

July, 2010

“Sharing the Cost of Global Warming”

***Etienne DE VILLEMEUR and
Justin LEROUX***

Sharing the Cost of Global Warming*

Etienne Billette de Villemeur[†] Justin Leroux[‡]

July 2010

Abstract

Due to meteorological factors, the distribution of the environmental damage due to climate change bears no relationship to that of global emissions. We argue in favor of offsetting this discrepancy, and propose a “global insurance scheme” to be financed according to countries’ responsibility for climate change. Because GHG decay very slowly, we argue that the *actual* burden of global warming should be shared on the basis of *cumulated* emissions, rather than sharing the expected costs of actual emissions as in a Pigovian taxation scheme. We characterize new versions of two well-known cost-sharing schemes by adapting the responsibility theory of Bossert and Fleurbaey (1996) to a context with externalities.

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

Keywords: Climate Change, Cost Sharing, Responsibility, Compensation.

*This paper was written while the first author was visiting the Université de Montréal, whose hospitality is gratefully acknowledged. We wish to thank Jérémy Laurent-Lucchetti and Yves Sprumont for stimulating discussions. We are also thankful to seminar assistance at the Ninth International Meeting of the Society of Social Choice and Welfare, the 2008 Canadian Resource and Environment Economics Annual Meeting, as well as the participants to seminars at Université Laval and Ottawa University, to the Montreal Natural Resources and Environmental Economics Workshop and to the Equity in Environmental and Resource Economics Conference (Toulouse, June 09) and to the Equipe seminar (Lille, May 10).

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

[‡]HEC Montréal, CIRANO and CIRPÉE. 3000 chemin de la Côte-Sainte-Catherine, Montréal, QC, H3T 2A7, Canada. E-mail: justin.leroux@hec.ca. *Corresponding author.*

1 Introduction

Nowadays, climate change is a notion pervading our collective human psyche, from policy design to everyday conversations, usually under the more casual designation of "global warming". Accompanying our awareness of climatic 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 resulting from climate change 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 climate change, climate change is a prime example of a global externality.

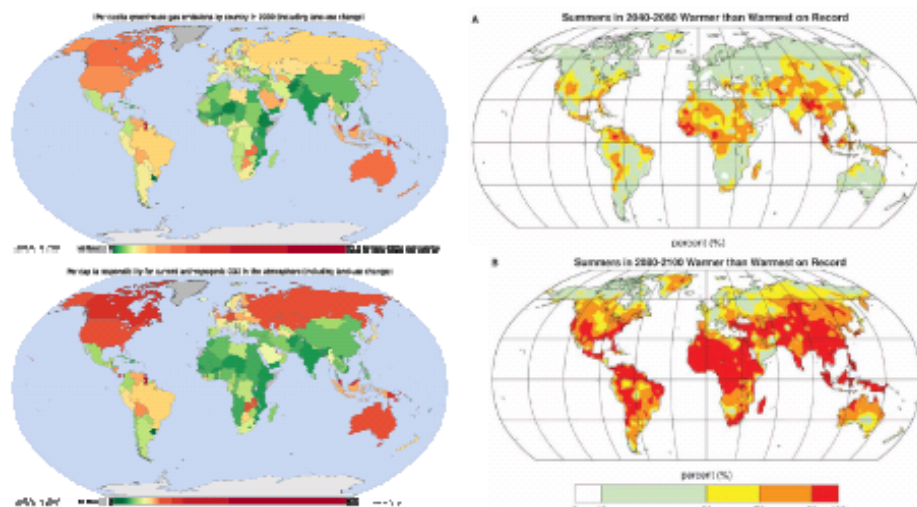


Figure 1. Left, up: per capita GHG emissions by country in 2000; left, down, per capita responsibility in cumulated emissions of GHG. Right, likelihood that future average summer temperatures exceed the highest temperature observed on record. Up, for the period 2040-2060; down, for 2080-2100. Sources: World Resources Institute, via Wikipedia, for emissions; and Science, vol. 323 (Jan. 9th, 2009) for temperatures.

Yet, because the discrepancy between the distribution of 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 aggregate environmental impact of climate change as well as emissions patterns of each country are observable, one can hope to solve this global cost-sharing problem of sorts.

We argue in favor of a global insurance scheme that washes out differences in the distributions of damages, for which countries are not responsible. Similarly, 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). While the "polluter pays" principle behind this approach is very appealing, Pigou taxation can only perform satisfactorily when the social marginal cost can be reasonably well approximated. However, given that the lifetime of GHG emissions may last more than hundreds or even thousands of years (see Archer, 2005, and Archer and Brovkin, 2008), Pigou taxation is based on expected costs of uncertain events far off in the future, which are very difficult to estimate in practice (Stern 2008). In addition, Pigou taxation is silent on the question of discounting these future costs over hundreds of years and on the normative question of intergenerational equity. Lastly, even if the first-best level of emissions is reached, *some* climate change will occur (as a result of past and current emissions), because *some* pollution is efficient. Pigou taxation does not explicitly handle the normative question of horizontal (or intra-generational) 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 the shortcomings identified above while remaining entirely compatible with the Pigovian outcome (i.e., a first-best pollution level). The key feature of our approach is that it does not rely on estimates of future costs, but rebalances the distribution of current (known) damages each period, thus also circumventing the issue of discounting. Within each period, 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 expliciting the responsibility of countries to the global damage according to 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 accomodate the most prominent competing ideologies, precisely because we let the planner decide explicitly the characteristics for which countries are to be held responsible.

