Voting on Prices *vs.* Voting on Quantities in a World Climate Assembly

Martin L. Weitzman[*]

January 25, 2015
Comments Appreciated

Abstract

This paper posits the conceptually useful allegory of a futuristic "World Climate Assembly" that votes on global carbon emissions via the basic principle of majority rule. Two variants are considered. One is to vote on a universal *price* (or tax) that is internationally harmonized, but the proceeds from which are domestically retained. The other is to vote on the overall *quantity* of total worldwide emissions, which are then distributed for free (via a pre-decided fractional subdivision formula) as individual allowance permits that are subsequently marketed in an international cap-and-trade system. The model of the paper suggests that the majority-voted price is likely to be less distortionary and easier to enact than the majority-voted total quantity of permits. While the study is centered on a formal model, the tone of the policy discussion resembles more an exploratory think piece.

JEL Codes: F51, H41, Q54

Keywords: Climate change, Global warming, International public goods, Prices versus quantities

1 Introduction: Global Warming Gridlock

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

---

[*]Department of Economics, Harvard University (mweitzman@harvard.edu).

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 the paper I blur the distinction between carbon dioxide and carbon, since the two are linearly related.[1]

The world is currently mired in what has aptly been called *global warming gridlock.*[2] 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 overcoming, by coordinated action, the problem of free riding.

For purposes of exposition, at first I blur the distinction between a price and a 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 currently more widely used throughout the world, and in that sense are perhaps more visible or more familiar than pollution taxes. My purpose here is merely to indicate that the perhaps less-familiar carbon tax already has some significant arguments in its favor – as a prelude to some new theoretical 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

---

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

[2] *Global Warming Gridlock* is the title of a book by David Victor (2011).

[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, Ockenfels and Stoft (2013), Nordhaus (2007, 2013), and the many further references cited in these works.

that include uncertainty, carbon taxes outperform tradeable 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 the entire idea of a market-based approach to the 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 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.

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 tradeable permits and against a carbon tax.[5]

A point in favor of tradeable 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. 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

---

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

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

significantly greater than the higher compliance costs of decarbonization that are incurred.[6] Firms and countries in a cap-and-trade regime will therefore struggle hard for a larger share of the stock of freely distributed emissions allocations. As will be shown and highlighted later, the political appeal of free tradeable permits is a double-edged sword. When voting on total emissions (or more generally during negotiations), 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 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 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, 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 tradeable 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

---

[6]See Goulder et al (2010) and the further cited references therein.

4

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 majority-voted aggregate worldwide quantity cap into individual quantity caps) would make it difficult to enact such a quantity-based approach.

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?

The inspiration for this paper is the perception of a desperate need for some radical rethinking of international climate policy. I ask a 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 the setup of this paper nations will 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, 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. Two variants are considered, according to which instrument a World Climate Assembly might vote on. One is to vote on a universal *price* (or tax) that is internationally harmonized, but the proceeds from which are domestically retained. The other is to vote on the overall *quantity* of total worldwide emissions, which are then distributed for free (via a pre-decided fractional subdivision formula) as individual allowance permits subsequently marketable in an international cap-and-trade system. The model of the paper, while highly specific, attempts to explain how the majority-voted price is likely to be less distortionary and easier to enact than the majority-voted total quantity of permits. I believe the model gives some insights, which suggest that negotiating a uniform price on carbon emissions might be closer to a social optimum than negotiating the total quantity of tradeable emissions permits.

What is the justification for a new international organization like the WCA? The ul-

5

timate justification is 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.

One might object 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 rule, such as majority rule, would render progress more likely, and is at least worth considering.

2   The Model

The formulation here is at a heroic level of abstraction. I wave away innumerable "practical" considerations to focus on a theoretical model. I beg the reader's indulgence for a willing suspension of disbelief while the basic argument is being developed.

The analysis is made clearest and most transparent when the fundamental unit is the person, so that everything is normalized per capita. In reality, of course, people belong to some larger entity, here called a "nation," that (hopefully or presumably) acts on their behalf. (The nation here is an elastic concept, since for the purposes of this paper it might be more appropriate to consider regional blocs like the European Union as if it comprised a single nation or maybe even consider regional sub-national states.) It is easiest to conceptualize that all of the people belonging to one nation are identical agents whose tastes and technology are representative of that nation. For an individual belonging to a nation, everything – emissions, costs, benefits – is expressed in per-capita terms for that nation. (Inversely, one could take costs and benefits on the national level as given primitives and impute to each citizen the corresponding per-capita costs and benefits as a function of per-capita emissions, being careful to ensure that the imputed per-capita costs and benefits aggregate consistently

to the given national costs and benefits.) The nation in this model is effectively an entity that votes or negotiates on behalf of its citizens in accordance with their preferences. It does not matter which comes first here – individual preferences or national preferences – since at this level of abstraction they are assumed to be essentially the same thing.

