On a World Climate Assembly and the Social Cost of Carbon

Martin L. Weitzman*

May 21, 2017 revision

Abstract

This paper argues that a uniform global tax-like price on carbon emissions, whose revenues each country retains, can provide a focal point for a reciprocal common climate commitment, whereas quantity targets, which do not nearly so readily present such a single focal point, tend to rely ultimately on individual quantity commitments. The paper postulates the conceptually useful allegory of a futuristic "World Climate Assembly" (WCA) that votes for a single worldwide price on carbon emissions via the basic democratic principle of one-person one-vote majority rule. A WCA-like uniform tax-price (whose proceeds are domestically retained) counters self-interest by incentivizing countries or agents to internalize the externality because each WCA-agent's higher abatement cost from a higher emissions price is counter-balanced by that agent's extra benefit from inducing all other WCA-agents to simultaneously lower their emissions in response to the higher price. The paper derives fresh insights and new simple formulas that relate each emitter's most-preferred world price of carbon to the world "Social Cost of Carbon" (SCC), and further relates the WCA-voted world price of carbon to the world SCC. I argue that the WCA-voted price and the SCC are unlikely to differ sharply. Some implications are discussed. The overall methodology of the paper is a mixture of mostly classical with some behavioral economics.

JEL Codes: F51, H41, Q54, Q58, K33

*Department of Economics, Harvard University (mweitzman@harvard.edu). For helpful comments on an earlier draft, but without implicating them in errors or interpretations, I am indebted to Joseph Aldy, Thomas Aronsson, Daniel Bodansky, Patrick Bolton, John Broome, Dallas Burtraw, Vincent Crawford, Duncan Foley, Roger Guesnerie, Louis Kaplow, Robert Keohane, Matthew Kotchen, N. Gregory Mankiw, Gilbert Metcalf, Juan Moreno-Cruz, Axel Ockenfels, Ian Parry, John Roemer, Richard Schmalensee, Joseph Shapiro, Thomas Sterner, Cass Sunstein, Massimo Tavoni, Gernot Wagner, Jorgen Weibull, David Weisbach, and two anonymous referees.

Keywords: Climate change, Global warming, International public good, Prices versus quantities, Social cost of carbon, World climate assembly, Climate club, Paris agreement.

1 Introduction

Climate change is a global public-goods externality whose formal 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 (CO_2), which is by far the most important greenhouse gas (GHG), but in principle the discussion could be extended to emissions of all relevant GHGs. Throughout the paper I blur the distinction between carbon dioxide and carbon because the two are linearly related.[1]

The core problem confronting the political economy of climate change is an inability to coordinate global social outcomes 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. A national government can (at least in principle) *impose* targets or policies on national public goods. With climate change there is no overarching global governance mechanism capable of coordinating the actions necessary to overcome the international problem of free-riding. Instead, instruments of control, such as prices and/or quantities, must be *negotiated* among sovereign nations.

My point of departure throughout this paper is the critical centrality of the international free-rider problem as the primary cause of negotiating difficulties on climate-change emissions. Negotiators here are playing a game in which self-interested strategies are a crucial consideration. Negotiating rules "frame" an important part of the game, and can thereby "frame" the form that self-interest takes, for better or for worse. The basic challenge, as I see it, is to construct a relatively simple, familiar, transparent, and acceptable one-dimensional international quid-pro-quo mechanism, which automatically aligns self interest with world interests by embodying the principle of "I will if you will."[2]

Throughout this paper I basically argue that a uniform global tax-like price on carbon emissions, whose revenues each country retains, can provide a focal point for a reciprocal

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

[2] For more about the intellectual coherence of this quid-pro-quo price mechanism, see MacKay, Cramton, Ockenfels and Stoft (2015) and the references they cite. They originated the phrase "I will if you will."

common commitment, whereas quantity targets, which do not nearly as readily present such a single focal point, have a tendency to deadlock on free-style individual commitments. As a consequence, negotiating a uniform minimum global carbon tax or price can help to solve the externality problem while individual quantity caps tend to incorporate it. I explain why negotiating or voting a uniform minimum carbon price embodies what I will call a "countervailing force" against narrow self-interest by automatically incentivizing all negotiating parties to internalize, at least approximately, the global warming externality. The paper formalizes the sense in which each agent's extra abatement cost from a higher uniform 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 uniform price.

The style of this paper is a sometimes awkward blend of classical with some behavioral economics. In a few places the paper reads more like a legal-scholarly argument based on a "preponderance of evidence" than a fully rigorous mathematical argument derived from axiomatic first principles. Such a mixture of modes of argument seems unavoidable in discussing actual attempts to resolve the free-rider problem associated with the international public good of climate change.

This paper focuses on the price of carbon as the standard for measuring and comparing mitigation efforts. The paper postulates and analyzes the conceptually useful allegory of a futuristic "World Climate Assembly" (WCA) that votes for a single worldwide minimum price on carbon emissions via the basic democratic principle of one-person one-vote majority rule. At a high level of abstraction, the WCA is very roughly patterned on how a representative democratic legislature within a democratic state might ultimately decide, by majority rule, the level of spending on within-state public goods. Taken less literally, the thought experiment of a WCA may help us to concentrate our thinking on what negotiations might be trying to accomplish. If the spirit of this WCA conceptual framework can be accepted in the first place (as a kind of a "worldwide plebiscite"), then voting or negotiating a single internationally-binding minimum carbon price, the proceeds from which are domestically retained, tends to counter self-interest by having the important property of incentivizing countries or agents to automatically internalize the externality. In the concluding section of the paper I discuss what might entice countries to entertain a WCA-voted majority-rule price of carbon and why countries might uphold the results of this WCA-voted outcome.[3]

Some of the themes presented here have been preliminarily explored in previous papers (Weitzman (2014, 2016)). This paper extends previous work by having a more coherent and comprehensive motivating discussion, followed by five new propositions that give some

[3]In particular, I discuss the possible relationship between this paper's proposed WCA and the "Climate Club" proposed by William Nordhaus (2015).

fresh insights and novel interpretations to the problem. New simple formulas are derived that relate each emitter's single-peaked most-preferred world price of carbon to the world "Social Cost of Carbon" (SCC), and that also relate the WCA-voted price of carbon to the SCC. Using these new results, the paper argues that the WCA-voted majority-rule carbon price and the SCC are unlikely to differ sharply. An extremely simplified numerical exercise roughly supports this conclusion.[4] Some implications of the paper's main results are discussed.

2 Brief Background History of Climate Negotiations

From the actual entering into force of the Kyoto Protocol in February 2005 to the Paris COP21 Agreement of December 2015 (and perhaps long afterwards), the world seems mired in what has been called "global warming gridlock."[5]

The Kyoto Protocol, negotiated in December 1997, began by dividing the world into two huge blocs. The "Annex I" bloc of countries included most of the world's high-income advanced industrial nations. The rest of the world, the "non-Annex I" bloc of countries, included most of the world's low-income developing nations. In a gesture towards the principle of top-down coordination, the Annex I countries agreed to "legally binding" average emissions reductions in 2008-2012 of approximately 5% relative to their baseline emissions of 1990. The non-Annex I countries were not constrained by "legally binding" emissions reductions, but otherwise agreed to cooperate.

In reality, the "legally binding" emissions reductions of the Kyoto Protocol were anything but, because there was no provision for a mechanism to enforce compliance. There was no provision for a mechanism to enforce compliance because, essentially, the parties did not really want to be bound by such a mechanism.

Almost from the beginning, the United States and Australia refused to ratify the Kyoto treaty (ostensibly on the grounds that the non-Annex I countries were unfairly exempt from responsibilities). Subsequently, Canada, Japan, and Russia pulled out of their part of the agreement and refused to take on future commitments.

[4] The numerical example is due to Kotchen (2016), one section of whose wide-ranging paper on the many possible interpretations of the SCC touches upon some themes of this more-narrowly-focused paper.

[5] *Global Warming Gridlock* is the title of a book by David Victor (2011), who popularized the phrase. For more information on the Kyoto Protocol, see the Wikipedia entry for "Kyoto Protocol" and the many other relevant references cited there. For more information on the Paris COP21 Agreement, see the Wikipedia entry for "Paris Agreement" and the many other relevant references cited there. A balanced evaluation of COP21 (and comparisons with the Kyoto Protocol) is given in Keohane and Oppenheimer (2016). They describe the COP21 Paris Agreement as "discretion and vagueness" replacing the Kyoto Protocol's "mandates and simplicity."

I think it is fair to say that the "spirit" of Kyoto was a top-down intended adherence to something like the following scenario. The Annex I countries would agree to show good faith first by voluntarily lowering their average emissions in 2008-2012 by about 5% relative to their 1990 emissions. Then, in a second stage, after around 10 years (approximately by 2010 or so), the hope was that the non-Annex I countries would be impressed by the good faith effort shown over the previous decade by the Annex I countries and would (hopefully) join the effort by pledging something like, say, an emissions reduction target of about 5% in 2020 (relative to 1990 emissions), while the Annex I countries would agree to a more stringent emissions reduction target of, say, about 10%. In reality, no such second stage of synchronized ratcheted-up commitments ever materialized.

The recently concluded Paris COP21 Agreement of December 2015 (by contrast with Kyoto) made no formal distinction between developed and developing countries. In principle, all nations were treated symmetrically. The Paris Agreement nominally covered countries currently accounting for some 95% of world carbon dioxide emissions. Countries agreed to make voluntary pledges by whatever (not necessarily comparable) formulas they chose, now named euphemistically "Intended Nationally Determined Contributions" (INDCs). COP21 committed countries to report INDC compliance every five years or so and to set new (and hopefully more ambitious) INDCs going forward to the future, a policy sometimes labeled "pledge and review."