More precisely, the responsibility/compensation approach we adopt is related to that in Bossert & Fleurbaey (1996) in that 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, and 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, according to the planner, a country which experiences cold winters, like Canada, should be held responsible for its emissions but at the same time should be compensated for the fact that its emissions are high due to heating needs. In fact, it has been established (Bossert, 1995) that holding economic agents fully responsible for some characteristics was incompatible with full compensation along the other characteristics unless the cost function is separable along the responsibility/compensation dimensions. Because one cannot reasonably assume such separability, the two taxation schemes that will stand out from our analysis (the *Egalitarian Equivalent* and the *Conditionally Equivalent* mechanisms) result from compromises between full compensation and full responsibility.

From a policy standpoint, our findings show that environmental taxation *à la* Pigou, which is often considered the epitome of responsabilization, popularized by its implementation of the first-best pollution levels, is compatible with a modicum of compensation for differences in irrelevant characteristics. In particular, the Conditionally Equivalent solution can be viewed as a Pigou tax followed by redistributive transfers. To the best of our knowledge, it provides 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.

1.1 A reverse approach to taxation

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 $D_t \equiv \sum_{i,t} d_{it}$.¹ Obviously, given the global nature of the issue, the financing of the levelling 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.

Two arguments are in order. First, given that we consider climate change to be closely related to GHG emissions, we argue that the damage-offsetting scheme should be financed in relation to the countries' emission levels, via a tax on emissions of sorts.² Ideally, this tax should be set so as to achieve efficient

¹ The dynamic nature of the problem and the relation between damage and carbon emissions will be made explicit below, as needed in the argumentation.

² Naturally, this point becomes moot if the planner believes climate change to be completely

emission levels. However, we take the view that 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 . Therefore, unlike the common Pigovian view that the emissions tax should be set equal to the discounted expected future social marginal damage (Nordhaus, 1992), the mechanism we propose repays the "debt", D_t , of *observed* environmental damage *each period*. Thus, our approach avoids the problematic issue of accurately evaluating expected future costs of uncertain future events, which proves extremely difficult in practice (see, e.g., Stern's AEA presidential address 2008). This task is all the more difficult as carbon emissions may persist in the atmosphere for up to thousands of years, thus making unrealistic the estimation of future damage so far in the future. Instead, by basing the emissions tax on current *observed* environmental damage, our mechanism circumvents this difficult estimation issue entirely³.

An additional benefit of our "reverse" approach compared to the forward-looking one is that it circumvents another very problematic issue related to the large lifetime of CO₂ emissions: intergenerational discounting. Indeed, discounting beyond several decades already poses the delicate question of how to consider future costs or benefits. In other words, it is not just the discount rate itself which is contentious, but the very nature of the discounting (e.g., exponential versus hyperbolic, see Henderson and Bateman, 1995). With the timespans involved by CO₂ emissions covering possibly hundreds or even thousands of years, it seems foolish to envision being able to reach any non-controversial present values of environmental damage. By contrast, our "reverse" approach bypasses the issue entirely due to the fact that period-per-period optimizing countries will

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, regardless of the causes.

³Evaluating current damages remains a difficult task, but a considerably less daunting one than predicting damage occurring in the distant future.

behave as if they had anticipated expected future damage. Indeed, we show by a simple dynamic programming argument in a companion paper (Billette de Vlameur and Leroux, 2010) that the usual, forward-looking Pigovian tax, under rational expectations, coincides with our "reverse" scheme. The intuition is that each country would correctly anticipate the impact of its current emissions on future damages and respond according to its own discount rate.

1.2 Compensation and responsibility

We now address 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". Hence, 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 resulting cost shares will be tailored to reflect this assignment of responsibilities, whether it considers that countries are fully responsible for all historical emissions (what we call the *Historical Responsibility* view) or whether one deems climate change to be independent of CO₂ emissions (what we call the *External Shock* view).

Our approach is related to that of Bossert (1995) and Fleurbaey and Bossert (1996) in that we separate country characteristics into two categories: that of "relevant" characteristics for which countries are deemed responsible and which are related to climate change (i.e., possibly past and present emissions), and "irrelevant" characteristics for which countries should not be held responsible or which are unrelated to climate change (i.e., aspects of geographical location like latitude and coastality; and possibly emissions, depending on the point of

view). Then, 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 considerations of responsibility and compensation, making them typically incompatible in their strong versions except for unreasonably simplistic case 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 and 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. Thus, while still taking differences in characteristics into account, cost allocation becomes mathematically feasible, at the expense of being able to charge strictly according to marginal costs. Hence, the possible solutions we offer will differ in how strongly they depart from marginal-cost pricing.

More specifically, one of the two solutions we propose, the *Egalitarian Equivalent mechanism*, splits the consequences of deviations from a reference vector of *irrelevant* characteristics while sharing equally the residual impact of global warming, once each country has paid for its *incremental* contribution⁴. 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

⁴The incremental cost is the additional cost imposed by the presence of an additional *agent*, as opposed to an additional *unit* of emissions (as is the case for marginal cost). Clearly, the smaller the emissions of a country, the closer the incremental cost is to the marginal cost.

expense of responsibility.

Symmetrically, the second solution we propose, the *Conditionally Equivalent mechanism* guarantees each agent the average payoff of a hypothetical situation in which all countries have the characteristics they are responsible for equal to a reference level. Each country bears the consequences of any deviation from this reference level. As such, the Conditionally Equivalent mechanism insists on countries' responsibility at the expense of compensation considerations. Because countries are (marginally) taxed according to marginal damage, it is akin to Pigou taxation, thus implementing the first-best level of emissions.

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 damage function. As a result, the analysis allows for applications in more general settings where externalities are present. Moreover, our results confirm and extend the appeal of the *Egalitarian Equivalent* and the *Conditionally Equivalent* solutions to the more general context of externalities.

2 The model

Let $S = \{1, \dots, m\} \subset \mathbb{N}$ be the set of countries⁵. We denote by n_i the population of country i and by $n = (n_1, \dots, n_m) \in \mathbb{N}^m$ 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

⁵We use the word "countries" for simplicity, but our analysis readily applies to regions, which is especially relevant for large countries.

