Sharing the Cost of Global Warming*

Etienne Billette de Villemeur, Toulouse School of Economics (IDEI & GREMAQ), 21 allée de Brienne, 31000 Toulouse, France. Email: etienne.devillemeur@TSE-fr.eu

Justin Leroux, HEC Montréal (CIRANO & CIRPEE), 3000 chemin de la Côte-Sainte-Catherine, Montréal, QC, H3T 2A7, Canada. Email: justin.leroux@hec.ca.

Abstract: Due to meteorological factors, the distribution of climate change damage bears no relationship to that of emissions across the globe. We argue in favor of offsetting this discrepancy, and propose a global insurance scheme of sorts, to be financed according to countries' responsibilities in the matter. We analyze this horizontal equity problem through the lens of the responsibility/compensation framework *à la* Bossert and Fleurbaey (1996), which we adapt to a context of externalities. We characterize new versions of two well-known cost-sharing schemes.

J.E.L. Classification: D62, D63, Q54.

Keywords: Climate Change, Responsibility, Egalitarian Liberalism.

*We thank Jérémy Laurent-Lucchetti and Yves Sprumont for stimulating discussions and two anonymous referees for comments leading to vast improvements in the original manuscript. This paper was written while de Villemeur was visiting the Université de Montréal, whose hospitality is gratefully acknowledged. Leroux is grateful for support from the SSHRC.

I. Introduction

Nowadays, climate change pervades our collective psyche, from policy design to everyday conversations, usually under the more casual designation of "global warming". Accompanying our awareness of climate change is the growing realization that the impacts of global warming are not uniformly distributed across the globe. Yet, by itself, the fact that countries are unequally affected by climate change does not warrant a cry against injustice. Indeed, if the regions most affected by climate change were also the ones contributing to it the most, the observation would be less shocking. However, when comparing maps of recent and cumulated emissions with that of temperature anomalies, one cannot help but notice that they do not coincide (Figure 1). Add to it the fact that the melting of icecaps will disproportionately impact coastal cities, and it becomes clear that some countries end up generating more harm than they endure, while others must absorb more damage than they cause. In other words, under the near-consensual assumption that the increase in human activity has contributed to it, climate change is a prime example of a global externality.

[FIGURE 1 ABOUT HERE]

Yet, because the discrepancy between the distribution of greenhouse gas (GHG) emissions in the atmosphere and their resulting impacts is due to "natural" phenomena (e.g., winds, currents, the melting of icecaps, etc.), and because it is impossible to trace back to its origin the damage borne by any given region, we argue that the distribution of damages lies beyond the responsibility of any country. Nonetheless, provided the impact of climate change and emissions patterns are observable, one can still hope to solve this global cost-sharing problem of sorts.

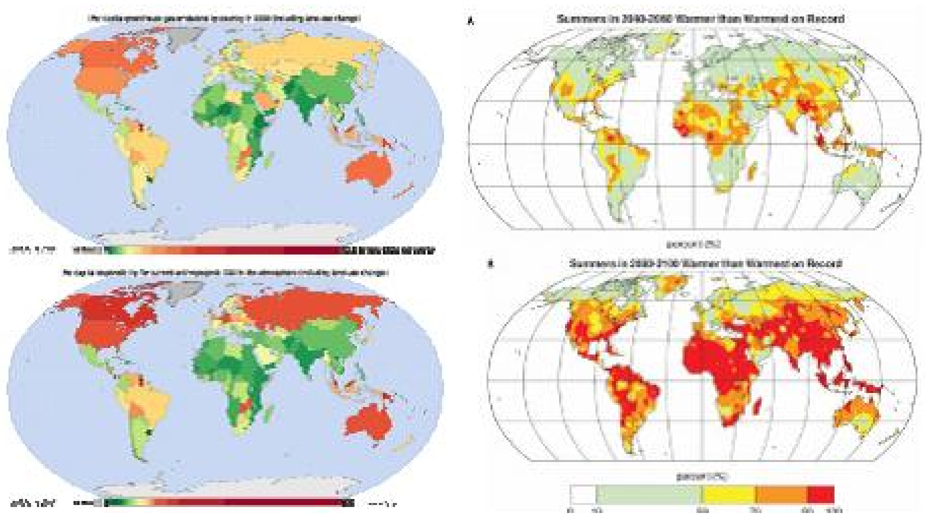

Figure 1: Left, up: per capita GHG emissions by country in 2000; left, down, per capita cumulated emissions. Right: likelihood that future average summer temperatures exceed highest temperature observed on record. Up, for the period 2040-2060; down, for 2080-2100. Sources: World Resources Institute, via Wikipedia, for emissions; Science, vol. 323 (01/09/09) for temperatures.

We argue in favor of a global insurance scheme that washes out differences in the distributions of damages, for which countries are not responsible. However, the financing of this scheme should hold countries responsible for the (global) damage for which they are indeed responsible.

A standard approach to implementing the first-best level of pollution is through Pigou taxes (or equivalent schemes), which succeed by making polluters internalize the social marginal cost of their externality (Pigou, 1932; Baumol, 1972; Nordhaus, 1992). However, even if the first-best level of emissions is reached, *some* climate change will occur, because *some* pollution is efficient, and Pigou taxation does not handle the normative question of horizontal (or spatial) equity raised by the imbalance between the distribution of emissions and the distribution of damages.

We propose an alternative approach to Pigou taxation which overcomes this shortcomings while remaining entirely compatible with the Pigovian outcome (i.e., a first-best pollution level). We treat the issue as a traditional cost-sharing problem, where damages depend on countries' past and current emissions. We address the issue of global equity by accounting for the responsibility of countries to the global damage through their past and current emissions as well as other characteristics (GDP, geographical location, population, etc.). We are fully aware that the debate on countries' responsibilities for past emissions is still raging, and it is a debate which is beyond the scope of this work. Nonetheless, our approach is flexible enough to accommodate the most prominent competing ideologies, precisely because we let the planner decide for which characteristics countries are to be held responsible.

More precisely, the responsibility/compensation approach we adopt is related to that in Bossert & Fleurbaey (1996). The planner's first decision is to identify for which characteristics countries should be held responsible and for which characteristics they should be held "non responsible"—and thus compensated. For instance, in our context, the planner may deem that current populations are responsible for their current GHG emissions and must therefore pay for all the resulting damages, but that they are not responsible for their geographical location and corresponding climate, so that their cost share should not depend on their geographical location, all else equal (emissions, in particular). Yet, a tension already arises from this seemingly innocent decision. For example, a country which experiences cold winters, like Canada, should be held responsible for its emissions but should at the same time be compensated for the fact that emissions are high due to heating needs. In fact, it has been established (Bossert, 1995) that holding economic agents fully responsible for some characteristics is incompatible with full compensation along the other characteristics unless the cost function is separable in the responsibility/compensation dimensions. Because one cannot reasonably assume such separability in practice, the two taxation schemes that stand out from our analysis (the *Egalitarian Equivalent* and the *Conditional Equality* mechanisms) result from compromises between full compensation and full responsibility.

From a policy standpoint, our findings reveal that environmental taxation *à la* Pigou, which is often considered the epitome of responsibilization, popularized by its implementation of the first-best pollution levels, is compatible with a modicum of compensation for differences in "irrelevant" characteristics[1]. In particular, the Conditional Equality solution can be viewed as a Pigou tax followed by redistributive transfers. Moreover, this solution offers the first axiomatic specification of how to allocate the Pigou tax revenues, both to wash out the damages suffered by each country, and for cross-country compensation.

A global insurance scheme

The leveling of environmental damage presented above requires possibly large funds, equal to the sum of all climate-change-related damage across the globe. Indeed, if we denote by d_{it} the environmental damage suffered by country i due to climate change over a given period, t, the total amount of funds necessary to cancel out the damage for all countries in that period adds up to $D_t \equiv \sum_{i,t} d_{it}$.[2] Obviously, given the global nature of the issue, the financing of the leveling compensation must be collectively borne by the very same countries which are receiving these compensations. Hence, the leveling compensation is, in essence, a redistribution mechanism designed to counter the arbitrary nature of the distribution of climate-change damage across the globe.

[1] Keeping with the language of Bossert & Fleurbaey (1996), *irrelevant* characteristics are those for which countries are not deemed responsible.

[2] The dynamic nature of the problem and the relation between damage and carbon emissions will be made explicit below, as needed in the argumentation.

Given that we consider climate change to be closely related to GHG concentration levels, we argue that the damage-offsetting scheme should be financed in relation to the countries' emission levels.[3] In a companion paper (Billette de Villemeur and Leroux, 2010), we show that taxing countries according to their emissions *stock* is an alternative to the standard Pigou taxation (which taxes emission *flows*) that leads to the same emissions patterns, and is thus first-best efficient. According to this view, emissions are like environmental debt issued by polluters at all times $s < t$, a portion of which will be collected over time in the form of the environmental damage D_t.

This has several advantages.[4] Particularly, unlike the tax on emissions *flows*, which relies on the computation of the *discounted* sum of *expected* future social marginal damage (Nordhaus, 1992), the mechanism we propose repays the "debt", D_t, of *observed* environmental damage in *each period.* This approach allows us to focus on the problem of spatial justice, namely, how to share the damage in each period.

Compensation and responsibility

Our focus is on the question of who will pay how much. Given our responsibility-based approach to fairness, this final discussion amounts to sorting out the responsibility of each country in the matter. We adopt the following general principle: "Countries should pay for damage caused for which they are responsible and be compensated for damage suffered for which they are not". Our approach allows one to make recommendations based on the planner's assignment of responsibilities (which is beyond the scope of this work). More precisely, the

[3] Naturally, this point becomes moot if the planner believes climate change to be completely unrelated to human activity. In that case, the rest of our analysis can be interpreted as tackling the issue of the arbitrariness of climate change—which is considered to be happening, with no identified cause.