All in all, it seems fair to say that the COP21 Accord, which is as yet not fully tried, represents movement in a direction that may slow GHG emissions. An important first step achieved in COP21 is the seeming acceptance of a repeated five-year cycle: pledge-verify-review-repledge-etc. At the end of the day, COP21 appears to me to be essentially a gamble that the relatively modest voluntary bottom-up reductions in emissions may buy enough time for the world to develop inexpensive future carbon-free technologies. This seems a risky bet. (Remember, it is stabilized atmospheric *stocks* of GHGs at low levels that matter for limiting climate change damages, which requires *zero net flows* of emissions – a vastly more ambitious goal than stabilizing GHG flows per se.) In any event, it will take maybe a decade or so to sort out the effectiveness of the Paris COP21 Agreement. Thus far, at least, the modest voluntary reductions in emissions seem not nearly enough to keep global warming on a track below the stated goal of no more than a worldwide average temperature increase of 2°C.[6]

The core weakness of the COP21 Paris Agreement is essentially the same as the core weakness of the Kyoto Protocol. Neither approach addresses the central problem of free-riding on an international public good of great importance. Under COP21, there is no penalty for voluntarily setting under-ambitious national targets, and furthermore there is

[6] Much less the virtually-impossible 1.5°C goal also mentioned in the Paris COP21 Agreement.

also no penalty for non-compliance by a country with its own voluntary self-announced targets. The only mechanism countering under-ambition or non-compliance is "blame and shame," which seems to me like a weak incentive for cutting back significantly on the free-riding associated with global carbon emissions.

I think the INDC label says a lot, even granting that the language was constrained by realpolitik diplomatic compromises. The *contributions* are chosen by each country. These COP21 *contributions* are *intended* and *nationally-determined*.[7] It is hard for me to envision how the labels could more strongly emphasize the strictly voluntary nature of the entire exercise. This does not seem likely to overcome free-riding in the international public goods problem that is central to climate change policy.

Consider the following hypothetical science-fiction-like thought experiment. In the U.S. clean air amendments of 1990, the U.S. Environmental Protection Agency (EPA) essentially *assigned* the initial caps on sulfur dioxide (SO_2) emissions to various power-plant emitters and allowed (or even encouraged) a cap-and-trade system. Suppose, instead of assigning initial caps on SO_2 emissions, the EPA had allowed power plants or companies or states or regions to *voluntarily* negotiate between themselves their own initial caps on SO_2 emissions. Suppose further that no penalties were imposed by the EPA for either under-ambitious voluntary targets or under-fulfillment of these under-ambitious voluntary targets. Everyone would conclude that this was a crazy idea that stood effectively no realistic chance of seriously curtailing U.S. SO_2 emissions. Yet this is essentially the way that COP21 proposed to deal with CO_2 emissions on a worldwide scale via the "Nationally Determined Contribution" (NDC) approach. Maybe this is an unfair comparison because COP21 is much better than no international agreement at all. It remains to be seen whether COP21 can evolve into a worldwide system with penalty-teeth that actually "bite" but, again, it seems to me like a risky bet.

If the Paris COP21 approach fails to halt "dangerous anthropogenic global warming" – in the form of a perception of an impending climate catastrophe that is felt on a worldwide grassroots level – then I think there may be more pressure on creating a top-down international mechanism with actual sanctions that actually work. Desperate times demand desperate measures. If climate change becomes sufficiently threatening to an "average" citizen of the world on a grassroots level, then public opinion may support relinquishing some national sovereignty in favor of the greater good. Opportunities for comprehensive top-down solutions will likely arise (probably in response to future perceptions of climate-linked disasters) and we should be ready beforehand by thinking through the consequences

[7] When a state formally submits its instrument of ratification to the UNFCCC Secretariat, it is supposed to include its "Nationally Determined Contribution" (NDC, no I). Even so, I think the term NDC is revealing.

now. This paper is futuristic (or, perhaps, even science-fiction-like) in the sense that it is targeted towards the eventuality of a meaningful top-down climate change treaty that goes well beyond the narrow volunteerism of COP21.

3 Parable of a World Climate Assembly

The inspiration behind the line of reasoning in this paper is the perception of a strong need for some radical rethinking of international climate policy. As a possibly useful conceptual guide for what negotiations might hope to accomplish, I ask the reader to temporarily suspend disbelief by considering what might happen in a futuristic "World Climate Assembly" (WCA) that votes for a *single* harmonized worldwide minimum price on carbon emissions, which will apply for everyone, via the basic democratic principle of one-person one-vote majority rule.[8] In this conceptualization, nations would vote along a single price dimension for their desired level of emissions stringency on behalf of their citizen constituents, but the votes are weighted by each nation's population. Applying the median voter theorem would then yield a weighted-median outcome, where the weights are the nation's population. An important part of this setup is that each nation retains internally the proceeds from the internationally harmonized price-tax, which in the model will be assumed to be revenue neutral for the nation as a whole.

The idea of a WCA should be viewed as an idealized attempted solution of the international free-riding problem of GHG emissions. It should be conceptualized as an alternative way of thinking about the unattainable first best. Right now, anything like a WCA seems hypothetical and hopelessly 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 that big new problems may require big new solutions.

In a future world searching for some effective solution to the important externality of climate change, perhaps it is at least worth considering establishing a new organization along the lines of a WCA. After all, even were the world to agree to focus on a one-dimensional harmonized carbon price, it is useful to have some concrete fallback decision mechanism behind vague "negotiations" because there are bound to be disagreements, whose resolution

[8] In principle, one could consider alternative WCA voting rules, such as one vote per country, or one vote per dollar of GDP, or so forth. In these alternatives there is no good analogue of the basic propositions of this paper. Furthermore, I think the inherent democracy of one-person one-vote is an attractive feature in and of itself, thereby perhaps easing acceptance of WCA participation.

is unclear, about what that common carbon price should be. I essentially assume that it is in the interest of enough nations to forfeit their free-rider 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 a 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 could be useful in indicating what might be the outcome of less-formal international negotiations on a harmonized price of carbon.[9]

It could be objected that a "consensus" voting rule, not a majority voting rule, is employed in negotiations under the UN Framework Convention on Climate Change (UNFCCC). 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 seems 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 is unlikely to be resolved without a binding agreement on, and enforcement of, some overall formula for dividing emissions responsibilities among nations. Voluntary altruism alone will almost surely 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 determining targets, verifying fulfillment, and punishing non-compliance. The question then becomes: *which* collective-commitment frameworks and formulas are more promising than which others? Again, any "answer" must come in an unavoidable mixture of behavioral and classical economic analysis.

4 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 worldwide binding price on carbon emissions, the proceeds from which are domestically retained. For expositional simplicity, I identify this single binding price

[9] As a possible example of tentative movement in this direction, the World Bank (2015) is currently attempting to apply a worldwide SCC of \$30/tCO_2 in its program evaluations, rising over time to \$80/tCO_2 in 2050. Several countries have adopted a carbon price for evaluating regulatory impacts of domestic policies – including Canada, Finland, France, Germany, Italy, Mexico, Netherlands, Norway, Sweden, United States, United Kingdom. The U.S. Government currently uses a carbon price of \$40/tCO_2 as a point estimate for cost-benefit analysis of environmental regulations, growing over time at approximately the appropriate discount rate (see Interagency Working Group on the Social Cost of Carbon (2015)). As of 2017, the ultimate future fate of this SCC=\$40/tCO_2 estimate seems uncertain.

on carbon as if it were an internationally-harmonized, nationally-collected carbon tax. At a theoretical level of abstraction, I blur the distinction between a carbon price (however it is attained) and a carbon tax. At first I merely assume the acquiescence by each nation to a common binding minimum price on carbon, the proceeds from which are domestically retained. In the concluding section of the paper I investigate mechanisms and sanctions that might be used to induce, or even compel, membership in a World Climate Assembly that votes for a single worldwide price on carbon emissions via the basic democratic principle of one-person one-vote majority rule.[10]

A system of uniform national carbon taxes with revenues kept in the taxing country is a relatively simple and transparent way to achieve an internationally-harmonized carbon price.[11] 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 whichever internal mechanism they choose – a tax, a cap-and-trade system, a hybrid system, or whatever else results in an observable price of carbon not below the internationally-agreed minimum. I elaborate further on this issue later.

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 a direct refund in the form of equal lump sum "carbon dividend" payments to each citizen.[12] This, I think, is a desirable property. I will assume that the net effect of taxing carbon and rebating the tax returns domestically is essentially revenue-neutral. By contrast, the revenues generated from an internationally harmonized cap-and-trade system flow as highly visible (and highly variable) external transfer payments across national borders, which might be politically intolerable for countries required to pay other countries very large (and highly variable) sums of taxpayer-financed money to buy permits.[13] We economists, I fear, have failed to sufficiently convince the public that there is an important distinction between a self-collected national tax-price on carbon emissions and an external cross-border price imposed

[10] In particular, my proposed WCA solution will be linked there to William Nordhaus's (2015) proposed idea of a "Climate Club."

[11] There is a fair-sized literature arguing for a carbon-tax (or carbon-price) approach. See, e.g., Metcalf and Weisbach (2009), Cooper (2010), Cramton, Ockenfels and Stoft (2015), Nordhaus (2013), and the many further references cited in these works. For some considerations that favor cap-and-trade see, e.g., Gollier and Tirole (2017) and the references they further cite.

[12] An attempt at an equitable redistribution was done in the Canadian province of British Columbia, which is widely viewed as having a successful carbon-tax program. For a balanced evaluation of the B.C. carbon tax experience, see Harrison (2013).