*current*⁶ per capita emissions, by $x^p = (x_1^p, \dots, x_n^p)$ and $x^c = (x_1^c, \dots, x_n^c)$ the *past* and *current* per capita GHG emissions profiles, respectively.⁷

Each country's per capita private *current* benefits are associated to its own emissions levels via a mapping $b_i : \mathbb{R}_+^2 \rightarrow \mathbb{R}$, $(x_i^p, x_i^c) \mapsto b_i(x_i^p, x_i^c)$,⁸ which is continuous and non-decreasing in each argument. We assume benefits to be fully transferable across countries. Let $b = (b_1, \dots, b_m)$ be the profile of per capita benefit functions, one per country. We interpret the differences in the benefit functions across countries to be essentially geographic in nature (latitude, altitude, coastality, etc.).⁹

In contrast with private benefits, which are solely dependent upon a country's own emissions, per capita country environmental damage will depend on total emissions. More precisely, let $X = \sum_{i=1, \dots, m} n_i x_i$ be the total level of emissions, where $x_i = x_i^p + x_i^c$ designates country i 's cumulated emissions¹⁰. Whenever convenient, we write $X_{-i} = \sum_{j \neq i} n_j x_j$. Formally, we denote by $d_i(X)$ the *current* per capita damage incurred by country i . We do not make any assumption on the functions $d_i(X)$ other than continuity. 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 $d_i(X) < 0$

⁶From a practical standpoint, current emissions can be interpreted as emissions in the very recent past, say, between 5 and 8 years. This time frame could coincide with the reevaluation periods of the major international protocols (Kyoto, Copenhagen, etc.).

⁷To 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 described as "current" will refer to a value at time t .

⁸While 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 .

⁹We 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.

¹⁰For 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. It will become clear that we view population as a "neutral" variable, i.e., belonging to neither category.

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 $d = (d_1, \dots, d_m)$ the profile of per capita damage functions. Finally, we call (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 damage due to global warming 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 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, $t : \mathcal{P} \rightarrow \mathbb{R}^n$ which (potentially) hold countries responsible for their past and current emissions levels, x_i^p and x_i^c , while 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 . Moreover, we consider a country's population, n_i , to be a "neutral" characteristic in the sense that it does not warrant compensation (it is not an irrelevant characteristic) nor reward (it is not a relevant characteristic). 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 GHG 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 $x^c = (x_1^c, \dots, x_m^c)$. We refer to it as **no Historical Responsibility** (hereafter **nHR**). Third, past emissions may be considered a natural benchmark to measuring countries' "needs". According to this so-called **Grand-Fathering** view (hereafter **GF**), countries are held responsible for variations between current and past emissions levels $x^{GF} = (x_1^{GF}, \dots, x_m^{GF})$, where $x_i^{GF} = x_i^c - \gamma x_i^p$, for some parameter γ possibly reflecting technical progress. Finally, some still argue that no causal link between human emissions and climate change can be ascertained. According to this fourth viewpoint, which we shall call the **External Shock** view (henceforth, **ES**), countries' emissions levels are irrelevant in redistributing the costs associated with climate change. Note that the **ES** view is not at odds with the desire to redistribute the impacts of climate change; it simply assumes that damages are not caused by emissions.¹¹

Prior to further investigating the concepts of responsibility and compensation in a global warming context, we impose two minimal fairness requirement. Anonymity requires countries with identical characteristics to be treated equally, while Solidarity asks that no country benefits from seeing the damage of other countries suddenly increase, all else equal.¹²

Axiom 1 (Anonymity) *For any $P = (n, b, d, x^p, x^c) \in \mathcal{P}$, and any $i, j \in S$,*

$$(b_i, x_i^p, x_i^c) = (b_j, x_j^p, x_j^c) \implies t_i(P) = t_j(P)$$

Axiom 2 (Solidarity) *For any $P = (n, b, d, x^p, x^c)$ and $P' = (n, b, d', x^p, x^c)$*

¹¹For readability, the remainder of the paper formulates the global warming problem according to the **HR** view. The corresponding results obtained under the other views of responsibility and can be found in the Appendix.

¹²See, e.g., Thomson (2003) for a comprehensive survey of the use of these standard axioms in the cost-sharing literature.

such that $d'_j \geq d_j$ for all $j \in S$,

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

for all $i \in S$.

3 Responsibility and compensation

3.1 Penalizing (or rewarding) for 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, for which they are considered responsible, irrespective of other, "irrelevant" emissions, if any. A first approach to responsibility consists in arguing that whatever the distribution of irrelevant characteristics, changes in one country's relevant characteristics, should affect only this country. This yields:

Axiom 3 (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 $(\hat{x}_i^p, \hat{x}_i^c)$, all else equal¹³, then:

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

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

¹³We 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 $j \in S \setminus \{i\}$, the following holds:..."

Full Marginal Responsibility is a very demanding axiom as it requires each country to pay the full marginal cost of its own emissions. In fact, it follows from a familiar argument in the cost-sharing literature (e.g., Leroux, 2004) that unless D is linear, no budget-balanced transfer function satisfies *FMR*:

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

Proof. See Appendix A.3.1 ■

This negative result is due to the fact that the notion of marginal damage becomes blurry for non-linear damage functions. This leads us to considering a less demanding axiom, which insists on assigning marginal responsibilities when damage is linear:

Axiom 4 (LMR) *Linear Marginal Responsibility:*

Suppose individual damage functions are linear. Consider a change from P to $\hat{P}$ where some country's emissions change from (x_i^p, x_i^c) to $(\hat{x}_i^p, \hat{x}_i^c)$, for some i , all else equal, then:

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

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

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

Proposition 2 *LMR and Solidarity imply Average Damage Pricing:*

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

Proof. See Appendix A.2.1. ■

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 is ill-suited to handling an efficient, or merely a realistic emissions level. 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 5 (FRR) *Full Reference Responsibility*:

Let $(\tilde{x}^p, \tilde{x}^c) \in \mathbb{R}_+^2$ be a reference vector of per capita emissions and define total reference emissions accordingly: $\tilde{X} = \left(\sum_{j=1,\dots,m} n_j\right) \tilde{x}$, with $\tilde{x} = \tilde{x}^p + \tilde{x}^c$.

