Internalizing the Climate Externality: Can a Uniform Price Commitment Help?

MARTIN L. WEITZMAN[a]

ABSTRACT

It is difficult to resolve the global warming free-rider externality problem by negotiating many different quantity targets. By contrast, negotiating a single internationally-binding minimum carbon price (the proceeds from which are domestically retained) counters self-interest by incentivizing agents to internalize the externality. In this contribution I attempt to sketch out, mostly with verbal arguments, the sense in which each agent's extra cost from a higher emissions price is counter-balanced by that agent's extra benefit from inducing all other agents to simultaneously lower their emissions in response to the higher price. Some implications are discussed. While the paper could be centered on a more formal model, here the tone of the discussion resembles more that of an exploratory think piece directed to policy-makers and the general public.

Keywords: Climate change, Global Warming, International public goods, Prices versus quantities, UN climate negotiations

http://dx.doi.org/10.5547/2160-5890.4.2.mwei

1. INTRODUCTION: GLOBAL WARMING GRIDLOCK

The world is currently mired in what has aptly been called *global warming gridlock*.[1] The core problem confronting the economics of climate change is an inability to overcome the obstacles associated with free riding on a very important international public good. The 'international' part is significant. Even within a nation, it can be difficult to resolve public goods problems. But at least there is a national government, with some governance structure, able to exert some control over externalities within its borders. With climate change there is no overarching international governance mechanism capable of coordinating the actions necessary to overcome the problem of free riding.

Throughout this paper I use the terms "climate change" and "global warming" interchangeably. The term "climate change" is currently in vogue and is a more apt description overall. But the term "global warming" is more evocative of this paper's main theme. Global warming is a *global* public-goods externality whose resolution requires an unprecedented degree of international cooperation and coordination. This international climate-change externality has frequently been characterized as the most difficult public goods problem that humanity has ever faced. I concentrate in this paper on carbon dioxide emissions, but in principle the discussion could be extended to emissions of all relevant greenhouse gases. Throughout

1. Global Warming Gridlock is the title of a book by David Victor (2011), who popularized the phrase.

[a] Department of Economics, Harvard University, U.S.A. E-mail: mweitzman@harvard.edu.

the paper I blur the distinction between carbon dioxide and carbon, since the two are linearly related.[2]

My point of departure throughout all of what follows is the critical centrality of the international free-rider problem as a cause (really *the* cause) of negotiating difficulties on climate change emissions. Negotiators here are playing a game in which self-interested strategies are a crucial consideration. It turns out that negotiating rules define an important part of the game, and can thereby change self-interest, for better or for worse.

In this paper I try to argue that a uniform global price on carbon emissions can provide a focal point for a common commitment, while quantity targets, which do not as readily present such a single focal point, have a tendency to rely ultimately on individual commitments. As a consequence, negotiating a global price helps to solve the externality problem while individual caps essentially incorporate it. I will try to explain why negotiating a uniform carbon price embodies what I call a "countervailing force" against narrow self-interest by automatically incentivizing all negotiating parties to (approximately) internalize the externality.

2. NEGOTIATING PRICES VS. NEGOTIATING QUANTITIES

At first, for simplicity of exposition I assume that a commitment to a global price of carbon will be implemented as an internationally harmonized, but nationally retained, carbon tax.

An internationally harmonized but nationally retained carbon tax (or price) has already been proposed as a potential solution to the global warming externality, and has been examined on its merits.[3] In what follows I very briefly summarize some of the possible virtues of an internationally-harmonized but nationally-collected carbon tax (or price) that have already been noted in the literature. My foil here is an internationally harmonized cap-and-trade system. This kind of global-design comparison is complicated and full of subjective judgements about what might or might not work better in practice and why or why not. Cap-and-trade systems are perhaps more widely used throughout the world to control pollution, and in that sense are perhaps more visible or more familiar than pollution taxes (although fossil-fuel taxes and subsidies are ubiquitous, if somewhat hidden, almost everywhere). My purpose here is merely to indicate that the perhaps less-familiar uniform carbon tax already has some significant arguments in its favor—as a prelude to some new arguments for negotiating a uniform price on carbon that I will later develop in this paper.

Both quantity-based and price-based controls are inherently uncertain for the period during which they apply (in between times of periodic review), but the uncertainty takes different forms. With cap-and-trade, total emissions are known but the price or (marginal) cost is uncertain. With a carbon tax, the price or (marginal) cost of carbon emissions is known, but total emissions are uncertain. On the basis of economic models of climate change that include uncertainty, carbon taxes outperform tradable permits, both theoretically and in numerical simulations.[4] In the real world, above and beyond theory and numerical simulations, I think that energy price volatility is very poorly tolerated by the general public. Swings in carbon prices, especially in extreme cases, could sour public opinion and discredit for some time thereafter (decades, generations?) the entire idea of a market-based approach to the

2. One ton of carbon equals 3.67 tons of carbon dioxide. My default unit is carbon dioxide (CO_2).

3. There is actually a fair-sized literature on a carbon-tax (or carbon-price) approach. See, e.g., Metcalf and Weisbach (2009), Cooper (2010), Cramton and Stoft (2012), Nordhaus (2007, 2013), and the many further references cited in these works.