[13] Goulder and Schein (2013), among others, discuss the potential for very large cross-border flows of money from nations buying allowances to nations selling allowances. In this sense, much more is at stake in negotiating a favorable initial assignment of allowances within an international cap-and-trade system than in negotiating a uniform price-tax that is nationally retained. I return to this important theme in the concluding section of the paper.

as if being taxed by an outside party, which siphons tax-like revenues away from the home country.[14]

There already exists a sizable "traditional" literature comparing an international carbon-emissions tax with an international cap-and-trade system.[15] Arguments can be made on both sides. In my opinion, one of the most important "traditional" advantages of carbon-emissions taxes over a cap-and-trade system is the elimination of price volatility, which is extremely poorly tolerated by businesses, politicians, and the public, all of whom desire to rely on a known stable price of CO_2 emissions. Borenstein (2016) surveys existing cap-and-trade systems for carbon emissions and concludes that price volatility is "a major flaw in cap-and-trade" because "a major predictor of variation in GHG emissions is [the state of] the economy" and "the probability of hitting a middle ground – where allowance prices are not so low as to be ineffective and not so high as to trigger a political backlash – is very low."[16]

In this paper I revisit the debate of "prices versus quantities" from a different, non-traditional, partially-behavioral, angle. Going back to basics, I would suggest that the instruments of negotiation for helping to resolve the global warming externality by curtailing GHG emissions should ideally possess three desirable properties.

1. *Induce cost-effectiveness.*

[14] Several economists have argued that retained carbon taxes offer a "double dividend" in the sense that overall distortions from taxes on labor and capital can be reduced enough that there is a net benefit after wisely recycling carbon tax revenues. Other things being equal, this would be an argument in favor of a higher price-tax. (For surveys on the literature of the double dividend, see Goulder (2002) or Jorgenson et al (2013).) One could also argue the opposite position – that increased tax revenues will be squandered by some governments, thereby yielding a "negative dividend." Other things being equal, this would be an argument in favor of a lower price-tax. In this paper I assume, as a starting point, net revenue neutrality of a carbon tax. Later I indicate how, if there were a constant net tax-offset revenue-recycling effect of the type described here, it could be incorporated into the relevant damages coefficient and all of the theory would go through.

[15] See, e.g., the survey by Goulder and Schein (2013) and the many further references they cite (including Weitzman (1974)). Relative price stability is a consideration that emerges from this literature. Ease of administration (and avoidance of stealing) is another. A third desideratum is minimal payments traversing international boundaries, because such payments potentially represent a politically explosive issue.

[16] On the price volatility of cap-and-trade systems, see also Aldy and Viscusi (2015). The influence of so-called "complementary policies" (such as aggressive sector-specific command-and-control targets like specific carbon-intensity energy standards) can play a role in making the price elasticity of demand for carbon allowances low, resulting in large price fluctuations within a cap-and-trade system. In the presence of a binding cap-and-trade regime (as opposed to a binding tax-price regime) aggressive sector-based targets can have the perverse effect of relocating CO_2 emissions to other sectors, but not reducing net total emissions. This would not happen with a tax-price system, where sector-based targets will add to total emissions abatement. See, e.g., the discussion in Goulder and Schein (2013). There is also a purely economic argument against price volatility. The present discounted cost-minimizing way to abate is to equalize present discounted marginal costs over time, which means a stable emissions price growing at the rate of interest and allowing quantities to vary over the business cycle. Thus, unless they are banked in an ideal way, permits are not as cost-effective as a tax. I owe this insight to N. Gregory Mankiw.

2. Be of *low – preferably one – dimension based on a "natural" focal point* to facilitate finding an agreement with relatively low transaction costs.

3. *Embody "countervailing force" against narrow self-interested free-riding by automatically incentivizing all negotiating parties to internalize the externality via a simple, familiar, transparent instrument and formula that incorporates a common climate commitment based on principles of reciprocity, quid-pro-quo, and I-will-if-you-will.*

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

On the first desirable property, in principle both a carbon price and tradable permits achieve decentralized cost-effectiveness (provided agreement can be had in the first place). In principle, either approach minimizes total compliance costs and is more comprehensive than, and superior to, a patchwork of command-and-control regulations on carbon emissions.

The second desirable property (low dimensionality) argues in favor of a one-dimensional harmonized tax-like carbon price over an n-dimensional harmonized cap-and-trade system among n nations. Alas, this "curse of dimensionality" argument is elusively difficult to formulate rigorously, or even to articulate coherently. My argument here is necessarily, at least in part, behavioral or psychological or cultural, and relies on empirical counterexamples. In this situation two important empirical counterexamples are the breakdown of the quantity-based top-down Kyoto approach and the under-ambitious quantity-based bottom-up "Nationally Determined Contributions" (NDCs) actually volunteered by most nations under the COP21 Paris Agreement.

For n different national entities, a quantity-based treaty involves assigning n different emissions quotas (with or without side payments and 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 had in practice arrived at a bad quantity-based top-down solution that essentially devolved into free-rider regional volunteerism. The ultimate outcome of the COP21 Paris Agreement remains to be seen, but so far the quantity-based bottom-up NDCs actually volunteered by the parties seem underwhelming, even leaving aside the near-impossibility of achieving the stated goal of keeping global warming below 2°C.

In 1739, David Hume (*A Treatise of Human Nature*, Section VII) outlined a basic argument favoring the success of low-n negotiations over high-n negotiations.[17]

[17] "Two neighbors may agree to drain a meadow, which they possess in common; because 'tis easy for them to know each other's mind; and each must perceive, that the immediate consequence of his failing in

Some two hundred years later, Ronald Coase (1937, 1960) introduced, and subsequently popularized, the concept of "transaction cost".[18] The basic idea, as applied here, is that n parties to a negotiation can be prevented from attaining a socially desirable Pareto-efficient outcome, with side payments, by the (search, information, coordination, bargaining, decision, monitoring, policing, enforcing, etc.) costs of transacting the agreement among themselves. Other things being equal, it seems eminently plausible that transaction costs increase monotonically with the number of parties n.[19]

In 1960, Thomas Schelling introduced, and subsequently popularized, the notion of a "focal point" in game-theoretic negotiations.[20] As applied to the setup of this paper, 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 by the parties 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 because, ultimately, a focal point is whatever people *believe* is a focal point. While I think that Schelling's focal-point argument is an important behavioral-psychological-cultural insight into how negotiations might work, it is difficult to model rigorously, and actual applications seem to be as much of an art as a science.[21]

Negotiating a one-dimensional uniform price with single-peaked preferences has the significant additional property of allowing a simple majority-rule voting equilibrium in the form of the median-voter result of Duncan Black. Importantly here, the one-dimensional case of a

his part, is, the abandoning of the whole project. But 'tis very difficult, and indeed impossible, that a thousand persons shou'd agree in any such action; it being difficult for them to concert so complicated a design, and still more difficult for them to execute it; while each seeks a pretext to free himself of the trouble and expence, and wou'd lay the whole burden on others." This quote is cited in Nordhaus (2017).

[18] See Coase (1937) and Coase (1960). Coase himself apparently did not invent or even use the term "transaction cost" but he prominently employed the concept. For an application of the transaction cost approach to controlling greenhouse gas emissions, see Libecap (2013).

[19] Dixit and Olson (2000) formalize a particular version of this argument.

[20] 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.

[21] David Weisbach suggested to me the following analogy for judging whether a focal point is more or less salient. Consider cap-and-trade compared with a price-tax. In each case, the natural focal point is equality for the salient feature of the respective regulatory system. For cap-and-trade, the natural focal point is equal per-capita assignment of allowances. This would involve very large taxpayer-financed trans-border revenue flows from the nations purchasing allowances to the nations selling them, which would generate massive conflict about what is "fair" and has no realistic chance of being accepted. For a price-tax, the natural focal point is "equal tax rate, keep the revenues," which by comparison seems much more tolerable. This is not a proof, but I believe that Weisbach's insight is a helpful behavioral-economic intuition.

single price (with single-peaked preferences) avoids the Arrow impossibility theorem, which states, loosely speaking, that no consistent social choice mechanism exists for making group decisions involving multiple dimensions.[22] Note that this part of the argument, concerning social choice mechanisms, has rigorous foundations and does not rely so directly on behavioral assumptions.

In the case of international negotiations on climate change, I believe that Hume's emphasis on the importance of low-number simplicity, Coase's concept of transaction costs, Schelling's notion of a salient focal point, Black's median-voter result, and Arrow's impossibility theorem can all be used as 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. The "law of one price," which for emissions abatement implies equal marginal effort, is already a familiar salient feature of competitive markets, whereas there is no such thing as a "law of one quantity." I cannot defend the salience of one price rigorously, other than to argue that it is roughly "in the spirit of" Hume, Coase, Schelling, Black, and Arrow. At the end of the day, this argument constitutes more of a plausible behavioral conjecture than a rigorous theorem. I believe, though, that the "preponderance of evidence" points strongly in this direction.

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 and thereby align self interest with society's interests. Such incentives should ideally take the form of a simple, reciprocal, common climate commitment based on a familiar transparent mechanism that embodies the quid-pro-quo principle of "I will if you will." 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 (regardless of whether or not my cap will be tradable as an allowance 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 this is a crucial distinction for the success or failure of a cap-and-trade regime. A quantity-based international system fails because

[22] Mas-Collell, Whinston, and Green (1995) contains a textbook treatment of the Black median-voter result and the Arrow impossibility theorem.

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.

A domestically-collected, internationally-harmonized carbon price is different. If the price were imposed on me alone, then 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 minimum price of carbon emissions has a built-in self-enforcing mechanism that countervails free-riding. This theme is investigated formally in the next sections of the paper.

In previous work (e.g., Weitzman (2014)), I tried to model the role of this third "countervailing force" property of an internationally-harmonized and nationally-collected carbon price, but the results were clumsy and incomplete. In this paper, new simple formulas are derived that relate each emitter's single-peaked most-preferred world price of carbon to the world efficiency-price of carbon emissions (aka SCC). Also new here is a detailed analysis of the relationship between the WCA-voted price of carbon and the world SCC. Some implications for majority-rule voting in a WCA are discussed.

5 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 cleanest 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 with respect to carbon price negotiations, enforcement, and revenue recycling. The nation here is an elastic concept, since for the purposes of this paper it might be more appropriate to consider a regional bloc like the European Union as if it comprised a single nation.[23] In this paper it is not really necessary to assume that all of the individuals in one nation (or jurisdiction) are the same.[24] However, it is easiest to start off by conceptualizing

[23] Also, for some purposes, it might be appropriate to consider states or provinces of a large country as the relevant entity.

