduce a more operational definition for Pareto efficiency (PE) than the one provided in Property 1. This can be accomplished via the technique of *scalarization*, a cornerstone of the multicriteria-optimization literature. Scalarization assigns a non-negative weight w_i to each agent i's utility and maximizes the resulting weighted sum of utilities. If all the weights are positive, the optimal solution will be PE. Moreover, since the problem is convex, all PE allocations implying positive emissions for

[7] This assessment is complicated by the fact that compensation will be received and disbursed by the current population, which is not responsible for the status-quo and historical emissions.

[8] Consider for example the benefits approach (iv) described earlier: we would be hard-pressed to consider a country's vulnerability to climate change, ostensibly due to geography or historical events, as a relevant characteristic for which it must be held responsible.

all agents can be produced in this way (see Section 4.7.4 of Boyd and Vandenberghe [9]).[9]

So let us use scalarization to study the set of positive PE allocations. Given a weight vector $\boldsymbol{w} > \mathbf{0}$ consider the following optimization problem:

$$\max_{\boldsymbol{e} \geq \mathbf{0}} \quad \sum_{i=1}^{I} w_i \left(b_i(e_i) - c_i(e) \right) \tag{1}$$

The strictly concave benefit functions b_i are increasing up to point $\bar{e}_i$, after which they become decreasing. If an agent i's benefit function b_i is increasing in addition to strictly concave, then this point is trivially defined to be infinity. The strict concavity of the b_i's and the fact that the c_i's are increasing imply that, at optimality,

$$e_i < \bar{e}_i, \quad i = 1, ..., I \tag{2}$$

Applying the KKT conditions to problem (1) and assuming an interior optimum yields the following necessary and sufficient condition for optimality:

$$b_i'(e_i) = \frac{\sum_{j=1}^{I} w_j c_j'(e)}{w_i}. \tag{3}$$

Assuming optimization problem (1) yields an interior optimal solution, this solution is unique and corresponds to a PE allocation.

The special structure of problem (1) allows us to go further and obtain a sharp characterization of PE allocations.

Proposition 1

(i) An allocation $\mathbf{e} > \mathbf{0}$ *is Pareto efficient if and only if*

$$\sum_{i=1}^{I} \frac{c_i'(e)}{b_i'(e_i)} = 1.$$

(ii) Given a Pareto efficient allocation $\mathbf{e} > \mathbf{0}$*, weight vectors* $\boldsymbol{w}$ *satisfying*

$$b_1'(e_1)w_1 = ... = b_I'(e_I)w_I$$

are such that problem (1) with this choice of weights has $\mathbf{e}$ *as its unique optimal solution.*

[9] Note that this procedure will not produce all PE allocations. Scalarizing with all *non-negative* weight vectors ($\boldsymbol{w} \geq \mathbf{0}$) will [9], but we are not interested in PE allocations that are derived by assigning zero weight to some agents –as they will imply zero emissions for those agents, and will thus tend to conflict with individual rationality and fairness.

Proof. First we prove (i). Suppose $\boldsymbol{e} > \mathbf{0}$ is Pareto efficient. Then, there must exist a a weight vector $\boldsymbol{w} > \mathbf{0}$ such that Eq. (3) is satisfied. Manipulating this equation yields:

$$b_i'(e_i) = \frac{\sum_{j=1}^{I} w_j c_j'(e)}{w_i}, \;\; \forall i = 1, ..., I \;\; \Rightarrow \;\; \frac{b_i'(e_i)}{c_i'(e)} = \frac{\sum_{j=1}^{I} w_j c_j'(e)}{c_i'(e) w_i}, \;\; \forall i = 1, ..., I$$
$$\Rightarrow \;\; \sum_{i=1}^{I} \frac{c_i'(e)}{b_i'(e_i)} = 1. \tag{4}$$

Suppose now $\boldsymbol{e} > \mathbf{0}$ satisfies $\sum_{i=1}^{I} \frac{c_i'(e)}{b_i'(e_i)} = 1$. Consider a weight vector $\boldsymbol{w} \geq \mathbf{0}$ that satisfies $b_i'(e_i)w_i = b_j'(e_j)w_j$ for all pairs of agents i, j. Then, for every $i = 1, ..., I$:

$$\sum_{j=1}^{I} w_j c_j'(e) = w_i b_i'(e_i) \frac{c_i'(e)}{b_i'(e_i)} + \sum_{j \neq i} \frac{b_i'(e_i) w_i}{b_j'(e_j)} c_j'(e) = w_i b_i'(e_i) \underbrace{\sum_{j=1}^{I} \frac{c_j'(e)}{b_j'(e_j)}}_{=1} = w_i b_i'(e_i). \tag{5}$$

Thus, Eq. (3) is satisfied for all $i = 1, ..., I$, for this combination of $\boldsymbol{e}$ and $\boldsymbol{w}$. Hence, $\boldsymbol{e}$ is the unique optimal solution of problem (1) for this choice of positive weights and thus Pareto efficient. Part (ii) follows. ■

Proposition 1 provides a compact and easily verifiable condition for Pareto efficiency. Given any allocation $\boldsymbol{e} > \mathbf{0}$, to determine whether it is PE it is sufficient to compute $\sum_{i=1}^{I} \frac{c_i'(e)}{b_i'(e_i)}$. If this quantity equals 1, then $\boldsymbol{e}$ is Pareto efficient; otherwise it is not. We will make use of this result extensively.

4 Main Results

As stated in the Introduction the main question we wish to answer is whether, and under what conditions, the three properties of Pareto efficiency (PE), individual rationality (IR), and fairness (F) are compatible. Note that this problem is well-posed since (i) PE and IR are compatible[10], (ii) IR and F are compatible by choosing $\boldsymbol{e} = \tilde{\boldsymbol{e}}$ and (iii) PE and F are seen to be compatible by considering the allocation that maximizes the utility of the agent with the lowest D_i (i.e., agent 1) and setting $e_j = 0$ for all $j \neq 1$.[11]

We make the following, arguably reasonable, assumption for technical convenience. It is likely to hold in most practical instances of common-resource management.

Assumption 1 *If allocation* $\mathbf{e}$ *is individually rational, then* $\mathbf{e} > \mathbf{0}$.

[10] if $\tilde{\boldsymbol{e}}$ is undominated it is PE, if not then there exists a PE allocation that dominates it and thus is IR.

[11] It is possible, if slightly more involved, to produce allocations satisfying PE and F such that $e_i > 0$ for all i. Details available upon request.

In light of Assumption 1 and Proposition 1, the existence of an allocation satisfying PE, IR and F is equivalent to the existence of a positive solution to the following system of nonlinear inequalities and equalities:

$$u_i(\boldsymbol{e}) \geq u_i(\tilde{\boldsymbol{e}}), \quad i = 1, ..., I \tag{6}$$

$$\frac{e_{i_1}}{\tilde{e}_{i_1}} \geq ... \geq \frac{e_{i_I}}{\tilde{e}_{i_I}} \tag{7}$$

$$\sum_{i=1}^{I} \frac{c_i'(e)}{b_i'(e_i)} = 1. \tag{8}$$

The constraints of Eqs. (6)-(7), defining IR and F allocations, are convex and thus theoretically and computationally tractable. By contrast, the equality constraint (8) defining positive PE allocations will, with the exception of some special cases, be nonlinear and nonconvex. This means that we cannot embed it as a constraint in an optimization problem and maintain tractability [9].

4.1 Conditions for existence

We begin by demonstrating that, in general, the three properties of PE, IR and F are incompatible. This impossibility holds even if we restrict $\tilde{\boldsymbol{e}}$ to be a Nash equilibrium.