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

$$t_i(P) - t_i(\tilde{P}) = (x_i - \tilde{x}) \left[\frac{D(X) - D(\tilde{X})}{X - \tilde{X}} \right],$$

for all $i \in S$.

Remark 1 Clearly, if the damage function, D , is linear, FMR and FRR coincide.

A complementary interpretation of responsibility is that each country must pay the *incremental* damage it imposes onto the rest of society:¹⁴

$$t_i(P) - t_i(P|_{x_i=(0,0)}) = \frac{D(X) - D(X_{-i})}{n_i}.$$

Holding countries responsible only for their relevant characteristics implies that the transfer $t_i(P|_{x_i=(0,0)})$ should not depend on characteristics for which countries are not deemed responsible:

Axiom 6 (ECEIC) *Equal Contribution for Equal Irrelevant Characteristics*¹⁵

$$[b_i = b_j] \implies$$

$$t_i(P) - \frac{D(X) - D(X_{-i})}{n_i} = t_j(P) - \frac{D(X) - D(X_{-j})}{n_j}.$$

A weaker version only requires an equal contribution when all countries' irrelevant characteristics are identical to a reference.

Axiom 7 (ECRIC) *Equal Contribution for Reference Irrelevant Characteristics*.

Consider a reference benefit function, $\tilde{b}$, then:

If $b_i = \tilde{b}$ for all $i \in S$, then:

$$t_i(P) - \frac{D(X) - D(X_{-i})}{n_i} = t_j(P) - \frac{D(X) - D(X_{-j})}{n_j}$$

for all $i, j \in S$.

Remark 2 Clearly, *FMR* is a more demanding axiom than *FRR*, and *ECEIC*

¹⁴The shorthand notation $P|_{x_i=(0,0)}$ designates a global warming problem identical to P in every way except for country i 's emissions, which are zero. Recall that we are adopting the **HR** view of responsibility. For instance, under the **nHR** view, the analog would be $P|_{x_i^e=0}$.

¹⁵This axiom, and others considered here, result from adapting axioms found in Bossert (1995) and Bossert and Fleurbaey (1996) to a context with externalities.

is more demanding than ECRIC. However, no such relationship exists between FRR and ECEIC.¹⁶

3.2 Compensating for 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 their welfare. In other words, only differences in relevant characteristics should matter.

The above argument can be interpreted to mean that all should equally suffer—or benefit—as a result of a change in one country’s irrelevant characteristic:

Axiom 8 (GSIC) *Group Solidarity towards Irrelevant Characteristics.*

Consider a change from b_i to $\hat{b}_i$, for some i , all else equal. Denote $P = (n, b, d, x^p, x^c)$ and $\hat{P} = (n, \hat{b}, d, x^p, x^c)$, then:

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

for all $j \neq i$.

Remark 3 *Under GSIC, budget balance implies*

$$\begin{aligned} t_i(P) - t_i(\hat{P}) &= (1 - \frac{n_i}{N})(b_i(x_i^p, x_i^c) - \hat{b}_i(x_i^p, x_i^c)), \quad \text{and} \\ t_j(P) - t_j(\hat{P}) &= \frac{n_i}{N}(b_i(x_i^p, x_i^c) - \hat{b}_i(x_i^p, x_i^c)) \end{aligned}$$

for all $j \neq i$.

Another interpretation of compensation consists in requiring that citizens

¹⁶In the absence of externalities, FMR coincides FRR and implies ECEIC (Bossert and Fleurbaey, 1996).

of two countries with identical *relevant* characteristics should end up with the same payoff:

Axiom 9 (EPER) *Equal Payoff for Equal Responsibility.*

For all $i, j \in S$, $[x_i = x_j] \implies$

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

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 10 (EPRR) *Equal Payoff for Reference Responsibility.*

Let $(\hat{x}^p, \hat{x}^c) \in \mathbb{R}_+^2$ be a reference vector of per capita emissions. If $(x_i^p, x_i^c) = (\hat{x}^p, \hat{x}^c)$ for all $i \in S$, then:

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

for all $i, j \in S$.

Remark 4 Clearly, *EPER* is more demanding than *EPRR*. However, no such relationship between *GSIC* and the other two axioms.¹⁷

3.3 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 $D(X)$ is linear, *FMR* (or *FRR*) and *GSIC* are incompatible

¹⁷If population had been considered an "irrelevant" characteristic, then *GSIC* would be stronger than *EPER*, as in Fleurbaey and Bossert (1996). The proof is available upon request.

unless the benefit function is additively separable in countries' relevant and irrelevant characteristics.

Proposition 3 *Suppose D is linear. GSIC and FMR are incompatible unless all benefit functions are identical up to a constant; i.e., unless there exists a vector $\alpha \in \mathbb{R}^n$ and a function $h : \mathbb{R}_+^2 \rightarrow \mathbb{R}$ such that*

$$b_i(x_i^p, x_i^c) = \alpha_i + h(x_i^p, x_i^c)$$

for all $(x_i^p, x_i^c) \in \mathbb{R}_+^2$.