[4] We refer the reader to Billette de Villemeur and Leroux (2011) for a detailed account of the properties of a retrospective carbon tax.

resulting cost shares will be tailored to reflect any possible assignment of responsibilities, from considering that countries are fully responsible for all historical emissions (what we call the *Historical Responsibility* view), to taking historical emissions as a rightful claim on today's atmosphere (the so-called *Grandfathering* view). Finally, we also allow for an intermediate viewpoint, where countries (and their inhabitants) are considered responsible for their current emissions, but not for their history (*no Historical Responsibility*).

Our work belongs to the growing literature on equal opportunity theories of justice (also called liberal egalitarianism) since its revival (Rawls, 1971; Sen, 1985, Roemer, 1998, Fleurbaey, 1995, Bossert and Fleurbaey, 1996, Cappelen and Tungodden, 2003, 2007; Tungodden, 2005).[5] The general message of liberal egalitarian ethics is that individuals should be rewarded for the result of differences in characteristics which are in their control (or "relevant characteristics"), all the while being compensated for the result of differences in characteristics beyond their control ("irrelevant characteristics").

In our context, the "relevant" characteristics are current emissions and, possibly, past cumulative emissions. By contrast, we consider the "irrelevant" characteristics to be aspects of geographical location (like latitude and coastality) and possibly past emissions (depending on the point of view on responsibility). We formulate axioms which any taxation rule must verify to be compatible with the desired view of responsibility adopted by the planner.

Ideally, one would want to hold countries fully responsible for differences in their relevant characteristics while being fully compensated for differences in their irrelevant characteristics. Unfortunately, a strong tension exists between these principles of responsibility and compensation, making them typically incompatible in their strong versions except for the unreasonably simplistic case

[5] We refer the reader to Johansson-Stenman and Konow (2010) for a review of theories of justice and their importance in environmental policy.

of a linear damage function. Therefore, when deciding on what sharing rule to implement, the planner faces a trade-off between relaxing the extent to which countries are held responsible for their relevant emissions or the extent to which they can be compensated for differences in irrelevant characteristics. The type of cost-sharing mechanisms that emerge from our analysis consists in holding countries responsible—or compensating them—for differences in characteristics, not among themselves *per se*, but relative to a reference level. This allows for taking differences in characteristics into account while reconciling the otherwise incompatible principles of responsibility and compensation. The solutions we offer will differ in how strongly they depart from marginal-damage pricing.

Specifically, one of the two solutions we propose, the *Egalitarian Equivalent mechanism*, redistributes welfare according to differences in *irrelevant* characteristics while sharing equally the residual impact of global warming, once each country has paid for its *incremental* contribution to total damage.[6] In other words, the Egalitarian Equivalent mechanism prices emissions at incremental cost while balancing the budget via lump-sum transfers based on countries' irrelevant characteristics. In the responsibility-compensation spectrum, one could argue that the Egalitarian Equivalent mechanism insists on compensation at the expense of responsibility.

Symmetrically, the *Conditional Equality mechanism* guarantees each agent the average payoff of a hypothetical situation in which all countries' relevant characteristics are equal to a reference level. Each country bears the consequences of any deviation from this reference level. As such, the Conditional Equality mechanism insists on countries' responsibility at the expense of compensation considerations. If the reference level is the first-best level of emissions, countries are taxed according to marginal damage, thus implementing the first-

[6] The incremental cost is the additional cost imposed by the presence of an additional *agent*, as opposed to an additional *unit* of emissions (as would be the case for marginal cost). Clearly, the smaller the emissions of a country, the closer its incremental cost is to the marginal cost.

best level of emissions. While the latter is akin to Pigou taxation, the mechanism also combines the taxation part with redistributive lump-sum transfers.

Our compensation-responsibility approach builds upon that developed in Bossert (1995) and Bossert and Fleurbaey (1996) after adapting it to a context of externalities. Indeed, their setting focuses on wealth redistribution in the absence of externalities, which would be tantamount to assuming a damage function which is separable in the countries' emissions levels. Such an assumption would be ill-adapted in our context. By contrast, our setup introduces interdependence between the countries' characteristics (emission levels) through the global damage function. As a result, the analysis allows for applications in more general settings where externalities are present.

It should be noted that we do not explicitly address the other two fundamental fairness issues raised by climate change that are intergenerational equity and fairness in the presence of uncertain consequences. Nonetheless, to some extent, the intergenerational problem is already present in the choice of the responsibility view; i.e., on whether current generations should be held responsible for their ancestors' actions. Also, while we do not surmise that our framework is rich enough to fully answer the intergenerational question, we contend that our recommendations will not interfere with possible additional transfers aimed at correcting intergenerational inequity. In fact, one can readily imagine that intergenerational policies be implemented within each country, perhaps through investments possibilities, on top of—and independently of—spatial considerations. Hence, transfer-wise, the intergenerational equity problem is somewhat orthogonal to our analysis.

The uncertainty problem is also related to the choice of the responsibility viewpoint. Indeed, the latter determines to what extent the ex post outcome warrants redistribution relative to ex ante decisions (Fleurbaey, 2008). This is most clearly seen under a fourth view of responsibility, which denies that climate change has anything to do with human activity. This view considers global warming to be simply due to bad luck. Thus, it clears everyone from any responsibility whatsoever by considering all characteristics "irrelevant", and recommends transfers accordingly.[7]

II. The model

Let $S = \{1, ..., m\} \subset \mathbb{N}$ be the set of countries[8]. We denote by n_i the population of country i and by $n = (n_1, ..., n_m) \in \mathbb{N}^m$ the population profile. We write $N = \sum_{i=1}^{m} n_i$. We denote by $x_i = (x_i^p, x_i^c) \in \mathbb{R}_+^2$ country i's vector of *past* and *current*[9] per capita emissions, and by $x^p = (x_1^p, ..x_m^p)$ and $x^c = (x_1^c, ..x_m^c)$ the *past* and *current* per capita emissions profiles, respectively.[10]

Each country's per capita private *current* benefits are associated to its own emissions levels via a mapping $b_i : \mathbb{R}_+^2 \to \mathbb{R}$, $(x_i^p, x_i^c) \mapsto b_i\,(x_i^p, x_i^c)$.[11] We assume benefits to be fully transferable across countries. Let $b = (b_1, .., b_m)$ be the profile of per capita benefit functions, one per country. Think of benefits as per capita GDP or some more elaborate wealth measure, possibly enhanced by some human development index, for instance. We interpret the differences in the benefit functions across countries to be essentially geographic in nature (latitude,

[7] For space considerations, we do not give details here of the analysis for this so-called External Shock view. They follow the same analysis as for the other views and can be found in a working paper version of this manuscript, which is available upon request.

[8] We use the word "countries" for simplicity, but our analysis readily applies to regions, which is especially relevant for large countries.

[9] From a practical standpoint, current emissions can be interpreted as emissions within a time frame coinciding with the reevaluation periods of the major international protocols (Kyoto, Copenhagen, etc.).

[10] To lighten notation, we do not index variables by their time subscript. It will be implicit throughout that all the analysis takes place at time t, and that $x^c = x_t^c$ while $x^p = \sum_{s<t} x_s^c$. More generally, any value decribed as "current" will refer to a value at time t.

[11] While b_i is not necessarily independent of the country's population, n_i, we do not explicitly consider variations in n_i. Moreover, we make no hypothesis on how b_i should vary with n_i.

altitude, coastality, etc.).[12] Likewise, past emissions can be thought of as a proxy for capital accumulation, so that $b_i(x_i^p, \cdot)$ represents current production possibilities resulting from past capital investment decisions.

Contrasting with private benefits, which are solely dependent upon a country's own emissions, per capita environmental damage depends on total emissions. More precisely, let $X^k = \sum_{i=1,\ldots,m} n_i x_i^k$, with $k = p, c$, be the total level of past and current emissions, respectively, and denote $X = X^p + X^c$.[13] We denote by $d_i(X)$ the *current* per capita damage incurred by country i. We make no assumptions on the functions $d_i(X)$. In particular, it may be the case that some countries actually benefit from global warming for some values of X. That is to say we do not exclude the possibility that $d_i(X) < 0$ for some countries. Nonetheless, we assume total damage,

$$D(X) = \sum_{i=1}^{m} n_i d_i(X),$$

to be positive and non-decreasing in X. We denote by $d = (d_1, ...d_m)$ the profile of per capita damage functions. Finally, we call $P = (n, b, d, x^p, x^c)$ a *global warming problem* and denote by $\mathcal{P}$ the class of such problems.

Our goal is to design a *transfer schedule* to correct the arguably uneven distribution of climate damage while providing incentives for countries to reduce their emissions, possibly up to inducing full efficiency. Formally, this amounts to compensating every country for the per capita damage it incurs, $d_i(X)$, while setting up vectors of per capita transfer payments, $t_i(n, b, d, x^p, x^c)$, to finance the total amount compensated: $\sum_i n_i t_i(n, b, d, x^p, x^c) = D(X)$. The per capita

[12] We formulate the model on a per-capita basis in order to account for the relative sizes of the various countries. Given the large heterogeneity in country size across the globe, per-country considerations would bias the analysis.

[13] For simplicity, we assume n_i to be constant over time. While this assumption is unrealistic, its consequence is mainly one of accounting. Our goal is to focus on the role of relevant and irrelevant characteristics. We refer the reader to our concluding discussion, where we argue that population is a "neutral" variable, i.e., belonging to neither category.

payoff of country i is then $b_i(x_i^p, x_i^c) - t_i(n, b, d, x^p, x^c)$. We are interested in transfer functions, $t : \mathcal{P} \to \mathbb{R}^n$ which (potentially) hold countries responsible for their past and current emissions levels, x_i^p and x_i^c, after fully washing out the damage suffered, d_i.