Theorem 1 *There exist problem instances in which Pareto efficiency, individual rationality and fairness are incompatible. This holds even if we constrain $\tilde{\boldsymbol{e}}$ to be a Nash equilibrium.*

Proof. Let $I = 2$, $b_i(e_i) = \ln(e_i)$ for $i = 1, 2$, and $c_1(e) = 5e^2$ and $c_2(e) = e^2$. Thus, both agents perceive the same benefits from emissions but agent 1 experiences five times higher damages. In addition, suppose $D_1 \leq D_2$.

The status-quo allocation is assumed to be the Nash equilibrium of the associated game. The equilibrium conditions are given by

$$\frac{1}{\tilde{e}_1} = 10\tilde{e}$$
$$\frac{1}{\tilde{e}_2} = 2\tilde{e},$$

yielding

$$\tilde{e}_1 = 0.1291, \quad \tilde{e}_2 = 5\tilde{e}_1 = 0.6455.$$

Suppose $\boldsymbol{e}$ satisfies fairness (F) and individual rationality (IR).

By F,

$$\frac{e_1}{\tilde{e}_1} \geq \frac{e_2}{\tilde{e}_2} \Rightarrow \frac{e_1}{\tilde{e}_1} \geq \frac{e_2}{5\tilde{e}_1} \Rightarrow e_1 \geq \frac{e_2}{5}. \tag{9}$$

By IR,

$$
\begin{aligned}
& \log(e_2) - e^2 \geq \log(\tilde{e}_2) - \tilde{e}^2 \\
\Rightarrow \quad & \log(e_2) \geq \log(\tilde{e}_2) + e^2 - \tilde{e}^2 \\
\overset{(9)}{\Rightarrow} \quad & \log(e_2) \geq \log(\tilde{e}_2) + \left(\frac{6}{5}e_2\right)^2 - \tilde{e}^2 \\
\Rightarrow \quad & \log(e_2) - \frac{36}{25}e_2^2 \geq \log(\tilde{e}_2) - \tilde{e}^2 = -1.0377 \\
\Rightarrow \quad & e_2 \in [0.5348, 0.6454] \Rightarrow e_2 \geq 0.5348. \qquad (10)
\end{aligned}
$$

Eqs. (9)- (10) imply

$$e_1 \geq 0.1070. \qquad (11)$$

Consequently,

$$\sum_{i=1}^{2} \frac{c_i'(e)}{b_i'(e_i)} = 10e \cdot e_1 + 2e \cdot e_2 \overset{(10)-(11)}{\geq} 1.3726.$$

By Proposition 1, $\boldsymbol{e}$ fails Pareto efficiency. ■

The proof of Theorem 1 illustrates how PE, IR and F can occasionally lead to an irreconcilable tension. In the example that is explored, one agent experiences much lower damages than the other. In equilibrium, this leads to them emitting much more (five times more) and enjoying much higher utility than the other agent. Meanwhile, fairness requires that the low-damage/high-polluting agent undertake a greater proportional reduction in emissions with respect to the Nash equilibrium status-quo.

The main driver behind the impossibility result is agent 2's IR constraint combined with the fairness requirement. Put together, they imply a lower bound on the emissions of the low-damage agent that is quite high. Combining this lower bound with the fairness requirement yields another lower bound, this time on the emissions of the other agent. Putting these bounds together yields emissions that are collectively too high to satisfy Pareto efficiency.

While it is not possible to guarantee that a PE, IR and F allocation always exists, there are many instances in which it does. However, providing a better understanding of this point is complicated by the nonlinearity of Eq. (8).

Fortunately, there is a way to address the existence question that bypasses the challenges of Eq. (8). Let us introduce the following optimization problem[12], which we denote by ENV:

$$
\begin{aligned}
ENV = \quad & \min_{\boldsymbol{e}} \quad e \\
& \text{s.t.} \quad (6) - (7)
\end{aligned}
$$

[12] Recall that, for an allocation $\boldsymbol{e}$, the scalar e denotes aggregate emissions: $e = \sum_{i=1}^{I} e_i$.

The problem ENV searches for the IR-F allocation that minimizes aggregate emissions. As such, this allocation represents a IR-F constrained "environmental" optimum. ENV is a convex optimization problem and so is analytically and computationally easy to solve. Moreover, it has a structure that allows us to make sharp statements about its optimum. The result of Lemma 1 will prove very useful later on.

Lemma 1 *Optimization problem* ENV *has a unique solution* $\boldsymbol{e^{min}}$. *Moreover,* $\boldsymbol{e^{min}} \leq \boldsymbol{e}$ *for all* $\boldsymbol{e}$ *satisfying IR and F.*

Proof. For ease of exposition, and without loss of generality, suppose $D_1 \leq D_2 .... \leq D_I$. Suppose $\boldsymbol{e^{min}}$ is an optimal solution of ENV. It will therefore satisfy the bounds (2) for all $i = 1, ..., I$. Suppose $\hat{\boldsymbol{e}}$ is another feasible solution of ENV such that there exists a set of indices $\mathcal{I}^- \subset \{1, 2, ..., I\}$ such that $\hat{e}_j < e_j^{min}$ for all $j \in \mathcal{I}^-$.

Let $i = \max\{k : k \in \mathcal{I}^-\}$. Since $\hat{\boldsymbol{e}}$ is not necessarily optimal for ENV, $\hat{e} \geq e^{min}$, which in turn implies $c_i(\hat{e}) \geq c_i(e^{min})$. Moreover, $\hat{e}_i < e_i^{min} < \bar{e}_i$ implies $b_i(\hat{e}_i) < b_i(e_i^{min})$, so that

$$u_i(\tilde{\boldsymbol{e}}) \leq b_i(\hat{e}_i) - c\left(\hat{e}\right) < b_i(e_i^{min}) - c\left(e^{min}\right).$$

Consider the allocation $\boldsymbol{e'} = (\boldsymbol{e^{min}_{-i}}, e_i^{min} - \epsilon)$, where $0 < \epsilon \leq e_i^{min} - \hat{e}_i$. For any choice of ϵ in that range, $\boldsymbol{e'}$ will satisfy IR and will result in a lower objective function value than $\boldsymbol{e^{min}}$. To avoid a contradiction, there must exist a fairness constraint involving i that prevents it from becoming smaller: i.e., there must exist $j > i$ such that

$$\frac{e_i^{min}}{\tilde{e}_i} = \frac{e_j^{min}}{\tilde{e}_j}.$$

This means that

$$\frac{\hat{e}_i}{\tilde{e}_i} < \frac{e_i^{min}}{\tilde{e}_i} = \frac{e_j^{min}}{\tilde{e}_j}. \tag{12}$$

On the other hand, fairness implies

$$\frac{\hat{e}_i}{\tilde{e}_i} \geq \frac{\hat{e}_j}{\tilde{e}_j}. \tag{13}$$

Eqs. (12)-(13) together yield

$$\frac{\hat{e}_j}{\tilde{e}_j} < \frac{e_j^{min}}{\tilde{e}_j} \Rightarrow \hat{e}_j < e_j^{min},$$

a contradiction since $j > i$ and $i = \max\{k : k \in \mathcal{I}^-\}$. ■

We now make the following assumption regarding status-quo emissions.

Assumption 2 *The status-quo allocation* $\tilde{\boldsymbol{e}}$ *satisfies* $\sum_{i=1}^{I} \frac{c_i'(\tilde{e})}{b_i'(\tilde{e}_i)} > 1$.

Assumption 2 implies that the status-quo allocation is Pareto inefficient because it results in emissions that are too high. It is a natural assumption to make in the context of common resource management, where self-interested rational behavior often leads to over-exploitation. From an economic-theory standpoint, if $\tilde{\boldsymbol{e}}$ is an interior Nash equilibrium, then by definition $b_i'(\tilde{e}_i) - c_i'(\tilde{e}) = 0$ for all $i = 1, ..., n$, implying that $\sum_{i=1}^{I} \frac{c_i'(\tilde{e})}{b_i'(\tilde{e}_i)} = n > 1$. The inefficiency of such equilibria is often referred to as the "tragedy of the commons" [23].