4. See Hoel and Karp (2002), Pizer (1999), and Weitzman (1974).

climate change problem. On the other hand, it is difficult for me to imagine the broad public getting quite so upset because total emissions fluctuate.

It has been argued, I think convincingly, that a carbon tax is more easily administered and is more transparent than a cap-and-trade system. This consideration is especially important in a comprehensive international context that would include all major emitting countries. Under international cap-and-trade, governments will allocate valuable emissions permits to their nation's firms and residents. In some places, under some circumstances, there may be a great temptation for kleptocrats to effectively steal these valuable emissions permits and sell them on the international market.

The collected revenues from an internationally harmonized carbon tax remain within each country, and could be used to offset other taxes or even be redistributed internally as lump sum payments. This, I think, is a desirable property. By contrast the revenues generated from an internationally harmonized cap-and-trade system flow as highly visible external transfer payments across national borders, which might be less easily tolerated by countries required to pay other countries large sums of taxpayer-financed money to buy permits.[5]

This extremely brief, and perhaps somewhat biased, discussion of the advantages of an internationally harmonized carbon tax (compared to cap-and-trade) is not intended to be comprehensive. There are also legitimate arguments in favor of internationally harmonized tradable permits and against a carbon tax.[6]

A point in favor of tradable permits, frequently emphasized by its advocates, is the political appeal of giving free allowance permits to carbon-intensive industry groups (as contrasted with taxing them directly on their carbon emissions). As was pointed out, carbon taxes that are internally-levied and collected by a national government could be used to reduce other, more distortionary, taxes—or they could even be distributed directly to the citizenry as lump-sum payments. But this redistribution aspect of a carbon tax is hidden, behind the scenes as it were. Individual firms will prefer, and typically strongly prefer, what they perceive as the lesser burden of freely allocated permits over the greater perceived burden of pollution taxes. Indeed, studies show that the market value of the free allowances is typically significantly greater than the higher compliance costs of decarbonization that are incurred.[7] Firms and countries in a cap-and-trade regime will therefore struggle hard for a larger share of the total amount of freely distributed emissions allocations. The political appeal of freely distributed tradable permits is a double-edged sword. When negotiating emissions caps, a serious income distortion is introduced because a nation is much more concerned with the revenues from its own free quota allocations than it is concerned with overall international social optimality. Auctioning off the allowances would eliminate this income-effect distortion on the individually desired level of free permit allocations, but then we are effectively back in a tax-like system.

Both approaches (an internationally harmonized but domestically collected carbon price, and freely distributed marketable permits) are subject to immense—sometimes seemingly

5. Of course, persuading nations to commit to negotiating a uniform price of carbon in the first place might well involve some "green-fund" equity transfers. Because the imposed "carbon tax" is internally retained within each nation, then, at least for small changes, the green-fund transfers needed to offset increased costs of compliance for price changes are deadweight-loss second-order Harberger triangles of the relatively modest form $(\Delta P \times \Delta Q)/2$. The corresponding international transfers in a cap-and-trade system (which can be either positive or negative, depending, among other things, on initial cap assignments) are first-order immodest rectangles of the form $P \times \Delta Q$.

6. For a critical review of carbon taxes vs. cap-and-trade, see Goulder and Schein (2013) and the many further references they cite.

7. See Goulder et al (2010) and the further cited references therein.

overwhelming—criticisms. In both cases there are innumerable practical details that must be attended to and worked out. In both cases an effective international treaty needs to be binding, which raises uncomfortable issues of enforcement mechanisms and international sanctions. Additionally, there might be mixed hybrid systems. I merely want to establish a level playing field where the idea of an internationally harmonized carbon tax already commands at least as much intellectual respect as an internationally harmonized cap-and-trade system.

The Kyoto approach to global warming was inspired by the ultimate vision of a top-down worldwide treaty limiting the output of each nation's carbon dioxide emissions. It had been wishfully hoped that the highly incomplete Kyoto quantity assignments might have grown over time into a comprehensive binding system of national emissions caps. If these comprehensive caps were freely traded internationally as emissions permits, and if every country had implemented a comprehensive cap-and-trade system internally, it would have caused there to be one uniform worldwide price of carbon emissions, thereby guaranteeing cost effectiveness.

As events played out, Kyoto did not come close to its inspirational vision of an internationally harmonized binding system of emissions caps. By now, the quantity-based Kyoto-type approach has pretty much broken down, leaving the world with a patchwork of sporadic regional volunteerism that does not address centrally how to efficiently correct the critical international externality of global warming.