Throughout the paper, we consider that countries are not responsible for their damage function, d_i, nor their benefit function b_i. We justify the former by the fact that the distribution of damage depends on natural phenomena about which nothing can be done. Also, we argue that even though the location of population settlements are endogenous, the decision to locate to potentially vulnerable areas (should the sea level rise significantly, for instance) were made long before climate change was ever a concern. Similarly, it can be argued that the vast majority of differences in productivity, the b_i's, are also due to geographical factors (the presence of natural resources, climate, etc.). Also, while migrations are possible, one may argue that no one is responsible for where they are born. Thus we shall consider differences in the b_i's to be outside of any individuals' responsibility.[14]

We contrast several views of responsibility regarding emissions. First, since the cost of global warming depends upon total emissions, one may argue that countries should be held responsible for all of their emissions, both past and current. We shall call this view *Historical Responsibility* (hereafter **HR**). Second, one may argue instead that countries should not be held responsible for emissions that go back to a time when the impact of emissions on climate change had not been suspected. According to this view, past emissions are irrelevant and countries should be held responsible only for current emissions levels $x^c = (x_1^c, ..., x_m^c)$. We refer to it as *no Historical Responsibility* (here-

[14] Likewise, today's production function, $b_i(x_i^p, \cdot)$, also partly results from temporal factors (decisions made by previous generations). Hence, views which do not hold individuals responsible for past emissions, x_i^p, also take the stance that individuals are not responsible for *when* they are born.

after **nHR**). Third, past emissions may be considered a natural benchmark to measuring countries' "needs". According to this so-called *Grandfathering* view (hereafter **GF**), countries are held responsible for variations between current and past emissions levels $x^{GF} = (x_1^{GF}, ..., x_m^{GF})$, where $x_i^{GF} = x_i^c - \gamma x_i^p$, for some time-rescaling parameter γ. For expositional purposes, the main body of the text is framed according to the **nHR** viewpoint.[15]

Prior to further investigating the concepts of responsibility and compensation, we impose a minimal fairness requirement: Solidarity asks that no country benefits from seeing the total damage suddenly increase, all else equal.[16]

Axiom 1 (Solidarity) *For any* $P = (n, b, d, x^p, x^c)$ *and* $P' = (n, b, d', x^p, x^c)$ *such that* $D' \geq D$*, and all* $i \in S$*:*

$$b_i(x_i^p, x_i^c) - t_i(P') \leq b_i(x_i^p, x_i^c) - t_i(P).$$

III. Responsibility and compensation[17]

Penalizing (or rewarding) differences in relevant characteristics

If countries are considered to be responsible for—at least some of—their emissions, differences in these emissions should affect their final payoffs. In fact, a strong interpretation of responsibility consists in holding countries fully responsible for the damage they contribute to causing via "relevant" emissions, x_i^c, for which they are considered responsible, irrespective of past emissions, x_i^p, for which they are not deemed responsible, and other "irrelevant" characteristics

[15] Note that this does not imply that we consider this viewpoint to be more reasonable or normatively more appealing than the others. The corresponding results obtained under **HR** and **GF** views of responsibility can be found in Section V..

[16] See, e.g., Thomson (2003) for a comprehensive survey of the use of these standard axioms in the cost-sharing literature.

[17] The terminology often used in the liberal egalitarianism literature contrasts the notion of compensation with that of reward. In order to avoid confusion, due to the negative connotation associated with carbon emissions, we chose to avoid the term "reward" and to abuse terminology slightly by using the term "responsibility" instead.

as summarized by the function b_i. Thus, a first approach to responsibility consists in arguing that whatever the distribution of irrelevant characteristics, changes in a country's relevant characteristics should affect only this country. This yields:

Axiom 2 (FMR) ***Full Marginal Responsibility:***
Consider a change from P to $\hat{P}$ where some country i's emissions change from (x_i^p, x_i^c) to $(x_i^p, \hat{x}_i^c)$, all else equal[18]*, then:*

$$n_i(t_i(\hat{P}) - t_i(P)) = D(\hat{X}) - D(X),$$

and $t_j(\hat{P}) = t_j(P)$ for all $j \neq i$.

Full Marginal Responsibility is a very demanding axiom as it requires each country to pay the full damage of its own emissions. In fact, Proposition 1 establishes that it is not possible to disentangle one country's responsibility from that of the others. The presence of nonlinearity forbids full separation of responsibilities because the notion of marginal contributions to damage becomes blurry.[19]

Proposition 1 *No cost-sharing mechanism satisfies* FMR *unless the damage function, D, is linear in total emissions.*

Proof. Let $P = (n, b, d, x^p, x^c) \in \mathcal{P}$ and let $\alpha, \beta \neq 0$ such that $\alpha + \beta \neq 0$. Consider the global warming problem $\hat{P} = (n, b, d, x^p, \hat{x}^c)$ with $\hat{x}^c = x^c + (\alpha, \beta, 0, ..., 0)$. Denote by $P_\alpha = \left(n, b, d, x^p, x^c_{[\alpha]}\right)$ and $P_\beta = \left(n, b, d, x^p, x^c_{[\beta]}\right)$ the "interim" global warming problems such that $x^c_{[\alpha]} = x^c + (\alpha, 0, ..., 0)$ and

[18] We opted for an informal statement for the sake of readability. Formally, the statement should read: "For any $i \in S$, and any P, $\hat{P} \in \mathcal{P}$ such that $(\hat{n}, \hat{b}, \hat{d}) = (n, b, d)$ and $(\hat{x}_j^p, \hat{x}_j^c) = (x_j^p, x_j^c)$ for all $j \in S \backslash \{i\}$, the following holds:..."

[19] This is also true of traditional Pigovian taxation, where the tax rate also depends on total emissions, through marginal damage, when D is not linear.

$x^c_{[\beta]} = x^c + (0, \beta, 0, ..., 0)$, respectively.

By *FMR* applied to P and P_α,

$$n_1(t_1(P_\alpha) - t_1(P)) = D(X + n_1\alpha) - D(X),$$

and $t_i(P_\alpha) = t_i(P)$ for all $i \neq 1$. Next, by applying *FMR* to P_α and $\hat{P}$, it follows that $t_i\left(\hat{P}\right) = t_i(P)$ all $i \neq 1, 2$ and:

$$\begin{aligned} n_1(t_1(\hat{P}) - t_1(P)) &= D(X + n_1\alpha) - D(X), \\ n_2(t_2(\hat{P}) - t_2(P)) &= D(X + n_1\alpha + n_2\beta) - D(X + n_1\alpha). \end{aligned}$$

Similarly, by applying *FMR* from P to $\hat{P}$ via P_β, we get $t_i\left(\hat{P}\right) = t_i(P)$ all $i \neq 1, 2$ and:

$$\begin{aligned} n_1(t_1(\hat{P}) - t_1(P)) &= D(X + n_1\alpha + n_2\beta) - D(X + n_2\beta), \\ n_2(t_2(\hat{P}) - t_2(P)) &= D(X + n_2\beta) - D(X). \end{aligned}$$

The above implications are only compatible if

$$D(X + n_1\alpha) - D(X) = D(X + n_1\alpha + n_2\beta) - D(X + n_2\beta).$$

Linearity of D follows. ■

This leads us to considering a less demanding axiom, which insists on assigning marginal responsibilities only when damage is linear:

Axiom 3 (LMR) ***Linear Marginal Responsibility:***

Suppose the damage function D is linear. Consider a change from P to $\hat{P}$ where some country's emissions change from (x^p_i, x^c_i) to $(x^p_i, \hat{x}^c_i)$, all else equal. Then

$t_j(\hat{P}) = t_j\left(P\right)$ *for all* $j \neq i$*, and:*

$$n_i(t_i(\hat{P}) - t_i\left(P\right)) = D(\hat{X}) - D\left(X\right).$$

Solidarity and *LMR* together imply that one should charge countries a per-unit cost equal to the average global damage.

Proposition 2 LMR *and* Solidarity *imply average damage pricing:*

$$t_i\left(P\right) = x_i \frac{D\left(X\right)}{X}.$$

Proof. Let $P = (n, b, d, x^p, x^c) \in \mathcal{P}$, define $AD : \mathbb{R}_+ \to \mathbb{R}$, $\hat{X} \mapsto \left[D(X)/X\right]\hat{X}$, the linear function determined by the average damage at X, and consider the following functions:

$$\begin{aligned}\hat{D}(\cdot) &= \sup\{D, AD\}, \quad \text{and} \\ \check{D}(\cdot) &= \inf\{D, AD\}.\end{aligned}$$

By construction, $\check{D} \leq D, AD \leq \hat{D}$ with $\check{D}(0) = \hat{D}(0) = 0$ and $\check{D}(X) = \hat{D}(X) = D(X)$. By budget balance and *Solidarity*, these inequalities imply that transfers should be the same whether the damage function is D or AD. Lastly, the result follows by applying *LMR* to the fact that AD is linear. ∎

Hence, it would seem that the rather weak axiom of *LMR* points to a strong characterization of average damage pricing. While formally correct, we argue that this result relies heavily on a reference level (zero emissions), which we deem inappropriate in our context: granting such a special role to the unattainable (and undesirable) outcome of zero emissions seems ill-suited to building a theory meant to handle realistic emissions levels. Instead, we allow for the planner to decide on the appropriate reference emissions level $\tilde{X}$. In practice, this reference level, $\tilde{X}$, can be thought of as a target emissions level. In that case, the ratio $\frac{D(X)-D(\tilde{X})}{X-\tilde{X}}$ approximates the notion of marginal damage *around* the reference level $\tilde{X}$. The corresponding axiom, *Full Reference Responsibility* asks that countries be held responsible for departures from the reference level on a per capita basis:

Axiom 4 (FRR) ***Full Reference Responsibility:***

Let $\tilde{x}^c \in \mathbb{R}_+$ be a reference level of current per capita emissions and define global reference emissions accordingly: $\widetilde{X}^c = N\tilde{x}^c$.