We are now ready to prove the main existence result.

Theorem 2 *Suppose Assumptions 1-2 hold. There exists an allocation satisfying Pareto efficiency, individual rationality, and fairness if any only if the (unique) optimal solution of problem ENV, $\boldsymbol{e^{min}}$, satisfies*

$$\sum_{i=1}^{I} \frac{c_i'(e^{min})}{b_i'(e_i^{min})} \leq 1.$$

Proof. Define the function $g : [0, \bar{e}_1] \times ... \times [0, \bar{e}_I] \mapsto \Re$ such that

$$g(\boldsymbol{x}) = \sum_{i=1}^{I} \frac{c_i'\left(\sum_{i=1}^{I} x_i\right)}{b_i'(x_i)}. \tag{14}$$

This function is positive, continuous and increasing in x_i for all i. By Lemma 1, the allocation $\boldsymbol{e^{min}}$ uniquely attains the minimum of $g(\cdot)$ over the intersection of IR and F allocations:

$$g(\boldsymbol{e^{min}}) < g(e), \ \ \forall \boldsymbol{e} \text{ that satisfy Eqs.(6)-(7).}$$

Thus

$$g(\boldsymbol{e^{min}}) > 1 \Rightarrow g(\boldsymbol{e}) > 1, \ \ \forall \boldsymbol{e} \text{ that satisfy Eqs.(6)-(7),}$$

which, together with Proposition 1, implies that there exists no individually rational and fair allocation satisfying Pareto efficiency. This establishes necessity.

Suppose now that $g(\boldsymbol{e^{min}}) \leq 1$. If $g(\boldsymbol{e^{min}}) = 1$ then by Proposition 1 $\boldsymbol{e^{min}}$ is Pareto efficient and we are done. Suppose instead that $g(\boldsymbol{e^{min}}) < 1$. By Assumption 2, the status-quo allocation, $\tilde{\boldsymbol{e}}$, satisfies $g(\tilde{\boldsymbol{e}}) > 1$. Consider the following parametric set of allocations

$$\boldsymbol{e}(\alpha) = (1 - \alpha)\boldsymbol{e^{min}} + \alpha\tilde{\boldsymbol{e}}, \ \ \alpha \in [0, 1].$$

The constraints of Eqs. (6)-(7), defining IR and F allocations respectively, give rise to convex sets. Thus, their intersection will also be convex. Consequently, since both $\tilde{\boldsymbol{e}}$ and $\boldsymbol{e^{min}}$ satisfy IR and F, so will $\boldsymbol{e}(\alpha)$ for any $\alpha \in [0, 1]$. Moreover, as g is continuous and increasing in every argument, the function $h : [0, 1] \mapsto \Re_+$ such that

$$h(\alpha) = g(\boldsymbol{e}(\alpha))$$

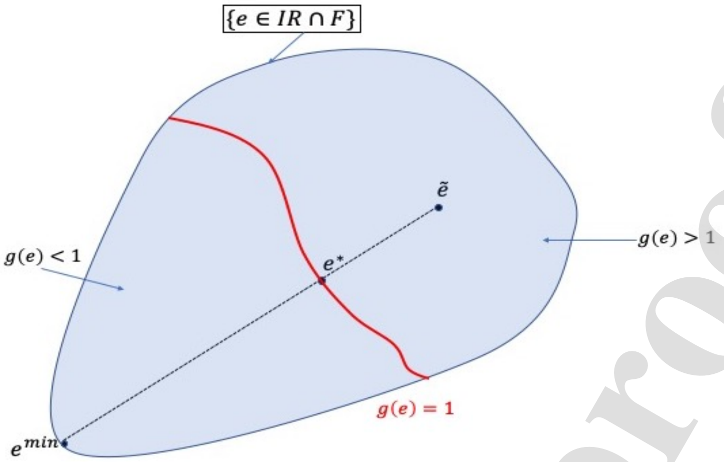

Figure 1: Illustration of the proof of Theorem 2. Note how Lemma 1 implies that $\boldsymbol{e}^{\boldsymbol{min}} \leq \boldsymbol{e}$ for all $\boldsymbol{e}$ satisfying individual rationality and fairness.

is continuous and increasing (recall that, by Lemma 1, $\boldsymbol{e}^{\boldsymbol{min}} \leq \tilde{\boldsymbol{e}}$). Since $h(0) < 1$ and $h(1) > 1$, by the Intermediate Value Theorem there exists a unique $\alpha^* \in (0,1)$ such that $h(\alpha^*) = 1$. Hence, the allocation

$$\boldsymbol{e}^* \equiv \boldsymbol{e}(\alpha^*) = \alpha^* \tilde{\boldsymbol{e}} + (1-\alpha^*)\boldsymbol{e}^{\boldsymbol{min}} \tag{15}$$

will be PE, in addition to satisfying IR and F. Moreover, it will result in aggregate emissions that are strictly lower than those of the status-quo allocation:

$$e^* = \sum_{i=1}^{I} e_i^* = (1-\alpha^*)e^{min} + \alpha^* \tilde{e} < \tilde{e}.$$

■

The proof of Theorem 2 is such that, when allocations satisfying the desired properties exist, we actually exhibit one. This allocation, summarized in Eq. (15), has a simple structure: it is a convex combination of the status-quo allocation and the allocation that minimizes aggregate emissions subject to IR and F. The theorem's assumptions imply that there will exist exactly one allocation in the convex hull of $\tilde{\boldsymbol{e}}$ and $\boldsymbol{e}^{\boldsymbol{min}}$ that achieves PE.

The proof of Theorem 2 establishes the following corollary.

Corollary 1 *Suppose Assumptions 1-2 hold and the optimal solution of problem ENV,* $\boldsymbol{e}^{\boldsymbol{min}}$*, satisfies* $\sum_{i=1}^{I} \frac{c_i'(e_i^{min})}{b_i'(e_i^{min})} \leq 1$*. There exists an allocation* $\boldsymbol{e}$ *satisfying Pareto efficiency, Individual Rationality, and Fairness which satisfies* $e < \tilde{e}$*.*

Remarks. The allocation $\boldsymbol{e^{min}}$ plays a prominent role in the existence of PE-IR-F allocations. As mentioned earlier, this allocation uniquely minimizes aggregate emissions subject to IR and F. It represents a constrained first-best outcome from an environmental standpoint, and will almost always violate Pareto efficiency. It turns out that the *way* in which this violation occurs is crucial to the existence question. Existence will hold if and only if $\boldsymbol{e^{min}}$ implies emissions that are inefficiently *low*, in a precise sense.

Let us elaborate on the above statement. Consider an allocation $\boldsymbol{e}$ satisfying IR and F and the quantity $g(\boldsymbol{e}) = \sum_{i=1}^{I} \frac{c_i'(e)}{b_i'(e_i)}$, which, similarly to the status-quo outcome $\tilde{\boldsymbol{e}}$, is assumed to be greater than 1. Focus on agent i and the corresponding term in the summation of $g(\boldsymbol{e})$ and suppose (to avoid trivialities) that $\frac{c_i'(e)}{b_i'(e_i)} \leq 1$. This ratio ranges from 0 to 1 and provides a measure of the closeness of agent i from her best-response level of emissions, given the emissions of all other agents. The greater this ratio, the closer agent i is to her best-response emissions. To reach PE, some (though not necessarily all) of these ratios must become smaller so that they all together sum to 1. Thus, we need to move away from $\boldsymbol{e}$ to a different allocation where at least one agent is emitting less.

One way of accomplishing this, suggested by the proof of Theorem 2, is to move in the direction of $\boldsymbol{e^{min}}$, the least polluting IR-F allocation. If $g(\boldsymbol{e^{min}}) < 1$, then, as we get closer and closer to $\boldsymbol{e^{min}}$, we are guaranteed to cross the Pareto frontier. Conversely, if $g(\boldsymbol{e^{min}}) > 1$, then this procedure will not work: we will keep approaching the Pareto frontier but never actually reach it. More importantly, starting from any IR-F allocation, if $g(\boldsymbol{e^{min}}) > 1$, then there is simply no way to lower and rearrange emissions to a Pareto efficient level without at some point violating IR or F.