Throughout this paper I argue that it is very difficult to resolve the global warming externality problem by directly assigning individual quantity targets. A meaningful comprehensive quantity-based treaty involves specifying as many different binding emissions quotas (whether in the form of tradable permits or not) as there are national entities. Each national entity has a self-interested incentive to negotiate for itself a high cap on carbon emissions—much higher than would be socially optimal. The resulting free-rider problem plagues a quantity-based approach. Even if there were a collective commitment to negotiate or vote on a second-stage worldwide total emissions cap, which I will later assume for the sake of argument, disagreements over the first-stage fractional subdivision formula (for disaggregating the negotiated or majority-voted aggregate worldwide quantity cap into individual quantity caps) would make it difficult to enact such a quantity-based approach.[8]

The inspiration for this paper is the perception of a desperate need for some radical rethinking of international climate policy. As a possibly useful conceptual guide for what negotiations might accomplish, I sometimes ask the reader to temporarily suspend disbelief by considering what might happen in a "World Climate Assembly" (WCA) that votes on global carbon emissions via the basic principle of one-person-one-vote majority rule. In this conceptualization, nations would vote along a single dimension for their desired level of emissions stringency on behalf of their citizen constituents, but the votes are weighted by each nation's population.

Right now, anything like a WCA seems hypothetical and futuristic. It presumes a state of mind where the climate change problem has become sufficiently threatening on a grassroots level that world public opinion is ready to consider novel governance structures which involve relinquishing some national sovereignty in favor of the greater good. What might be the justification for a new international organization like the WCA? The ultimate justification is

8. One could try to argue that binding green-fund equity payments are required to get n countries to agree in the first place to negotiate a uniform carbon price, also representing an n-dimensional problem. However, footnotes 5 and 7 suggest that the required green-fund payments may be smaller than the absolute value of the (positive or negative) transfers involved in a cap-and-trade regime that starts off, say, with equal per-capita permit assignments.

that big new problems may require big new solutions. For a world desperately wanting new solutions to the important externality of climate change, perhaps it is at least worth considering establishing a new organization along the lines of the WCA. After all, it is useful to have some concrete fallback decision mechanism behind vague "negotiations" because even with the focus on a one-dimensional harmonized carbon price (or with the focus on a one-dimensional quantity of total emissions), there are bound to be disagreements whose resolution is unclear. I merely assume that it is in the interest of enough nations to forfeit their rights to pollute in favor of a WCA voting solution of the global warming externality. This is truly a heroic assumption at the present time because the WCA does not correspond to any currently-existing international body. Taken less literally, the thought experiment of a hypothetical WCA can still help us to concentrate our thinking and intuition on what negotiations should be trying to accomplish. In other words, I am hoping that the fiction of a WCA might be useful in indicating what might be the outcome of less-formal international negotiations.

It might be objected that a "consensus" voting rule, not a majority voting rule, is employed in negotiations under the United Nations Framework on Climate Change. This "consensus" voting rule has been widely interpreted as requiring near-unanimity. With such a restrictive voting rule, significant progress on resolving the global warming externality is virtually impossible. Surely, a less restrictive voting-like rule, such as majority rule, would render progress more likely, and is at least worth considering.

One aspect should perhaps be emphasized above all others at the outset. The global warming externality problem cannot be resolved without a binding agreement on some overall formula for dividing emissions responsibilities among nations. Volunteer altruism alone will not solve this international public-goods problem. Of necessity there must be some impingement on national sovereignty in the form of an international mechanism for coordinating targets, verifying fulfillment, and punishing non-compliance. The question then becomes: *Which* collective-commitment frameworks and formulas are more promising than which others?

3. THEORY OF NEGOTIATING A UNIFORM CARBON PRICE

In this paper I examine the theoretical properties of a natural one-dimensional focus on negotiating a single binding price on carbon emissions, the proceeds from which are domestically retained. As was previously mentioned, for expositional simplicity, I identify this single binding price on carbon as if it is a harmonized carbon tax. At a theoretical level of abstraction, I blur the distinction between a carbon price and a carbon tax. However, in actuality the important thing is acquiescence by each nation to a binding minimum price on carbon emissions, not the particular mechanism by which this binding minimum price is attained by a particular nation.

A system of uniform national carbon taxes with revenues kept in the taxing country is a relatively simple and transparent way to achieve harmonized carbon prices. But it is not necessary for the conclusions of this paper. Nations or regions could meet the obligation of a minimum price on carbon emissions by whatever internal mechanism they choose—a tax, a cap-and-trade system, a hybrid system, or whatever else results in an observable price of carbon. I elaborate further on this issue in my concluding remarks.

At a theoretical level, I would suggest that the instruments of negotiation for helping to resolve the global warming externality should ideally possess three desirable properties.

1. *Induce cost effectiveness.*
2. Be of *one dimension centered on a "natural" focal point* to facilitate finding an agreement with relatively low transactions costs.
3. *Embody "countervailing force" against narrow self interest by automatically incentivizing all negotiating parties to internalize the externality.*

Using these three desirable theoretical properties as criteria, I now compare and contrast an idealized binding harmonized price with an idealized binding cap-and-trade system.

On the first desirable property, in principle both a carbon price and tradable permits achieve cost effectiveness (provided agreement can be had in the first place).

The second desirable property (low dimensionality) argues in favor of a one-dimensional harmonized carbon price over an *n*-dimensional harmonized cap-and-trade system among *n* nations. Alas, this argument is elusively difficult to formulate rigorously, or even to articulate coherently. My argument here is necessarily intuitive or behavioral and relies on empirical counter-examples. In this case a primary empirical counter-example is the breakdown of the quantity-based Kyoto approach.