Proof. In Appendix A.3.2. ■

Consequently, the only way to reconcile the concepts of compensation and responsibility is to weaken at least one of the two axioms. We discuss these weakenings in turn and characterize the corresponding mechanisms in the next section.

4 The Egalitarian Equivalent and the Conditionally Equivalent mechanisms

4.1 The Egalitarian Equivalent mechanism

The *Egalitarian Equivalent mechanism* splits the consequences of deviations from a reference vector of *irrelevant* characteristics while sharing equally the residual impact of global warming, after each country has paid for its incremental contribution:

Definition 1 *Egalitarian Equivalent transfer:*

Let $P \in \mathcal{P}$ and consider a reference vector of irrelevant characteristics $\tilde{b}$. Then,

for any $i \in S$,

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

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*

$$t = t^{HR-EE}.$$

Proof. See the Appendix A.2.2. ■

Remark 5 *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.*

4.2 The Conditionally Equivalent mechanism

The *Conditionally Equivalent 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 *Conditionally Equivalent (CE) transfer:*

Let $P \in \mathcal{P}$ and consider a reference vector of relevant characteristics $(\tilde{x}^p, \tilde{x}^c) \in \mathbb{R}_+^2$.

For any $i \in S$,

$$\begin{aligned} t_i^{HR-CE}(P) &= b_i(\tilde{x}^p, \tilde{x}^c) + (x_i - \tilde{x}) \left(\frac{D(X) - D(\tilde{X})}{X - \tilde{X}} \right) \\ &\quad - \frac{1}{N} \left(\sum_{j=1}^m n_j b_j(\tilde{x}^p, \tilde{x}^c) - D(\tilde{X}) \right), \end{aligned} \quad (2)$$

where $\tilde{X} = \left(\sum_{j \in N} n_j \right) \tilde{x}$.

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

Theorem 2 *A transfer schedule, t , satisfies FRR and EPRR if and only if*

$$t = t^{HR-CE}.$$

Proof. See Appendix A.2.3. ■

Remark 6 *The above characterization is tight: the strengthening of EPRR into EPER yields an impossibility. Indeed, the reader can check that the CE solution does not satisfy EPER.*

5 Conclusion

The following table 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.¹⁸ This is due to the fact that we consider population to be a "neutral" characteristic. Yet, population is a crucially important characteristic of the problem at hand due to the fact that our approach to responsibility is at the per capita level while data on the characteristics considered (emissions and benefits in particular) are likely to only be available at the aggregate level. This dichotomy inevitably places special emphasis on the population characteristic. Yet, while we do not deem individuals responsible for the population of the country they belong to (i.e., population is not a relevant characteristic), we do not believe the taxation scheme should compensate or penalize them for it (i.e., population is not an irrelevant characteristic either). These practical considerations illustrate 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, no general theory has been established that considers neutral characteristics.

¹⁸It suffices to show that ECRIC and EPRR are incompatible. (Appendix A.3.3)

References

- [1] Archer, D. (2005) "Fate of fossil-fuel CO₂ in geologic time", *J Geophysical Res*, **110** C09S05.
- [2] Archer, D. and V. Brovkin (2008) "The millennial atmospheric lifetime of anthropogenic CO₂", *Climatic Change*, **90** 283-297.
- [3] Baumol, W.J. (1972) "On Taxation and the Control of Externalities", *Amer. Econ. Rev.*, **62** 307-322.
- [4] Billette de Villemeur, E. and J. Leroux (2010) "Looking forward versus looking backward in environmental taxation", mimeo.
- [5] Bossert, W. (1995) "Redistribution mechanisms based on individual characteristics", *Math. Soc. Sci.*, **29** 1-17.
- [6] Bossert, W. and M. Fleurbaey (1996) "Redistribution and Compensation", *Soc. Choice Welfare*, **13** 343-355.
- [7] Henderson, N. and Bateman, I. (1995), "Empirical and Public Choice Evidence for Hyperbolic Social Discount Rates and the Implications for Intergenerational Discounting", *Environmental and Resource Economics*, **5**, 413-423.
- [8] Leroux, J. (2004), "Strategyproofness and efficiency are incompatible in production economies", *Economics Letters*, **85**, 335-340.
- [9] Nordhaus, W.D. (1992) "The ecology of markets", *Proc. Natl Acad. Sci.*, **89** 843-850.
- [10] Pigou, A.C., *The Economics of Welfare*, 4th ed., London, 1932.
- [11] Stern, N. (2008) "The Economics of Climate Change", *Amer. Econ. Rev.: Papers and Proceedings*, **98** 1-37.

- [12] Thomson, W. (2003) "Axiomatic and game-theoretic analysis of bankruptcy and taxation problems: a survey," *Math. Soc. Sci.*, **45** 249-297.

A Appendix

A.1 Other views on responsibility

As we mention in the body of the paper, several views of responsibility can be considered to tackle the global warming problem. We describe below how the analysis would be affected by considering the **nHR**, the **GF** or the **ES** view.

A.1.1 Penalizing (or rewarding) for differences in relevant characteristics

Axiom 3 (FMR) *Full Marginal Responsibility:*

nHR-FMR *Consider a change from P to $\hat{P}$ where some country's emissions change from (x_i^p, x_i^c) to $(x_i^p, \hat{x}_i^c)$, for some i , all else equal, then:*

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

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

GF-FMR *Let $x_i^{GF} = x_i^c - \gamma x_i^p$. Consider a change from P to $\hat{P}$ where some country's emissions change from (x_i^p, x_i^{GF}) to $(x_i^p, \hat{x}_i^{GF})$, for some i , all else equal, then:*

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

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

Remark 7 *The **ES** version of FMR is not well defined because no country bears any responsibility in the matter.*

Axiom 5 (FRR) *Full Reference Responsibility:*

Let $(\hat{x}^p, \hat{x}^c) \in \mathbb{R}_+^2$ be a reference vector of per capita emissions. Define the grandfathering per capita emission reference $\hat{x}^{GF} = \hat{x}^c - \gamma \hat{x}^p$ accordingly. Similarly, let $\tilde{X} = \left(\sum_{j=1, \dots, m} n_j \right) \tilde{x}$, where $\tilde{x} = \hat{x}^p + \hat{x}^c$, denote the total emission reference level.

nHR-FRR For any $P = (n, b, d, x^p, x^c) \in \mathcal{P}$,
and any $\tilde{P} = (n, b, d, x^p, \tilde{x}^c \cdot \mathbf{1}^m) \in \mathcal{P}$,

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

for all $i \in S$.