4.2 (Non)uniqueness and the selection problem

Now let us examine the uniqueness of PE-IR-F allocations. The following result suggests that uniqueness will very rarely hold.

Proposition 2 *Suppose Assumptions 1-2 hold and consider the optimal solution of problem ENV,* $\mathbf{e^{min}}$. *There exists a unique allocation satisfying Pareto efficiency, individual rationality, and fairness if and only if* $\mathbf{e^{min}}$ *is Pareto efficient, i.e.,* $\sum_{i=1}^{I} \frac{c_i'(e^{min})}{b_i'(e_i^{min})} = 1$.

Proof. Before we begin recall the notation $g(\boldsymbol{e}) = \sum_{i=1}^{I} \frac{c_i'(e)}{b_i'(e_i)}$.

First we show sufficiency. If $\boldsymbol{e^{min}}$ is Pareto efficient then $g(\boldsymbol{e^{min}}) = 1$. By Lemma 1, all other allocations satisfying IR and F will be such that $g(\boldsymbol{e}) > 1$ and so will fail PE. Thus, $\boldsymbol{e^{min}}$ is the unique PE-IR-F allocation.

Now we address necessity. Suppose there exists a unique allocation satisfying PE-

IR-F. Thus, there exists a unique IR-F allocation $\boldsymbol{e}^*$ satisfying $g(\boldsymbol{e}^*) = 1$. By Lemma 1, $g(\boldsymbol{e}^{\boldsymbol{min}}) \leq g(\boldsymbol{e}^*) = 1$. The convexity of the set of IR-F allocations implies that all other IR-F allocations $\boldsymbol{e} \neq \boldsymbol{e}^*$ satisfy either (i) $g(\boldsymbol{e}) < 1$ or (ii) $g(\boldsymbol{e}) > 1$. Case (i) contradicts Assumption 2. Case (ii) implies $\boldsymbol{e}^* = \boldsymbol{e}^{\boldsymbol{min}}$. ■

Assuming existence holds, Proposition 2 establishes that, unless the IR-F environmental optimum $\boldsymbol{e}^{\boldsymbol{min}}$ is PE, there will be a multiplicity of PE-IR-F solutions. The likely non-uniqueness of PE-IR-F allocations introduces the problem of how to select from a set of such allocations. Unfortunately, the nonconvexity of the PE constraint precludes neat analytical approaches. We thus need to explore alternative paths.

The proof of Theorem 2 suggests a way forward by focusing on function $g(\cdot)$ as defined in Eq. (14). It is the following. First, we generate two sets of IR-F allocations, let's call them $\mathcal{S}^-$ and $\mathcal{S}^+$. An allocation $\boldsymbol{e}$ belonging to $\mathcal{S}^-$ (resp. $\mathcal{S}^+$) is IR-F and satisfies $g(\boldsymbol{e}) < 1$ (resp. > 1). Subsequently, for every pair of allocations $(\boldsymbol{e}^-, \boldsymbol{e}^+) \in \mathcal{S}^- \times \mathcal{S}^+$ we compute $a \in [0, 1]$ such that $g\left(a\boldsymbol{e}^- + (1-a)\boldsymbol{e}^+\right) = 1$. By the Intermediate Value Theorem such a value of a must exist. The allocation $a\boldsymbol{e}^- + (1-a)\boldsymbol{e}^+$ will thus satisfy PE-IR-F. Provided there are no duplicates, this procedure, graphically depicted in Figure 2, will produce $|\mathcal{S}^-| \cdot |\mathcal{S}^+|$ PE-IR-F allocations.

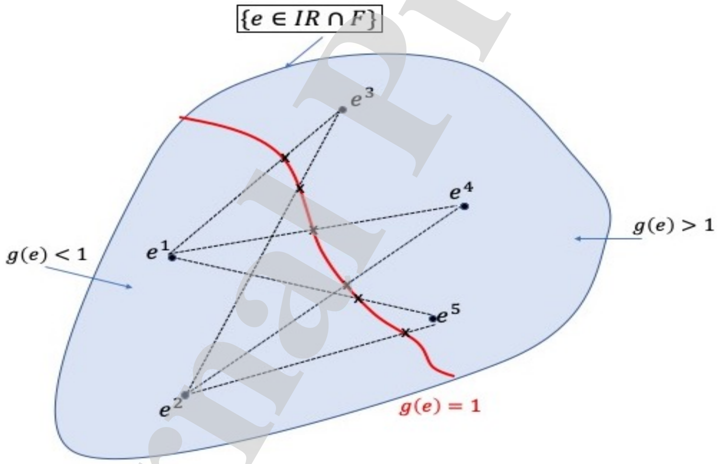

Figure 2: Generating PE-IR-F allocations. Here, $\mathcal{S}^- = \{\boldsymbol{e}^1, \boldsymbol{e}^2\}$ and $\mathcal{S}^+ = \{\boldsymbol{e}^3, \boldsymbol{e}^4, \boldsymbol{e}^5\}$, leading to $2 * 3 = 6$ PE-IR-F allocations, each indicated by an x.

4.3 Comments

Asymmetry. Setting the technical details aside, the take-home message of Theorem 2 is that, if even the most environmentally favorable IR-F allocation results in emissions that are collectively too high, then there is simply no way of reconciling PE with IR and F. What factors might compel $\boldsymbol{e^{min}}$ to have this feature? It is difficult to provide a sharp answer, but the proof of Theorem 1 offers some possible, albeit speculative, clues.

Recall that the example that was used featured significant asymmetry in agents' damage functions. One agent was simply much more affected by the pollution externality than the other. This lead to a lopsided status-quo in which the low-damage agent emitted much more, and had much higher utility, than the high-damage one. Within this setting, fairness (quite reasonably) required that the high-damage agent bear a smaller share of the mitigation needed to move to an efficient outcome. This requirement complicated the participation of the low-damage agent, setting a lower bound on the emissions that made joining the agreement profitable for her. Consequently, fairness implied a lower bound on the high-damage agent's emissions. Combining these two bounds implied that all IR-F allocations would yield inefficiently high amounts of pollution.

In this specific case, the driver behind the incompatibility of PE-IR-F is the asymmetry in damages. This can be seen by varying the damage function of the high-damage agent and applying the machinery of Theorem 2. To wit, consider the exact same setting as that of Theorem 1 with the only difference that the damage function of agent 1 is parameterized to $c_1(e) = c \cdot e^2$, where $c > 0$. The parameter c functions as a measure of asymmetry: the closer it is to 1, the more similar agent 1's damages are to agent 2's. Varying c and solving the resulting ENV optimization problem, we observe that smaller levels of asymmetry allow us to get closer and closer to existence.

This is illustrated in Table 1, where we list the values of $\tilde{\boldsymbol{e}}$, $\boldsymbol{e^{min}}$, and $g(\boldsymbol{e^{min}})$ for different values of c.[13] Starting from the case of $c = 5$ that was used in the proof of Theorem 1, we see that $g(\boldsymbol{e^{min}})$ is decreasing in c. In addition, values of $c \in \{3, 4, 5\}$ yield $g(\boldsymbol{e^{min}}) > 1$ and so exclude the compatibility of PE-IR-F, whereas the opposite is true when $c \in \{1, 2\}$. When $c = 2.59$ we have $g(\boldsymbol{e^{min}}) = 1.0003$ and so, for all practical intents and purposes, we can say that the cutoff between existence and non-existence occurs at that level of asymmetry. Moreover, when $c = 2.59$, Proposition 2 implies that $\boldsymbol{e^{min}}$ will uniquely satisfy PE-IR-F.

