Harvard Environmental Economics Program

On a World Climate Assembly and the Social Cost of Carbon

Martin L. Weitzman
Harvard University

HARVARD Kennedy School
JOHN F. KENNEDY SCHOOL OF GOVERNMENT

heep@harvard.edu
http://heep.hks.harvard.edu

On a World Climate Assembly and the Social Cost of Carbon

Martin L. Weitzman

Harvard University

The Harvard Environmental Economics Program

The Harvard Environmental Economics Program (HEEP) develops innovative answers to today's complex environmental issues, by providing a venue to bring together faculty and graduate students from across Harvard University engaged in research, teaching, and outreach in environmental, energy, and natural resource economics and related public policy. The program sponsors research projects, convenes workshops, and supports graduate education to further understanding of critical issues in environmental, natural resource, and energy economics and policy around the world. For more information, see the Program's website: http://heep.hks.harvard.edu.

Acknowledgements

Citation Information

On a World Climate Assembly
and the Social Cost of Carbon

Martin L. Weitzman[*]

October 28, 2016

Abstract

This 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. If this WCA framework can be accepted in the first place, then voting on a single internationally-binding minimum carbon price (the proceeds from which are domestically retained) tends to counter self-interest by incentivizing countries or agents to internalize the externality. I attempt to sketch out the sense in which each WCA-agent's extra 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 first proposition of this paper derives a relatively simple formula relating each emitter's single-peaked most-preferred world price of carbon emissions to the world "Social Cost of Carbon" (SCC). The second and third propositions relate 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.

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 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 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.

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 common commitment, whereas quantity targets, which do not nearly as readily present such a single focal point, have a tendency to rely ultimately on individual commitments.[2] 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 will try to 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

---

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

[2] 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.

negotiating parties to internalize, at least approximately, the global warming externality. 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 world interests by embodying the principle of "I will if you will."[3]

The style of this paper is a sometimes awkward blend of classical with some behavioral economics, which seems unavoidable in discussing actual attempts to resolve the free-rider problem associated with the international public good (or bad) of climate change. 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. If this WCA framework can be accepted in the first place (as a kind of a "worldwide plebiscite"), then voting on a single internationally-binding minimum carbon price, the proceeds from which are domestically retained, tends to counter self-interest by incentivizing countries or agents to internalize the externality. Toward the end 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.[4]

Some of the themes presented here have been preliminarily explored in previous papers.[5] The principal novelty of these notes is to develop relatively simple closed-form expressions relating each emitter's single-peaked most-preferred world price of carbon to the world "Social Cost of Carbon" (SCC) and also relating the WCA-voted price of carbon to the SCC. Using some 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. 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 afterwards), the world seemed mired in what has been called "global warming gridlock."[6]

---

[3]For more about the intellectual coherence of this quid-pro-quo price mechanism, see MacKay, Cramton, Ockenfels and Stoft (2015).

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

[5]See Weitzman (2014) and Weitzman (2015).

[6]*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 Accords, see the Wikipedia entry for "Paris Agreement" and the many other relevant references cited there. A balanced evaluation of COP21 is Keohane and Oppenheimer (2016).

The Kyoto agreement, negotiated in December 1997, began by dividing the world into two huge blocs under the so-called "principle of common but differentiated responsibility and respected capabilities." 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, at the end of the day, 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.

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 emissions in 2010 by about 5% relative to their 1990 emissions. Then, in a second stage, after around 10 years (approximately by 2012 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 an emissions reduction target of 5% in 2020 (relative to 1990 emissions), while the Annex I countries would agree to a more stringent emissions reduction target of 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 Accord nominally covered countries currently accounting for some 95% of world carbon dioxide emissions. Countries agreed to make voluntary pledges, now named euphemistically (in realpolitik diplomatic language) "Intended Nationally Determined Contributions" (INDCs). The INDCs aspired to be transparent in the sense that monitoring, reporting, and verification would in principle be subject to uniform standards. COP21 committed countries to report INDC compliance every five years or so and to set new (and hopefully more ambitious) INDCs for the next five-year pe-

riod, a policy sometimes labeled "pledge and review." There was also provision for possible international linkages via the euphemistically named "Internationally Transferred Mitigation Outcomes" (ITMOs).