The total world population is m. Each person-agent is indexed by $i = 1, 2, ..., m$. In what follows I abstract away from dynamics in favor of a static-flow analysis. I assume agents can convert their preferences about desired stocks of atmospheric carbon dioxide into consistent preferences about corresponding emissions flows for the period under consideration.

Let E_i stand for the carbon dioxide emissions of agent i. The cost of emissions level E_i for agent i is given by the cost function $C_i(E_i)$, where $C_i'(E_i) < 0$ and $C_i''(E_i) > 0$.

Total world emissions are

$$E = \sum_{i=1}^{m} E_i. \tag{1}$$

The benefit to i of world total emissions level E is given by the benefit function $B_i(E)$ where, in the relevant range of E, it is postulated that $B_i'(E) < 0$ and $B_i''(E) \geq 0$. (Note that I could equivalently work with a "damages function" $D_i(E) \equiv -B_i(E)$.)

For any given $\{E_i\}$, world social welfare is

$$W = \sum_{i=1}^{m} B_i(E) - \sum_{i=1}^{m} C_i(E_i), \tag{2}$$

where E is given by (1). Note that W expressed by (2) is a particular form of a public goods setup where the producers of the public bad (pollution) coincide with the consumers of the public bad (pollution).

The socially optimal values of $\{E_i\}$, which maximize welfare W in (2), are denoted $\{E_i^*\}$. Then $E^* = \sum E_i^*$ is the corresponding level of optimal worldwide total emissions. Let P^* be the corresponding socially optimal price of carbon dioxide emissions. (Alternatively, P^* is the optimal SCC, standing for Social Cost of Carbon.) The Samuelson-Lindahl solution to maximizing (2) here takes the form

$$P^* = -\sum_{i=1}^{m} B_i'(E^*), \tag{3}$$

where, for all i,

$$P^* = -C_i'(E_i^*). \tag{4}$$

I now make two further simplifying assumptions that will allow closed-form expressions in the analysis that follows.

The first assumption is that, throughout the short-term period for which the analysis is intended to apply, benefits are linear in total emissions within the relevant range, having the reduced form

$$B_i(E) = \beta_i - b_i\,E. \tag{5}$$

Equation (5) means that the marginal benefit for each agent i is a constant

$$B_i'(E) = -b_i \tag{6}$$

for some positive coefficient b_i that is allowed to differ for different i. I feel that (6) is reasonably accurate for small time periods because emissions are a flow, whereas benefits (or damages) are a function of the accumulated stock, and for carbon dioxide the flow-stock ratio is small. Note that the marginal benefit coefficients $\{b_i\}$ can differ among different agents i, while still being constant for each particular agent i.

An immediate consequence of (6) combined with (3) is that

$$P^* = \sum_{i=1}^{m} b_i, \tag{7}$$

independent of the cost functions. This property greatly simplifies the analysis without, I hope, losing too much realism. An equivalent rewriting of (7), which will later prove useful, is

$$P^* = m\,\bar{b}, \tag{8}$$

where

$$\bar{b} = \frac{\sum_{i=1}^{m} b_i}{m} \tag{9}$$

is the *average* value of marginal benefits b_i.

The second simplifying assumption makes the reduced form of all cost functions quadratic in emissions with the same degree of curvature. (Again, this property on the cost side greatly simplifies the analysis and interpretation without, I hope, losing too much realism.) Then, within the relevant range, the cost function $C_i(E_i)$ for agent i is assumed to be of the quadratic form

$$C_i(E_i) = \alpha_i - c_i\,E_i + \frac{\gamma}{2}\,E_i^2, \tag{10}$$

where α_i, c_i, γ are positive constants and the corresponding marginal cost function for agent

i (which is what really interests us here) is the analytically convenient linear form[7]

$$C_i'(E_i) = -c_i + \gamma\, E_i. \tag{11}$$

Combining (7) and (4) with (11), and then inverting the resulting expression, the optimal values of E_i^* are immediately characterized for each i by the equation[8]

$$E_i^* = \frac{c_i - \sum b_i}{\gamma}. \tag{12}$$

Henceforth I refer to the "World Climate Assembly" by its initials: WCA.

Armed with the two key simplifying assumptions (5) and (10) of this section, we are ready to analyze the voting equilibrium of the WCA in price P and, after adding some more structure, in quantity E.

3   Voting on a Single Price in the WCA

The nation in this section is presumed to vote its preferred carbon price in the WCA for some particular time period on behalf of its citizens and in accordance with their preferences. The nation then cooperates internationally to enforce the domestic imposition of the value of the internationally harmonized carbon price that is voted by majority rule in the WCA, and it recycles internally the domestic revenues raised by the tax-like price. I assume that this domestic recycling is efficient, as if by lump sum internal transfers, so there is no net national loss from the internally-imposed carbon price per se. To the extent that majority voting in the WCA mimics the outcome of a negotiating process, the voted price might mimic the negotiated uniform price.