For any $P = (n, b, d, x^p, x^c) \in \mathcal{P}$, denote $\tilde{P} = (n, b, d, x^p, \tilde{x}^c \cdot \mathbf{1}^m) \in \mathcal{P}$, where $\mathbf{1}^m$ stands for the m-unit vector. For all $i \in S$:

$$t_i(P) - t_i\left(\tilde{P}\right) = (x_i^c - \tilde{x}^c)\left[\frac{D(X) - D\left(X^p + \tilde{X}^c\right)}{X - (X^p + \tilde{X}^c)}\right].$$

Another interpretation of responsibility is that each country must pay the *incremental* damage it imposes onto the rest of society:[20]

$$n_i(t_i(P) - t_i(P|_{x_i^c=0})) = D(X) - D(X - n_i x_i^c).$$

Holding countries responsible only for their relevant characteristics implies that the transfer $t_i(P|_{x_i^c=0})$ should not depend on characteristics for which countries are not deemed responsible. A natural way to do this is to require that countries that only differ in their relevant characteristics provide an equal net contribution to the financing of global damage $D(X)$.

Axiom 5 (ECEIC) ***Equal Contribution for Equal Irrelevant Characteristics***[21]

[20] The shorthand notation $P|_{x_i^c=0}$ designates a global warming problem identical to P in every way except for country i's current emissions, which are zero. Recall that we are adopting the **nHR** view of responsibility. For instance, under **HR**, the analog would be $P|_{x_i=(0,0)}$

[21] This axiom, and others considered here, result from adapting axioms found in Bossert (1995) and Bossert and Fleurbaey (1996) to a context with externalities.

For all $i,j \in S: \; [b_i = b_j, x_i^p = x_j^p] \implies$

$$t_i(P) - \frac{D(X) - D(X - n_i x_i^c)}{n_i} = t_j(P) - \frac{D(X) - D(X - n_j x_j^c)}{n_j}.$$

ECEIC will prove too strong an axiom in what follows. Hence, we consider a weaker version that only requires an equal net contribution when all countries' irrelevant characteristics are identical to a reference.

Axiom 6 (ECRIC) ***Equal Contribution for Reference Irrelevant Characteristics.***
Consider a reference benefit function, $\tilde{b}$*, and a reference level for past emissions,* $\tilde{x}^p$*. Denote* $\widetilde{X}^p = N\tilde{x}^p$*.*
If $b_i = \tilde{b}$ *and* $x_i^p = \tilde{x}^p$ *for all* $i \in S$*, then for all* $i,j \in S:$

$$t_i(P) - \frac{D\left(\tilde{X}^p + X^c\right) - D\left(\tilde{X}^p + X^c - n_i x_i^c\right)}{n_i} = t_j(P) - \frac{D\left(\tilde{X}^p + X^c\right) - D\left(\tilde{X}^p + X^c - n_j x_j^c\right)}{n_j}.$$

Remark 1 *Clearly,* FMR *is a more demanding axiom than* FRR*, and* $ECEIC$ *is more demanding than* $ECRIC$*. However, no such relationship exists between* FRR *and* $ECEIC$*.*[22]

Compensating differences in irrelevant characteristics

A first approach to dealing with the issue of compensation consists in arguing that differences in irrelevant characteristics should *not* drive individual welfare. In other words, only differences in relevant characteristics should matter.

The above argument can be interpreted to mean that all should equally

[22] In the case of linear damage, however, FMR coincides with FRR and implies ECEIC.

suffer—or benefit—as a result of a change in one country's irrelevant characteristic:

Axiom 7 (GSIC) ***Group Solidarity towards Irrelevant Characteristics.*** *Consider a change from* b_i *to* $\hat{b}_i$ *and from* x_i^p *to* $\hat{x}_i^p$*, for some* $i \in S$*, all else equal. Denote* $P = (n, b, d, x^p, x^c)$ *and* $\hat{P} = (n, \hat{b}, d, \hat{x}^p, x^c)$*, then for all* $j \in S \backslash \{i\}$ *:*

$$t_j(P) - t_j\left(\hat{P}\right) = \left[\hat{b}_i(\hat{x}_i^p, x_i^c) - t_i\left(\hat{P}\right)\right] - \left[b_i(x_i^p, x_i^c) - t_i(P)\right].$$

Another interpretation of compensation consists in requiring that citizens of two countries with identical *relevant* characteristics should end up with the same payoff:

Axiom 8 (EPER) ***Equal Payoff for Equal Responsibility.*** *For all* $i, j \in S:$ $[x_i^c = x_j^c \implies b_i(x_i^p, x_i^c) - t_i(P) = b_j(x_j^p, x_j^c) - t_j(P)]$.

Finally, a considerably weaker version of the above axiom requires final payoff equality only in when the relevant characteristics of all countries are equal to a given reference level.

Axiom 9 (EPRR) ***Equal Payoff for Reference Responsibility.*** *Let* $\tilde{x}^c \in \mathbb{R}_+$ *be a reference level for current per capita emissions. If* $x_i^c = \tilde{x}^c$*, for all* $i \in S$*, then, for all* $i, j \in S$ *:*

$$b_i(x_i^p, x_i^c) - t_i(P) = b_j(x_j^p, x_j^c) - t_j(P).$$

Remark 2 *Clearly, EPER is more demanding than EPRR. However, no such relationship between GSIC and the other two axioms. However, if population had been considered an "irrelevant" characteristic then GSIC would imply EPER, as in Fleurbaey and Bossert (1996).*

Tension between compensation and responsibility

As it turns out, it is generally impossible to compensate countries for differences in irrelevant characteristics while penalizing or rewarding them for differences in relevant characteristics, at least in the strong interpretation of these concepts. In fact, even when D is linear, *FRR* (or *FMR*) and *GSIC* are incompatible unless the benefit function is additively separable in countries' relevant and irrelevant characteristics.

Proposition 3 *Suppose D is linear.* GSIC *and* FRR *are incompatible unless the productivity of current emissions is identical across countries: for all $i, j \in S$,*

$$\frac{\partial b_i}{\partial x_i^c}(x_i^p, \cdot) \equiv \frac{\partial b_j}{\partial x_j^c}(x_j^p, \cdot).$$

Proof. The proof technique is adapted from that of Lemma 1 in Bossert (1995). Because D is linear, denote by δ its damage rate: $D(X) = \delta X$. Let $P = (n, b, d, x^p, x^c) \in \mathcal{P}$ and let $\tilde{x}^c \in \mathbb{R}_+$ be some reference emissions level. Throughout the proof, n and d will remain constant, so that we will identify P with its benefit-emissions profile, (b, x^p, x^c). Define

$$\begin{aligned} (b, x^p, x^c)^{1,1} &= [(b_1, x_1^p, x_1^c), (b_1, x_1^p, x_2^c), ..., (b_1, x_1^p, x_m^c)], \\ (b, x^p, x^c)^{1,m} &= (b, x^p, x^c), \end{aligned}$$

and, for all $k \in \{1, ..., m-1\}$:

$$(b, x^p, x^c)^{1,k} = [(b_1, x_1^p, x_1^c), (b_2, x_2^p, x_2^c), ...(b_k, x_k^p, x_k^c), (b_1, x_1^p, x_{k+1}^c), ..., (b_1, x_1^p, x_m^c)].$$

For each $k = 1, ..., m$, denote by $P^{1,k}$ the global warming problem corresponding to $(b, x^p, x^c)^{1,k}$. It follows from FRR that:

$$t_1(P^{1,1}) - t_1(\tilde{P}^1) = \delta(x_1^c - \tilde{x}^c),$$

with $\tilde{P}^1 = (b_1 \cdot \mathbf{1}^m, x_1^p \cdot \mathbf{1}^m, \tilde{x}^c \cdot \mathbf{1}^m)$. Next, applying GSIC for the shift from $P^{1,1}$ to $P^{1,2}$ yields:

$$t_j\left(P^{1,2}\right) - t_j\left(P^{1,1}\right) = \left[b_1(x_1^p, x_2^c) - t_2\left(P^{1,1}\right)\right] - \left[b_2(x_2^p, x_2^c) - t_2\left(P^{1,2}\right)\right], \quad (1)$$

for all $j \neq 2$. By budget balance, we know that

$$\begin{aligned}\sum_{j=1}^{m} n_j t_j\left(P^{1,2}\right) &= \delta X^{1,2} = \delta\left(N x_1^p + n_2\left(x_2^p - x_1^p\right) + X^c\right), \\ \sum_{j=1}^{m} n_j t_j\left(P^{1,1}\right) &= \delta X^{1,1} = \delta\left(N x_1^p + X^c\right),\end{aligned}$$

Thus, by subtracting:

$$\begin{aligned}\sum_{j=1}^{m} n_j\left[t_j\left(P^{1,2}\right) - t_j\left(P^{1,1}\right)\right] &= \delta n_2\left(x_2^p - x_1^p\right) \\ &= (N - n_2)\left[b_1(x_1^p, x_2^c) - b_2(x_2^p, x_2^c)\right] + N\left[t_2\left(P^{1,2}\right) - t_2\left(P^{1,1}\right)\right],\end{aligned}$$

with the last equality following from summing up Expression (1). Hence,

$$t_2\left(P^{1,2}\right) - t_2\left(P^{1,1}\right) = \delta\frac{n_2}{N}\left(x_2^p - x_1^p\right) - \left(1 - \frac{n_2}{N}\right)\left[b_1(x_1^p, x_2^c) - b_2(x_2^p, x_2^c)\right].$$

Expression (1) applied to $j = 1$ yields:

$$\begin{aligned}t_1\left(P^{1,2}\right) - t_1\left(P^{1,1}\right) &= \left[b_1(x_1^p, x_2^c) - b_2(x_2^p, x_2^c)\right] + \delta\frac{n_2}{N}\left(x_2^p - x_1^p\right) \\ &\quad - \left(1 - \frac{n_2}{N}\right)\left[b_1(x_1^p, x_2^c) - b_2(x_2^p, x_2^c)\right] \\ &= \frac{n_2}{N}\left[b_1(x_1^p, x_2^c) - b_2(x_2^p, x_2^c)\right] + \delta\frac{n_2}{N}\left(x_2^p - x_1^p\right).\end{aligned}$$