GF-FRR For any $P = (n, b, d, x^p, x^c) \in \mathcal{P}$,
and any $\tilde{P} = (n, b, d, \tilde{x}^p \cdot \mathbf{1}^m, \tilde{x}^c \cdot \mathbf{1}^m) \in \mathcal{P}$,

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

for all $i \in S$.

Remark 8 Again, the **ES** version of FRR is not well defined because no country bears any responsibility in the matter.

Axiom 6 (ECEIC) Equal Contribution for Equal Irrelevant Characteristics.

nHR-ECEIC $[b_i = b_j \text{ and } x_i^p = x_j^p] \implies$

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

for all $i, j \in S$.

GF-ECEIC $[b_i = b_j \text{ and } x_i^p = x_j^p] \implies$

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

for all $i, j \in S$.

ES-ECEIC $[b_i = b_j, x_i^p = x_j^p, \text{ and } x_i^c = x_j^c] \implies$

$$t_i(P) = t_j(P),$$

for all $i, j \in S$.

Axiom 7 (ECRIC) *Equal Contribution for Reference Irrelevant Characteristics.*

Consider a reference benefit function, $\tilde{b}$, and a reference emissions vector, $(\tilde{x}^p, \tilde{x}^c)$.

nHR-ECRIC If $[b_i = \tilde{b}, \text{ and } x_i^p = \tilde{x}^p]$ for all $i \in S$, then:

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

for all $i, j \in S$.

GF-ECRIC If $[b_i = \tilde{b} \text{ and } x_i^p = \tilde{x}^p]$ for all $i \in S \implies$

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

for all $i, j \in S$.

ES-ECRIC If $[b_i = \tilde{b}, x_i^p = \tilde{x}^p, \text{ and } x_i^c = \tilde{x}^c]$ for all $i \in S \implies$

$$t_i(P) = t_j(P),$$

for all $i, j \in S$.

A.1.2 Compensating for differences in irrelevant characteristics.

Axiom 8 (GSIC) *Group Solidarity towards Irrelevant Characteristics.*

For any $P = (n, b, d, x^p, x^c) \in \mathcal{P}$, any $i \in S$,

nHR-GSIC Consider a change from b_i to $\hat{b}_i$ and from x_i^p to $\hat{x}_i^p$, for some i , 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:

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

for all $j \neq i$.

GF-GSIC Consider a change from b_i to $\hat{b}_i$ and from (x_i^p, x_i^c) to $(\hat{x}_i^p, \hat{x}_i^c)$, for some i , all else equal, such that $x_i^{GF} = \hat{x}^{GF}$. Denote $P = (n, b, d, x^p, x^c)$ and $\hat{P} = (n, \hat{b}, d, \hat{x}^p, \hat{x}^c)$, then:

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

for all $j \neq i$.

ES-GSIC Consider a change from b_i to $\hat{b}_i$ and from (x_i^p, x_i^c) to $(\hat{x}_i^p, \hat{x}_i^c)$, for some i , all else equal. Denote $P = (n, b, d, x^p, x^c)$ and $\hat{P} = (n, \hat{b}, d, \hat{x}^p, \hat{x}^c)$, then:

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

for all $j \neq i$.

Axiom 9 (EPER) Equal Payoff for Equal Responsibility¹⁹.
For any $P = (n, b, d, x^p, x^c) \in \mathcal{P}$,

nHR-EPER $x_i^c = x_j^c \implies$

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

GF-EPER $x_i^{GF} = x_j^{GF} \implies$

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

ES-EPER In all cases:

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

for all $i, j \in S$.

Axiom 10 (EPRR) Equal Payoff for Reference Responsibility.
Let $(\tilde{x}^p, \tilde{x}^c) \in \mathbb{R}_+^2$ be a reference vector of per capita emissions.

nHR-EPRR If $x_i^c = \tilde{x}^c$ for all $i \in S$, then:

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

for all $i, j \in S$.

GF-EPRR If $x_i^{GF} = \tilde{x}^{GF}$ for all $i \in S$, then:

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

for all $i, j \in S$.

¹⁹This axiom, and others considered here, results from the reinterpretation of an axiom found in Bossert and Fleurbaey (1996).

ES-EPRR In all cases:

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

for all $i, j \in S$.

Remark 9 For each interpretation of the above three axioms, the most demanding is **GSIC** while the least demanding is **EPRR**.²⁰

A.2 Proof of propositions and theorems

A.2.1 Proof of Proposition 2

Proposition: *LMR* and *Solidarity* imply Average Damage Pricing.

Proof. Let $P = (n, b, d, x^p, x^c) \in \mathcal{P}$, define $AD : \mathbb{R}_+ \rightarrow \mathbb{R}$, $\tilde{X} \mapsto [D(X)/X] \tilde{X}$ 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. ■

A.2.2 The Egalitarian Equivalent mechanism

Definition 3 *Egalitarian Equivalent transfer:*

For any $P \in \mathcal{P}$ and any reference vector of irrelevant characteristics $(\tilde{b}, \tilde{x}^p, \tilde{x}^c) :$

nHR-EE For any $i \in N$,

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

²⁰See Fleurbaey and Bossert [6].