[24] For example, the theory would go through if the nation were subdivided into independent voting blocs with population-weighted voting power in the WCA.

that all of the people belonging to one nation are identical individuals whose tastes and technology are representative of that nation. Thus, for clarity, I am effectively assuming that a representative agent stands in for the nation. For an individual belonging to a nation everything – emissions, costs, damages – is expressed in per-capita terms for that nation. (Inversely, one could take costs and damages on the national level as given primitives and impute to each citizen the corresponding per-capita costs and damages as a function of per-capita abatement or emissions, being careful to ensure that the imputed per-capita costs and damages aggregate consistently to the given national costs and damages.[25])

The nation here is effectively an entity that enforces the imposition of an internationally-harmonized minimum price on the CO_2 it emits and recycles internally the domestic revenues raised by the tax-like price. I assume that this recycling is efficient, as if by revenue-neutral lump sum internal transfers, so there is zero net national loss (or gain) from the internally-imposed carbon price per se. (The only real cost of a carbon price is the increased cost of emitting less carbon.) Additionally, when it comes to voting or negotiating a carbon price for some particular time period, the nation effectively votes or negotiates on behalf of its citizens in accordance with their preferences. These assumptions are vulnerable, but they may make sense as an abstraction and can serve as a point of departure for further discussion.

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 wishes about desired stock levels of GHGs into wishes about corresponding GHG flows for the period under consideration. Thus, I am presuming that a static flow model can give insights here that carry over to a dynamic stock-flow formulation.[26]

Let E_i stand for the level of carbon emissions of person-agent i. The cost of attaining emissions level E_i for person i is given by the function $C_i(E_i)$, where $C_i'(E_i) < 0$ and $C_i''(E_i) > 0$. If the universal price on carbon emissions is P, then the profit-maximizing response (or reaction function, or "demand for carbon emissions") of agent-person i is $E_i(P)$, where, for each $i = 1, 2, ...m$, agent i is minimizing over E_i the expression $PE_i + C_i(E_i)$, resulting in m

[25] All of the results to be presented in this paper go through if the fundamental unit is the nation because both per-capita costs and per-capita damages of the representative agent are merely scaled up by the nation's population, leaving the fundamental analysis and conclusions unchanged. When it comes to quantifying benefits or costs, it is unclear to me whether it is easier to think primarily in terms of representative agents converted to nations or nations converted to representative agents. In any event, the two approaches are ultimately equivalent. However, as noted, the analysis and notation are made cleanest and most transparent when the fundamental unit is the person and everything is expressed in per-capita terms.

[26] This is not a trivial assumption. Intuitively, it seems to me to be OK for purposes of simplified modeling to initially consider the static case as representing a single period in a string of periods. However, I must confess that I do not know exactly how to rigorously convert a dynamic multi-period multigenerational analysis into a static reduced-form one-period analysis.

first-order conditions reflecting that everyone's marginal abatement cost equals P, or

$$-C_i'(E_i(P)) = P. \tag{1}$$

The total worldwide emissions level corresponding to (1) is

$$E(P) = \sum_{i=1}^{m} E_i(P). \tag{2}$$

Condition (1) holding for all i at the same price P guarantees worldwide cost-effectiveness, meaning that the total world emissions $E(P)$ are being produced at least total cost.

The damage of total worldwide emissions level E for agent i is given by the damages function $D_i(E)$, where $D_i'(E) > 0$ and $D_i''(E) \geq 0$.

The loss to agent i of an imposed carbon price of P is

$$L_i(P) = D_i(E(P)) + C_i(E_i(P)), \tag{3}$$

where the price-tax of P does not appear *directly* in (3) because it is assumed to be recycled in a revenue-neutral fashion.

The total world social loss of imposing a uniform carbon price P is

$$L(P) = \sum_{i=1}^{m} L_i(P) = \sum_{i=1}^{m} [D_i(E(P)) + C_i(E_i(P))]. \tag{4}$$

The world "Social Cost of Carbon" (SCC) here is understood to be the efficiency price P^* that minimizes the world social loss function (4).[27] The corresponding first-order condition $L'(P^*) = 0$ can be expressed as an analogue of the classic Samuelson public-goods condition for a situation where everyone is simultaneously a consumer and a producer of the public bad. This analogue of the Samuelson Pareto-efficiency formula appropriate to the setup here is

$$P^* = \sum_{i=1}^{m} D_i'(E(P^*)), \tag{5}$$

where (1) holds simultaneously for each i at $P = P^*$. I assume that the above analogue of the Samuelson first-order public goods condition is sufficient, as well as necessary.[28]

It should be noted that the world SCC (or, equivalently, the world efficiency price of

[27] Note that I am defining the SCC here as what more accurately should be called the "efficiency price of carbon" in the Pareto-optimal solution. Within a dynamic context, sometimes alternative definitions of the SCC are based on given non-optimal trajectories, such as "business as usual."

[28] This is readily shown when $E_i''(P) = 0$, but it holds much more generally,

carbon P^*) loses much of its welfare justification in the case of climate change because it is difficult to argue, for such a unique one-off event involving present and future generations, that the winners might actually compensate the losers by lump-sum side-payment transfers, which, hypothetically, could have the potential to ensure that the Pareto-efficient solution is actually attained and everyone is made better off. Nevertheless, such a Pareto-efficient price P^* has an almost iconic status within economics and it will be fruitful to compare it, e.g., with the majority-rule voting outcome of the WCA. Thus, I am thinking of the world SCC of P^* as a benchmark or point of departure for what follows.

6 Two General Propositions

Consider next what is the optimal level of an internationally-harmonized carbon price *from the narrow perspective of agent i*. Preferences of i for a worldwide price P are given by the loss function (3). Because revenues from the carbon price-tax are nationally collected and assumed to be efficiently recycled by the nation to which i belongs, there is presumed to be no net tax burden per se. (The only real burden to i here is the cost C_i incurred by obeying condition (1)). The worldwide emissions-price level that i would most prefer solves the problem of minimizing over P the loss function $L_i(P)$ given by expression (3). The solution P_i satisfies the first-order condition $L_i'(P_i) = 0$. I assume this solution is sufficient, as well as necessary,[29] which implies that the preference of i for P is single peaked at $P = P_i$.

The next proposition expresses i's most-preferred price P_i in a particularly useful form.

Proposition 1

$$P_i = D_i'(E(P_i)) \times \left(\frac{E'(P_i)}{E_i'(P_i)} \right). \tag{6}$$

Proof. *From (3), the first-order condition $L_i'(P_i) = 0$ translates into*

$$D_i'(E(P_i))\, E'(P_i) + C_i'(E_i(P_i))\, E_i'(P_i) = 0. \tag{7}$$

Use equation (1) to substitute P_i for $-C_i'(E_i(P_i))$ in (7) and rewrite the resulting expression as (6). ■

Equation (6) is a basic result that serves as a point of departure for the rest of the analysis in this paper. From (6), the factor $E'(P_i)/E_i'(P_i) = dE/dE_i$ is in the form of a multiplier indicating the ratio of the change in total global emissions dE/dP divided by the change in agent i's emissions dE_i/dP. For each unit positive change in its most-preferred price

[29] This is readily shown when $E_i''(P) = 0$, but it holds much more generally,

P_i, agent i is "spending" the cost consequences of abating an extra amount $-dE_i/dP$ but it is "purchasing" the benefit of worldwide extra emissions abatement $-dE/dP$. Equation (6) signifies that agent i is here reacting by applying a multiplier dE/dE_i that scales up the effect of its own narrow marginal damages $D_i'(E(P_i))$ by however many times greater is the value of the world's total marginal emissions response (to a price change) than i's own marginal emissions response (to a price change).

Equation (6) conveys an exact sense in which an internationally harmonized but nationally retained carbon price is internalizing the global warming externality for agent i. The basic underlying idea is that, at its preferred worldwide price $P_i\,(= -C_i'(E_i(P_i))$, each agent i's extra cost from a higher uniform emissions price is counterbalanced, via (6), by that same agent's lessened damage (times the multiplier dE/dE_i) from inducing all other agents to simultaneously lower their emissions in response to that higher price.

Notice from (6) what agent i is *not* doing here. Agent i is *not* equating its marginal cost of abatement $-C_i'(E_i(P_i))\,(= P_i)$ to the narrow marginal damages from one more unit of its own emissions $D_i'(E(P_i))$, which would be the analogue here to the condition for a narrowly-self-interested Nash-equilibrium solution, and which would result in a free-riding way-too-low provision of the public good.[30] Instead, as will be explained, agent i is making some kind of a partial-golden-rule-like imputation of what would be the corresponding world efficiency-price of carbon if all other agents had the same marginal damages function as i, namely $D_i'(E)$, and the same marginal cost function as i, namely $C_i'(E_i)$.

If agent i is *not* behaving like a narrowly-self-interested Nash-equilibrium free rider, as just described, then what *does* characterize i's behavior in preferring P_i to any other world price of carbon emissions? To further enrich our understanding of condition (6), let us perform the following thought experiment. Pick *any particular* person-agent i. Imagine, hypothetically, that all *other* agents $j = 1, 2, ..., m$ have the *same* marginal damages function as i, namely $D_j'(\cdot) = D_i'(\cdot)$, and the *same* marginal cost function as i, namely $C_j'(\cdot) = C_i'(\cdot)$. Let P_i^* stand for *what would be the social cost of carbon* in this hypothetical world where every agent $j = 1, 2, ..., m$ is identical (in marginal damages and marginal costs) with agent i.

The following non-trivial theorem tightly identifies the general relationship between P_i and P_i^*.

[30] Note the following interpretation: if agent i erroneously treats the emissions E_j of all *other* agents $j \neq i$ as exogenously fixed, then $E_j^{'} = 0$ for $j \neq i$ and the multiplier $E'/\ E_i^{'}$ on the right hand side of (6) is equal to 1, thereby yielding the conventional free-riding Nash result that $-C_i' = D_i'$. I owe this interpretation to Thomas Aronsson.