As a result,

$$t_1\left(P^{1,2}\right) - t_1(\tilde{P}^1) = \delta(x_1^c - \tilde{x}^c) + \delta\frac{n_2}{N}\left(x_2^p - x_1^p\right) + \frac{n_2}{N}\left[b_1(x_1^p, x_2^c) - b_2(x_2^p, x_2^c)\right].$$

Applying iteratively over the shift from $P^{1,k}$ to $P^{1,k+1}$ yields the following:

$$t_1(P) = t_1(\tilde{P}^1) + \delta(x_1^c - \tilde{x}^c) + \delta\sum_{k=2}^{m}\frac{n_k}{N}\left(x_k^p - x_1^p\right) - \sum_{k=2}^{m}\frac{n_k}{N}\left[b_k(x_k^p, x_k^c) - b_1(x_1^p, x_k^c)\right]. \tag{2}$$

Likewise, define the following:

$$\begin{aligned}
(b, x^p, x^c)^{2,1} &= [(b_1, x_1^p, x_1^c), (b_2, x_2^p, x_2^c), ..., (b_2, x_2^p, x_m^c)], \\
(b, x^p, x^c)^{2,m} &= (b, x^p, x^c), \qquad \text{and, for all } k \in \{1, ..., m-1\}: \\
(b, x^p, x^c)^{2,k} &= [(b_1, x_1^p, x_1^c), (b_2, x_2^p, x_2^c), ...(b_k, x_k^p, x_k^c), (b_2, x_2^p, x_{k+1}^c), ..., (b_2, x_2^p, x_m^c)].
\end{aligned}$$

A similar reasoning as above yields the following expression:

$$t_1(P) = t_1(\tilde{P}^2) + \delta(x_1^c - \tilde{x}^c) + \sum_{k=3}^{m}\delta\frac{n_k}{N}\left(x_k^p - x_2^p\right) - \sum_{k=3}^{m}\frac{n_k}{N}\left[b_k(x_k^p, x_k^c) - b_2(x_2^p, x_k^c)\right], \tag{3}$$

with $\tilde{P}^2 = (b_2 \cdot \mathbf{1}^m, x_2^p \cdot \mathbf{1}^m, \tilde{x}^c \cdot \mathbf{1}^m)$.. Combining expressions (2) and (3) leads to:

$$\frac{n_2}{N}\left[b_2(x_2^p, x_2^c) - b_1(x_1^p, x_2^c)\right] = t_1(\tilde{P}^1) - t_1(\tilde{P}^2) + \delta\left(1 - \frac{n_1}{N}\right)\left(x_2^p - x_1^p\right) + \sum_{k=3}^{m}\frac{n_k}{N}\left[b_1(x_1^p, x_k^c) - b_2(x_2^p, x_k^c)\right].$$

The right-hand side of the equation being independent of x_2^c, it follows that $\frac{\partial b_1}{\partial x_2^c}(x_1^p, x_2^c) - \frac{\partial b_2}{\partial x_2^c}(x_2^p, x_2^c) = 0$ for all values of x_2^c. The result follows from repeating the argument with all pairs of countries. ■

Consequently, the only way to reconcile the concepts of compensation and responsibility is to combine GSIC with the other responsibility axioms and FRR with the other compensation axioms. It will turn out that only the combination with the weaker versions of these axioms are viable. We characterize the corresponding mechanisms in the next section.

IV. Two mechanisms

The Egalitarian Equivalent mechanism

The *Egalitarian Equivalent mechanism* redistributes along a reference vector of *irrelevant* characteristics and splits equally the residual climate damage, after each country has paid for its incremental contribution:

Definition 1 *Egalitarian Equivalent (EE) transfer:*
Let $P \in \mathcal{P}$ and consider a reference benefit function, $\tilde{b}$, and a reference level of historical emissions, $\tilde{x}^p$. Then, for any $i \in S$:

$$t_i^{nHR-EE}(P) = b_i\left(x_i^p, x_i^c\right) - \tilde{b}\left(\tilde{x}^p, x_i^c\right) - \sum_{j=1}^m \frac{n_j}{N}\left[b_j\left(x_j^p, x_j^c\right) - \tilde{b}\left(\tilde{x}^p, x_j^c\right)\right]... \qquad \text{(EE1)}$$

$$... + \frac{1}{n_i}\left[D\left(\tilde{X}^p + X^c\right) - D\left(X_{-i}\right)\right] - \frac{1}{N}\left(\sum_{j=1}^m\left[D\left(\tilde{X}^p + X^c\right) - D\left(X_{-j}\right)\right] - D(X)\right) \qquad \text{(EE2)}$$

where $X_{-i} = \tilde{X}^p + X^c - n_i^c x_i^c$.

The Egalitarian Equivalent mechanism is characterized by the combination of *GSIC* and *ECRIC*.

Theorem 1 *A transfer schedule, t, satisfies* GSIC *and* ECRIC *if and only if $t = t^{nHR-EE}$.*

Proof. It is easily checked that t^{nHR-EE} satisfies the required axioms. Conversely, let $P = (n, b, d, x^p, x^c) \in \mathcal{P}$, let $\tilde{b}$ be a reference benefit function and $\tilde{x}^p$ be a reference level of past emissions, and denote $\tilde{P} = (n, \tilde{b} \cdot \mathbf{1}^m, d, \tilde{x}^p \cdot \mathbf{1}^m, x^c)$.

For all $k = 1, ..., m-1$, define

$$P^k = (n, (b_1, ..., b_k, \tilde{b}, ..., \tilde{b}), d, (x_1^p, ..., x_k^p, \tilde{x}^p, ..., \tilde{x}^p), x^c),$$

and let $P^m = P$.

By *ECRIC,* we know that the contribution of country i,

$$t_i\left(\tilde{P}\right) - \frac{D\left(\tilde{X}^p + X^c\right) - D\left(\tilde{X}^p + X^c - n_i x_i^c\right)}{n_i}$$

should not depend upon country i in the global warming problem $\tilde{P}$. By budget balance, it follows that, for all $i \in S$:

$$\begin{aligned} t_i(\tilde{P}) &= \left[D\left(\tilde{X}^p + X^c\right) - D\left(\tilde{X}^p + X^c - n_i x_i^c\right)\right]/n_i ... \\ &... + \frac{1}{N}\left(D(X) - \sum_{j=1...m}\left[D\left(\tilde{X}^p + X^c\right) - D\left(\tilde{X}^p + X^c - n_j x_j^c\right)\right]\right). \end{aligned}$$

Next, switching from global warming problems $\tilde{P}$ to P^1, *GSIC* writes

$$t_i\left(P^1\right) - t_i\left(\tilde{P}\right) = \left[\tilde{b}\left(\tilde{x}^p, x_1^c\right) - t_1\left(\tilde{P}\right)\right] - \left[b_1\left(x_1^p, x_1^c\right) - t_1\left(P^1\right)\right],$$

for all $i \in S\backslash\{1\}$. This yields:

$$\begin{aligned} t_1\left(P^1\right) - t_1\left(\tilde{P}\right) &= \left(1 - \frac{n_1}{N}\right)\left[b_1(x_1^p, x_1^c) - \tilde{b}\left(\tilde{x}^p, x_1^c\right)\right], \quad and\ for\ all\ i \in S\backslash\{1\}: \\ t_i\left(P^1\right) - t_i\left(\tilde{P}\right) &= \frac{n_1}{N}\left[\tilde{b}\left(\tilde{x}^p, x_1^c\right) - b_1\left(x_1^p, x_1^c\right)\right]. \end{aligned}$$

Moving up from P^1 to P^2 and applying again *GSIC* gives

$$\begin{aligned} t_1\left(P^2\right) - t_1\left(P^1\right) &= \frac{n_2}{N}\left[\tilde{b}\left(\tilde{x}^p, x_2^c\right) - b_2\left(x_2^p, x_2^c\right)\right], \\ t_2\left(P^2\right) - t_2\left(P^1\right) &= \left(1 - \frac{n_2}{N}\right)\left[b_2\left(x_2^p, x_2^c\right) - \tilde{b}\left(\tilde{x}^p, x_2^c\right)\right], \end{aligned}$$

so that:

$$t_1\left(P^2\right)-t_1\left(\tilde{P}\right)=b_1\left(x_1^p,x_1^c\right)-\tilde{b}\left(\tilde{x}^p,x_1^c\right)+\sum_{j=1}^{2}\frac{n_k}{N}\left[\tilde{b}\left(\tilde{x}^p,x_j^c\right)-b_j\left(x_j^p,x_j^c\right)\right].$$

Successively applying *GSIC* while moving up to $P^m=P$ yields the result:

$$t_i\left(P\right)-t_i\left(\tilde{P}\right)=b_i\left(x_i^p,x_i^c\right)-\tilde{b}\left(\tilde{x}^p,x_i^c\right)-\sum_{j=1}^{m}\frac{n_j}{N}\left[b_j\left(x_j^p,x_j^c\right)-\tilde{b}\left(\tilde{x}^p,x_j^c\right)\right].$$

■

Remark 3 *The above characterization is tight: the strengthening of ECRIC into ECEIC yields an impossibility. Indeed, the reader can check that the EE solution does not satisfy ECEIC.*

The Conditional Equality mechanism

The *Conditional Equality mechanism* guarantees each agent the average payoff of a hypothetical situation in which all countries' relevant characteristics are equal to a reference level. Each country bears the consequences of any deviation from this reference level. Formally,

Definition 2 *Conditional Equality (CE) transfer:*
Let $P\in\mathcal{P}$ and consider a reference vector of relevant *characteristics $\tilde{x}^c\in\mathbb{R}_+$. For any $i\in S$:*

$$t_i^{nHR-CE}(P)=b_i(x_i^p,\tilde{x}^c)-\tfrac{1}{N}\textstyle\sum_{j=1}^{m}n_jb_j(x_j^p,\tilde{x}^c)... \qquad \text{(CE1)}$$
$$...+\left(x_i^c-\tilde{x}^c\right)\left(\tfrac{D(X)-D(\tilde{X})}{X-\tilde{X}}\right)+\tfrac{1}{N}D(\tilde{X}). \qquad \text{(CE2)}$$

where $\tilde{X}=X^p+N\tilde{x}^c$.