GF-EE For any $i \in N$,

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

ES-EE For any $i \in N$,

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

For each of the four views, the **E**galitarian **E**quivalent mechanism is characterized by the appropriate combination of **GSIC** and **ECRIC**.

Theorem 3 A mechanism, t , satisfies **nHR-GSIC** and **nHR-ECRIC** if and only if

$$t = t^{nHR-EE}.$$

A mechanism, t , satisfies **GF-GSIC** and **GF-ECRIC** if and only if

$$t = t^{GF-EE}.$$

A mechanism, t , satisfies **ES-GSIC** and **ES-ECRIC** if and only if

$$t = t^{ES-EE}.$$

Proof. We prove the result under the **HR** viewpoint, but the proof technique is similar for the other "views".

It is easily checked that t^{HR-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 denote $\tilde{P} = (n, (\tilde{b}, \dots, \tilde{b}), d, x^p, x^c)$. For all $k = 1, \dots, m-1$, define

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

and let $P^m = P$.

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

$$t_i(\tilde{P}) - \frac{D(X) - D(X_{-i})}{n_i}$$

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

$$\begin{aligned} t_i(\tilde{P}) &= [D(X) - D(X_{-i})] / n_i \\ &+ \frac{1}{N} \left(D(X) - \sum_{j=1 \dots m} [D(X) - D(X_{-j})] \right), \end{aligned}$$

for all $i \in N$.

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

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

for all $i \neq 1$. This yields:

$$t_1(P^1) - t_1(\tilde{P}) = \left(1 - \frac{n_1}{N}\right) [b_1(x_1^p, x_1^c) - \tilde{b}(x_1^p, x_1^c)]$$

and

$$t_i(P^1) - t_i(\tilde{P}) = \frac{n_1}{N} [\tilde{b}(x_1^p, x_1^c) - b_1(x_1^p, x_1^c)],$$

for all $i \neq 1$. Moving up from P^1 to P^2 and applying again *GSIC* gives

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

so that

$$\begin{aligned} t_2(P^2) - t_2(\tilde{P}) &= b_2(x_2^p, x_2^c) - \tilde{b}(x_2^p, x_2^c) \\ &+ \sum_{k=1}^2 \frac{n_k}{N} [\tilde{b}(x_k^p, x_k^c) - b_k(x_k^p, x_k^c)]. \end{aligned}$$

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

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

Note that the damage function, D , does not enter in this proof because the **HR** view on responsibility considers countries to be responsible of all emissions—past and current—and their resulting damage. By contrast, the Egalitarian Equivalent mechanism redistributes along the dimensions for which countries are *not* responsible. ■

A.2.3 The Conditionally Equivalent mechanism

Definition 4 *The Conditionally Equivalent transfer:*

For any $P \in \mathcal{P}$ and any reference vector of relevant characteristics $\tilde{x}^c \in \mathbb{R}_+$:

nHR-CE For any $i \in N$,

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

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

GF-CE For any $i \in N$,

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

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

ES-CE For any $i \in N$,

$$t_i^{ES-CE}(P) = b_i(x_i^p, x_i^c) - \frac{1}{\sum_j n_j} \left(\sum_j n_j b_j(x_j^p, x_j^c) - D(X) \right).$$

In fact, in all four interpretations, the **Conditionally Equivalent** mechanism is characterized by the appropriate combination of **FRR** and **EPRR**.

Theorem 4 *A mechanism, t , satisfies **nHR-FRR** and **nHR-EPRR** if and only if*

$$t = t^{nHR-CE}.$$

*A mechanism, t , satisfies **GF-FRR** and **GF-EPRR** if and only if*

$$t = t^{GF-CE}.$$

*A mechanism, t , satisfies **ES-EPRR** if and only if*

$$t = t^{ES-CE}.$$

Proof. We prove the result using the **HR** viewpoint, but the proof technique is similar for the other "views".

It is easily checked that t^{HR-CE} satisfies the required axioms. Conversely, let $P = (n, b, d, x^p, x^c) \in \mathcal{P}$ consider a reference emissions schedule $(\tilde{x}^p, \tilde{x}^c) \in \mathbb{R}_+^2$.

Denote $\tilde{P} = (n, b, d, \tilde{x}^p \cdot \mathbf{1}^m, \tilde{x}^c \cdot \mathbf{1}^m)$.

For all $k = 1, \dots, m-1$, define

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

and let $P^m = P$.

Let $\tilde{X}$, X^k and X be the emissions levels associated with $\tilde{P}$, the P^k 's and P^m respectively.

By *Anonymity*, budget balance, and *EPRR*,

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

for all $i \in N$.

Next, switching from global warming problem $\tilde{P}$ to P^1 , *FRR* yields:

$$t_1(P^1) - (x_1 - \tilde{x}) \left(\frac{D(X^1) - D(\tilde{X})}{X^1 - \tilde{X}} \right) = t_1(\tilde{P}),$$

and $t_i(P^1) = t_i(\tilde{P})$, for all $i > 1$.

Next, switching from P^k to P^{k+1} yields

$$t_j(P^{k+1}) - (x_j - \tilde{x}) \left(\frac{D(X^{k+1}) - D(\tilde{X})}{X^{k+1} - \tilde{X}} \right) = t_j(\tilde{P}),$$

for all $j \leq k$ and $t_j(P^{k+1}) = t_j(\tilde{P})$ otherwise.

The result follows:

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

■

A.3 Additional Appendix (not intended for final publication)

A.3.1 Proof of Proposition 1

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

Proof. We prove the result along the **HR** viewpoint, but the proof technique is similar for the other views on responsibility.