All in all, it seems fair to say that the COP21 Accord, which is as yet not fully tried, represents genuine progress. A critically important first step achieved in COP21 is the near-universal acceptance of the 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. So 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.[7]

The core weakness of the COP21 Paris Accord 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. With COP21, there is no penalty for voluntarily setting under-ambitious national targets, and furthermore there is 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*.[8] 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 to me like a realistic strategy for overcoming the free-riding problem associated with an international public good of great importance.

Consider the following hypothetical 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

---

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

[8]When a state formally submits its instrument of ratification to the UNFCC Secretariat, it is supposed to include its NDC (no I). Even so, I think the term NDC is revealing.

emissions, the EPA allowed power plants or companies or states 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 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 INDC 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 catastrophes) and we should be ready beforehand by thinking through the consequences now. This paper is futuristic 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.[9]

3   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 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.[10] In this

---

[9] As will be explained in the concluding section, the WCA can be made to cohere with a top-down "climate club" along the lines of Nordhaus (2015).

[10] In principle, one could consider alternative 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.

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. 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.

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 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 might be useful in indicating what might be the outcome of less-formal international negotiations.[11]

It could be objected that a "consensus" voting rule, not a majority voting rule, is employed in negotiations under the UN Framework on Climate Change. This "consensus" voting rule has been widely interpreted as requiring near-unanimity. With such a restrictive voting rule, significant progress on resolving the global warming externality 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

---

[11]As a possible example of 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. EPA currently uses a carbon price of $40/tCO_2$ as a point estimate for cost-benefit analysis of environmental regulations (see Interagency Working Group on the Social Cost of Carbon (2015)).

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.

I try to explain why voting on (or, perhaps hopefully, even just negotiating) a uniform minimum carbon price empowers what I will call a "countervailing force" against narrow self-interest by automatically incentivizing all parties to internalize, at least approximately, the global warming externality. Again, 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 embodies the principle of "I will if you will."

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 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 and a carbon tax. However, in actuality, the important thing is acquiescence by each nation to a common binding minimum price on carbon emissions, not the particular mechanism by which this common binding minimum price is attained by a particular nation. 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. Toward the end of this 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.[12]

A system of uniform national carbon taxes with revenues kept in the taxing country is a relatively simple and transparent way to achieve internationally-harmonized carbon prices.[13]

---

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

[13]There is actually a fair-sized literature on a carbon-tax (or carbon-price) approach. See, e.g., Metcalf and Weisbach (2009), Cooper (2010), Cramton, Ockenfels and Stoft (2015), Nordhaus (2013), and the many further references cited in these works. In many of these references the question of exactly who collects the

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 equal lump sum payments to each citizen.[14] 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.[15] We economists, I fear, have failed to sufficiently convince the public that there is an important distinction between a self-collected national tax on carbon emissions and an external tax-like price imposed as if by a third party, which siphons tax-like revenues away from the home country.[16]

There already exists a sizable "traditional" literature comparing carbon-emissions taxes with a cap-and-trade system.[17] Arguments can be made on both sides. In my opinion,

carbon tax or price is left somewhat up in the air.

[14] Such kind of equal redistribution was done in the Canadian province of British Columbia, which is widely viewed as having a highly successful carbon-tax program. For a balanced evaluation of the B.C. carbon tax experience, see, e.g., Harrison (2013).

[15] Of course, persuading nations to commit to negotiating a uniform price of carbon in the first place might well involve some "greenfund" equity transfers. Because the imposed "carbon tax" is internally retained within each nation, then, at least for small changes, the green-fund transfers needed to offset increased costs of compliance for price changes are deadweight-loss second-order Harberger triangles of the relatively modest form $(\Delta P \times \Delta Q)/2$. The corresponding international transfers in a cap-and-trade system (which can be either positive or negative, depending, among other things, on initial cap assignments) are first-order immodest rectangles of the form $P \times \Delta Q$. I am unsure about the significance of this argument for non-small changes. Goulder and Schein (2013) discuss the potential for very large revenue flows from the nations purchasing allowances to the nations selling them.

[16] 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.

[17] See, e.g., the survey by Goulder and Schein (2013) and the many further references they cite. Ease of administration is one consideration that emerges from this literature. Relative price stability is another.