The CE mechanism is characterized by *FRR* and *EPRR*.

Theorem 2 *A transfer schedule, t, satisfies* FRR *and* EPRR *if and only if $t = t^{nHR-CE}$.*

Proof. It is easily checked that t^{nHR-CE} satisfies the required axioms. Conversely, let $P = (n, b, d, x^p, x^c) \in \mathcal{P}$ and consider a reference emissions schedule $\tilde{x}^c \in \mathbb{R}_+$. Denote $\tilde{P} = (n, b, d, x^p, \tilde{x}^c \cdot \mathbf{1}^m)$ and let $\tilde{X} = X^p + N\tilde{x}^c$ be the total emissions level associated with $\tilde{P}$.

By budget balance and *EPRR*, for all $i \in S$:

$$t_i(\tilde{P}) = b_i(x_i^p, \tilde{x}^c) - \frac{1}{N}\left(\sum_{j=1}^{m} n_j b_j(x_j^p, \tilde{x}^c) - D\left(\tilde{X}\right)\right).$$

Next, switching from global warming problem $\tilde{P}$ to P, *FRR* yields:

$$t_i(P) = t_i(\tilde{P}) + (x_i^c - \tilde{x}^c)\left(\frac{D(X) - D\left(\tilde{X}\right)}{X - \tilde{X}}\right),$$

for all $i \in S$. ■

Remark 4 *This characterization is also tight: the reader can check that the CE solution does not satisfy EPER.*

IV..1

Comments on the two schemes

Due to the fact that our model distinguishes damage from benefits, the **EE** and **CE** mechanisms lend themselves very well to interpretation. Specifically, each scheme is made up of a redistribution component (Expressions (EE1) and (CE1), respectively) and a cost-sharing portion (Expressions (EE2) and (CE2), respectively).

In the cost-sharing part of the **EE** mechanism, (EE2), each individual pays the (per capita) incremental damage actually caused by its country's current emissions, $\left[D\left(\tilde{X}^p + X^c\right) - D\left(\tilde{X}^p + X^c - n_i x_i^c\right)\right]/n_i$, plus a (possibly negative) contribution, identical for all, so as to cover total damage, $D(X)$. The implicit reference in computing the incremental damage of a country's current emissions, $n_i x_i^c$, is the situation $\tilde{X}^p + X^c$. The part that is shared equally across individuals carries a particularly easy interpretation when D is linear. Indeed, it is equal to $D\left(\tilde{X}^p\right)$, the damage caused by past emissions, for which, according to the **nHR** view, countries are not considered to be responsible. If D is convex, then $\sum_{j=1}^{m}\left[D\left(\tilde{X}^p + X^c\right) - D\left(\tilde{X}^p + X^c - n_j x_j^c\right)\right] > D\left(\tilde{X}^p + X^c\right) - D\left(\tilde{X}^p\right)$ so that the portion of damage that is equally shared is smaller than $D\left(\tilde{X}^p\right)$. If D is concave, the converse holds true. Note also that, when D is convex, average incremental damage $\left[D\left(\tilde{X}^p + X^c\right) - D\left(\tilde{X}^p + X^c - n_i x_i^c\right)\right]/\left(n_i x_i^c\right)$ is decreasing with $n_i x_i^c$ so that big emitters tend to pay less by unit of emissions. Again, the converse holds true if D is concave.

The redistributive part of the **EE** mechanism, (EE1), also has a clear interpretation. Each country is asked to contribute the difference between its actual benefit and what it would get if it had the reference characteristics $(\tilde{b}, \tilde{x}^p)$ instead of its own. The collected amount of money is redistributed equally across individuals. The nature of the mechanism depends very much upon the implied reference productivity of current emissions, $\tilde{b}(\tilde{x}^p, \cdot)$. If $\tilde{b}\left(\tilde{x}^p, .\right) > b_i\left(x_i^p, .\right)$ for all $i \in S$, all countries receive some money to achieve this ambitious target, a financial support toward equality which is financed by an equal tax burden. On the contrary, if $\tilde{b}\left(\tilde{x}^p, .\right) < b_i\left(x_i^p, .\right)$ for all $i \in S$, all countries provide some money from their "surplus" (as compared to the low reference) which is redistributed equally. In practice, a reference level for $\tilde{b}(\tilde{x}^p, \cdot)$ might be GDP or some more elaborate wealth measure, enhanced by a target Human Development Index, for instance. Alternatively, one may chose a statistical average of worldwide benefits as measured by GDP: $\tilde{b}(\tilde{x}^p, \cdot) = \frac{1}{N}\sum n_i b_i(x_i^p, \cdot)$, which is one of the views we consider in an ongoing empirical project.

IV..1.1

In the cost-sharing part of the **CE** mechanism, (CE2), individuals share equally $D(\tilde{X})$**,** with $\tilde{X}$ now defined as $\tilde{X} = X^p + N\tilde{x}^c$**.** Any departure from the target, $\tilde{x}^c$, is taxed at rate $\frac{D(X)-D(\tilde{X})}{X-\tilde{X}}$. As compared to the **EE** mechanism, the **CE** mechanism tends to share equally a greater fraction of the damage, at least in the case where D is linear. If D is convex, $\left[D(X) - D\left(\tilde{X}\right)\right] / \left(X - \tilde{X}\right)$ is decreasing with $\left(X - \tilde{X}\right)$. Thus, if

$$X - \tilde{X} = \sum_{j=1}^{m} n_j \left(x_j^c - \tilde{x}^c\right) > n_i x_i^c$$

then

$$\left(x_i^c - \tilde{x}^c\right)\left(\frac{D(X) - D\left(\tilde{X}\right)}{X - \tilde{X}}\right) < \frac{1}{n_i}\left[D(X) - D(X_{-i})\right].$$

This says that, if we are far from the target, that is $X >> \tilde{X}$, countries specific (taxation) part tend to be smaller with the **CE** mechanism than with the **EE** mechanism, especially for small emitters. The converse holds true if $X \simeq \tilde{X}$.

A particular case worth considering is the limiting case $X \to \tilde{X}$. In that case $\left[D(X) - D\left(\tilde{X}\right)\right] / \left(X - \tilde{X}\right)$ equals marginal damage. This is why the family of **CE** mechanisms can be seen as encompassing the Pigou Tax. More precisely, if the target $\tilde{X}$ is fixed at the efficient level, any deviation from that efficient level is taxed a its marginal damage and $t_i^{nHR-CE}(P)$ can be seen as a Pigou tax plus a lump-sum redistributive transfer.

V. Other views on responsibility

So far, our analysis has focused on the **nHR** viewpoint. Two other views on responsibility are often put forth in the climate change debate. On the one hand, the **HR** view recognizes one's responsibility for past emissions, and therefore requires that countries having historically emitted the most bear the brunt of the burden of climate policies. On the other hand, the **GF** view considers past emissions as a reflection of a country's needs. In what follows, we unveil the formulation of our main results along these two views.[23]

Historical Responsibility (HR)

By holding countries responsible for both past and current emissions, the **HR** view requires that historically low emitters be compensated for the damage caused by the historically high emitters. In the climate change debate, it is often the argument put forth by advocates of more lenient policies for developing countries. However, it would be an oversimplification to say that the **HR** view unambiguously favors poor countries. Indeed, if historical emissions are viewed as a measure for historical investment, as is often the case in the debate, the **HR** view asks that countries be rewarded for past investments, as opposed to seeing their returns redistributed equally like in the **nHR** framework.

Definition 3 *Egalitarian Equivalent transfers. For any $P \in \mathcal{P}$, any reference benefit function $\tilde{b}$, and any $i \in S$:*

$$\begin{aligned} t_i^{HR-EE}(P) \quad = \quad & \frac{1}{n_i}\left[D\left(X\right) - D\left(X - X_{-i}\right)\right] - \frac{1}{N}\left(\sum_{j=1}^{m}\left[D\left(X\right) - D\left(X - X_{-j}\right)\right] - D(X)\right) ... \\ & ... + b_i\left(x_i^p, x_i^c\right) - \tilde{b}\left(x_i^p, x_i^c\right) - \sum_{j=1}^{m}\frac{n_j}{N}\left[b_j\left(x_j^p, x_j^c\right) - \tilde{b}\left(x_i^p, x_j^c\right)\right], \end{aligned}$$

[23] All of the axioms presented in the **nHR** framework translate readily to the **HR** and **GF** view. Therefore, we only present the corresponding expressions for the EE and CE mechansims. The proofs techniques are similar to those in the **nHR** framework and are omitted for space considerations.

where $X_{-j} = X - n_j x_j$ for all j.

Theorem 3 *A mechanism, t, satisfies **HR-GSIC** and **HR-ECRIC** if and only if $t = t^{HR-EE}$.*

Definition 4 *Conditional Equality transfers. For any $P \in \mathcal{P}$, any reference vector of emissions $(\tilde{x}^p, \tilde{x}^c) \in \mathbb{R}_+^2$, and any $i \in S$:*

$$\begin{aligned} t_i^{HR-CE} &= (x_i - \tilde{x}) \left(\frac{D(X) - D\left(\tilde{X}\right)}{X - \tilde{X}} \right) - \frac{D(\tilde{X})}{N} ... \\ &\quad ... + b_i(\tilde{x}^p, \tilde{x}^c) - \frac{1}{N} \sum_j n_j b_j(\tilde{x}^p, \tilde{x}^c), \end{aligned}$$

where $\tilde{x} = \tilde{x}^p + \tilde{x}^c$ and $\tilde{X} = N\tilde{x}$.