In this section I examine the theoretical properties of a natural one-dimensional focus on voting (or negotiating) a single binding price on carbon emissions, the proceeds from which are domestically retained. For 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, whose level is voted upon, not the particular mechanism by which this binding minimum price is attained

---

[7]The relevant range of E_i for which (11) holds is $0 < E_i < c_i/\gamma$. Note that (11) is essentially the same marginal cost functional form used in my original "Prices *vs.* Quantities" article of 1974, where the specification (11) corresponded to a family of linear supply-reaction schedules having the same slope that are shifted up or down (or right or left) to varying degrees for various states of the world i.

[8]I assume the interior solution condition $c_i > \sum b_i$ holds for all i.

by a particular nation.[9] I elaborate further on this issue in my concluding remarks.

Let $E_i(P)$ denote the chosen emissions level of agent i as a parametric function of an imposed price P on emissions. Agent i is choosing $E_i(P)$ to minimize over E_i the expression $P E_i + C_i(E_i)$, resulting in the first order condition

$$-C_i'(E_i(P)) = P. \tag{13}$$

Plugging (11) into (13) and then inverting the resulting expression yields the formula

$$E_i(P) = \frac{c_i - P}{\gamma}. \tag{14}$$

Total emissions as a function of the imposed price P are $E(P) = \sum E_i(P)$, which, when combined with (14), gives the formula

$$E(P) = \frac{\sum_{i=1}^m c_i - mP}{\gamma} \tag{15}$$

Consider next what is the most preferred level of an internationally harmonized carbon price from the narrow perspective of agent i. This is the price that agent i would pairwise vote for in the WCA against any alternative price. Because revenues from the carbon price are collected and recycled by the nation to which i belongs, there is no tax burden per se. (The only real burden to i here is the higher compliance cost incurred by obeying condition (13)). The emissions-price level P_i that i would most prefer solves the problem

$$\max_P \{B_i(E(P)) - C_i(E_i(P))\}, \tag{16}$$

which satisfies the first-order condition

$$B_i'(E(P_i))E'(P_i) = C_i'(E_i(P_i)) E_i'(P_i). \tag{17}$$

It is readily shown that the second derivative with respect to P of the expression being maximized in (16) is negative, meaning that the first-order condition (17) is necessary and sufficient for a maximum, and also signifying that preferences over P for each i are single-peaked.

---

[9] 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.

From (6), (15), (13), (14), we have in (17) that $B_i' = -b_i$, $E' = -m/\gamma$, $C_i' = -P_i$, $E_i' = -1/\gamma$, which substitutions allows (17) to be rewritten as

$$P_i = m\, b_i. \tag{18}$$

What is the message of equation (18)? Compare the price that agent i most prefers, (18), with the socially optimal price, (8). The two expressions are close. Equation (18) is identical with equation (8), *except that* b_i substitutes for $\bar{b}$. In other words, agent i is picking what *would be* the socially optimal price *if* every other agent had identical benefits to agent i. It is as if agent i is attempting to find the social optimum consistent with its own benefit function being universalized (as if the total social benefit is $m\, B_i(E)$), in place of society's true total benefit function $\sum_{i=1}^{m} B_i(E)$.

In the context of voting on a uniform price, expression (18) immediately suggests an interpretation. Because the preferences of each agent are single peaked in price, the median voter theorem applied here to (18) predicts the outcome of a voting equilibrium to be the price

$$\overleftrightarrow{P} = m\, \overleftrightarrow{b}, \tag{19}$$

where $\overleftrightarrow{P}$ is the median of the sample $\{P_i\}$ while $\overleftrightarrow{b}$ is the median of the sample $\{b_i\}$. Now the symmetry between (8) and (19) is complete. The socially optimal price P^* given by (8) is m times the *mean* value of $\{b_i\}$, whereas the majority-voted price $\overleftrightarrow{P}$ given by (19) is m times the *median* value of $\{b_i\}$. Thus, the majority-voted price is close here to the socially optimal price to the extent that the median marginal benefit $\overleftrightarrow{b}$ is close to the mean marginal benefit $\bar{b}$. Although based on a particular specification, even so I consider this result to be an outcome in favor of majority voting for a uniform price on carbon, the proceeds from which are internally retained.

The model is indicating a sense in which population-weighted majority rule for an internationally harmonized carbon price might 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, agents representing various nations are, to a greater or lesser degree, internalizing the externality. Loosely speaking, a "representative" agent-nation is fully internalizing the externality 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.