The role of asymmetry in hindering the existence of PE-IR-F allocations is consistent with basic results in the IO literature on cartel formation. Indeed, it is well-known that collusive agreements tend to be easier to sustain when the parties are more similar to each other, in their utility functions or otherwise (Cabral [10]). At the same time, it is worth noting that some papers in the strategic IEA literature reach different conclusions on the effect of asymmetry. For example, McGinty [31] and Finus and McGinty [18] find that, in certain contexts, the presence of asymmetric agents can facilitate the formation of environmental agreements by increasing the gains to cooperation.

[13] Recall that $\tilde{\boldsymbol{e}}$ is taken to be the Nash equilibrium.

c	$(\tilde{e}_1, \tilde{e}_2)$	(e_1^{min}, e_2^{min})	$g(\boldsymbol{e^{min}})$	Existence?
5	(0.1291,0.6455)	(0.1070,0.5348)	1.3726	No
4	(0.1581,0.6325)	(0.1254,0.5014)	1.2573	No
3	(0.2041,0.6124)	(0.1508,0.4523)	1.0912	No
2	(0.2887,0.5774)	(0.1865,0.3729)	0.8344	Yes
1	(0.5000,0.5000)	(0.2254,0.2254)	0.4064	Yes

Table 1: Examining the existence of PE-IR-F allocations for different levels of damage asymmetry c (in the setting of the proof of Theorem 1). Computations performed in Matlab.

As far as our framework is concerned, we suspect that agent heterogeneity will tend to complicate the existence of PE-IR-F agreements. However, finding a measure of asymmetry that allows for a systematic investigation of the above claim is not straightforward.

Implementation. An implicit assumption that drives the analysis of this Section is perfect information on agent benefit and cost functions. While this assumption is commonplace in the literature on international environmental agreements, it is far from trivial. Indeed, precise knowledge regarding the structure of agent benefits and costs is unlikely to be available to a central planner, and so must either be estimated or elicited. Notable exceptions to this trend are Helm and Wirl [26] (working with a two-agent, principal-agent model), and especially Martimort and Sand Zantman [30] who apply mechanism design theory to the IEA context. These papers assume that certain parameters of the benefit and damage functions are unknown and need to be elicited. A central planner offers contracts that assign levels of mitigation and transfers as a function of declared agent types. The contracts are designed to satisfy constraints on incentive compatibility, participation and budget balance. In a linear-quadratic setting, Martimort and Sand Zantman [30] show that the optimal mechanism can be approximated by a simple two-item menu, consisting of combinations of upfront contributions to a climate fund and linear subsidies on mitigation. This positive result is however compromised by limitations involving enforcement and commitment.

5 Application to Global Climate Policy

In this section we provide a proof of concept of our theoretical findings. Our application focus is global climate-change policy and we work within the framework of Nordhaus [32].

In [32] Nordhaus developed a static version of his well-known integrated assessment model DICE. This new model was named C-DICE and its main function was to study the effect of climate clubs and trade tariffs as a way of mitigating emissions. For the purposes of our work, we will focus squarely on Nordhaus's model and ignore the trade and climate-club dimension.

In Nordhaus's framework, there are 15 global regions. Each region i's utility is given by the function:

$$u_i(\boldsymbol{e}) = Q_i - A_i(e_i) - d_i(e) \tag{16}$$

where Q_i is region i's output (i.e., GDP), A_i is its abatement cost and d_i its climate damages. As before, the initial allocation of emissions is denoted by $\tilde{\boldsymbol{e}}$. Taking this into account, abatement costs are given by the expression

$$A_i(e_i) = \alpha_i Q_i \left(\frac{e_i - \tilde{e}_i}{\tilde{e}_i} \right)^2 .$$

Here, $\alpha_i > 0$ is a parameter measuring the costliness of emissions reductions. Climate damages are assumed to be linear in aggregate emissions so that

$$d_i(e) = \gamma_i e.$$

Here $\gamma_i > 0$ is region i's social cost of carbon. The linearity of the damage functions is justified in Nordhaus [32] by appealing to the static nature of the model. Translating the above functions into the notation of the previous section, we write

$$b_i(e_i) = Q_i - \alpha_i Q_i \left(\frac{e_i - \tilde{e}_i}{\tilde{e}_i} \right)^2 \tag{17}$$

$$c_i(e) = \gamma_i e \tag{18}$$

Table 2 summarizes information on all parameter values.

We apply the results of Sections 3 and 4 to derive a Pareto efficient, individually rational and fair allocation in this setting. In doing so, we employ a criterion of fairness that is a special case of Property 3, in which we assume that D_i is equal to GDP per capita of region i (an instance of the capability approach). Thus, regions with higher GDP per capita are required to undertake higher relative emissions reductions with respect to the status-quo. This fairness criterion, whereby richer countries are asked to undertake a proportionally greater amount of abatement, is commonly discussed in global climate-change negotiations [38, 27].

We proceed by considering optimization problem ENV, with benefit and cost functions given by Eqs. (17)-(18), and all parameter values as indicated in Table 2. We solve the resulting version of ENV in Matlab using the nonlinear solver fmincon.[14]

[14] We need to try a few different starting points before the solver converges to an optimum. Specifically, we solve the problem 4-5 times, each time inserting the current local optimum as the solver's updated starting point.

Region i	$\tilde{e}_i$ (ton CO_2)	Q_i (GDP$_i$) (US \$)	α_i (scalar)	γ_i (\$/CO_2)
South Saharan Africa	234,646,913	2,075,769,196,150	0.00587	0.572
India	2,049,561,902	5,962,906,305,677	0.03011	1.643
Rest of World	1,552,161,782	6,235,641,996,372	0.01048	1.718
China	8,293,771,000	13,496,409,330,000	0.05003	3.718
Eurasia	877,792,391	1,434,179,207,149	0.04923	0.395
South Africa	458,061,282	614,313,024,090	0.04142	0.169
Latin America	1,204,677,235	5,119,453,985,952	0.01701	1.410
Brazil	429,462,339	2,816,369,351,334	0.00599	0.776
South East Asia	1,764,016,979	5,787,020,468,419	0.01665	1.594
Middle East	2,007,397,016	5,733,919,629,177	0.04246	1.580
Russia	1,737,103,381	3,226,527,302,200	0.04764	0.889
European Union	3,718,923,879	16,906,105,087,184	0.01924	4.657
Japan	1,172,544,223	4,386,177,677,532	0.02523	1.208
Canada	499,877,528	1,419,490,125,740	0.03512	0.391
United States	5,444,142,792	15,533,948,728,220	0.02721	4.279

Table 2: Data and calibrated parameter values used in Nordhaus [32]. Status-quo emissions and GDP data refer to year 2011.

The allocation $\boldsymbol{e^{min}}$ is exhibited in Table 3. Applying Eq. (14) to the benefit and damage functions of Nordhaus yields

$$g(\boldsymbol{e}) = \sum_{i=1}^{15} \frac{\tilde{e}_i^2 \gamma_i}{2(\tilde{e}_i - e_i)\alpha_i Q_i}.$$

This leads to values of $g(\boldsymbol{e^{min}}) = 0.8200 < 1.$ and $g(\tilde{\boldsymbol{e}}) = \infty$. Thus, Assumption 2 is satisfied as is the condition of Theorem 2. Assumption 1 is also satisfied. Thus, Theorem 2 is applicable and the existence of PE-IR-F allocations ensured. In addition, Proposition 2 implies that such allocations will not be unique.