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."[18]

In this paper I revisit the debate of "prices versus quantities" from a different, non-traditional, 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.*

2. Be of *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 transparent formula that embodies 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 will 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 cost-effectiveness (provided agreement can be had in the first place).

The second desirable property (low dimensionality) argues in favor of a one-dimensional harmonized tax-like carbon price over an n-dimensional harmonized cap-and-trade system among n nations. Alas, this argument is elusively difficult to formulate rigorously, or even to articulate coherently. My argument here is necessarily, at least in part, behavioral or psychological, 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 "intended nationally determined contributions"

A third desideratum is minimal payments traversing international boundaries, which represents a politically explosive issue.

[18] The influence of so-called "complementary policies" (such as aggressive sector-based targets like low carbon fuel 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, aggressive sector-based targets can have the perverse effect of relocating CO_2 emissions to other sectors, but not reducing net total emissions. See, e.g., the discussion in Goulder and Schein (2013).

actually volunteered by most nations under the COP21 Paris Accord.

With n different national entities, a quantity-based treaty involves assigning n different emissions quotas (whether tradable or not). Treaty making can be viewed as a coordination game with n different players. Such a game can have multiple solutions, often depending delicately on the setup, what is being assumed, and, most relevant here, the choice of negotiating instrument. In the case of Kyoto, the world had in practice arrived at a bad quantity-based free-rider solution that essentially devolved to regional volunteerism. The ultimate outcome of the COP21 Paris Accord remains to be seen, but so far the INDCs 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.

Thomas Schelling introduced and popularized the notion of a focal point in game-theoretic negotiations.[19] 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 a very important behavioral-psychological insight into how the world works, it is difficult to model rigorously, and actual applications seem to be as much of an art as a science.[20]

The concept of "transaction cost" is associated with the work of Ronald Coase.[21] The basic idea is that n parties to a negotiation can be prevented from attaining a socially desirable outcome by the costs of transacting the agreement among themselves. One could argue behaviorally that, other things being equal, transaction costs increase monotonically

---

[19]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.

[20]David Weisbach suggested to me the following analogy for judging whether a focal point is more or less "natural." 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 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.

[21]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).

with the number of parties n.

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 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]

In the case of international negotiations on climate change, I believe that both Schelling's concept of a salient focal point and Coase's concept of transaction costs (as well as Black's median-voter result and Arrow's impossibility theorem) can be used as informal arguments to support negotiating a single harmonized carbon price whose proceeds are nationally rebated. Put directly, it is easier to negotiate one price than n quantities. The "law of one price" is already a familiar salient feature of competitive markets, whereas there is no such thing as a "law of one quantity."[23] I cannot defend the salience of one price rigorously. At the end of the day, this is more of a plausible behavioral conjecture than a rigorous theorem.

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. Such incentives should ideally take the form of a simple, reciprocal, common climate commitment based on the quid-pro-quo principle of "I will if you will." I believe this third property is arguably the most important property of all. This "countervailing force" property is inherently built into a price-based harmonized system of emissions charges, but it is absent from a quantity-based international cap-and-trade system, at least as traditionally formulated.

If I am assigned a cap on emissions, then it is in my own narrow free-riding self-interest to want my cap to be as large as possible (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 no one has an incentive to internalize the externality and everyone has the self-interested

---

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

[23]One could also try to argue that one price is "fair" in the sense that it induces equal (marginal) effort among the parties. I am unsure how convincing this argument is, but it might carry some weight.

incentive to free-ride. What remains is essentially an erratic pattern of altruistic individual volunteerism that is far from a socially optimal resolution of the problem.

An internationally-harmonized, domestically-collected, carbon price is different. If the price were imposed on me alone, 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.

In previous work, I have tried to model formally the role of this third "countervailing force" property of an internationally-harmonized but nationally-collected carbon price.[24] I constructed a basic model indicating an exact sense in which each agent's extra cost from a higher international minimum emissions price is counter-balanced by that agent's extra benefit from inducing all other agents to simultaneously lower their emissions via the higher international minimum price. The key insight from this way of looking at things is that in voting on (or more generally negotiating) a universal minimum carbon price, various nations are, to a greater or lesser degree, internalizing the externality. Loosely speaking, an "average" nation is fully internalizing the externality because its extra cost from a higher emissions price is exactly offset by its extra benefit from inducing all other nations to simultaneously lower their emissions via the higher price.