Theorem 4 *A mechanism, t, satisfies **HR-FRR** and **HR-EPRR** if and only if $t = t^{HR-CE}$.*

Grandfathering (GF)

The **GF** view is usually considered as the other extreme stance on responsibility Indeed, it is often associated with favoring developed countries (the historically high emitters) because, by viewing past emissions as a measure of a country's needs, it effectively grants them larger claims on today's atmosphere. Once again, the issue is not so simple. In fact, considering past emissions as a measure of a country's needs amounts to admitting that one is not responsible for them, and that differences in this dimension should be washed out. However, because past emissions are also a measure of technological investment, **GF** also implies that developed countries should not be granted the full return on their technological investment. Formally, **GF** asks that countries are responsible only for variations between past and current emissions: $x_i^{GF} = x_i^c - \gamma x_i^p$, for some time-scaling parameter γ.

Definition 5 *Egalitarian Equivalent transfers. For any $P \in \mathcal{P}$, any reference benefit function, $\tilde{b}$, any reference level of historical emissions, $\tilde{x}^p$, and any $i \in S$:*

$$\begin{aligned} t_i^{GF-EE} &= b_i\left(x_i^p, x_i^c\right) - \tilde{b}\left(\tilde{x}^p, \gamma\left(\tilde{x}^p - x_i^p\right) + x_i^c\right) - \sum_{j=1...m} \frac{n_j}{N}\left[b_j\left(x_j^p, x_j^c\right) - \tilde{b}\left(\tilde{x}^p, \gamma\left(\tilde{x}^p - x_j^p\right) + x_j^c\right)\right]... \\ &\quad ... + \frac{1}{n_i}\left[D\left(\tilde{X}^p + \gamma\left(\tilde{X}^p - X^p\right) + X^c\right) - D\left(X_{-i}\right)\right]... \\ &\quad ... - \frac{1}{N}\left(\sum_{j=1...m}\left[D\left(\tilde{X}^p + \gamma\left(\tilde{X}^p - X^p\right) + X^c\right) - D\left(X_{-j}\right)\right]\right), \end{aligned}$$

where $\tilde{X}^p = N\tilde{x}^p$ and $X_{-i} = \tilde{X}^p + \gamma\left(\tilde{X}^p - X^p\right) + X^c - n_i\left(x_i^c - \gamma x_i^p\right)$.

Theorem 5 *A mechanism, t, satisfies **GF-GSIC** and **GF-ECRIC** if and only if $t = t^{GF-EE}$.*

Definition 6 *Conditional Equality transfers. For any $P \in \mathcal{P}$, any reference "Grand-fathering" emissions level $\tilde{x}^{GF} \in \mathbb{R}_+$, and any i,*

$$\begin{aligned} t^{GF-CE} &= \left[x_i^{GF} - \tilde{x}^{GF}\right]\left(\frac{D\left(X\right) - D\left(\tilde{X}\right)}{X - \tilde{X}}\right) - \frac{D(\tilde{X})}{N}... \\ &\quad ... + b_i(x_i^p, \gamma x_i^p + \tilde{x}^{GF}) - \frac{1}{N}\left(\sum_{j=1...m} b_j(x_j^p, \gamma x_j^p + \tilde{x}^{GF}) - D\left(\tilde{X}\right)\right), \end{aligned}$$

where $x_i^{GF} = x_i^c - \gamma x_i^p$ and $\tilde{X} = \left(1+\gamma\right)X^p + N\tilde{x}^{GF}$.

Theorem 6 *A mechanism, t, satisfies **GF-FRR** and **GF-EPRR** if and only if $t = t^{GF-CE}$.*

VI. Conclusion

The following table, where x stands for incompatibility, summarizes the relationship between the axioms considered thus far.

Table 1			
Axioms	GSIC	EPER	EPRR
FMR	x	x	x
FRR	x	x	t^{CE}
ECEIC	x	x	x
ECRIC	t^{EE}	x	x

As in the theory on responsibility and compensation formalized by Bossert (1995) and Bossert and Fleurbaey (1996), the Egalitarian Equivalent and the Conditional Equivalent solutions play a key role. These findings confirm the importance of these solutions, even in settings where externalities are present.

However, unlike in Bossert and Fleurbaey (1996), the (equivalent of) axioms ECEIC and ECRIC are generally incompatible with both EPER and EPRR.[24] This is due to the fact that we have not applied our responsibility/compensation axioms to all variables. Indeed, we do not deem individuals responsible for the population of the country they belong to (i.e., population is not a relevant characteristic), nor do we believe that the consequences of differences in populations warrant any compensation (i.e., population is not an irrelevant characteristic either). As such, population can be thought of as a "neutral" characteristic. Yet, population is a crucially important characteristic for the problem at hand because, in practice, most of the per capita data available are computed from aggregate country data. These practical considerations point to the necessity of formally introducing a third type of characteristics—"neutral" characteristics—in the axiomatic analysis to responsibility and compensation. To the best of our knowledge, this has yet to be done.

[24]It suffices to show that ECRIC and EPRR are incompatible. (Appendix A)

References

[1] Baumol, W.J. (1972), On Taxation and the Control of Externalities, *American Economic Review 62* 307-322.

[2] Billette de Villemeur, E. and Leroux, J. (2010), Looking ahead versus looking back: Revisiting the carbon tax, mimeo.

[3] Bossert, W. (1995), Redistribution mechanisms based on individual characteristics, *Mathematical Social Sciences 29* 1-17.

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

[5] Cappelen, A. and Tungodden, B.(2003), Reward and responsibility:how should we be affected when others change their effort?, *Politics Philosophy & Economics 2* 191-211.

[6] Cappelen, A, and Tungodden, B. (2007), Local autonomy and interregional equality, *Social Choice and Welfare 28* 443–460.

[7] Fleurbaey, M. (1995), Equality and Responsibility, *European Economic Review 39* 683-689.

[8] Fleurbaey, M. (2008), *Fairness, Responsibility and Welfare*, Oxford University Press.

[9] Johansson-Stenman, O., and Konow, J.(2010), Fair Air: Distributive Justice and Environmental Economics, *Environmental and Resource Economics 46* 147-166.

[10] Leroux, J. (2004), "Strategyproofness and efficiency are incompatible in production economies", *Economics Letters 85*, 335-340.

[11] Nordhaus, W.D. (1992), The ecology of markets, *Proceedings of the. National Academy of Science 89* 843-850.

[12] Pigou, A.C. (1932), *The Economics of Welfare*, 4th ed., London.

[13] Rawls, J. (1971) *A Theory of Justice*, Harvard U. Press, Cambridge.

[14] Roemer, J.E. (1998) *Equality of Opportunity*, Harvard U. Press, Cambridge.

[15] Sen, A. (1985) *Commodities and Capabilities*, North Holland, Amsterdam.

[16] Stern, N. (2008), The Economics of Climate Change, *American Economic Review: Papers and Proceedings 98* 1-37.

[17] Thomson, W. (2003), Axiomatic and game-theoretic analysis of bankruptcy and taxation problems: a survey, *Mathematical Social Sciences.45* 249-297.

[18] Tungodden, B. (2005),Responsibility and redistribution: The case of first-best taxation, *Social Choice and Welfare 24* 33-44.

A Incompatibility between ECRIC and EPRR

Proposition 4 *EPRR and ECRIC are incompatible unless D is linear.*

Proof. Suppose, t, satisfies both EPRR and ECRIC. Let $\tilde{x}^p$, $\tilde{x}^c$ and $\tilde{b}$ be the reference emissions levels and the reference benefit function, respectively. Consider a global warming problem where $x^p = \tilde{x}^p \cdot \mathbf{1}^m$, $x^c = \tilde{x}^c \cdot \mathbf{1}^m$ and $b = \tilde{b} \cdot \mathbf{1}^m$. Invoking EPRR and ECRIC together yields, for all $i, j \in S$:

$$b_i(x_i^p, x_i^c) - \frac{D(\tilde{X}^p + X^c) - D(\tilde{X}^p + X^c - n_i x_i^c)}{n_i}$$
$$= b_j(x_j^p, x_j^c) - \frac{D(\tilde{X}^p + X^c) - D(\tilde{X}^p + X^c - n_j x_j^c)}{n_j},$$

which, upon recalling that $b_i(x_i^p, x_i^c) = b_j(x_j^p, x_j^c) = \tilde{b}(\tilde{x})$ yields:

$$\frac{D(\tilde{X}) - D(\tilde{X} - n_i\tilde{x}^c)}{n_i} = \frac{D(\tilde{X}) - D(\tilde{X} - n_j\tilde{x}^c)}{n_j}.$$

Unless D is linear, the above equality generically does not hold when $n_i \neq n_j$.

■

B Comparison with Bossert-Fleurbaey (Additional appendix not intended for final publication)

The theoretical framework of Bossert and Fleurbaey (1996, hereafter BF96) divides agents' characteristics, $a = (a^R, a^S)$, into those for which agents are responsible, a^R, and those which are considered irrelevant, a^S. Their setting is one of redistribution of individual wealth, $f(a_i)$, into post-transfer wealth, $F_i(a^R, a^S) = f(a_i) - t_i(a)$ such that the transfers sum to zero: $\sum_i t_i(a^R, a^S) = 0$.

By contrast, our environment is one where transfers do not sum up to zero, but include contributions towards paying the costs of a damage resulting from the combined actions of the agents. Specifically, $\sum_i t_i(P) = D(X)$. The fact that D is not constant and not separable in the x_i's is the key ingredient of the externality problem.

In this appendix, we compare our framework with that of BF96. Our approaches are very similar in spirit, and we shall highlight the similarities between our two settings. While our net payoff after redistribution, $b_i(x_i^p, x_i^c) - t_i(P)$, clearly corresponds to their $F_i(a)$, there exists no obvious equivalent for their $f(a_i)$ in our setting, because of the externality problem. In other words, the presence of externalities does not allow for a simple rewriting of the problem for an appropriately chosen f, adapted to take damage into account. In fact, as will become clear, the "natural" extensions of the axioms in BF96 imply different expressions for f, depending on the axiom, which is a clear sign of the added complexity brought about by the externality problem.