As noted, the majority-rule carbon price $\overleftrightarrow{P}$ is close to the optimal carbon price P^* when the median marginal benefit $\overleftrightarrow{b}$ is close to the mean marginal benefit $\bar{b}$. This is as

good a result as one might hope for from a voting solution. The mean and the median are both measures of central tendency. At this level of abstraction I find it difficult to argue whether the mean marginal benefit of abatement per capita should be greater or less than the median marginal benefit of abatement per capita. If the two are equal, then majority voting obtains the optimal solution. If the two are unequal, the analysis provides a measure of how far away from optimality is majority rule. Of course this is just a model with quite restrictive assumptions, but in a post-Kyoto world of stalemated negotiations I find attractive the image of a WCA-style population-weighted median carbon price as being a useful point of departure that perhaps holds out some prospect of coming "close enough" to an optimal solution. Of course, readers are free to form their own conclusions from what is but a highly simplified model with specific functional forms.

This is about as far as theory can take us here. When the model is tightly structured with the specifications and assumptions applying to this section of the paper, the main result here indicates an exact sense in which majority rule for a harmonized national carbon price can come close to fully and completely internalizing the global warming externality. As was previously indicated, I think that the formal WCA voting result of the model of this section of the paper may perhaps be interpreted somewhat less formally as indicating that negotiating a uniform national carbon price may have a desirable property that favors supplying a near-optimal level of emissions. If the median marginal benefit (per capita) is close to the mean marginal benefit (per capita), then the socially optimal carbon price has the property that roughly half of the world's population wants the price to be higher, while roughly the other half of the world's population wants the price to be lower. This might be interpreted as a desirable feature even without the formal mechanism of majority-rule voting in the WCA.

Why does the marginal cost coefficient c_i not appear anywhere in the formula (19)? The underlying reason is that from (14), the agent has already performed a mental calculation of what would happen to its desired emissions E_i if the price P_i were very slightly changed. The fixed marginal cost coefficient c_i in (14) is irrelevant to a weighting by agent i of the incremental benefit against the incremental cost of very slightly changing the price P_i.

The analysis has participants in the WCA voting on a *single* universal price (or tax) that is internationally harmonized, but the proceeds from which are domestically retained. I think that a single price is a natural focal point for voting or negotiations. Thomas Schelling introduced and popularized the notion of a focal point in game theory.[10] Generally speaking,

---

[10]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.

a focal point of a multi-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 solution. 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. I am trying to argue here that a single universal price (or tax) that is internationally harmonized, but the proceeds from which are domestically retained, is a "natural" focal point for voting or negotiations.

I might also try to defend the idea of a single universal carbon price by relying on the higher transactions cost of voting or negotiating along multiple dimensions, as opposed to voting or negotiating along a single dimension. With respect to voting on more than one dimension, Arrow's impossibility theorem[11] indicates that there may not exist a voting equilibrium over multiple dimensions, which always exists in the case of single-peaked preferences along one dimension. With respect to negotiations over multiple dimensions, one might argue that the higher the number of dimensions, the more costly is it to negotiate a mutually satisfactory resolution. The concept of "transactions cost" is associated with the work of Ronald Coase.[12] The basic idea is that multiple 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 dimensionality of the target to be agreed upon.

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 multiple quantities – perhaps 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 assume that the essential contrast is between one binding price assignment versus multiple binding quantity assignments on the allocation of tradeable permits as a function of the total quantity of tradeable permits – and I then proceed to examine the consequences in terms of one-dimensional voting in a WCA.

---

[11]My description is a crude characterization of one aspect of multi-dimensional voting. See the Wikipedia entry on Arrow's impossiblity theorem and the many further references cited therein.

[12]Coase himself apparently 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).

I next proceed to analyze one-dimensional voting in the WCA on the total quantity of tradeable permits, which are then sub-allocated to each agent on the basis of a previously agreed upon predetermined formula.

4   Problems With Assigning Quantities Among Sovereign Entities

This paper now considers a system of voting (or negotiating) on the quantity side one world-wide aggregate emissions cap, *given* a previous-round subdivision formula for disaggregating the total cap into individual fractional caps. Such a formula might be set, for example, by a preceding agreement on various individual target reductions from various baselines. Unfortunately, there can be severe negotiating problems in agreeing on a first-round set of national fractional quantity targets that would indicate how to sub-divide an overall world-wide total cap on emissions. But, for the sake of comparing a majority voting rule for a one-dimensional price with a majority voting rule for a one-dimensional quantity, I assume for quantities that there is already in place an agreed-upon formula for decomposing the single quantity of total emissions into its component sub-assignment of individual caps.