Proposition 2 ***(The "Partial Golden Rule" Theorem)***

$$P_i = P_i^*. \tag{8}$$

Proof. *Because every agent is identical with i, equation (2) implies*

$$E'(P_i) = m\, E_i'(P_i), \tag{9}$$

and (6) becomes

$$P_i = D_i'(E(P_i)) \times m. \tag{10}$$

But condition (10) is exactly condition (5) for the special case where all agents $j \neq i$ have the same marginal damages function as i, namely $D_i'(\cdot)$. (Condition (1) is automatically satisfied for $P = P_i$ because each identical agent $j \neq i$ has the same marginal cost function as i, namely $C_i'(\cdot)$). Since P_i satisfies the same necessary and sufficient optimality conditions as P_i^*, the two are equal. ■

Proposition 2 may look almost like a tautology, but it is not at all tautological. P_i^* is describing what the efficiency-price of carbon (aka SCC) *would be* in a hypothetical world where everyone has the *same* marginal damage function as i, namely $D_i'(\cdot)$, and the *same* marginal cost function as i, namely $C_i'(\cdot)$. Proposition 2 states that the favorite world price of agent i, namely P_i, is *as if* i is behaving like a partially-benevolent dictator by imposing its own damages and costs on the rest of the world and calculating what the world efficiency-price (aka SCC) *would then be*. Effectively, equation (10) means that the most-preferred world price P_i of partially-benevolent dictator i is its own marginal damage $D_i'(E(P_i))$ scaled up by m into a kind of partial-golden-rule-like imputation of what would then be the corresponding world efficiency-price of carbon if everyone had marginal damage $D_i'(E(P_i))$. The overarching idea of Proposition 2 is that the WCA is putting people in a position where their interest is to vote not just for their own narrow self interest, but for a version of their self interest that internalizes a substantial part of the externality (by imagining a world of people just like themselves).

It follows almost immediately from Proposition 2 that, in the strictly hypothetical case where all m agents are exactly identical, then $P_i = P^*$ for all i. In this identical-agent situation, therefore, a majority-voting WCA rule automatically resolves the social coordination problem (since every agent i would vote for $P_i = P^*$), resulting in a first-best Pareto optimum.

For the more general case where agents differ, majority-rule voting does not necessarily yield the first-best Pareto optimum. In the next section I will make two basic linearity

assumptions that seem to me to be relatively innocuous. These linearity assumptions will allow for a closed-form expression of (6) that is relatively easy to interpret in terms of the linear parameters. This closed-form expression will clarify the relationship of the world price that i would most prefer, P_i, to the world SCC= P^*, and will further clarify the relationship between the WCA-voted price and the world SCC=P^*.

7 Two Linearity Assumptions

To make further progress on understanding most clearly the expression (6), we need to put some more structure on the problem. I now make two simplifying linearity assumptions that will allow a closed-form expression in place of (6). These two linearity assumptions might be accepted at face value, or treated as Taylor-theorem approximations that hold increasingly accurately in the neighborhood of small changes. In any event, I believe the two linearity assumptions represent only a little sacrifice of generality relative to the clarification and understanding they bring to the more general expression (6).

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

$$D_i(E) = \alpha_i + d_i\, E, \tag{11}$$

where $d_i > 0$. Equation (11) means that the marginal damage for each agent i is a constant

$$D_i'(E) = d_i \tag{12}$$

for some positive coefficient d_i that is allowed to differ for different i.[31] I feel that (12) is reasonably accurate for small time periods because emissions are a flow, whereas damages are a function of the accumulated stock of GHGs, and for carbon dioxide the flow-stock ratio is small over a five or ten year period.

An immediate consequence of (12) combined with (5) is that

$$P^* = \sum_{i=1}^{m} d_i, \tag{13}$$

independent of the cost functions $\{C_i(E_i)\}$. This decomposition property (that the global

[31] If there were a constant net tax-offset revenue-recycling enhancement (or diminution) per unit of tax for i, it could be incorporated into the marginal damages coefficient d_i and all of the theory would go through. The "double dividend" effectively makes d_i higher, whereas the "negative dividend" effectively makes d_i lower. Otherwise, the theory goes through.

efficiency-price of carbon is independent of emissions costs) greatly simplifies the analysis without, I hope, losing too much realism.

Another immediate consequence of (12), this time from combining it with (6), is that

$$P_i = d_i \times \left(\frac{E'(P_i)}{E_i'(P_i)} \right), \tag{14}$$

for all cost functions $\{C_i(E_i)\}$.

The reaction function $E_i(P)$ relating emissions of agent i to an imposed carbon price of P is given implicitly by condition (1). The second simplifying linearity assumption is that, throughout the period for which the analysis is intended to apply, this reaction (or "demand for emissions") function $E_i(P)$ is of the linear reduced form

$$E_i(P) = \beta_i - s_i P, \tag{15}$$

where β_i and the reaction coefficient s_i are both positive and the relevant range[32] of P here is $0 \leq P \leq \beta_i/s_i$.

The linearity assumption (15) is essentially ad hoc, but it might be defended as a simplifying approximation that gives some useful insights. Note from (15) that

$$E_i'(P) = -s_i, \tag{16}$$

so that the reaction of E_i to a unit change in price is conveyed by $-s_i$, which, while constant for each given i, is allowed to differ for different i. The reaction coefficient s_i is a measure of the (price) *sensitivity* of emissions $\Delta E_i(\Delta P)$ to a change in price ΔP.

Use (1) to substitute the marginal cost of abatement $-C_i'(E_i)$ for P in the linear reaction formula (15). After rearrangement, this yields a marginal cost of abatement function $-C_i'(E_i)$ that is linear in emissions level E_i, of the form

$$-C_i'(E_i) = \frac{\beta_i - E_i}{s_i} \tag{17}$$

throughout the relevant range $0 \leq E_i \leq \beta_i$. Comparing (15) with (17) shows that the coefficient s_i does double duty. In (15), s_i is interpretable as a sensitivity-reaction coefficient for price changes. In (17), other things being equal, $1/s_i$ is interpretable as a measure of the change in the marginal cost of emissions abatement per unit change of emissions abatement.

A third, and here the most important, interpretation of the price-sensitivity coefficient s_i

[32] For simplicity I rule out corner solutions. I believe that the framework here could be extended to cover corner solutions, but the paper is complicated enough (and long enough) without this extension.

is that it represents the distortionary deadweight loss to i from a positive carbon price-tax change. If the price-tax change is $\Delta P > 0$, then, from (15), the induced emissions change is $\Delta E_i = -s_i \, \Delta P$. The associated deadweight loss (DWL) here is $-(\Delta E)(\Delta P)/2$, or

$$DWL_i(\Delta P) = s_i \times \left(\frac{(\Delta P)^2}{2} \right). \tag{18}$$

Equation (18) is the appropriate version here of the famous Ramsey-type principle that raising taxes on a relatively elastically-demanded good is more distortionary and more damaging than raising taxes on a relatively inelastically-demanded good.

The worldwide *average* sensitivity-reaction coefficient to a change in price is

$$\overline{s} \equiv \frac{\sum_{i=1}^{m} s_i}{m}. \tag{19}$$

The worldwide *average* marginal damage coefficient is

$$\overline{d} \equiv \frac{\sum_{i=1}^{m} d_i}{m}. \tag{20}$$

This concludes the description of the linearity assumptions, which will be used to simplify greatly the analysis of (6) by expressing P_i in terms of a readily-interpretable closed-form equation.

8 The "Most Preferred" Price of i Under Linearity

A new main result of this paper is the following proposition.

Proposition 3 *Under the linearity assumptions (11) and (15), the relationship between the world carbon-price level P_i that i would most prefer and the world SCC of P^* is given by the expression*

$$P_i = P^* \times \left[\left(\frac{d_i}{\overline{d}} \right) \div \left(\frac{s_i}{\overline{s}} \right) \right]. \tag{21}$$

Proof. From (2), it follows for all P (including $P = P_i$) that

$$E'(P) = \sum_{i=1}^{m} E_i'(P). \tag{22}$$

Substitute (16) and (19) into the right hand side of (22), which yields, for $P = P_i$, the equation

$$E'(P_i) = -m \, \overline{s}. \tag{23}$$

Combining (20) with (13) implies

$$\overline{d} = \frac{P^*}{m}. \tag{24}$$

Multiply numerator and denominator in the right hand side of (14) by $\overline{d}$. For $\overline{d}$ in the numerator, substitute the right hand side of (24). Then use (16) (with $P = P_i$) to replace $E_i'(P_i)$ and (23) to replace $E'(P_i)$ in the right hand side of the resulting expression. This yields a main new result of this paper, expressing P_i in terms of the closed-form equation (21). ■

Equation (21) is a relatively crisp and simple expression. Basically, the world carbon-price level P_i that i would most prefer is the world efficiency-price of carbon P^* scaled up or down by the multiplication factor $d_i/\overline{d}$ (representing i's proportional deviation of its own marginal damages from average marginal damages) and the division factor $s_i/\overline{s}$ (representing i's proportional deviation of its price sensitivity from average price sensitivity). We analyze the comparative-statics roles of these two multiplication/division factors in turn.

As a point of departure for analyzing expression (21), note that if there is an "average" agent j with $d_j = \overline{d}$ and $s_j = \overline{s}$, then, for this "average" agent, $P_j = P^*$.

The multiplication factor $d_i/\overline{d}$ in equation (21) means that, other things being equal, the ratio P_i/P^* is proportional to the ratio $d_i/\overline{d}$. Here P_i is "adjusted" from being equal to P^* by the multiplication scaling factor $d_i/\overline{d}$, which makes sense. Thus, P_i differs multiplicatively from P^* in (21) by the extent to which d_i differs multiplicatively from $\overline{d}$. Other things being equal, higher (lower) values of i's marginal damage d_i – relative to the worldwide average value of marginal damages $\overline{d}$ – cause agent i to want relatively higher (lower) values of its most-preferred world price P_i.