This paper extends previous work by deriving a relatively simple formula that relates 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.[25] 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

---

[24]See Weitzman (2014, 2015).

[25]The general formulation of this section is an adaptation and expansion of the framework in Weitzman (2014) to this paper. The really novel contribution of these notes comes in the following sections of the paper, which obtain neat, intuitive, closed-form expressions for the case of linear approximations.

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.[26] In this paper it is not really necessary to assume that all of the individuals in one nation (or jurisdiction) are the same.[27] However, it is easiest to start off by conceptualizing 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.[28])

The nation here is effectively an entity that enforces the imposition of an internationally-harmonized minimum carbon 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 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.[29]

---

[26] For some purposes it might be appropriate to consider states or provinces of a large country as the relevant entity.

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

[28] 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 essentially 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.

[29] This is not a trivial assumption. Intuitively, it seems to me to be OK for purposes of simplified modeling

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 price on carbon emissions is P, then the profit-maximizing response (or reaction function, or "demand for carbon emissions") of individual 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 the first-order condition

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

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 for 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 world social loss of imposing a 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 SCC is the price P^* that minimizes the world social loss function (4).[30] The corresponding first-order condition 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

---

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 analysis into a static reduced-form one-period analysis.

[30]Note that I am defining the SCC as the shadow price of carbon in the Pareto-optimal solution. Sometimes alternative definitions of the SCC are based on given trajectories, such as "business as usual."

appropriate to the setup here is

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

where (1) holds simultaneously for each i at $P = P^*$.

It should be noted that the world SCC (or, equivalently, the world efficiency price of 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 transfers, which would ensure that the Pareto-efficient solution is actually attained. 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 a WCA. Thus, I am thinking of the world SCC of P^* as a benchmark or point of departure for what follows.

Consider next what is the optimal level of an internationally-harmonized carbon price from the narrow perspective of agent i. 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 $L_i'(P_i) = 0$, which, from (3), can be rewritten as

$$D_i'(E(P_i))\, E'(P_i) = -C_i'(E_i(P_i))\, E_i'(P_i).$$ (6)

Preferences of i for a worldwide price P are given by the convex loss-function expression (3). It can readily be shown (from convexity of $D_i(E)$ and $C_i(E_i)$) that preferences of i for price P are single peaked with peak value $P = P_i$.

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

$$P_i = D_i'(E(P_i)) \times \left(\frac{E'(P_i)}{E_i'(P_i)} \right).$$ (7)

If all m agents are identical, then $E'(P_i) = mE_i'(P_i)$ for all i and (7) becomes $P_i = mD_i'(P_i)$ or $D_i'(P_i) = P_i/m$ for all i. Furthermore, with identical agents, all $D_i'(P)$ are equal for all i and all P_i are equal for all i. Some reflection reveals that these symmetry conditions, when plugged into (5), imply that $P_i = P^*$ for all i. In this identical-agent situation, therefore, a majority-voting rule automatically resolves the social coordination problem

16

(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 previous work (Weitzman (2014)) I attempted to elaborate further the meaning and implications of condition (7). Unfortunately, the exposition there was more clumsy and less complete than it needed to be. 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 (7) that is more elegant and easier 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 SCC=P^*.

6 Two Linearity Assumptions

To make further progress on understanding most clearly the expression (7), 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 (7). 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 (7).

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

$$D_i(E) = \alpha_i + d_i\, E, \tag{8}$$

where $d_i > 0$. Equation (8) means that the marginal damage for each agent i is a constant

$$D_i'(E) = d_i \tag{9}$$

for some positive coefficient d_i that is allowed to differ for different i.[31] I feel that (9) 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. Note that the marginal damage coefficients $\{d_i\}$ can

---

[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.

differ among different agents i, while still being constant for each particular agent i.

An immediate consequence of (9) combined with (5) is that

$$P^* = \sum_{i=1}^{m} d_i, \tag{10}$$

independent of the cost functions $\{C_i(E_i)\}$. This decomposition property (that the global efficiency-price of carbon is independent of emissions costs) greatly simplifies the analysis without, I hope, losing too much realism.