With different national entities, a quantity-based treaty involves assigning different emissions quotas (whether tradeable or not). Quantity-based treaty making can be viewed as a coordination game with a number of different players. Such a game can have multiple solutions, often depending delicately on the setup and what is being assumed. In the case of Kyoto, the world has in practice arrived at a bad solution that has essentially devolved into regional volunteerism.

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 tradeable 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. I believe that such a multi-dimensional negotiating problem associated with quantity allocations is very difficult to resolve.

Within a nation, the government *assigns* caps. But *among* sovereign nations, 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.

With the following example, I try to illustrate some of the problems involved in setting the initial decomposition formula. After having voted in the WCA by majority rule the one-dimensional total quantity of allocated caps E in a cap and trade system, the decomposition formula sub-allocates the total E into individual tradeable caps $\{\widehat{E}_i\}$ where $E = \sum \widehat{E}_i$.

Let E_i^0 be the "pledged" or "targeted" or "base" level of emissions for agent i. I assume that the levels of pledged emissions E_i^0 are internationally accepted as a base case, although there is a severe free rider problem here because everyone wants their "negotiated volunteered" target level of E_i^0 to be as large as possible. The worldwide base level of total emissions is then $E^0 = \sum E_i^0$. Suppose the world is contemplating cutting back further on total emissions than E^0. Suppose that agent i agrees to a proportionate cutback of emissions from E_i^0 by the positive fraction λ_i of total world cutbacks, where $\sum \lambda_i = 1$. (In reality there is a severe free rider problem here too because everyone wants their "negotiated volunteered" decomposition coefficient of λ_i to be as large as possible.)

Let E represent the total quantity of worldwide emissions, which the WCA votes on by majority rule. Let $\widehat{E}_i(E)$ represent the emissions cap assigned to agent i as a function of the total quantity of emissions E that the WCA votes for. Then the proposed rule for decomposing E into $\widehat{E}_i(E)$ is

$$E_i^0 - \widehat{E}_i(E) = \lambda_i \left(E^0 - E \right) \tag{20}$$

whose reduced form is

$$\widehat{E}_i(E) = a_i + \lambda_i E, \tag{21}$$

where $a_i \equiv E_i^0 - \lambda_i E^0$. Then in (21) we have the property $\sum a_i = 0$ and $\sum \lambda_i = 1$.

From now on, in this section of the paper I work directly with the linear reduced form (21) as if the coefficients $\{a_i\}$ and $\{\lambda_i\}$ are given (always with $\sum a_i = 0$, $\lambda_i > 0$, and $\sum \lambda_i = 1$). If fact I regard the determination of the coefficients $\{a_i\}$ and $\{\lambda_i\}$ as being likely subject to severe free riding problems because it is in the narrow self interest of everyone to have their own values of a_i and λ_i be as large as possible. I do not believe that there are values here of the coefficients $\{a_i\}$ and $\{\lambda_i\}$ that correspond to the natural symmetry of assuming a majority vote in the WCA on the "natural" single uniform price of last section.

If each party would accept as *given* the assigned distributional coefficients $\{a_i, \lambda_i\}$ and the decomposition formula (21), one might then imagine voting for the total emissions E in a WCA. *Contingent* upon $\{a_i, \lambda_i\}$ being accepted as given, this system seemingly possesses the desirable property of having a one-dimensional locus of negotiations or voting (here on

E). And there is also some countervailing force against negotiating or voting for a high value of E. Although i's automatic assignment of a high emissions target $\widehat{E}_i(E)$ when E is high (via (21) with given $\lambda_i > 0$) helps i directly by lowering its emissions costs, this domestic effect is counteracted to some extent by the benefits that i would lose from high E because then everyone else would also be emitting more.

But now follow the thought experiment further by asking : Where do the distributional coefficients $\{a_i, \lambda_i\}$ come from in the first place? They are presumably the result of a multi-party negotiating process where there is no countervailing force to the selfish desire of country i to make its own a_i and λ_i as high as possible. With many different agents, there will be the usual difficult bargaining over many different distributional coefficients $\{a_i, \lambda_i\}$, 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 analogous to $\widehat{E}_i$ (or the fractions analogous to a_i and λ_i).[13] In this intra-national case there is no voting and there exists a kind of natural symmetric duality between a one-dimensional price P and a one-dimensional total quantity of tradeable permits E. But there is no international governance organization that has the unilateral power to assign caps or fractions. These caps or fractions must be *negotiated* among sovereign national entities. This breaks the one-dimensional symmetry because now one price P is contrasted with the asymmetry of many vested sovereign interests jockeying for initial distributions of the form $\widehat{E}_i$ or $\{a_i, \lambda_i\}$. There is thus a critical distinction between intra-national and international cap-and-trade systems. In the international case the initial distribution of cap permits 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.