With *n* different national entities, a quantity-based treaty involves assigning *n* different binding emissions quotas (whether tradable or not). Treaty making can be viewed as a coordination game with *n* different players. Such a game can have multiple solutions, often depending delicately on the setup, what is being assumed, and, most relevant here, the choice of negotiating instrument. In the case of Kyoto, the world has in practice arrived at a bad quantity-based solution that has essentially devolved to regional volunteerism.

Thomas Schelling introduced and popularized the notion of a focal point in game theory.[9] Generally speaking, a focal point of an *n*-party coordination game is some salient feature that reduces the dimensionality of the problem and simplifies the negotiations by limiting bargaining to some manageable subset, hopefully of one dimension. The basic idea is that by limiting bargaining to a salient focus, there may be more hope of reaching a good outcome. In a somewhat circular definition, a focal point is anything that provides a focus of convergence. The "naturalness" or "salience" of a focal point is an important aspect of Schelling's argument that is difficult to define rigorously and is ultimately intuitive.

The concept of "transactions cost" is associated with the work of Ronald Coase.[10] The basic idea is that *n* parties to a negotiation can be prevented from attaining a socially desirable outcome by the costs of transacting the agreement among themselves. One could try to argue that, other things being equal, transactions costs increase at least proportionally with the number of parties *n*.

In the case of international negotiations on climate change, I believe that both Schelling's concept of a salient focal point and Coase's concept of transactions costs can be used as informal arguments to support negotiating a single harmonized carbon price whose proceeds are nationally rebated. Put directly, it is easier to negotiate one price than *n* quantities—especially when the one price can be interpreted as "fair" in terms of equality of marginal effort. I cannot defend this claim rigorously. At the end of the day, this is more of a plausible conjecture than a rigorous theorem. Whether justly or not, throughout this paper I basically

9. Schelling (1960). See also the special 2006 issue of the Journal of Economic Psychology devoted to Schelling's psychological decision theory, especially the introduction by Colman (2006). Three of the seven articles in this issue concerned aspects of focal points, testifying to the lasting influence of the concept.

10. Coase himself did not invent or even use the term "transactions cost" but he prominently employed the concept. See Coase (1960). For an application of the transactions cost approach to controlling greenhouse gas emissions, see Libecap (2013).

assume that the essential contrast is between one binding price assignment versus *n* binding quantity assignments—and I then proceed to examine the consequences.

The third desirable property is that the instrument or instruments of negotiation should embody a "countervailing force" against narrow free-riding self-interest by incorporating incentives that automatically internalize the externality. I believe this third property is arguably the most important property of all. This "countervailing force" property is inherently built into a price-based harmonized system of emissions charges, but it is absent from a quantity-based international cap-and-trade system, at least as traditionally formulated.

If I am assigned a cap on emissions, then it is in my own narrow free-riding self-interest to want my cap to be as large as possible (whether or not my cap will be tradable as a permit). The self-interested part of me wants maximal leniency for myself. Other than altruism, there is no countervailing force on the other side encouraging me to lower my desired emissions cap because of the externality benefits I will be bestowing on others.

Within a nation, the government *assigns* binding caps. But *among* sovereign nations, binding caps must be *negotiated*. I believe that this is a crucial distinction for the success or failure of a cap-and-trade regime. A Kyoto-type quantity-based international system fails because no one has an incentive to internalize the externality and everyone has the self-interested incentive to free ride. What remains is essentially an erratic pattern of altruistic individual volunteerism that is far from a socially optimal resolution of the problem.

An internationally-harmonized domestically-collected carbon price is different. If the price were imposed on me alone, I would wish it to be as low as possible so as to limit my abatement costs. But when the price is uniformly imposed, it embodies a countervailing force that internalizes the externality for me. Counterbalancing my desire for the price to be low (in order to limit my abatement costs) is my desire for the price to be high so that other nations will restrict their emissions, thereby increasing my benefit from worldwide total carbon abatement. A binding uniform price of carbon emissions has a built-in self-enforcing mechanism that countervails free riding.[11]

In previous work, I have tried to model formally the role of this third "countervailing force" property of an internationally-harmonized but nationally-collected carbon price.[12] I constructed a basic model indicating an exact sense in which each agent's extra cost from a higher international emissions price is counter-balanced by that agent's extra benefit from inducing all other agents to simultaneously lower their emissions via the higher international price.

With further restrictions, the model showed that population-weighted majority rule for an internationally harmonized carbon price can come as close to an optimal price on emissions as the median per-capita marginal benefit is close to the mean per-capita marginal benefit. The key insight from this way of looking at things is that in voting (or more generally negotiating) a universal carbon price, various nations are, to a greater or lesser degree, internalizing the externality. Loosely speaking, an "average" nation is fully internalizing the exter-

11. Later I discuss negotiating one worldwide aggregate emissions cap (contingent upon a previous-round subdivision formula for n fractional targets, set, for example, by a preceding agreement on various target reductions from various baselines). A system based on negotiating aggregate emissions (given a subdivision formula) could, in principle, embody countervailing force against the global warming externality. But, again, I will conclude that negotiating the extra layer of n first-round Kyoto-like fractional subdivision target reductions will likely founder politically when applied on a worldwide scale.