Another immediate consequence of (9), this time from combining it with (7), is that

$$P_i = d_i \times \left(\frac{E'(P_i)}{E'_i(P_i)} \right), \tag{11}$$

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 $E_i(P)$ in P is of the linear reduced form

$$E_i(P) = \beta_i - s_i P, \tag{12}$$

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

The linearity assumption (12) is essentially ad hoc, but it might be defended as a simplifying approximation that gives some valuable insights. Note from (12) that

$$E'_i(P) = -s_i, \tag{13}$$

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.

Comparing (12) with (1), linearity in P of the reaction function (12) is equivalent to a marginal cost function $C'_i(E_i)$ that is linear in emissions level E_i, of the form

$$C'_i(E_i) = \frac{E_i - \beta_i}{s_i} \tag{14}$$

throughout the relevant range $0 \le E_i \le \beta_i$. Comparing (12) with (14) shows that the

coefficient s_i does double duty. In (12), s_i is interpretable as a sensitivity-reaction coefficient to price changes. In (14), $1/s_i$ is interpretable as a measure of the marginal cost of emissions reductions. Other things being equal, when s_i is large (small) the marginal cost of emissions reduction is small (large).

A third, and perhaps the most important, interpretation of the price-sensitivity coefficient s_i is that it represents the distortionary deadweight loss associated with a positive carbon price-tax change. If the price-tax change is ΔP, then, from (12), the induced emissions change is $\Delta E_i = -s_i \Delta P$. The associated deadweight loss here is $-(\Delta E)(\Delta P)/2$, or

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

$$(15)$$

Equation (15) 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

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

$$(16)$$

The worldwide *average* marginal damage coefficient is

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

$$(17)$$

This concludes the description of the linearity assumptions, which will be used to simplify greatly the analysis of (7) by expressing P_i in terms of a readily-interpretable closed-form equation.

7 A First Main Result

A new result in this paper is the following proposition.

Proposition 1 *Under the linearity assumptions (8) and (12), 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*

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

$$(18)$$

Proof. Start by rewriting terms in the right hand side of equation (11) as functions of the linearity parameters.

19

From (13), it follows that

$$E'_i(P_i) = -s_i,\qquad(19)$$

From (2), it follows that

$$E'(P_i) = \sum_{i=1}^{m} E'_i(P_i).$$

which, when combined with (16) and (19), can be rewritten as

$$E'(P_i) = -m\,\overline{s}.\qquad(20)$$

Combining (17) with (10) implies

$$P^* = m\,\overline{d}.\qquad(21)$$

Multiply numerator and denominator in the right hand side of (11) by $\overline{d}$. Then use (19), (20), and (21) to replace symbols 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 (18). ∎

Equation (18) is a relatively clean 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 two multiplicative factors $d_i/\overline{d}$ (representing i's proportional deviation of its damages from average damages) and $\overline{s}/s_i$ (representing one over i's proportional deviation of its price sensitivity from average price sensitivity).

If there is an "average" agent j with $d_j=\overline{d}$ and $s_j=\overline{s}$, then (18) implies that $P_j=P^*$ From this it follows immediately that in the hypothetical extreme case where all agents are identical, then $P_i=P^*$ for all i. In the general case, P_i is "adjusted" from P^* by the two multiplicative scaling factors $(d_i/\overline{d})$ and $(\overline{s}/s_i)$ that appear in (18).

Note from (18) what agent i is *not* doing here. Agent i is *not* equating its marginal cost of abatement $-C'_i$ to the narrow marginal benefit from one less unit of its own emissions d_i, which would be the analogue here to the condition for a narrowly-self-interested voluntary provision of public goods, and which would result in a free-riding way-too-low provision of the public good. 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 parameter values as i, namely d_i and s_i.