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 $\hat{P} = (n, b, d, \hat{x}^p, \hat{x}^c)$ with $\hat{x}^p + \hat{x}^c = x^p + x^c + (\alpha, \beta, 0, \dots, 0)$. Denote by $P_\alpha = (n, b, d, x_{[\alpha]}^p, x_{[\alpha]}^c)$ and $P_\beta = (n, b, d, x_{[\beta]}^p, x_{[\beta]}^c)$ the "interim" global warming problems such that $x_{[\alpha]}^p + x_{[\alpha]}^c = x^p + x^c + (\alpha, 0, \dots, 0)$ and $x_{[\beta]}^p + x_{[\beta]}^c = x^p + x^c + (0, \beta, 0, \dots, 0)$, respectively. Similarly, we denote by X_α and X_β the total emissions associated with P_α and P_β , respectively.

By *FMR* applied to P and P_α , it must be that

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

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:

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

and $t_i(\hat{P}) = t_i(P)$ all $i \neq 1, 2$.

Similarly, by applying *FMR* from P to $\hat{P}$ via P_β , we get:

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

and $t_i(\hat{P}) = t_i(P)$ all $i \neq 1, 2$.

The above implications are only compatible if

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

Linearity of D follows. ■

A.3.2 Proof of Proposition 3

Proposition: Suppose D is linear. *GSIC* and *FMR* are incompatible unless all benefit functions are identical up to a constant; i.e, unless there exists a vector $\alpha \in \mathbb{R}^n$ and a function $h : \mathbb{R}_+^2 \rightarrow \mathbb{R}$ such that

$$b_i(x_i^p, x_i^c) = \alpha_i + h(x_i^p, x_i^c)$$

for all $(x_i^p, x_i^c) \in \mathbb{R}_+^2$.

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 $P = (n, b, d, x) \in \mathcal{P}$ and let $(\tilde{x}^p, \tilde{x}^c) \in \mathbb{R}_+^2$ 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

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

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

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

For each $k = 1, \dots, m$, denote by $P^{1,k}$ the global warming problem corresponding to $(b, x^p, x^c)^{1,k}$. Recalling that $x_i = x_i^p + x_i^c$ and that $\tilde{x} = \tilde{x}^p + \tilde{x}^c$, it follows from *FMR* that:

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

with $\tilde{P}^1 = (n, (b_1, \dots, b_1), d, \tilde{x}^p \cdot \mathbf{1}^m, \tilde{x}^c \cdot \mathbf{1}^m)$. Next, applying *GSIC* iteratively yields

$$t_1(P) = t_1(\tilde{P}^1) + (x_1 - \tilde{x})\delta - \frac{1}{m} \sum_{k=2}^m (b_k(x_k^p, x_k^c) - b_1(x_k^p, x_k^c)). \quad (3)$$

Similarly, define

$$\begin{aligned} (b, x^p, x^c)^{2,1} &= [(b_2, x_1^p, x_1^c), (b_2, x_2^p, x_2^c), \dots, (b_2, x_m^p, x_m^c)], \\ (b, x^p, x^c)^{2,m} &= (b, x^p, x^c) \end{aligned}$$

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

$$(b, x^p, x^c)^{2,k} = [(b_1, x_1^p, x_1^c), (b_2, x_2^p, x_2^c), \dots, (b_k, x_k^p, x_k^c), (b_2, x_{k+1}^p, x_{k+1}^c), \dots, (b_m, x_m^p, x_m^c)].$$

For each $k = 1, \dots, m$, denote by $P^{2,k}$ the global warming problem corresponding to $(b, x^p, x^c)^{2,k}$. It follows from *FMR* that:

$$t_1(P^{2,1}) - t_1(\tilde{P}^2) = (x_1 - \tilde{x})\delta$$

with $\tilde{P}^2 = (n, (b_2, \dots, b_2), d, \tilde{x}^p \cdot \mathbf{1}^m, \tilde{x}^c \cdot \mathbf{1}^m)$. Next, applying GSIC iteratively yields

$$t_1(P) = t_1(\tilde{P}^2) + (x_1 - \tilde{x})\delta - \frac{1}{m} \sum_{k=1,3,\dots,m} (b_k(x_k^p, x_k^c) - b_2(x_k^p, x_k^c)) \quad (4)$$

Combining expressions (3) and (4) leads to

$$\begin{aligned} b_2(x_2^p, x_2^c) - b_1(x_2^p, x_2^c) &= m(t_1(\tilde{P}^1) - t_2(\tilde{P}^2)) + \sum_{k=1,3,\dots,m} (b_k(x_k^p, x_k^c) - b_2(x_k^p, x_k^c)) \\ &\quad - \sum_{k=3}^m (b_k(x_k^p, x_k^c) - b_1(x_k^p, x_k^c)). \end{aligned}$$

The right-hand side of the equation being independent of (x_2^p, x_2^c) , it follows that b_1 and b_2 are identical up to a constant. Repeating the argument yields the result. ■

A.3.3 Incompatibility between ECRIC and EPRR

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

Proof. We establish the proof by contradiction under the **HR** view. Consider a transfer scheme, t , satisfying both EPRR and ECRIC. Let $\tilde{x}$ and $\tilde{b}$ be the reference emissions levels and the reference benefit function, respectively. Consider any global warming problem P where $x = \tilde{x} \cdot \mathbf{1}^m$ and $b = \tilde{b} \cdot \mathbf{1}^m$. Invoking EPRR and ECRIC together yields, for all $i, j \in S$:

$$b_i(x_i) - t_i(P) + t_i(P) - \frac{D(X) - D(X_{-i})}{n_i} = b_j(x_j) - t_j(P) + t_j(P) - \frac{D(X) - D(X_{-j})}{n_j},$$

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

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

Unless D is linear, one can construct a global warming problem with $n_i \neq n_j$ such that the above expression does not hold. ■