12. See Weitzman (2014).

nality because its extra cost from a higher emissions price is exactly offset by its extra benefit from inducing all other nations to simultaneously lower their emissions via the higher price.

On the price side, a uniform carbon price automatically has the desirable property that cost effectiveness is guaranteed. I think that the formal voting result of the model might perhaps be interpreted somewhat less formally as indicating that negotiating an internationally harmonized (but nationally collected) carbon price may have an important desirable property on the quantity side as well. If the median marginal benefit (per capita) equals the mean marginal benefit (per capita), then the socially optimal carbon price has the property that, roughly speaking, half of the world's population wants the price to be higher, while the other half of the world's population wants the price to be lower. In this situation, the desirable quantity-side property is that the total worldwide output of all emissions might be "close" to being optimal to the extent that the outcome of negotiations mimics the outcome of majority voting. Although the real world is a far more complicated and nuanced place than the restrictive theoretical model that was constructed, I think this voting result is trying to indicate something positive (even if only at an abstract level) about how a negotiated uniform carbon price might possess some overall potential to counteract via internalization the externality of global warming.

4. MIGHT A MODIFIED CAP-AND-TRADE WORK AS WELL?

Previously I listed three desirable features that instruments for negotiating climate change should ideally possess: (1) cost effectiveness; (2) a natural one-dimensional focal point; (3) a built-in self-enforcement mechanism that internalizes the externality. I then explained that an internationally-harmonized but nationally-retained carbon price possesses all three properties, whereas an n-dimensional quantity-based cap-and-trade system at best (if it can be negotiated in the first place) possesses only the first property of cost effectiveness. With n different nations, there will be difficult bargaining over n different caps with no force other than altruism countervailing each nation's selfish desire to be a free rider and secure for itself a large cap on emissions.

But maybe I am being unfair to tradable permits. Suppose we imagine trying to convert the n-dimensional problem of allocating carbon emissions permits into some one-dimensional quantity analogue of a uniform price on carbon emissions. We might imagine a thought experiment where the cap-and-trade negotiators are sitting around a negotiating table and limiting themselves to simple linear formulas for allocating individual emissions caps as a fraction of total world emissions.[13]

Suppose the cap-and-trade negotiators must decide the total amount of emissions E, given a sub-allocation formula for deciding the fraction of emissions permits allotted to each nation. A standard way of conceptualizing this allocation problem for each country is in terms of an assigned fractional emissions reduction from an assigned baseline level. Here I think it is most instructive to view the essence of such an assignment process in terms of a simple linear reduced form that allots emissions permits $E_i(E) = a_i + b_i E$ to nation i (where $\sum a_i = 0$, $b_i > 0$, and $\sum b_i = 1$).

If each nation i would accept as *given* the assigned distributional coefficients (a_i, b_i) and

13. This approach is spelled out in more mathematical detail in Weitzman (2014).

the above sub-allocation formula $E_i(E) = a_i + b_i E$, one might then imagine negotiating over (or even voting for) the total emissions E. *Contingent* upon the distribution of coefficients being accepted as given, this system would seemingly possesses the desirable property of having a one-dimensional locus of negotiations (here the level of total worldwide emissions E). And there is also countervailing force against negotiating for a higher value of worldwide total emissions E. Although each nation i's automatic assignment of a higher individual emissions cap E_i when total emissions E are higher helps directly by lowering its emissions costs, this domestic effect is counteracted by the benefits that each nation would lose from a higher total worldwide emissions level, because then everyone else would also emit more. It appears that such a cap-and-trade system might in principle have desirable focal-point and countervailing-force properties *if* the assigned distribution coefficients were accepted and bargaining were restricted to negotiating total emissions.

But now follow the thought experiment further by asking: Where do the distributional sub-allocation coefficients (a_i, b_i) come from in the first place? They are presumably the result of an n-party negotiating process where there is no countervailing force to the selfish desire of each country to make its own fractional allocation coefficients as high as possible. With n different nations, there will be the usual difficult bargaining over n different distributional coefficients, with no externality-internalizing incentive countervailing each nation's desire to secure for itself a high fraction of emissions—again presumably resulting in a Kyoto-like breakdown.

When a cap-and-trade system is used to control pollution *within* a nation, the government of that nation *assigns* the caps (or the fractions of emissions).[14] In this intra-national case there is a natural symmetry between a one-dimensional price p and a one-dimensional total quantity of emissions E. But there is no international government that has the unilateral power to assign caps or fractions. These caps or fractions must be *negotiated* among sovereign nations. This breaks the one-dimensional symmetry because now one price p is contrasted with the asymmetry of n vested sovereign interests jockeying for the n initial fractional distributions. There is thus a critical distinction between intra-national and inter-national cap-and-trade systems. In the international case the initial distribution of caps is explicitly distributive, resulting in a war of words about who caused the global-warming problem and who should bear the burden of remedying it, who is rich and who is poor, what is fair and what is unfair, and so forth and so on. There could also be a war of words about the green-fund transfers required to induce participation in a uniform-price treaty, but, for reasons elaborated in footnotes 5 and 8 having to do with the difference between first-order and second-order transfers, I think that an internally-retained price treaty takes a lot of pressure off the green-fund payments.