For the sake of decomposing the meaning of (18), suppose hypothetically that all of the price-sensitivity coefficients are identical, so that $\overline{s}=s_i$ for all i. Then equation (18)

simplifies to

$$P_i = P^* \times \left(\frac{d_i}{\overline{d}}\right). \tag{22}$$

Expression (22) means that when all price-sensitivity coefficients s_i of the agents are the same, then the ratio P_i/P^* is exactly equal to the ratio $d_i/\overline{d}$. Here P_i is "adjusted" from P^* by the multiplicative scaling factor $d_i/\overline{d}$, which makes sense. Thus, P_i differs multiplicatively from P^* in (22) by the extent to which d_i differs multiplicatively from $\overline{d}$. Higher (lower) values of marginal damages d_i – relative to the worldwide average value $\overline{d}$ – cause agent i to want relatively higher (lower) values of its most-preferred world price P_i.

Making use of (21), equation (22) can alternatively be expressed as

$$P_i = md_i. \tag{23}$$

Equation (23) is describing what the efficiency-price of carbon *would be* in a hypothetical world where everyone has the *same* marginal damage as i, namely d_i. The favorite world price of agent i is as if i is behaving like a partially-benevolent dictator by imposing its own marginal damage d_i on the rest of the world and calculating what the world efficiency-price md_i would then be. Effectively, the favorite price P_i of agent i is its own marginal damage 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.

Next, and again for the sake of decomposing the meaning of (18), this time suppose hypothetically that all of the marginal damages of the agents are identical, so that $d_i = \overline{d}$ for all i. Then expression (18) simplifies to

$$P_i = P^* \times \left(\frac{\overline{s}}{s_i}\right). \tag{24}$$

Expression (24) means that when all of the marginal damages of all of the agents are the same, then the ratio P_i/P^* must be adjusted by the ratio $\overline{s}/s_i$. Even though marginal damages are identical, here P_i is "adjusted" from $P^* = md_i = m\overline{d}$ by the multiplicative scaling factor $\overline{s}/s_i$. Thus, P_i differs multiplicatively from P^* here by the extent to which $\overline{s}$ differs multiplicatively from s_i. An interpretation of (24) (in light of (15)) is along the following lines. When the price-sensitivity coefficient s_i is relatively low, 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 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.

Plug (21) and (16) into (24) (along with $d_i = \overline{d}$ for all i). This yields the expression

$$P_i = \bar{d} \times \left(\frac{\sum_{i=1}^m s_i}{s_i} \right) \left[= \bar{d} \times \left(\frac{dE}{dE_i} \right) \right]. \tag{25}$$

From (25), the factor $(\sum s_i / s_i)$ 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, along the lines of equation (11). For each unit positive change in its preferred price, agent i is "spending" the cost consequences of abating an extra amount $-dE_i/dP$ but it is "buying" the benefit of worldwide total emissions abatement $-dE/dP$. Equation (25) signifies that agent i is here reacting by applying a multiplier that scales up the effect of its own narrow marginal cost by however many times greater is the value of the world's marginal emissions response (to a price change) than i's own marginal emissions response (to a price change).

This concludes the discussion of the meaning of the basic formula (18). 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 (aka efficiency price of carbon) P^*. Equation (18) 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, each agent i's extra cost from a higher uniform emissions price is counterbalanced, via (18), by that same agent's lessened damage from inducing all other agents to simultaneously lower their emissions in response to that higher price.

The next section discusses some aspects of the application of (18) to the outcome of majority-rule voting in a World Climate Assembly.

8   Relating the WCA-voted Price to the SCC

Let Z represent a collection of m values of $\{Z_i\}$ for $i = 1, 2, ..., m$. The *median* value of $\{Z_i\}$ will be denoted (using the 'tilde' symbol) as $\widetilde{Z}$. Thus,

$$\widetilde{Z} \equiv median\left(\{Z_i\}\right).$$

In what follows, I will make use of the mathematical property that if $Z_i > 0$ for all i, then

$$\widetilde{\left(\frac{1}{Z}\right)} = \frac{1}{\widetilde{Z}}. \tag{26}$$

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 (18).[32]   It is not actually necessary to do multiple pairwise voting on multiple binary price alternatives. A shortcut is available. If each agent i submits 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 centrally calculated 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.[33]

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

Proposition 2 *The WCA-voted majority-rule price $\widetilde{P}$ can be expressed as*