The EE mechanism

In BF96, the egalitarian equivalent mechanism is defined by:

$$F_i^{BF-EE}(a) := f\left(a_i^R, \widetilde{a}^S\right) - \frac{1}{n}\sum_{j=1}^{n}\left[f\left(a_j^R, \widetilde{a}^S\right) - f\left(a_j\right)\right].$$

For the ECRIC axiom, which characterizes the EE mechanism together with GSIC, we posited that:
If $b_i = \tilde{b}$ and $x_i^p = \tilde{x}^p$ for all i, then, for all $i, j \in S$:

$$t_i\left(P\right) - \frac{D\left(\tilde{X}^p + X^c\right) - D\left(X_{-i}\right)}{n_i} = t_j\left(P\right) - \frac{D\left(\tilde{X}^p + X^c\right) - D\left(X_{-j}\right)}{n_j}, \quad (4)$$

where $X_{-i} = \tilde{X}^p + X^c - n_i^c x_i^c$ and X_{-j} is defined accordingly. The equivalent axiom in BF96 is the ETRS axiom:
If $a_i^S = \widetilde{a}^S$ for all i, then:

$$F_i(a) = f\left(a_i\right).$$

Clearly, this requirement would not be feasible in our setting because the transfers would not cover the damage, $D(X)$. However, a less demanding requirement yet in the same spirit as ETRS is the following. If $a_i^S = \widetilde{a}^S$ for all $i \in S$, then:

$$F_i\left(a\right) - f\left(a_i\right) = F_j\left(a\right) - f\left(a_j\right), \quad \text{for all } i, j \in S$$

With our notation, this expression can be rewritten as:
If $b_i = \tilde{b}$ and $x_i^p = \tilde{x}^p$ for all $i \in S$, then:

$$b_i\left(x_i^p, x_i^c\right) - t_i(P) - f_i\left(P\right) = b_j\left(x_i^p, x_i^c\right) - t_j(P) - f_j\left(P\right) \qquad \text{for all } i, j \in S$$

which, in turn, can be rewritten as:

$$t_i(P) - \left[b_i\left(x_i^p, x_i^c\right) - f_i\left(P\right)\right] = t_j(P) - \left[b_j\left(x_i^p, x_i^c\right) - f_j\left(P\right)\right].$$

Comparing with Expression (4), our definition of ECRIC implies defining f,up to a constant, as follows:

$$f\left(P\right)=b_{i}\left(x_{i}^{p},x_{i}^{c}\right)-\frac{D\left(\tilde{X}^{p}+X^{c}\right)-D\left(X_{-i}\right)}{n_{i}},$$

where $X_{-i}=\tilde{X}^{p}+X^{c}-n_{i}^{c}x_{i}^{c}$. Note that the reference level of past emissions enters in this reconstructed production function.

Now, if we apply this definition for the individual production function in the egalitarian equivalent mechanism of BF96, we get the following net payoff:

$$\begin{aligned} F_{i}^{BF-EE}\left(P\right) &= \tilde{b}\left(\tilde{x}^{p},x_{i}^{c}\right)-\frac{1}{n_{i}}\left[D\left(\tilde{X}^{p}+X^{c}\right)-D\left(X_{-i}\right)\right] \\ &\quad -\sum_{j=1}^{m}\frac{n_{j}}{N}\left[\tilde{b}\left(\tilde{x}^{p},x_{j}^{c}\right)-\frac{D\left(\tilde{X}^{p}+X^{c}\right)-D\left(X_{-j}\right)}{n_{j}}\right] \\ &\quad +\sum_{j=1}^{m}\frac{n_{j}}{N}\left[b_{j}\left(x_{j}^{p},x_{j}^{c}\right)-\frac{D\left(X\right)-D\left(X-n_{j}x_{j}^{c}\right)}{n_{j}}\right], \qquad (5) \end{aligned}$$

which is to be compared to:

$$\begin{aligned} b_{i}\left(x_{i}^{p},x_{i}^{c}\right)-t_{i}^{nHR-EE}(P) &= \tilde{b}\left(\tilde{x}^{p},x_{i}^{c}\right)-\frac{1}{n_{i}}\left[D\left(\tilde{X}^{p}+X^{c}\right)-D\left(X_{-i}\right)\right] \\ &\quad -\sum_{j=1}^{m}\frac{n_{j}}{N}\left[\tilde{b}\left(\tilde{x}^{p},x_{j}^{c}\right)-\frac{D\left(\tilde{X}^{p}+X^{c}\right)-D\left(X_{-j}\right)}{n_{j}}\right] \\ &\quad +\sum_{j=1}^{m}\frac{n_{j}}{N}b_{j}(x_{j}^{p},x_{j}^{c})-\frac{D(X)}{N}. \qquad (6) \end{aligned}$$

The "distributive part" (i.e., the terms in b) are identical under both schemes. The only difference between the net payoffs is related to the "cost-sharing part" (i.e., the terms in D), where $\sum_{j=1}^{m}[D\left(X\right)-D(X-n_{j}x_{j}^{c})]$ in Expression (5) becomes $D(X)$ in our setting (Expression 6). This discrepancy distinguishes the importance of the externality in at least two ways. First, the profile of net payoffs, $F_i^{BF-EE}(P)$, is not budget balanced whereas the transfers $t_i^{nHR-EE}(P)$ sum up to global damage, by construction. More precisely, the BF96 transfers do not take into consideration the part of damage due to irrelevant emissions. This is more easily seen in the case of linear damage, say $D(X) = \delta X$, where $\sum_{j=1}^{m} [D(X) - D(X - n_j x_j^c)] = \delta X^c$. Transfers sum to the "relevant" damage, that are shared equally.

Second, the redistributive part rests upon an "individualization" of the production function *à la* BF96, which attributes to each country the damage $D\left(\tilde{X}^p + X^c\right) - D\left(\tilde{X}^p + X^c - n_i^c x_i^c\right)$. More technically, one should also observe that ECRIC is *not* a simple reformulation of the "equivalent" axiom (ETRS) in BF96. Indeed, the first entails a relationship between transfers of different countries while the second specifies circumstances in which there should be no redistribution. In other words, while we share the spirit and the approach of BF96, our axioms are different, and, unless $D(\cdot)$ is separable in the x_i's, there is no simple relationship that link their work with ours due to the presence of the externality.

The CE mechanism

In BF96, the Conditional Equality mechanism is defined by:

$$F_i^{BF-CE}(a) := f(a_i) - f\left(\widetilde{a}^R, a_i^S\right) + \frac{1}{n} \sum_{j=1}^{n} f\left(\widetilde{a}^R, a_j^S\right).$$

In our context, the FRR axiom which (together with EPRR) characterizes the CE mechanism posits that, given $\tilde{x}^c \in \mathbb{R}_+$:
For any $P = (n, b, d, x^p, x^c) \in \mathcal{P}$, define $\tilde{P} = (n, b, d, x^p, \tilde{x}^c \cdot \mathbf{1}^m) \in \mathcal{P}$, where $\mathbf{1}^m$ stands for the m-unit vector. It must be the case that:

$$t_i(P) - t_i\left(\tilde{P}\right) = (x_i^c - \tilde{x}^c)\left[\frac{D(X) - D\left(\tilde{X}\right)}{X - \tilde{X}}\right], \quad \text{for all } i.$$

The equivalent axiom in BF96 is EIRR:

If $a_i^R = \widetilde{a}^R$, for all i, then

$$F_i(a) = F_j(a), \quad \text{for all } i, j.$$

In our setting, this rewrites as follows:

If $x_i^c = \tilde{x}^c$ for all i, then

$$b_i(x_i^p, x_i^c) - t_i(P) = b_j\left(x_j^p, x_j^c\right) - t_j(P), \quad \text{for all } i, j.$$

Note that this axiom is based on intra-profile comparisons, while FRR is based on comparing outcomes with those of a reference profile. The above expression coincides with our definition of FRR for the following individual production function:

$$f_i(P) = b_i(x_i^p, x_i^c) - (x_i^c - \tilde{x}^c)\left[\frac{D(X) - D\left(\tilde{X}\right)}{X - \tilde{X}}\right].$$

Observe that this individual production function differs from that obtained in the discussion of the EE mechanism. If we apply this definition to $F_i^{BF-CE}(a)$ we find:

$$\begin{aligned} F_i^{BF-CE}(P) \quad : \quad &= b_i(x_i^p, x_i^c) - b_i(x_i^p, \tilde{x}^c) + \sum_{j=1}^{m} \frac{n_j}{N} b_j\left(x_j^p, \tilde{x}^c\right) \\ &- (x_i^c - \tilde{x}^c)\left[\frac{D(X) - D\left(\tilde{X}\right)}{X - \tilde{X}}\right], \end{aligned}$$

where $\tilde{X} = X^p + N\tilde{x}^c$.

In our setting the Conditional Equality net payoff writes as follows:

$$\begin{aligned} b_i(x_i^p, x_i^c) - t_i^{nHR-CE}(P) &= b_i(x_i^p, x_i^c) - b_i(x_i^p, \tilde{x}^c) + \sum_{j=1}^{m} \frac{n_j}{N} b_j(x_j^p, \tilde{x}^c) \\ &\quad - (x_i^c - \tilde{x}^c) \left(\frac{D(X) - D\left(\tilde{X}\right)}{X - \tilde{X}} \right) - \frac{1}{N} D(\tilde{X}). \end{aligned}$$

The only difference between the net payoffs is the absence of the term $D(\tilde{X})/N$, where $X^p + N\tilde{x}^c$. Observe that, given the rewriting of f, the damage for which a country is not considered to be responsible (i.e., $D(\tilde{X})$) are not accounted for in their "production function", i.e.. $b_i(x_i^p, x_i^c) - (x_i^c - \tilde{x}^c)\left[\frac{D(X)-D(\tilde{X})}{X-\tilde{X}}\right]$. The redistribution process bears upon $D(X) - D\left(\tilde{X}\right)$ and the remaining part, $D\left(\tilde{X}\right)$, is shared equally.