But perhaps a formulation of this generality is biased against cap-and-trade. We might try to imbue the distribution coefficients with dimensionality-reducing salient qualities by imagining "naturally symmetric" focal allocations of the fractional coefficients. One such seemingly symmetric formula might be that each country is assigned the same fractional reduction of emissions from some agreed-upon baseline year. The Kyoto Protocol of 1997 adopted just a little of the spirit of this idea for developed countries alone, with the hope that some variant of it might later be extended to developing countries. The high-income industrialized countries (Annex I) agreed to "binding" commitments (but without any enforcement

14. Admittedly, this is often done in a way that eases special-interest acceptance, such as being allocated for free or almost for free based on something like a uniform reduction of previous pollution levels.

mechanism!) to reduce greenhouse gas emissions in 2012 by an average of 5% relative to 1990 levels (although allowing some individually-negotiated variations around that 5% average). Developing countries were exempt from any "binding" commitments. Overall, the Kyoto Protocol did not come close to fulfilling its initial aspirations. The U.S. and Australia did not ratify, Canada and Japan eventually dropped out, and individual compliance was at best spotty.[15] Furthermore, and perhaps most distressingly, non-Annex I countries have not formally agreed to any actual future "binding" commitments going forward from 2012. The Kyoto experience is subject to multiple interpretations. For me, it largely testifies to the great difficulty of negotiating binding international quantity caps on the major emitters. In the language that I have employed here, it has been overwhelmingly problematic to assign binding quantity-like distributional coefficients on a worldwide basis.

Other seemingly symmetric quantity formulas might also be examined. For example, one might entertain the idea of assigning the same worldwide emissions level per capita. This is a symmetric formula that embodies a certain concept of worldwide fairness, but a cap-and-trade system based on such an initial distribution of caps would involve massive transfers from the developed to the developing countries, which would likely prove politically unacceptable. Besides, even this formula does not address concerns regarding historical responsibility for the cumulative stock of emissions, which would surely be raised. Alternatively, one might imagine negotiating (or even voting on) an identical percentage reduction from some base case of emissions. In this situation, I think, everyone would first argue about the fairness of the baseline emissions that they were initially assigned.

I abstain from further speculation. My point is that no matter what quantity-like initial allocation mechanism I can imagine, an attempt to modify an international cap-and-trade system by making it one dimensional seems likely to founder for essentially the same reasons that an unmodified international cap-and-trade system founders. In a quantity-based system with n different sovereign nations I fear there will be intractable negotiations for n different distributional assignments (a_i, b_i), with no force countervailing each nation's free-riding desire to secure for itself a selfishly lenient emissions fraction of the total emissions E.[16]

Here is what I think is the essence of the one-price vs. n-quantities negotiation problem as elaborated in this section. A quantity-type system based on a formula like $E_i(E) = a_i + b_i E$ involves *two* layers of negotiations. First, the n parties must agree on the quantity-like distributional coefficients (a_i, b_i). Then, second, the parties must agree on the single worldwide aggregate level of emissions E. By contrast, a price-based system involves only *one* layer of negotiation, focused on agreeing to a single one-dimensional uniform price p. This latter is not an easy task, but it would seem generally easier to negotiate one price layer than two quantity layers (whose first layer involves assigning n quantity-like distributional coefficients). Admittedly this argument depends upon a particular way of framing the issue, but it seems to me that, in international negotiations among n sovereign nations, there may be an irreducible asymmetry between one price instrument vs. n quantity instruments.

15. The one bright spot might be considered the European Union, whose emissions trading system could perhaps be interpreted as evolving towards an EU-wide cap (declining annually) with member-state shares increasingly being determined by auctioning permits. I am unsure and somewhat skeptical about the extent to which this EU model might be extended to the world as a whole. For a generally favorable assessment of this possibility, see Ellerman (2010).

16. Bosetti and Frankel (2012) propose a constructive and imaginative allocation formula for emissions permits, but it still looks complicated and contentious to me.

Even while acknowledging that it only involves one layer of negotiations (as opposed to two on the quantity side), one could ask on the price side what might induce *n* countries to agree to a single harmonized charge for carbon emissions. We have been over this ground before. It all begins with the recognition that any resolution of the global warming free-rider problem requires a collective commitment to some binding restriction on the sovereign right of nations to freely emit as much carbon dioxide as they wish. Why might nations restrict their own sovereignty by collectively committing to a common price regime for resolving the global warming externality? Perhaps because enough of them come to realize (or are made to realize) that the international climate-change public good is sufficiently important to outweigh national rights to pollute the global commons—and that a radical collective problem may call for a radical collective solution. Without such a realization and the will to act upon it, progress on resolving the global warming externality will be limited to voluntary altruism, which seems not nearly enough to overcome the free rider problem.