For the division factor $s_i/\overline{s}$, equation (21) means that, other things being equal, the ratio P_i/P^* is inversely proportional to the ratio $s_i/\overline{s}$. Here P_i is "adjusted" from being equal to P^* by the division scaling factor $s_i/\overline{s}$. Thus, P_i differs from P^* division-wise in (21) by the extent to which s_i differs from $\overline{s}$. Other things being equal, higher (lower) values of of i's sensitivity coefficient s_i – relative to the worldwide average value of the sensitivity coefficient $\overline{s}$ – cause agent i to want relatively lower (higher) values of its most-preferred world price P_i. An interpretation of this result is along the following lines. Recall from (18) that the price-sensitivity coefficient s_i represents the distortionary deadweight loss to i from a positive carbon price-tax change. When the elasticity-like price-sensitivity coefficient s_i is low relative to $\overline{s}$, the self-imposed carbon tax is less distortionary with lower deadweight loss and, other things being equal, agent i therefore favors a relatively higher value of P_i. Conversely, when the elasticity-like reaction coefficient s_i is relatively high, the self-imposed

carbon tax is more distortionary with higher deadweight loss and, other things being equal, agent i therefore favors a relatively lower value of P_i. This relationship is the appropriate reflection here of the well-known Ramsey-type principle that raising taxes on a relatively elastically-demanded good is more distortionary and more damaging than raising taxes on a relatively inelastically-demanded good.

This concludes the discussion of the meaning of the basic formula (21). Hopefully the reader now has some sense of what determines the preferred worldwide price P_i of i and its relationship to the world SCC of P^*. The next section will discuss some aspects of the application of (21) to the outcome of majority-rule voting in a World Climate Assembly.

9 Relating the WCA-voted Price to the SCC

Note: the next two paragraphs apply to the general case of *any* $\{C_i(E_i)\}$, $\{D_i(E)\}$.

The World Climate Assembly (WCA) votes on pairwise alternatives for the desired level of a uniform carbon price, based on the principle of one person, one vote. By the median voter theorem, the outcome of WCA voting is the median value of $\{P_i\}$, denoted $\widetilde{P}$, where P_i is given by (6).[33] It is not actually necessary to do multiple pairwise voting on multiple binary price alternatives. A shortcut is available. If each agent i submits only its most-preferred price P_i, then $\widetilde{P}$ can be centrally calculated as the median of the centrally collected values of $\{P_i\}$. This value of $\widetilde{P}$ will defeat by majority vote any other proposed price P.

The median voter result signifies that *half of the world's population wants a uniform price of carbon greater than the WCA-majority-voted* $\widetilde{P}$*, whereas the other half of the world's population wants a uniform price of carbon less than the WCA-majority-voted* $\widetilde{P}$. Intuitively or heuristically, this might be considered as a not-bad stand-alone outcome in and of itself for a world where we are unsure in the first place what is the best actual welfare measure.[34] If majority-rule voting in a democratic national legislature is adjudged to be a "good enough" compromise as a way of deciding the level of spending on national public goods, then perhaps it can also be adjudged a "good enough" compromise mechanism for deciding the level of an international public good like GHG abatement. To obtain a tighter relationship between $\widetilde{P}$ and P^* we turn to the linear case.

[33] It was already noted that preferences of i for price P are single peaked with peak value $P = P_i$.

[34] Recall that the world efficiency-price of carbon P^* loses much of its welfare justification anyway in the case of climate change because it is difficult to argue (for such a unique one-off intergenerational event) that the winners will actually compensate the losers by lump-sum transfers, which would ensure that the Pareto-efficient solution P^* is actually attained. One could attempt to argue on a heuristic basis that, in such a situation, the median voting-equilibrium price $\widetilde{P}$ is approximately as good a welfare measure as the SCC of P^* – just because of the attractive symmetry that half of the world wants a higher price and the other half wants a lower price.

In what comes next for the linear case, the following notation will be used. If $Z = \{Z_i\}$ is a collection of m values of Z_i for $i = 1.2, ..., m$, then $\widetilde{Z}$ will stand for the *median* value of $\{Z_i\}$, while $\overline{Z}$ will stand for the *mean* value of $\{Z_i\}$.

The following proposition helps elucidate the relationship between $\widetilde{P}$ and P^* in the linear case.

Proposition 4 *The WCA-voted majority-rule price $\widetilde{P}$ can be expressed as*

$$\widetilde{P} = P^* \times \left[\widetilde{\left(\frac{d}{s} \right)} \div \left(\frac{\overline{d}}{\overline{s}} \right) \right]. \tag{25}$$

Proof. *From (21) and the median voter theorem, the outcome of WCA majority rule $\widetilde{P}$ can be expressed as*

$$\widetilde{P} = P^* \times \left(\frac{\overline{s}}{\overline{d}} \right) \times \widetilde{\left(\frac{d}{s} \right)}. \tag{26}$$

Then (25) is just a rewriting of (26). ■

In general, the value of the WCA-voted outcome $\widetilde{P}$ given by (25) depends on the distribution of the coefficients $\{d_i\}$ and $\{s_i\}$ and how they interact. In principle, almost anything could emerge. Nevertheless, I think there is some "hint" from (25) that $\widetilde{P}$ might be tolerably close to P^*. After all, $\widetilde{(d/s)}$ is *some* (imperfect) measure of the central tendency of $\{d_i/s_i\}$, while $\overline{d}/\overline{s}$ is also *some* (imperfect) measure of the central tendency of $\{d_i/s_i\}$. With a bit of wishful thinking, these two (imperfect) measures of central tendency might almost equal each other, in which case they almost cancel each other in (25) and $\widetilde{P}$ is unlikely to be too-too sharply different from P^*.[35]

I now look at some special cases of the distribution of $\{d_i\}$ and $\{s_i\}$ that will give more precise outcomes.

Consider a situation where there is more variability in $\{d_i\}$ relative to the variability of $\{s_i\}$. As an extreme case, suppose that all price-sensitivity response coefficients $\{s_i\}$ are the same, so that $s_i = \overline{s}$ for all i. In this case (26) becomes

$$\widetilde{P} = P^* \times \left(\frac{\widetilde{d}}{\overline{d}} \right). \tag{27}$$

[35] I have found it very difficult to say much more analytically about the expression (25) because it is difficult to decompose it further in the general case. As noted, there is some "hint" from (25) that $\widetilde{P}$ might be tolerably close to P^* because the two (imperfect) measures of central tendency might partially cancel each other. To make further progress on understanding condition (25), I turn to special cases and a very crude numerical example.

From (27), the majority-rule carbon price $\widetilde{P}$ is close to the world SCC (aka world efficiency price of carbon) P^* when the median marginal damage $\widetilde{d}$ is close to the mean marginal damage $\overline{d}$. 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 damage of emissions per capita should be greater or less than the median marginal damage of emissions per capita. If the two are equal in (27), then majority-rule voting for $\widetilde{P}$ obtains the world SCC (or efficiency price of carbon) P^*. If the two are unequal, the analysis provides a measure of how far away from the Pareto-optimal efficiency price is majority rule. Of course this is just a model with quite restrictive assumptions, but in a 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 holds out some prospect of coming "close enough" to the world SCC (aka world efficiency price of carbon).

Back to the case where $\{s_i\}$ are not all the same, one may still argue that $\widetilde{P}$ is close to P^* when the variability in $\{d_i/s_i\}$ is small across agents.[36] To formalize an extreme version of this argument, suppose that $d_i/s_i = k$ for some constant k and all i. Then it is straightforward to show that $\overline{d}/\overline{s} = k$ and $\widetilde{(d/s)} = k$ and, from (25), it follows that $\widetilde{P} = P^*$ in this extreme case.

A statistical generalization of the above extreme-case result is available. Suppose the ratios $\{d_i/s_i\}$ are independently identically distributed (iid) random variables. Strictly speaking, I am abusing terminology here because the result of the following proposition depends on the law of large numbers and would only hold in the limit of a large sample. Define

$$k_i = \frac{d_i}{s_i}, \tag{28}$$

meaning that k_i is the ratio of d_i over s_i, which will be treated as an iid random variable.

Proposition 5 *Suppose the following data generating process (dgp). The random variables k_i and s_i are each (separately) iid and*

$$d_i = s_i \times k_i. \tag{29}$$

Then (in the limit of a large sample size)

[36] There is some anecdotal evidence that in certain circumstances s_i and d_i are positively correlated. Other things being equal, those nations that are most sensitive to the damages from climate change tend also to be poorer and have an arguably higher price sensitivity of emissions to a change in energy price. This argument was suggested by a referee.

$$\widetilde{P} = P^* \times \left(\frac{\widetilde{k}}{\overline{k}} \right). \tag{30}$$

Proof. *From the independence of the random variables k_i and s_i, and from (29),*

$$\overline{d} = \overline{s} \times \overline{k},$$

which can be rewritten as

$$\frac{\overline{d}}{\overline{s}} = \overline{k}. \tag{31}$$

From (28),

$$\widetilde{\left(\frac{d}{s} \right)} = \widetilde{k}. \tag{32}$$

Plugging (31) and (32) into (25) yields the desired conclusion (30). ■

The result (30) indicates that, with the postulated data generating process, $\widetilde{P}/P^*$ is the ratio of the median of k ($=\widetilde{k}$) divided by the mean of k ($=\overline{k}$). Once again here, outcomes of $\widetilde{P}/P^*$ are whittled down to a possible proportional difference between two measures of central tendency. I see no reason offhand to believe here that the mean $\overline{k}$ of d/s is substantially different from the median $\widetilde{k}$ of s/d.

This is about as far as pure theory can take us. I think it is fair to say that the formal WCA voting model might be "hinting" that there may be some tendency for the majority-voted price $\widetilde{P}$ to be "close" to the world efficiency price of carbon P^* (=SCC) – or at least that the WCA-voted price $\widetilde{P}$ and the world SCC of P^* are unlikely to differ sharply.