In line with the proof of Theorem 2, we examine allocations of the form $\boldsymbol{e}(\alpha) = \alpha\tilde{\boldsymbol{e}} + (1-\alpha)\boldsymbol{e^{min}}$ for $\alpha \in [0, 1]$ and $\boldsymbol{e}(\alpha^*)$ for $\alpha^* = 0.18$ yields $g(\boldsymbol{e}(\alpha^*)) = 1.00005 \approx 1$. We thus conclude that the allocation

$$\boldsymbol{e^*} = \alpha^* \tilde{\boldsymbol{e}} + (1 - \alpha^*)\boldsymbol{e}^{min} = .18\tilde{\boldsymbol{e}} + .82\boldsymbol{e^{min}}$$

is Pareto efficient, individually rational and fair.[15] This allocation appears in the fifth

[15] For the curious reader, the fact that $g(\boldsymbol{e^{min}}) = 0.82 = 1 - 0.18 = 1 - \alpha^*$ is not a coincidence. This is because the utility functions of Nordhaus imply $g(\alpha\tilde{\boldsymbol{e}} + (1-\alpha)\boldsymbol{e}) = \frac{1}{1-\alpha} g(\boldsymbol{e})$ for any allocation $\boldsymbol{e}$.

column of Table 3.

For reasons that will become clear very soon, we also compute the allocation which maximizes total welfare subject to IR and F. We denote this by $\hat{\boldsymbol{e}}$ and summarize it in the last column of Table 3. This allocation is such that $g(\hat{\boldsymbol{e}}) \approx 1.1$ and so is Pareto inefficient, albeit mildly so. Its aggregate emissions are about 2% higher than those of PE-IR-F allocation $\boldsymbol{e}^*$.

Region i	GDPpc$_i$	$\tilde{e}_i$	e_i^{min}	e_i^*	$\hat{e}_i$
	(US$)	(ton CO$_2$)	(ton CO$_2$)	(ton CO$_2$)	(ton CO$_2$)
S. Saharan Africa	2,673	234,646,913	189,914,156	197,966,052	199,703,388
India	4,883	2,049,561,902	1,658,836,309	1,729,166,915	1,744,341,960
Rest of World	6,310	1,552,161,782	1,256,259,847	1,309,522,195	1,312,497,243
China	10,041	8,293,771,000	6,712,658,182	6,997,258,489	7,013,155,262
Eurasia	10,061	877,792,391	709,097,098	739,462,251	742,255,161
South Africa	11,910	458,061,282	368,573,493	384,681,295	387,333,445
Latin America	13,003	1,204,677,235	969,329,026	1,011,691,703	1,018,666,719
Brazil	14,301	429,462,339	345,561,698	360,663,813	363,150,377
SE Asia	15,768	1,764,016,979	1,419,395,013	1,481,426,967	1,491,640,529
Middle East	17,022	2,007,397,016	1,615,227,828	1,685,818,282	1,697,441,001
Russia	22,570	1,737,103,381	1,397,739,311	1,458,824,843	1,468,882,577
EU	33,409	3,718,923,879	2,872,737,351	3,025,050,926	3,144,696,251
Japan	34,316	1,172,544,223	905,749,000	953,772,140	991,495,266
Canada	41,333	499,877,528	386,137,736	406,610,898	422,692,972
United States	49,855	5,444,142,792	4,036,827,821	4,290,144,515	4,567,632,836
WORLD	13,284	31,444,140,642	24,844,043,867	26,032,061,284	26,565,584,987

Table 3: Numerical results using the model, data, and calibrated parameters of Nordhaus [32]. Status-quo emissions and GDP data refer to year 2011. Regions are listed in increasing GDP per capita.

Generating and selecting from a set of PE-IR-F allocations. As mentioned earlier, by Proposition 2, there will be a multiplicity of PE-IR-F allocations. We thus proceed to generate a set of suitable allocations from which we can select from.

Following the procedure laid out in Section 4, we begin by generating two sets of IR-F allocations, $\mathcal{S}^-$ and $\mathcal{S}^+$. All elements of $\mathcal{S}^-$ (resp. $\mathcal{S}^+$) will satisfy $g(\boldsymbol{e}) < 1$ (resp. > 1). The set $\mathcal{S}^-$ consists of 40 allocations, including: $\boldsymbol{e^{min}}$, nineteen allocations that are perturbations[16] of $\boldsymbol{e^{min}}$, and twenty allocations that are perturbations of $\boldsymbol{e}^*$ in the direction of decreasing $g(\cdot)$. The set $\mathcal{S}^+$ also consists of 40 allocations, including: $\tilde{\boldsymbol{e}}$,

[16] Here and elsewhere we make sure that the perturbations still satisfy IR and F.

nineteen allocations that are perturbations of $\tilde{\boldsymbol{e}}$, $\hat{\boldsymbol{e}}$, and nineteen allocations that are perturbations of $\hat{\boldsymbol{e}}$.

For every pair of allocations $(\boldsymbol{e}^-, \boldsymbol{e}^+) \in \mathcal{S}^- \times \mathcal{S}^+$ we compute $a \in [0, 1]$ such that $g\left(a\boldsymbol{e}^- + (1-a)\boldsymbol{e}^+\right) = 1$[17], record the allocation $a\boldsymbol{e}^- + (1-a)\boldsymbol{e}^+$, and include it the set of PE-IR-F allocations. Performing this operation for all $40 * 40 = 1600$ pairs of allocations in $\mathcal{S}^- \times \mathcal{S}^+$ leads to 1600 allocations satisfying PE-IR-F.

It is worth noting that the 1600 solutions generated present little variation. Table 4 includes some relevant descriptive statistics. For all regions i, we denote the average emissions ratios, i.e. the average of the quantity $\frac{e_i}{\tilde{e}_i}$ across all 1600 allocations, by $\mu_{\left(\frac{e_i}{\tilde{e}_i}\right)}$. The corresponding standard deviation is denoted by $\sigma_{\left(\frac{e_i}{\tilde{e}_i}\right)}$. We see that this standard deviation is is very small, considerably less than $0.01 \times \mu_{\left(\frac{e_i}{\tilde{e}_i}\right)}$.

Region i	$\mu_{\left(\frac{e_i}{\tilde{e}_i}\right)}$	$\sigma_{\left(\frac{e_i}{\tilde{e}_i}\right)}$
S. Saharan Africa	0.8557	0.0036
India	0.8537	0.0028
Rest of World	0.8504	0.0024
China	0.8483	0.0022
Eurasia	0.8454	0.0020
South Africa	0.8410	0.0018
Latin America	0.8393	0.0019
Brazil	0.8376	0.0021
SE Asia	0.8356	0.0026
Middle East	0.8340	0.0029
Russia	0.8320	0.0031
EU	0.8070	0.0031
Japan	0.8051	0.0034
Canada	0.8032	0.0036
United States	0.7784	0.0063

Table 4: Results based on the setting of Nordhaus [32]. Descriptive statistics of the emissions ratios $\frac{e_i}{\tilde{e}_i}$ of the 1600 simulated PE-IR-F allocations.

For the purposes of this exercise, we are interested in two dimensions of performance: aggregate welfare and aggregate emissions. Figure 3 depicts the performance of the 1600 simulated allocations along these two dimensions with respect to the status-quo $\tilde{\boldsymbol{e}}$. On the horizontal axis appear percentage reductions in aggregate emissions (i.e., the quantity $1 - \frac{e}{\tilde{e}}$), whereas, on the vertical axis appear percentage gains in welfare (i.e., the quantity $\frac{\sum_i u_i(\boldsymbol{e})}{\sum_i u_i(\tilde{\boldsymbol{e}})} - 1$).

Figure 3 demonstrates that relative welfare gains are quite small and exhibit very

[17] Allowing for an error of < 0.001.