But perhaps a formulation of this generality is biased against cap and trade. We might try to imbue the $\{a_i, \lambda_i\}$ with dimensionality-reducing salient qualities by imagining "naturally symmetric" focal allocations of $\{a_i, \lambda_i\}$. One such seemingly symmetric formula might be that each agent is assigned the same fractional reduction of emissions from some agreed-upon baseline year in the spirit of (20) with $\lambda_i = 1/m$. 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 mechanism!) to reduce greenhouse gas emissions in 2012 by an average of about

---

[13]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.

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.[14] Furthermore, and perhaps most distressingly, non-Annex I countries have not yet agreed to any actual future "binding" commitments on λ_i going forward from 2013. 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 of (21), it has been overwhelmingly problematic to assign binding quantity-like distributional coefficients $\{a_i, \lambda_i\}$ 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 so that $a_i = 0$ and $\lambda_i = 1/m$ in (21). 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, as previously mentioned, one might imagine negotiating (or even voting on) an identical per-capita fractional reduction $\lambda_i = 1/m$ from some base case of emissions E_i^0 in formula (20). In this situation, I think, everyone would first argue about the fairness of the baseline emissions E_i^0 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 multiple different sovereign nations I fear there will be intractable negotiations for multiple different distributional assignments, with no force countervailing each nation's free-riding desire to secure for itself a selfishly lenient emissions fraction.[15]

Here is what I believe is the essence of the one-price vs. multiple-quantities negotiation problem as elaborated in this section. A quantity-type system based on a formula like (21) involves *two* layers of negotiations. First, the various parties must agree on the various

---

[14]The one bright note 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).

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

quantity-like distributional coefficients $\{a_i, \lambda_i\}$ (with $\sum a_i = 0$ and $\sum \lambda_i = 1$). Then, second, the parties must agree on the single 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 makes sense to me that it is generally easier to negotiate one price layer than two quantity layers (whose first layer involves assigning multiple quantity-like distributional coefficients $\{a_i, \lambda_i\}$). Admittedly this argument depends upon a particular way of framing the issue, but it seems to me that, in international negotiations among many sovereign nations, there may be an irreducible asymmetry between one price instrument vs. multiple quantity instruments. In the next section I wave away the problems with assigning quantity decomposition fractions among sovereign entities enumerated in this section, in favor of pushing further the analysis of a single voted total quantity.

5 Voting on the Total Quantity of Tradeable Permits in the WCA

Despite the caveats of the last section of the paper, I begin this section by accepting that, somehow or other, quantity-like distributional coefficients are given and accepted as $\{a_i, \lambda_i\}$ in (21). I then proceed in this section to analyze WCA voting on total emissions E, which are subsequently decomposed further into individual caps $\widehat{E}_i(E)$ by the previously-agreed-upon formula (21). The algebra here is extremely messy, and I proceed quickly to wade through it by taking many shortcuts. What will emerge is that, compared with voting on a single price, voting on a single quantity of total emissions is made more complicated and more remote from social optimality by the distorting influence of agents taking account, in their WCA voting, of the value of their initial free assignment of caps in a cap-and-trade system. This distortion of an endowment-income effect is present in a quantity-based voting system but is absent from the price-based voting system already analyzed.

Begin with formula (15). If E is the total amount of permits issued, then it will also represent the total emissions after cap and trade. Expression (15) then gives $E(P) = (\sum c_i - mP)/\gamma$. Invert this expression to obtain

$$P(E) = \frac{\sum c_i - \gamma E}{m}. \tag{22}$$

Let $\widetilde{E}_i$ represent, in this quantity-based system, the amount of the equilibrium after-cap-and-trade actual physical emissions of agent i (expressed as a function of total emissions E).

Then $\widetilde{E}_i(E) = E_i(P(E))$ and, from (14), $E_i(P(E)) = (c_i - P(E))/\gamma$. Combining these two expressions and then substituting from (22) gives the formula

$$\widetilde{E}_i(E) = \frac{c_i}{\gamma} - \frac{\sum c_i}{m\gamma} + \frac{E}{m}. \tag{23}$$

The most preferred value of total emissions by agent i maximizes over all E the expression

$$B_i(E) - C_i(\widetilde{E}_i(E)) + P(E)[\widehat{E}_i(E) - \widetilde{E}_i(E)], \tag{24}$$

where the last term $P(E)[\widehat{E}_i(E) - \widetilde{E}_i(E)]$ represents the distortionary endowment-income effect of selling off permits $\widehat{E}_i(E) - \widetilde{E}_i(E)$ (alternatively buying permits $\widetilde{E}_i(E) - \widehat{E}_i(E)$) to other buyers (from other sellers) of tradeable permits.