10 A Crude Numerical Exercise

The empirical evidence on the world SCC (or P^*) comes almost exclusively from so-called Integrated Assessment Models (IAMs). The U.S. Interagency Working Group estimated a value of the SCC to be used in U.S. regulatory impact assessments as \$40 per ton of CO_2 (in 2014 dollars).[37] The number \$40/tCO_2 is the mean-value outcome of highly-variable results from three different IAMs with various parameter settings, evaluated at a 3 percent annual discount rate.

[37] See Interagency Working Group on Social Cost of Carbon, United States Government Technical Update of the Social Cost of Carbon for Regulatory Impact Analysis Under Executive Order 12866, revised July 2015.

It is difficult to deny that there is a very high degree of fuzziness in this estimate of SCC ($=P^*$). At one extreme, Pindyck (2017) argues that IAMs have flaws that render them close to useless as tools for policy analysis. Even without such an extreme position, estimates of P^* (equivalently SCC) from various other studies can easily range from less than \$10/tCO$_2$ to over \$100/tCO$_2$.[38] Defenders of the \$40/tCO$_2$ estimate of SCC ($=P^*$) typically do not deny the extreme uncertainty of this number, but defend it on the basis that the reality of the political process requires *some* defensible number, however fuzzy, over *no* number.[39]

From (13), $P^* = \sum d_i$. Estimated values of $\{d_i\}$ represent another exercise with fuzzy numbers. Nordhaus (2015) makes a constructive effort by first dividing the world into 15 regions, including the largest countries and aggregates of the other countries. He then effectively estimates the 15 values of $\{d_i\}$ corresponding to his 15 regional divisions, while admitting that evidence is sparse to nonexistent outside of high-income regions. He demonstrates substantial differences in the 15 values of $\{d_i\}$ from the 3 different IAMs used in the U.S. Interagency Working Group "average" estimate of P^*. Because the national estimates are so poorly determined, for many of his central national estimates Nordhaus effectively assumes that national values of damages are proportional to national GDPs.

The marginal abatement cost functions that appear in (17) are of the reduced linear form $-C_i'(E_i) = (\beta_i - E_i)/s_i$. Nordhaus (2015) effectively estimates the relevant parameters in (17) by combining a global estimate from his DICE-2013 model with detailed regional estimates from an engineering model by the consulting company McKinsey (2009). My impression here is that these numbers are also a bit fuzzy because the McKinsey regional estimates are based on a built-up micro-engineering approach that may not be terribly reliable.

Using the Nordhaus (2015) numbers (scaled up by a factor of two), in an imaginative exercise Kotchen (2016) attempts to compare (in the notation of this paper) P^* with $\widetilde{P}$. His P^* is selected as \$40/tCO$_2$. The $\widetilde{P}$ of this paper corresponds to his "population weighted majority voting rule," which he calculates as $\widetilde{P}$ = \$51/tCO$_2$.[40] Readers are free to make their own interpretation of this estimated difference between P^* and $\widetilde{P}$. Considering the very large degree of uncertainty in the underlying numbers, I would interpret this result as indicating that the WCA majority-voted value $\widetilde{P}$=\$51/tCO$_2$ is essentially indistinguishable from the world SCC estimate of P^*=\$40/tCO$_2$ – in the sense that the difference between the two numbers is considerably smaller than the scope of measurement error. Thus, I think there is license here to pretend from this very rough numerical exercise that $\widetilde{P} \approx P^*$.

[38] This range is cited in Nordhaus (2015).

[39] See, for example, Metcalf and Stock (2017).

[40] See Kotchen (2016) for more details and results of some other interesting numerical exercises. What I am calling here (and in previous work Weitzman (2014)) i's "most preferred price of carbon" – namely P_i – Kotchen (2016) labels as i's "strategic social cost of carbon."

11 Concluding Remarks

At the end of the day, there is no airtight logic in favor of a negotiated worldwide carbon-emissions price over negotiated worldwide carbon-emissions quantities, only a series of partial arguments. A desirable feature, I have argued, is the natural focal salience and the relatively low transaction costs of negotiating one price as against negotiating multiple quantities, which, while somewhat imprecise, in my opinion constitutes an important behavioral-psychological distinction. As was pointed out, negotiating a one-dimensional uniform price with single-peaked preferences has the significant additional property of allowing a majority-rule voting equilibrium, thereby avoiding the Arrow impossibility theorem (which casts a negative shadow on the ability of a social decision rule to resolve differences involving multiple dimensions).

A key argument in favor of a price over quantities is the self-enforcement mechanism that constitutes a main theme of this paper, namely the built-in "countervailing force" of a uniform price of carbon against narrow self-interested free-riding. *There is simply no politically-acceptable one-dimensional emissions-quantity analogue to* $\widetilde{P}$ *that has this important "countervailing force" property!*

In past papers (Weitzman (2014, 2016)) I discussed in torturous detail negotiating one worldwide aggregate emissions target or aggregate cap *contingent upon* a previous-round linear subdivision formula with $2n$ linear coefficients, set, for example, by a preceding agreement among the n countries on various target reductions from various baselines. (Think, e.g., of negotiated percentage reductions of emissions from negotiated base levels, where the parties vote on the aggregate emissions level and then disaggregate it according to the previously-agreed-upon linear formula.) A system based on voting for aggregate emissions (*given* linear subdivision formulas) could, in principle, embody some countervailing force against the global warming externality. But, I concluded that negotiating the extra layer of $2n$-dimensional first-round linear subdivision coefficients would very likely founder politically when applied on a worldwide scale for reasons similar to the multiple-quantity arguments already made in this paper.

I further concluded on the quantity side that, even with "seemingly symmetric" formulas for the initial cap allotments, a quantity-based system seems far more complicated, baroque, and objectionable in the international context than an internationally-harmonized carbon price. At one extreme, equal percentage reductions of emissions from a status-quo existing level in some particular base year would be utterly unacceptable to the emerging-economy countries like China or India, who would be placed at an unjust historical disadvantage. At the other extreme, equal per-capita initial assignment of caps would be utterly unac-

ceptable to developed countries like the U.S. or Japan, who would be forced to transfer huge amounts of taxpayer-financed money to purchase allowances from abroad. In-between quantity allotments of caps would likewise stumble on intractable issues of "fairness."

By contrast with multiple quantities, the countervailing force of a single price automatically incentivizes all negotiating parties to internalize the externality via a simple understandable formula that embodies a common climate commitment based on principles of reciprocity, quid-pro-quo, and I-will-if-you-will. This tendency towards aligning self interest with social interest by internalizing the externality gives national negotiators an incentive to offset their natural impulse to otherwise bargain for a low price. The model of a WCA in this paper tried to formalize this aspect. I think the paper is suggesting that the majority-rule-voted WCA price on carbon emissions of $\widetilde{P}$ might come tolerably close to the world SCC of P^*. Several special cases supported this tentative conclusion, as did a very rough numerical example.

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 one-dimensional countervailing-force argument that deserves to be highlighted. The purpose of this paper is primarily expository and exploratory. *Any* serious 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 suggest that negotiating a uniform minimum price on carbon emissions 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-length treatment, 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.[41] 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.

The international political realities of implementing a single worldwide price of carbon emissions are truly daunting. Even within a country, the domestic politics of imposing a uniform carbon tax are difficult. The U.S. for example has been dithering with climate policy for decades. But there may indeed be comprehensive strategies that can resonate across the political spectrum. A serious and much-noted recent proposal by a group of prominent credentialed conservatives argues strongly for a uniform U.S. carbon tax coupled

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

with equal per-capita lump-sum rebates. (See Climate Leadership Council (2017), "The Conservative Case for Carbon Dividends.") Maybe this proposal can form the nucleus of a comprehensive approach to climate change on which liberals and conservatives might agree and – who knows – some of it may even be generalizable to the international arena.

Because the internationalist formulation of this paper 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. An international system of equal national carbon taxes with revenues kept in the taxing country is a relatively simple and transparent way to achieve internationally-harmonized carbon prices. But it is not absolutely necessary for the conclusions of this paper. In principle, nations or regions could meet the obligation of a uniform minimum price on carbon emissions by whatever internal mechanism they choose – a tax, a cap-and-trade system with a price floor, some other hybrid system, or whatever else results in an observable price of carbon not below the uniform minimum.[42]

Of course any nation or region could choose to impose a carbon tax or price above the international minimum, for reasons of public health, traffic congestion, or something else.[43] The hope is that even a low positive initial value of a universal minimum carbon tax or price could be useful for gaining confidence and building trust in this price-based international architecture. Note that the act of confirming that each member of the WCA is charging the binding agreed-upon minimum price on carbon emissions would presumably require that nations open their internal tax books to careful external auditing.[44]

It might be argued that the real problem with a WCA is getting parties to agree to participate in it in the first place. If n transfer side-payments are required to get n countries to agree in the first place to vote on a uniform carbon price, then it might be argued that this constitutes an n-dimensional negotiating problem akin in complexity to assigning n allowance

[42] A worldwide uniform minimum carbon price could theoretically be attained in a worldwide cap-and-trade system by setting it as a floor, which could be enforced by making it a reserve price of permits actualized by a hypothetical international agency that buys up excess permits whenever the price falls below the floor. Alas, such a mechanism invites its own complicated free-rider problem, because each nation has an incentive not to spend its own money, but for *other* nations to spend *their* money to buy up excess permits. Alternatively, a hypothetical worldwide consignment auction for carbon permits with a uniform reserve price might work in theory but seems highly impractical in practice. Again here, there is a marked distinction between the simplicity of a one-dimensional price-tax and the complexity of negotiating a multidimensional quantity-based binding agreement among many different nations.

[43] Parry (2016) argues that the national efficiency price of carbon emissions, even without accounting for climate change, is substantial for many countries.

[44] The price-tax would be levied within the country that actually burns the carbon and releases the CO_2 emissions into the atmosphere. For convenience, this carbon price-tax should probably be levied as far upstream as possible within the country that actually combusts the carbon and produces the CO_2 emissions.

caps (with n transfer payments) in a worldwide cap-and-trade system. Considering the many behavioral-psychological and classical-economic arguments favoring a single carbon price that have been made throughout this paper, I simply find it difficult to accept the line of reasoning that transfer payments are likely to be as complicated to negotiate under a uniform price-tax system as transfer payments (along with cap assignments) under a cap-and-trade system.