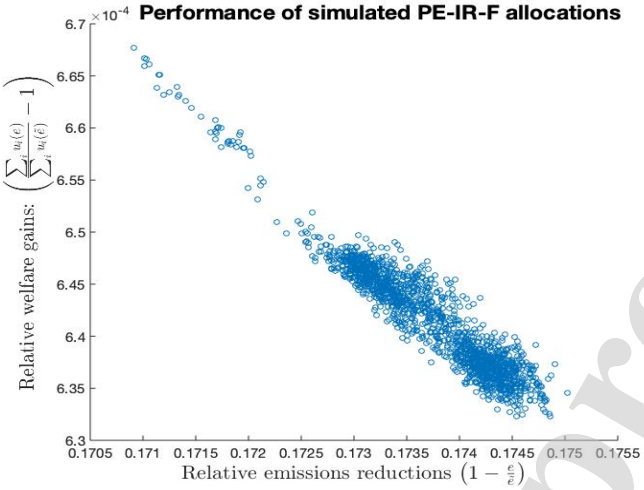

Figure 3: Results based on the setting of Nordhaus [32]. The correlation coefficient between emissions reductions and welfare gains is $\rho \approx -0.95$.

little variation across the set of simulated allocations. They range from a minimum of a little more than 0.063% to a maximum of a little less than 0.067%. The situation is different for emissions reductions, which are quite substantial, ranging from a minimum of about 17.1% to a maximum of 17.5%.

One interesting finding that emerges from Figure 3 is that, along the simulated PE-IR-F frontier, there is a tradeoff between welfare gains and emissions reductions. That is, once one achieves PE-IR-F, it is generally not possible to improve total welfare while also decreasing total emissions. In fact, the correlation coefficient between emissions reductions and welfare gains along the frontier is negative and very close to -1 ($\rho \approx -.95$).

We don't wish to make too much of the above result given the low magnitude and limited variation of welfare gains. That being said, it is worth pointing out that the strong negative correlation between welfare gains and emissions reductions of Figure 3 is driven almost entirely by the United States. Allocations with lower aggregate emissions tend to imply lower welfare gains for the United States, the richest and second-most polluting region (the most polluting in per capita terms). If we exclude the United States from the welfare calculation, we obtain a positive correlation between welfare gains and emissions reductions of $\rho \approx 0.78$. This can be readily seen in Figure 4 where, to be clear, the y-axis maps the quantities $\frac{\sum_{i \neq US} u_i(\boldsymbol{e})}{\sum_{i \neq US} u_i(\tilde{\boldsymbol{e}})} - 1$ and the x-axis is the same as before. The previous tradeoff between welfare and aggregate emissions has disappeared.

Given the above, we argue that it is sensible to set aside the welfare dimension and

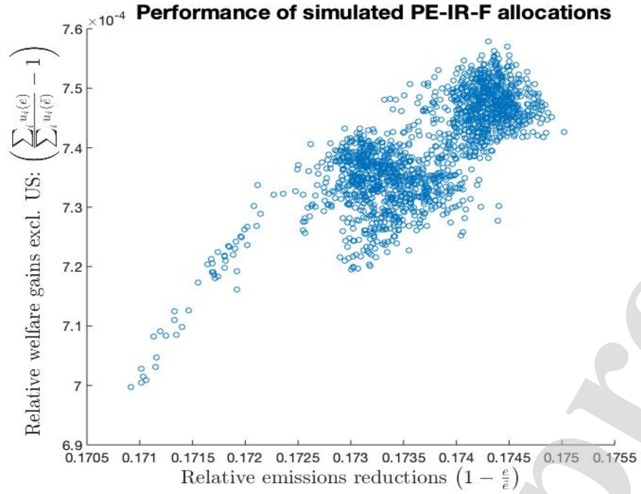

Figure 4: Results based on the setting of Nordhaus [32]. The y-axis measures relative aggregate welfare gains excluding the United States. The correlation coefficient between emissions reductions and welfare gains is $\rho \approx 0.78$.

select from the set of PE-IR-F allocations the one which minimizes aggregate emissions. We denote it by $\boldsymbol{e^{**}}$ and list it in the third column of Table 5. The fifth column of the same Table lists the emissions ratios for each region with respect to the status-quo, demonstrating that the fairness criterion is met. The last column of Table 5 lists the implied weights of each region according to $\boldsymbol{e^{**}}$, normalized to sum to 1, as stipulated in part (ii) of Proposition 1. As we see, poorer countries with less stringent mitigation targets tend to receive greater weight but the relation is not perfectly monotonic.[18] Finally, looking at Table 4, it is also worth noting that $\frac{e_i^{**}}{\bar{e}_i}$ is very close to $\mu_{\left(\frac{e_i}{\bar{e}_i}\right)}$, for all regions i.

It is interesting to compare the allocation $\boldsymbol{e^{**}}$ to the allocation that maximizes aggregate welfare, call it $\mathbf{e}^{\boldsymbol{W}}$. This allocation is necessarily Pareto efficient as it corresponds to the optimal solution of problem (1) for the case of equal weights (i.e. when $w_1 = ... = w_I$). We list it in the fourth column of Table 5.[19] The same Table illustrates how $\boldsymbol{e^W}$ violates both individual rationality and fairness.

Allocation $\mathbf{e}^{\boldsymbol{W}}$'s violations of fairness are stark, as evidenced in the sixth column

[18] This is a result of part (ii) of Proposition 1 applied to the Nordhaus benefit functions. Evidently, if a region's (i) costliness of emissions reductions weighted by its GDP is much lower and/or (ii) its status-quo emissions are much higher compared to those of another, then it could get assigned a higher implicit weight even though it has a more stringent mitigation target. This is the case, for instance, with India and the Rest of World regions.

[19] Consistent with Proposition 1, we verify that $g(\boldsymbol{e^W}) = 1.00046 \approx 1$.

of Table 5. For instance, South Saharan Africa, the region with the lowest GDP per capita, is required to lower its emissions by roughly 24%, whereas the United States, the region with the highest GDP per capita, by 16%. Similar imbalances occur for a great number of region pairs. To cite another, particularly extreme, one: the Rest of World region (third lowest GDP per capita) is asked to reduce its emissions by roughly 30%, whereas Canada (second highest GDP per capita) by roughly 12.5%. It stands to reason that allocation $\boldsymbol{e^W}$ would be hard to accept for many regions on account of its perceived unfairness.

Conversely, the allocation $\boldsymbol{e^W}$ comes very close to achieving individual rationality. The only region which experiences lower utility under $\boldsymbol{e^W}$ compared to the status-quo is South Africa.

Region i	GDPpc$_i$	e_i^{**}	e_i^W	$\frac{e_i^{**}}{\bar{e}_i}$	$\frac{e_i^W}{\bar{e}_i}$	w_i^{**}
	(US$)	(ton CO$_2$)	(ton CO$_2$)	(scalar)	(scalar)	(scalar)
S. Saharan Africa	2,673	201,185,309	178,027,398	0.8574	0.7587	0.1003
India	4,883	1,752,283,843	1,756,600,484	0.8550	0.8571	0.0585
Rest of World	6,310	1,323,892,859	1,091,520,101	0.8529	0.7032	0.1200
China	10,041	7,032,235,820	7,022,760,184	0.8479	0.8468	0.0600
Eurasia	10,061	742,586,187	741,168,748	0.8460	0.8444	0.0600
South Africa	11,910	386,152,172	354,996,130	0.8430	0.7750	0.0852
Latin America	13,003	1,014,106,993	995,786,478	0.8418	0.8266	0.0650
Brazil	14,301	360,937,647	292,700,648	0.8404	0.6816	0.1186
SE Asia	15,768	1,479,870,749	1,356,973,111	0.8389	0.7693	0.0845
Middle East	17,022	1,673,091,287	1,799,722,243	0.8335	0.8965	0.0368
Russia	22,570	1,442,595,059	1,494,081,495	0.8305	0.8601	0.0495
EU	33,409	2,987,614,092	3,192,033,618	0.8034	0.8583	0.0432
Japan	34,316	938,757,246	1,017,573,880	0.8006	0.8678	0.0395
Canada	41,333	398,709,911	437,294,402	0.7976	0.8748	0.0368
United States	49,855	4,206,691,063	4,563,458,120	0.7727	0.8382	0.0421
WORLD	13,284	25,940,710,237	26,294,697,040	0.8250	0.8362	N/A

Table 5: Numerical results using the model, data, and calibrated parameters of Nordhaus [32]. GDP data refer to year 2011.