Agent i most desires a *total* quantity of tradeable permits (equals total emissions) E that maximizes (24). Let this most-desired level by agent i of worldwide *total* tradeable permits (equals *total* emissions) be denoted E^i. Then E^i satisfies the first order condition

$$B_i'(E^i) - C_i'(\widetilde{E}_i(E^i))\,\widetilde{E}_i'(E^i) + P(E^i)[\widehat{E}_i'(E^i) - \widetilde{E}_i'(E^i)] + P'(E^i)[\widehat{E}_i(E^i) - \widetilde{E}_i(E^i)] = 0. \tag{25}$$

Into (25), substitute values $B_i'(E^i) = -b_i$, $C_i'(\widetilde{E}_i(E^i)) = -P(E^i)$, $\widetilde{E}_i'(E^i) = 1/m$, $P(E^i) = (\sum c_i - \gamma E^i)/m$, $\widehat{E}_i'(E^i) = \lambda_i$, $P'(E^i) = -\gamma/m$, $\widehat{E}_i(E^i) = a_i + \lambda_i E^i$, $\widetilde{E}_i(E^i) = c_i/\gamma - \sum c_i/m\gamma + E^i/m$. After canceling terms and inverting (25) to solve for E^i, tedious algebra yields the messy closed-form expression

$$E^i = \frac{-b_i(m/\gamma)^2 + \lambda_i m \sum c_i/\gamma^2 - a_i m/\gamma + mc_i/\gamma^2 - \sum c_i/\gamma^2}{(2\lambda_i m - 1)/\gamma}. \tag{26}$$

There is a one-to-one mapping between the total quantity of emissions E and the equilibrium price $P(E)$ corresponding to what is observed in equilibrium after trading the marketable permits. When agent i is voting for E^i, it is 'as if' agent i is voting for the equilibrium price $P^i = P(E^i)$. Substituting (26) into (22), one obtains, after rearrangement of terms, the 'as if' price that agent i votes for is

$$P^i = \frac{mb_i + \gamma a_i + \lambda_i \sum c_i - c_i}{2\lambda_i m - 1}. \tag{27}$$

Compare the preferred 'as if' price of agent i (obtainable indirectly from the preferred total quantity of tradeable permits in the WCA), P^i given by expression (27), with the preferred price of agent i when voting directly on a single price in the WCA, P_i given by expression (18). Expression (27) for P^i is far messier and more complicated than expression

(18) for P_i. It is much harder to argue that voting indirectly on P^i might be close to a social optimum than to argue (as was already done) that voting directly on P_i might be close to a social optimum. The reason that majority voting on P^i is likely further from an optimum than majority voting on P_i is because of the distortionary endowment-income effects having to do with the initial assignment of caps in a cap-and-trade system. This can be seen in expression (27) by examining, for example, the role of a_i. The bigger is a_i, the bigger is the 'as if' price P^i desired by agent i. The reason does not have to do with efficiency or social optimality. The reason is that with bigger a_i in the quantity system, agent i wants a higher price so that the market value of its initial assignment of tradeable permits is worth more. This distortionary endowment-income effect permeates formula (27) and indicates the source of the overall inefficiency in the formula. In choosing its preferred 'as if' price P^i, agent i is reacting too much to its initial endowment and not enough to what is socially optimal. By contrast, the price P_i preferred in formula (5) of the price-based system avoids altogether the distortionary influence of the initial allocation of quotas.

Contingent on everyone accepting as given the assigned distributional coefficients $\{a_i, \lambda_i\}$ and formula (21), we have derived agent i's preferred post-trade 'as if' price of emissions P^i in the quantity-voting system given by equation (27), which is generally distorted by an endowment-income effect. (The other side of the same coin is that it may not be so easy to get an agreement on formula (21) in the first place because agent i will fight to make its assignment of a_i and λ_i in the permit allocation process as high as possible, in order to obtain extra income from selling permits (or to lose less income from buying permits)).

Suppose, simply for the sake of argument, that the world accepts the symmetric principle of equal per-capita assignment of initial quotas, as given by the specification $a_i = 0$, $\lambda_i = 1/m$ in (21) for all i. As was indicated, I do not think this is terribly realistic because the revenues generated from a cap-and-trade system would flow as visible external transfer payments across national borders, which might be less easily tolerated by nations required to pay other nations large sums of taxpayer-financed money to buy permits. By contrast, a harmonized price of emissions that is internally retained does not involve any transfers of payments across national borders.

Suspending disbelief, then, suppose that all agents i accept the symmetric formula $a_i = 0$ and $\lambda_i = 1/m$. (I have already indicated why there might be resistance to accepting this formula, but here I want to examine the consequences for P^i as given by equation (27)). Therefore, substitute into (27) the further specification $a_i = 0$ and $\lambda_i = 1/m$. This substitution simplifies equation (27) into

$$P^i = mb_i + (\bar{c} - c_i), \tag{28}$$

where $\bar{c} = \sum c_i/m$ is interpretable as a measure of worldwide average cost.