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

Proof. *From (18) and the median voter theorem, the outcome of WCA majority rule $\widetilde{P}$ can be expressed as*

$$\widetilde{P} = P^* \times \left(\frac{\overline{s}}{\overline{d}} \right) \times \widetilde{\left(\frac{d}{s} \right)}. \tag{28}$$

Making use of (26) yields the following equation

$$\widetilde{\left(\frac{d}{s} \right)} = 1 \div \widetilde{\left(\frac{s}{d} \right)}. \tag{29}$$

Plug (29) into (28), yielding (27) ∎

In the most general case, the value of the WCA-voted outcome $\widetilde{P}$ given by (27) 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 (27) that $\widetilde{P}$ might be tolerably close to P^*. After all, $(\overline{s}/\overline{d})$ is *some* (imperfect) measure of the central tendency of $\{s_i/d_i\}$, while $\widetilde{(s/d)}$ is also *some* (imperfect) measure of the central tendency of

---

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

[33]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.

$\{s_i/d_i\}$. 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 (27) and $\widetilde{P}$ is unlikely to be too-too sharply different from P^*.[34]

I now look at some special cases of the distribution of $\{d_i\}$ and $\{s_i\}$ that will give more precise outcomes. The easiest case is where all agents are identical, in which case (27) implies that $P_i = \widetilde{P} = P^*$.

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 $\overline{s} = s_i$ for all i. In this case (28) becomes

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

From (30), 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 (30), then majority 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 Pareto optimality 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 more general case, I now attempt to examine the outcome of systematic randomness in this framework. Suppose the ratios $\{s_i/d_i\}$ 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

$$x_i = \frac{s_i}{d_i}, \tag{31}$$

---

[34]I have found it very difficult to say much more analytically about the expression (27) because it is difficult to decompose it further in the general case. As noted, there is some "hint" from (27) that $\widetilde{P}$ might be tolerably close to P^* because the two (imperfect) measures of central tendency might cancel each other. To make further progress on understanding condition (27), I turn to special cases and a very crude numerical example.

meaning that x_i is the ratio of s_i over d_i, which will be treated as a random variable.

Proposition 3 *Suppose the following data generating process (dgp). The random variables x_i and d_i are each (separately) iid and*

$$s_i \equiv d_i \times x_i. \tag{32}$$

Then (in the limit of a large sample size)

$$\widetilde{P} = P^* \times \left(\frac{\overline{x}}{\widetilde{x}} \right). \tag{33}$$

Proof. *From the independence assumption and (32),*

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

which can be rewritten as

$$\frac{\overline{s}}{\overline{d}} = \overline{x}, \tag{34}$$

From (31),

$$\widetilde{\left(\frac{s}{d} \right)} = \widetilde{x}, \tag{35}$$

Plugging (34) and (35) into (27) yields the desired conclusion (33). ∎

The result (33) indicates that, with the postulated dgp, $\widetilde{P}/P^*$ is the ratio of the mean of x ($=\overline{x}$) divided by the median of x ($=\widetilde{x}$). Once again here, outcomes of $\widetilde{P}/P^*$ are whittled down to a possible fractional difference between two measures of central tendency. I see no reason offhand to believe here that the mean $\overline{x}$ is substantially different from the median $\widetilde{x}$.

This is about as far as pure theory can take us. I think it is fair to say that the formal WCA voting model is 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.

9 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).[35] 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.

It is difficult to deny that there is a very high degree of fuzziness in this estimate of SCC $(=P^*)$. At one extreme, Pindyck (2015) 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 $100/tCO_2$.[36] 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.[37]

From (10), $P^* = \sum d_i$. 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 $P^*=\$40/tCO_2$. Because the national estimates are so poorly determined, for his central national estimates Nordhaus effectively assumes that national values of damages are proportional to national GDPs.

The marginal cost functions that appear in (14) are of the reduced linear form $C_i'(E_i) = (E_i - \beta_i)/s_i$. Nordhaus (2015) effectively estimates the relevant parameters in (14) by combining a global estimate from his DICE-2013 model with detailed regional estimates from an engineering model by McKinsey Company (2009). My impression here is that these numbers are a bit fuzzy because the McKinsey regional estimates are based on an engineering approach that may not be terribly reliable.

Using the Nordhaus (2015) numbers (scaled up by a factor of two), 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 $\widetilde{P} = \$51/tCO_2$.[38] 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-

---

[35] 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.