5. CONCLUDING REMARKS

At the end of the day, there is no airtight logic in favor of a negotiated price over negotiated quantities, only a series of partial arguments. One argument is that the revenues from a carbon price are nationally collected, so that the contentious distributional side is somewhat hidden and there is at least the appearance of fairness as measured by equality of marginal effort. A second desirable feature, I have argued, is the natural salience and relatively low transaction costs of negotiating one price as against negotiating *n* quantities, which, while somewhat imprecise, is in my opinion an important distinction. A third argument is the self-enforcement mechanism that constitutes the main theme of this paper, namely the built-in countervailing force of an imposed uniform price of carbon, which tends to internalize the externality and gives national negotiators an incentive to offset their natural impulse to otherwise bargain for a low price.

Of necessity, my argument has been sprinkled with subjective judgements. This, unfortunately, is the nature of the subject. To repeat yet again, this time after examining somewhat more carefully the alternatives, I judge it difficult to escape the conclusion that, in the context of an international treaty that covers all major emitters, it is more politically acceptable and it comes closer to a social optimum to negotiate one binding price than *n* binding quantities or quantity-like distributional coefficients.

My argument here is sufficiently abstract that it is open to enormous amounts of criticism on many different levels. There are so many potential complaints that it would be incongruous to list them all and attempt to address them one by one. These many potential criticisms notwithstanding, I believe the argument here is exposing a fundamental countervailing-force argument that deserves to be highlighted.

Because the formulation is at such a high level of abstraction, it has blurred the distinction between a carbon price and a carbon tax. As was previously noted, the important thing is acquiescence by each nation to a *binding minimum price* on carbon emissions, not the particular internal mechanism by which this obligation is met. A system of national carbon taxes with revenues kept in the taxing country is a relatively simple and transparent way to achieve harmonized carbon prices. But it is not necessary for the conclusions of this paper. Nations or regions could meet the obligation of a minimum price on carbon emissions by whatever internal mechanism they choose—a tax, a cap-and-trade system, a hybrid system, or whatever

else results in an observable price of carbon.[17] And any nation or region could choose to impose a carbon price above the international minimum. The hope is that even a low positive initial value of a universal minimum carbon price could be useful for gaining confidence and building trust in this price-based international architecture.

The purpose of this paper is primarily expository and exploratory. *Any* proposal to resolve the global warming externality will face a seemingly overwhelming array of practical administrative obstacles and will need to overcome powerful vested interests. That is the nature of the global warming externality problem. The theory of this paper seems to indicate that negotiating a uniform minimum price on carbon can have several desirable properties, including, especially, helping to internalize the global warming externality. To fully defend the relative "practicality" of what I am proposing would probably require a book, not an article. In any event, *this* article is not primarily about practical considerations of international negotiations. I leave that important task mostly to others.[18] However, I do want to mention just a few real-world considerations that have been left out of my mental model yet seem especially pertinent.

An example of a relatively small practical issue that I am waving aside is just where in the production chain a carbon price should be collected. I think the presumption would be that the carbon price should be collected by the country in which the carbon dioxide is actually released into the atmosphere. One might try to argue that a carbon price should be collected downstream as close as possible to the point where the carbon is burned. But this would involve an impractically large number of collection points. It is much easier to collect the price upstream, at various chokepoints where the carbon is first introduced into the carbon-burning economy.[19]

A truly critical issue is that a binding international agreement on a uniform minimum carbon price requires some serious compliance mechanism. To begin with, the carbon price must be observable. For enforcement, perhaps there is no practical alternative to using the international trading system for applying tariff-based penalties on imports from non-complying nations. Nordhaus (2015) advocates such an approach with uniform border tariffs on imports from non-member countries imposed by a "climate club" of member nations who agree to impose on themselves a harmonized carbon price. Cooper (2010) has argued for an expansive interpretation whereby the internationally agreed charge on carbon emissions would be considered a cost of doing business, such that failure to pay the charge would be treated as a subsidy that is subject to countervailing duties under existing provisions of the World Trade Organization.[20]

An efficient carbon price naturally produces more winners than losers (by the metric of the modified Pareto criterion). In the case of the global warming externality, which has been characterized as the greatest public goods problem of all time, it seems reasonable to suppose that there might be many times more winners than losers from imposing a uniform carbon price. Because countries here get to keep their own carbon-price-generated revenues, then welfare-compensating transfers, to the extent they are made at all, ought, at least for small

17. A minimum carbon price could be attained in a cap-and-trade system by setting it as a floor, which could be enforced, e.g., by making it a reserve price on the auctioning of permits.

18. See, e.g., Bodansky (2010) or Barrett (2005).

19. This set of issues and its distributional consequences (including references to other literature) is discussed extensively in Asheim (2012).

20. See also the discussion of the legality of such sanctions under WTO provisions in Metcalf and Weisbach (2009).

changes, to be relatively modest second-order deadweight-loss triangles instead of the relatively immodest first-order rectangle transfers associated with tradable permits from, say, an initial assignment of caps that are equal per-capita.[21]

I close by noting again that global warming is an extremely serious as-yet-unresolved international public goods problem. With the failure of a Kyoto-style quantity-based approach, the world has seemingly given up on a comprehensive global design, settling instead for sporadic national, sub-national, and regional measures. These partial measures seem far from constituting a socially efficient response to the global warming externality. Perhaps, as was previously suggested, the Kyoto-style quantity-based focus on negotiating emissions caps embodies a bad design flaw. The arguments of this paper indicate a way in which negotiating a binding internationally-harmonized nationally-collected minimum price on carbon emissions might help to internalize the global warming externality.