6 Conclusion

This paper has addressed the existence of allocations in the commons satisfying Pareto efficiency, individual rationality and a novel concept of fairness that holds pragmatic appeal. While these properties are not always compatible it is possible to obtain a sharp necessary and sufficient condition for existence to hold. This condition is theoretically and computationally tractable. Uniqueness will not in general hold and so a simulation-based procedure was proposed to generate sets of PE-IR-F allocations from which one can subsequently select. A proof of concept of the theoretical analysis based on the climate-change setting of Nordhaus [32] was provided, demonstrating that large emissions reductions are consistent with the three properties.

A fruitful avenue for future research involves the introduction of strategic considerations into the model. In particular, it would be interesting to enhance the individual rationality property with ideas from cooperative game theory such as belonging to the core. Instead of assuming that agents fall back onto a status-quo allocation if they don't reach an agreement, one could explore alternative participation concepts that allow for a degree of coalition formation. Doing so would deepen the axiomatic analysis and refine the set of candidate allocations in a meaningful way.

References

[1] Ambec, S., and Sprumont, Y. (2002). Sharing a river. *Journal of Economic Theory*, 107(2), 453-462.

[2] Ambec, S., and Ehlers, L. (2008). Sharing a river among satiable agents. *Games and Economic Behavior*, 64(1), 35-50.

[3] Ambec, S., and Ehlers, L. (2016). Regulation via the Polluter-pays Principle. *Economic Journal*, 126, 884-906.

[4] Ansink, E., and Weikard, H. P. (2012). Sequential sharing rules for river sharing problems. *Social Choice and Welfare*, 38, 187-210.

[5] Ansink, E., and Weikard, H. P. (2015). Composition properties in the river claims problem. *Social Choice and Welfare*, 44, 807-831.

[6] Barrett, S. (1994). Self-enforcing international environmental agreements. *Oxford Economic Papers*, 878-894.

[7] Barrett, S. (2003). *Environment and statecraft: The strategy of environmental treaty-making*. OUP Oxford.

[8] Bossert, W., and Fleurbaey, M. (1996). Redistribution and compensation. *Social Choice and Welfare*, 13, 343-355.

[9] Boyd, S., and Vandenberghe, L. (2008). *Convex Optimization*. New York, NY: Cambridge University Press.

[10] Cabral, L. M. (2017). *Introduction to industrial organization*. MIT press.

[11] Carraro, C., and Siniscalco, D. (1993). Strategies for the international protection of the environment. *Journal of Public Economics*, 52, 309-328.

[12] Chander, P., and Tulkens, H. (1995). A core-theoretic solution for the design of cooperative agreements on transfrontier pollution. *International Tax and Public Finance*, 2(2), 279.

[13] Chander, P., and Tulkens, H. (1997). The Core of an Economy with Multilateral Environmental Externalities. *International Journal of Game Theory*, 3(26), 379-401.

[14] De Villemeur, É. B., and Leroux, J. (2011). Sharing the cost of global warming. *Scandinavian Journal of Economics*, 113, 758-783.

[15] du Pont, Y. R., Jeffery, M. L., Gütschow, J., Rogelj, J., Christoff, P., and Meinshausen, M. (2017). Equitable mitigation to achieve the Paris Agreement goals. *Nature Climate Change*, 7, 38-43.

[16] Dutta, P. K., and Radner, R. (2004). Self-enforcing climate-change treaties. *Proceedings of the National Academy of Sciences*, 101(14), 5174-5179.

[17] Finus, M. (2008). Game theoretic research on the design of international environmental agreements: insights, critical remarks, and future challenges. *International Review of environmental and resource economics*, 2(1), 29-67.

[18] Finus, M., and McGinty, M. (2019). The anti-paradox of cooperation: Diversity may pay!. *Journal of Economic Behavior and Organization*, 157, 541-559.

[19] Fleurbaey, M. (2008). Fairness, responsibility, and welfare. Oxford University Press.

[20] Fleurbaey, M., and Maniquet, F. (2011). Compensation and Responsibility. In Handbook of Social Choice and Welfare, Volume 2, edited by Kenneth J. Arrow, Amartya Sen, and Kotaro Suzumura, 507–604. Amsterdam and Boston: Elsevier, North-Holland.

[21] Fleurbaey, M., and Peragine, V. (2013). Ex ante versus ex post equality of opportunity. *Economica*, 80(317), 118-130.

[22] Gampfer, R. (2014). Do individuals care about fairness in burden sharing for climate change mitigation? Evidence from a lab experiment. *Climatic Change*, 124, 65-77.

[23] Hardin, G. (1968). The tragedy of the commons. *Science*, 162, 1243-1248.

[24] Harstad, B. (2012). Buy coal! A case for supply-side environmental policy. *Journal of Political Economy*, 120, 77-115.

[25] Harstad, B. (2016). The dynamics of climate agreements. *Journal of the European Economic Association*, 14(3), 719-752.

[26] Helm, C., and Wirl, F. (2014). The principal–agent model with multilateral externalities: An application to climate agreements. *Journal of Environmental Economics and Management*, 67(2), 141-154.

[27] Höhne, N., Den Elzen, M., and Escalante, D. (2014). Regional GHG reduction targets based on effort sharing: a comparison of studies. *Climate Policy*, 14, 122-147.

[28] Lange, A., Vogt, C., and Ziegler, A. (2007). On the importance of equity in international climate policy: An empirical analysis. *Energy Economics*, 29, 545-562.

[29] Lange, A., Loschel, A., Vogt, C., and Ziegler, A. (2010). On the self-interested use of equity in international climate negotiations. *European Economic Review*, 54(3), 359-375.

[30] Martimort, D., and Sand-Zantman, W. (2016). A mechanism design approach to climate-change agreements. *Journal of the European Economic Association*, 14(3), 669-718.

[31] McGinty, M. (2007). International environmental agreements among asymmetric nations. *Oxford Economic Papers*, 59(1), 45-62.

[32] Nordhaus, W. (2015). Climate clubs: Overcoming free-riding in international climate policy. *American Economic Review*, 105, 1339-70.

[33] Öztürk, Z. E. (2020). Fair social orderings for the sharing of international rivers: A leximin based approach. *Journal of Environmental Economics and Management*, 101, 102302.

[34] Raupach, M. R., Davis, S. J., Peters, G. P., Andrew, R. M., Canadell, J. G., Ciais, P., Friedlingstein, P. , Frank Jotzo, F., Van Vuuren, D. and Le Quere, C. (2014). Sharing a quota on cumulative carbon emissions. *Nature Climate Change*, 4, 873-879.

[35] Roemer, J. E. (1998). Equality of opportunity. Harvard University Press.

[36] Roemer, J. E. (1998). Theories of distributive justice. Harvard University Press.

[37] Roemer, J. E., and Trannoy, A. (2016). Equality of opportunity: Theory and measurement. Journal of Economic Literature, 54(4), 1288-1332.

[38] Sheriff, G. (2019). Burden sharing under the Paris climate agreement. *Journal of the Association of Environmental and Resource Economists*, 6, 275-318.

[39] van den Brink, R., van der Laan, G., and Moes, N. (2012). Fair agreements for sharing international rivers with multiple springs and externalities.*Journal of Environmental Economics and Management*, 63, 388-403.

[40] Weikard, H. P., Finus, M., and Altamirano-Cabrera, J. C. (2006). The impact of surplus sharing on the stability of international climate agreements. *Oxford Economic Papers*, 58(, 209-232.

[41] World Bank. (2005). World Development Report 2006: equity and development. The World Bank.

[42] Young, H. P. (1994). Equity In Theory and Practice. Princeton University Press.

The author declares that he has no relevant or material financial interests that relate to the research described in this paper