Expression (28) should be compared with the majority-voted uniform price result (18), where $P_i = m\, b_i$. As was pointed out, the socially optimal price P^* given by (8) is m times the *mean* value of $\{b_i\}$, whereas the majority-voted price $\overleftrightarrow{P}$ given by (19) is m times the *median* value of $\{b_i\}$. Thus, the majority-voted price (19) is close to the socially optimal price to the extent that the median marginal benefit $\overleftrightarrow{b}$ is close to the mean marginal benefit $\bar{b}$. As has been noted, this seems like a strong result, even though a particular specification lies behind it.

By contrast, even after the additional simplifying assumption on the quantity side that $a_i = 0$ and $\lambda_i = 1/m$ for all i, the preferred 'as if' price of agent i (via preferred total emissions E^i), namely P^i in (28), still retains a distortionary endowment-income effect having to do with the initial assignment of caps in a cap-and-trade system, which shows up as the "extra" term $(\bar{c} - c_i)$ in (28). In order for (28) to deliver social optimality requires that the *mean of* $\{b_i\}$ equals the *median of* $\{b_i + (\bar{c} - c_i)/m\}$, which is a less likely condition to be met than the *mean of* $\{b_i\}$ equaling the *median of* $\{b_i\}$.

6   Concluding Remarks

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 different 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-framework. 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.

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 many different 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 low tax rates for themselves.

Of necessity, this paper has been laced with subjective judgements. This, unfortunately, is the nature of the subject. To repeat yet again, 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 multiple binding quantities or quantity-like distributional coefficients. In this context, it seems that WCA voting on a carbon emissions price is less distortionary from an optimal outcome than WCA voting on a total quantity of emissions in a cap-and-trade system.

The model of this paper is so abstract, so over-simplified, and so removed from reality 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 model here is exposing a fundamental countervailing-force argument for a majority-voted uniform price that deserves to be highlighted.

Because the model is at such a high level of theoretical 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.[16] 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 theoretical 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

---

[16]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.

the nature of the global warming externality problem. The theory of this paper suggests 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 a paper. In any event, *this* paper is not primarily about practical considerations of international negotiations. I leave that important task mostly to others.[17] However, I do want to mention just a few real-world considerations that have been left out of the model yet seem especially pertinent.

A binding international agreement on a uniform minimum carbon price presumably 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 for a price-based system. 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.[18]

An efficient carbon price naturally produces more winners than losers by 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 to be relatively modest second-order deadweight-loss triangles instead of the relatively immodest first-order rectangle transfers associated with tradeable permits.

Another practical issue 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]

---

[17]See, e.g., Bodansky (2010) or Barrett (2005).

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

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

I close by noting again that global warming is an extremely serious as-yet-unresolved externality 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 model of this paper is indicating 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

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

[2] Barrett, Scott (2005). *Environment and Statecraft: The Strategy of Environmental Treaty Making.* Oxford: Oxford University Press.

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

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

[5] Colman, A. M. (2006). "Thomas C. Shelling's psychological decision theory: Introduction to a special issue." *Journal of Economic Psychology*, 27: 603-608.

[6] 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.

[7] Cramton, Peter, Axel Ockenfels, and Steven Stoft (2013). "How to Negotiate Ambitious Global Emissions Abatement." Mimeo, 30 May 2013.

[8] 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.

[9] Goulder, LH, MAC 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.

[10] Goulder, Lawrence H., and Andrew R. Schein (2013). "Carbon Taxes vs. Cap and Trade: A Critical Review." *Climate Change Economics* 4(3): 1-28.

[11] Hoel, Michael, and Larry Karp (2002). "Taxes vs. quotas for a stock pollutant." *Resource and Energy Economics* 24: 367-384.

[12] Libecap, Gary D. (2013). "Addressing Global Environmental Externalities: Transaction Costs Considerations." *Journal of Economic Literature*, 52(2): 424-79.

[13] Metcalf, Gilbert E., and David Weisbach (2009). "The Design of a Carbon Tax." *Harvard Environmental Law Review* 33.2: 499-556.

[14] 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.

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

[16] Nordhaus, William D. (2015). "Climate Clubs: Designing a Mechanism to Overcome Free-riding in International Climate Policy." Presidential Address delivered to the American Economic Association January 4, 2015.

[17] Pizer, William (1999). "Optimal Choice of Policy Instrument and Stringency under Uncertainty: the Case of Climate Change." *Resource and Energy Economics*, 12: 255-287.

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

[19] Victor, David (2011). *Global Warming Gridlock.* Cambridge: Cambridge University Press.

[20] Weitzman, Martin L. (1974). "Prices vs. Quantities." *Review of Economic Studies* 41, 4: 477-491.