References

Asheim, Geir B. (2012). "A Distributional Argument for Supply-Side Climate Policies." *Environ Resource Econ*, published online: 11 August 2012.

Barrett, Scott (2005). *Environment and Statecraft: The Strategy of Environmental Treaty Making*. Oxford: Oxford University Press. http://dx.doi.org/10.1093/0199286094.001.0001.

Bodansky, Daniel (2010). *The Art and Craft of International Environmental Law*. Cambridge: Harvard University Press.

Bosetti, Valentina, and Jeffrey Frankel (2012). "Sustainable Cooperation in Global Climate Policy: Specific Formulas and Emissions Targets." Mimeo dated August 2012.

Colman, A. M. (2006). "Thomas C. Shelling's psychological decision theory: Introduction to a special issue." *Journal of Economic Psychology*, 27: 603–608. http://dx.doi.org/10.1016/j.joep.2006.06.002.

Cooper, Richard N. (2010). "The Case for Charges on Greenhouse Gas Emissions." In Joe Aldy and Robert Stavins (eds), *Post-Kyoto International Climate Policy: Architectures for Agreement*, Cambridge University Press.

Cramton, Peter, and Steven Stoft (2012). "Global Climate Games: How Pricing and a Green Fund Foster Cooperation." *Economics of Energy & Environmental Policy*, 1(2): 125–136. http://dx.doi.org/10.5547/2160-5890.1.2.9.

Cramton, Peter, Axel Ockenfels, and Steven Stoft (2015). "An International Carbon-Price Commitment Promotes Cooperation." *Economics of Energy & Environmental Policy*, 4(2): 51–64.

Ellerman, A. Denny (2010). "The EU's Emissions Trading Scheme: A Prototype Global System?" In Joe Aldy and Robert Stavins (eds), *Post-Kyoto International Climate Policy: Architectures for Agreement*, Cambridge University Press.

Goulder, L. H., M.A.C. Hafstead, and M. Dworsky (2010). "Impacts of alternative emissions allowance allocation methods under a federal cap-and-trade program." *Journal of Environmental Economics and Management*, 60(3), 161–181. http://dx.doi.org/10.1016/j.jeem.2010.06.002.

Goulder, Lawrence H., and Andrew R. Schein (2013). "Carbon Taxes vs. Cap and Trade: A Critical Review." *Climate Change Economics* 4(3): 1–28. http://dx.doi.org/10.1142/S2010007813500103.

Hoel, Michael, and Larry Karp (2002). "Taxes vs. quotas for a stock pollutant." *Resource and Energy Economics* 24: 367–384. http://dx.doi.org/10.1016/S0928-7655(02)00014-3.

Libecap, Gary D. (2013). "Addressing Global Environmental Externalities: Transaction Costs Considerations." *Journal of Economic Literature*, 52(2): 424–79. http://dx.doi.org/10.1257/jel.52.2.424.

21. Cramton, Ockenfels, and Stoft (2015) make an analogous argument in the form of a numerical example indicating that committing to a price tends to be less risky than quantity targets. Thus, according to this reasoning, equity transfers under cap-and-trade would have to be larger than equity transfers under a uniform price because of the increased risk imposed by caps. In a separate argument, they also indicate that choosing a particular green-fund equity-payment formula to encourage participation in a uniform price regime can itself be reduced from a seemingly n-dimensional problem to a one-dimensional focal problem.

Metcalf, Gilbert E., and David Weisbach (2009). "The Design of a Carbon Tax." *Harvard Environmental Law Review* 33.2: 499–556. http://dx.doi.org/10.2139/ssrn.1324854.

Nordhaus, William D. (2007). "To tax or not to tax: Alternative approaches to slowing global warming." *Review of Environmental Economics and Policy* 1(1): 26–44. http://dx.doi.org/10.1093/reep/rem008.

Nordhaus, William D. (2013). *The Climate Casino: Risk, Uncertainty, and Economics for a Warming World*. New Haven: Yale University Press.

Nordhaus, William D. (2015). "Climate Clubs: Designing a Mechanism to Overcome Free-riding in International Climate Policy." *American Economic Review* 105 (4); 1339–1370. http://dx.doi.org/10.1257/aer.15000001.

Pizer, William (1999). "Optimal Choice of Policy Instrument and Stringency under Uncertainty: the Case of Climate Change." *Resource and Energy Economics*, 12: 255–287. http://dx.doi.org/10.1016/S0928-7655(99)00005-6.

Schelling, Thomas C. (1960). *The Strategy of Conflict*. Harvard University Press.

Victor, David (2011). *Global Warming Gridlock*. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511975714.

Weitzman, Martin L. (1974). "Prices vs. Quantities." *Review of Economic Studies* 41, 4: 477–491. http://dx.doi.org/10.2307/2296698.

Weitzman, Martin L. (2014). "Can Negotiating a Uniform Carbon Price Help to Internalize the Global Warming Externality?" *Journal of the Association of Environmental and Resource Economists* 1(1/2): 29–49. http://dx.doi.org/10.1086/676039.