An additional important consideration is that, unlike a nationally-retained carbon tax, assigned allowances in an international cap-and-trade system create property rights worth several hundreds of billions of dollars.[45] Thus, far more real money is involved with bargaining over initial allowances in a (cross border) cap-and-trade system than is involved with bargaining over a uniform (nationally retained) tax.[46] Other things being equal, this should make a uniform price-tax easier to negotiate (because there is less at stake) than negotiating the initial allowances in a cap-and-trade system (because there is more at stake).

My tentative (if non-airtight) conclusion: it is difficult to get nations to agree to *anything* seriously blocking free-riding on climate change, but negotiating one universal price is *relatively* easier than negotiating n quantities. Note also that the WCA proposal has a built-in voting mechanism for dealing automatically with changes, whereas n international emissions-allowance caps (which need to be further disaggregated into a great many within-country emissions allowances) would have to be renegotiated every time there is a significant change in perceived damages or costs.

A truly critical issue is that a binding international agreement on a WCA-voted uniform minimum carbon tax or price requires some serious mechanism to induce participation and compliance. Perhaps greenfund transfer payments might help (and there is no contradiction in having a WCA with such side payments), but these, like mandatory quantity-allowance targets, would likely involve tricky multidimensional multinational negotiations where it is difficult to avoid self-interested free-riding. For enforcement, to make sure the uniform WCA price-tax is actually imposed, perhaps there is no practical alternative to using the international trading system for applying tariff-based penalties on imports from non-complying nations. 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

[45] The EDGAR database constructed by the European Commission estimates total world CO_2 emissions in 2015 to be 36 billion tons. At a world SCC of \$40/tCO_2, this amounts to a total inventory value of 1.4 trillion dollars. A full accounting is much more complicated, but this crude calculation may give some rough idea of the magnitude of wealth creation involved.

[46] Because the imposed "carbon tax" is internally retained within each nation, then, in the linear case, increased costs of compliance for positive price changes are deadweight-loss second-order Harberger triangles of the relatively modest form $-(\Delta P \times \Delta E)/2$. The corresponding international transfers in a cap-and-trade system (which can be either positive or negative, depending, among other things, on initial cap allowances) are first-order immodest rectangles of the form $-(P \times \Delta E)$. Goulder and Schein (2013), among others, discuss the potential for very large revenue flows from the nations purchasing allowances to the nations selling them.

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.[47]

Remember, the top-down WCA approach of this paper is predicated in the first place on a future situation where the climate change problem has become sufficiently threatening on a worldwide grassroots level that world public opinion is ready to condone truly novel world governance structures. The ultimate justification of the WCA approach is that big new problems, like the grave threat of catastrophic climate change, may require big new solutions. Desperate times demand desperate measures. Tampering with free trade via tariff-based penalties on countries that refuse to participate in the WCA should be seen as a unique exception to the basic overarching principle of free trade, which exception is predicated on a widespread perception that climate change is edging towards bringing disastrous effects.

In a far-sighted paper, William Nordhaus (2015) advocates a uniform border tariff on imports from non-member countries imposed by a voluntary "Climate Club" coalition of willing member nations. Members of the Climate Club agree to impose on themselves a harmonized carbon price, along with free trade amongst themselves, accompanied by a sufficiently stiff ad valorem tariff on imports from outsiders. The Climate Club is thus a kind of customs union because this trade bloc is composed of a free trade area with a common external tariff on the rest of the world. Nordhaus argues empirically that a price of \$25 per ton of CO_2 along with an ad valorem border tariff of 5% achieves high participation rates where an overwhelming majority of emitting nations, acting in their own self-interest, will wish to join the Climate Club.[48]

The WCA proposal of this paper can fit well with the Nordhaus Climate-Club idea. A perhaps loose end in the Nordhaus approach concerns what should be the negotiated universal club price of carbon emissions. As has already been pointed out, 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, there are bound to be disagreements, whose resolution is unclear, about what that common price should be. A WCA addresses this issue concretely and allows for a flexible price by majority rule as conditions change and circumstances warrant.

I close by recapitulating here the basic premise of this paper: a uniform global tax-like price on carbon emissions, whose revenues each country retains, can provide a focal point for

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

[48] My intuition is that an ad valorem border tariff of 10% would be more of a fail-safe guarantor of high participation rates (and 10% constitutes a nicer round number than 5%), but I have no hard evidence to back up this assertion. A Climate Club with less than total world participation of major CO_2 emitters will under-supply abatement, but if Nordhaus (2015) is correct, the actual participation is likely to be close to complete.

a reciprocal common commitment, whereas quantity targets, which do not nearly so readily present such a single focal point, have a tendency to rely ultimately on free-style individual quantity commitments. After the perceived failure of a Kyoto-style top-down quantity-based approach, the world has seemingly given up on a comprehensive global design, focusing instead in the 2015 Paris COP21 Agreement on essentially voluntary bottom-up quantity-based "Nationally Determined Contributions." Perhaps, as this paper has emphasized, a quantity-based emphasis on negotiating multiple emissions caps embodies a bad design flaw. The arguments of this paper suggest that a uniform-price-based international negotiating or voting mechanism might thwart the free-rider climate change problem by empowering an "I will if you will" approach.

References

[1] Aldy, Joseph E., and W. Kip Viscusi (2014). "Environmental Risk and Uncertainty" in Machina and Viscusi (eds) *Handbook of the Economics of Risk and Uncertainty*; Oxford: North Holland, 601-649.

[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] Borenstein, Severin (2016). "Fixing a major flaw in cap-and-trade." https://energyathaas.wordpress.com/2016/08/15/fixing-a-major-flaw-in-cap-and-trade/

[5] Climate Leadership Council (2017). "The Conservative Case for Carbon Dividends." https://www.clcouncil.org/wp-content/uploads/2017/02/TheConservativeCaseforCarbonDividends.pdf

[6] Coase, Ronald (1937). "The Nature of the Firm." *Economica*, 4 (16), 386-405.

[7] Coase, Ronald (1960). "The problem of social cost." *The Journal of Law and Economics*, 3(1), 1-44.

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

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

[10] Cramton, Peter, Axel Ockenfels, and Steven Stoft (2015). "An International Carbon-Price Commitment Promotes Cooperation." *Economics of Energy & Environmental Policy*, 4(2): 37-50.

[11] Dixit, Avinash, and Mancur Olson (2000). "Does voluntary participation undermine the Coase Theorem?" *Journal of Public Economics*, 76, 309-335.

[12] Gollier, Christian, and Jean Tirole (2017). "Negotiating Effective Institutions Against Climate Change." Chapter 7 in Cramton, MacKay, Ockenfels, and Stoft (eds), *Global Climate Pricing*. Cambridge: MIT Press, forthcoming 2017.

[13] Goulder, Lawrence H. (2002). *Environmental Policy Making in Economies with Prior Tax Distortions*. Amherst MA: Edward Elgar.

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

[15] Harrison, Kathryn (2013). "The Political Economy of British Columbia's Carbon Tax." OECD *Environment Working Papers*, No. 63, OECD Publishing.

[16] Interagency Working Group on Social Cost of Carbon (2015). United States Government Technical Update of the Social Cost of Carbon for Regulatory Impact Analysis Under Executive Order 12866, revised July 2015.

[17] Jorgenson, D. W., R. J. Goettle, M. S. Ho, and P. J . Wilcoxen (2013). *Double Dividend: Environmental Taxes and Fiscal Reform in the United States*. Cambridge MA: MIT Press.

[18] Keohane, Robert O., and Michael Oppenheimer (2016): "Paris: Beyond the Climate Dead End through Pledge and Review?"; forthcoming in *Politics and Governance* (http://belfercenter.ksg.harvard.edu/files/dp85_keohane-oppenheimer.pdf)

[19] Kotchen, Matthew J. (2016). "Which Social Cost of Carbon? A Theoretical Perspective." NBER Working Paper No. 22246.

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

[21] MacKay, D. P., P. Cramton, A. Ockenfels and S. Stoft (2015). "Price carbon – I will if you will." *Nature* 526, 315-316.

[22] Mas-Collell, Andreu, Michael Whinston, and Jerry Green (1995). *Microeconomic Theory.* New York: Oxford University Press.

[23] McKinsey Company (2009). *Pathways to a Low-Carbon Economy: Version 2 of the Global Greenhouse-Gas Abatement Cost Curve.* Available at www.mckinsey.com.

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

[25] Metcalf, Gilbert E., and James Stock (2017). "Integrated Assessment Models and the Social Cost of Carbon: A Review and Assessment of U.S. Experience." *Review of Environmental Economics and Policy* 11(1): 80-99.

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

[27] Nordhaus, William D. (2015). "Climate Clubs: Designing a Mechanism to Overcome Free-riding in International Climate Policy." *American Economic Review* 105(4): 1339-1370.

[28] Nordhaus, William D. (2017). "Climate Clubs and Carbon Pricing." Chapter 4 in Cramton, MacKay, Ockenfels, and Stoft (eds), *Global Climate Pricing.* Cambridge: MIT Press, forthcoming 2017.

[29] Parry, Ian (2016). "Reflections on the International Coordination of Carbon Pricing." CESifo Working Paper No. 5975.

[30] Pindyck, Robert S. (2017). "The Use and Misuse of Models for Climate Policy." *Review of Environmental Economics and Policy* 11(1): 100-114.

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

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

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

[34] 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.

[35] Weitzman, Martin L. (2016). "Voting on Prices vs. Voting on Quantities in a World Climate Assembly." *Research in Economics.* In press, available online 29 October 2016.

[36] World Bank (2015). *Integrating Climate Concerns into World Bank Group Actions.* Available at worldbank.org/en/topic/climatechange/brief/integrating-climate-change-world-bank.