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

[37] See, for example, Metcalf and Stock (2016).

[38] What I am calling here (and in previous work Weitzman (2014)) i's "most preferred price of carbon" – namely P_i – Kochen (2016) relabels as i's "strategic social cost of carbon."

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 to pretend from this very crude numerical exercise that $\widetilde{P} \approx P^*$.

10   Concluding Remarks

At the end of the day, there is no airtight logic in favor of a negotiated carbon-emissions price over negotiated 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!*[39] 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 tends towards internalizing the externality and gives national negotiators an incentive to offset their natural impulse to otherwise

---

[39]In past papers (Weitzman (2014, 2015)) 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 status-quo 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* a linear subdivision formula) could, in principle, embody some countervailing force against the global warming externality. But, I concluded that negotiating the extra layer of $2n$ first-round linear subdivision coefficients would likely founder politically when applied on a worldwide scale.

I further concluded that, even with seemingly symmetric formulas for the initial quantity allotments, such a quantity-based system seemed far more complicated, baroque, and objectionable in the international context than an internationally-harmonized carbon price. For example, equal per-capita initial assignment of caps would generate huge (and highly variable) post-trade international transfers in a cap-and-trade system, which would be politically unacceptable to countries buying the permits. (On this aspect, see also footnote 20.) Goulder and Schein (2013) discuss the potential for very large trans-border revenue flows from the countries purchasing the allowances to the countries selling them.

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 crude 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 countervailing-force argument that deserves to be highlighted.

Because the formulation is at such a high level of abstraction, it has blurred the distinction between a carbon price and a carbon tax. As was previously noted, the important thing is acquiescence by each nation to a *binding minimum price* on carbon emissions, not the particular internal mechanism by which this obligation is met. 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.[40]

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.[41] 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.

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

---

[40]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 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.

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

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.[42] 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.

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 greenfund transfer 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 caps (with n transfer payments) in a worldwide cap-and-trade system. Considering the many behavioral-psychological-economic arguments favoring a single carbon price that have been made throughout this paper, I simply find it difficult to believe this argument. My tentative (if non-airtight) conclusion: it is difficult to get nations to agree to *anything* 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 emissions quotas with n transfer payments would have to be renegotiated every time there is a significant change. To recapitulate in one sentence 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 a reciprocal common commitment, whereas quantity targets, which do not nearly as readily present such a single focal point, have a tendency to rely ultimately on individual commitments.

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 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 that failure to pay the charge would be treated as a subsidy that is subject to countervailing duties under

---

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

existing provisions of the World Trade Organization.[43]

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 principle of free trade, which exception is predicated on a widespread perception that climate change is edging towards bringing catastrophic 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" of 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 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.[44]

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 noting again that global warming is an extremely serious as-yet-unresolved international public goods problem. With the failure of a Kyoto-style top-down approach, the world has seemingly given up on a comprehensive global design, focusing instead in the 2015 Paris COP21 Accord on essentially voluntary bottom-up nationally determined contributions. These partial measures are laudable, but they seem to me to be far from

---

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

[44]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 rate is likely to be close to 100%.

constituting a socially adequate response to the global warming externality. Perhaps, as this paper has emphasized, a quantity-based focus on negotiating emissions caps embodies a bad design flaw. The arguments of this paper suggest a way in which voting on a binding internationally-harmonized nationally-collected minimum tax or price on carbon emissions might help to internalize the global warming externality by empowering an "I will if you will" approach.

References

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

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

[3] 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/

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

[5] 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.

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

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

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

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

[10] 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.

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

[12] 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)

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

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

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

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

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

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

[19] Metcalf, Gilbert E., and James Stock (2015). "Integrated Assessment Models and the Social Cost of Carbon: A Review and Assessment." Available at http://works.bepress.com/gilbert_metcalf/114/.

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

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

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

[23] Pindyck, Robert S. (2015). "The Use and Misuse of Models for Climate Policy." NBER Working Paper No. 21097.

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

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

[26] 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.

[27] Weitzman, Martin L. (2015). "Voting on Prices vs. Voting on Quantities in a World Climate Assembly." NBER Working Paper No. 20925.

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