8 How a Minimum Carbon-Price Commitment Might Help to Internalize the Global Warming Externality

Martin L. Weitzman

Introduction

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

Downloaded from http://direct.mit.edu/books/chapter-pdf/269406/9780262340380_ccs.pdf by guest on 23 May 2022

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

In this chapter, I try to argue that a uniform minimum 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 as readily present such a single focal point, have a tendency to rely ultimately on individual commitments. As a consequence, negotiating a uniform minimum global carbon tax or price can help to solve the externality problem, whereas individual caps essentially incorporate it. I will try to explain why negotiating a uniform minimum carbon price embodies what I call a "countervailing force" against narrow self-interest by automatically incentivizing all negotiating parties to internalize, at least approximately, the global warming externality. The basic challenge, as I see it, is to construct a relatively simple and acceptable one-dimensional international quid-pro-quo mechanism, which automatically embodies the principle of "I will if you will."[2]

Some Brief History of Climate Negotiations

In the decade from the actual entering into force of the Kyoto Protocol in February 2005 to the Paris COP21 agreement of December 2015, the world seemed mired in what has aptly been called *global warming gridlock*.[3]

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. 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.

Downloaded from http://direct.mit.edu/books/chapter-pdf/269406/9780262340380_ccs.pdf by guest on 23 May 2022

In reality, the “legally binding”emissions reductions of the Kyoto Protocol were anything but legally binding 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 want to be bound by such a mechanism.

Almost from the beginning, the United States and Australia refused to ratify the Kyoto treaty (largely 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 agreed 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 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 quota of about 10%. In reality, no such second stage of ratcheted-up commitments ever materialized.

The recently concluded Paris COP21 agreement of December 2015 (by contrast with Kyoto) made no formal dichotomy between developed and developing countries. In principle, all nations were treated symmetrically. The Paris agreement covered countries currently accounting for some 95% of world carbon dioxide emissions. Countries agreed to make voluntary pledges, now named euphemistically “Intended Nationally Determined Contributions” (INDCs). The INDCs aspired to be transparent in the sense that monitoring, reporting, and verification would 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 period, a policy sometimes labeled “pledge-and-review.” There was also provision for possible international linkage via the euphemistically named “Internationally Transferred Mitigation Outcomes” (ITMOs).

All in all, the COP21 agreement seems like an improvement over the Kyoto Protocol. It appears to be essentially a gamble that the modest

Downloaded from http://direct.mit.edu/books/chapter-pdf/269406/9780262340380_ccs.pdf by guest on 23 May 2022

voluntary slowdowns in emissions may buy enough time to develop inexpensive future carbon-free technologies. This seems to be a risky bet. In any event, it will take maybe a decade or more to sort out the effectiveness of the Paris COP21 agreement.

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. There is no penalty for voluntarily setting underambitious national targets, and there is no penalty for noncompliance by a country with its own voluntary self-announced targets. Under COP21, the only mechanism for compliance is "blame and shame," which seems like a weak incentive for cutting back on global emissions.

I think the INDC label says it all. The *contributions* are chosen by each country. These COP21 contributions are *intended* and *nationally determined*. 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 formula for overcoming the free-riding problem associated with an international public good of great importance.

If the Paris COP21 approach fails to halt "dangerous anthropogenic global warming," which takes the form of a perception of an impending climate catastrophe that is felt on a grassroots level, then I think there may be more pressure on creating a top-down international mechanism that actually works. If climate change becomes sufficiently threatening to an "average"citizen of the world, public opinion may support relinquishing some national sovereignty in favor of the greater good. This chapter is targeted to such an eventuality.

I will try to explain why negotiating a uniform minimum carbon price empowers what I will call a "countervailing force" against narrow self-interest by automatically incentivizing all negotiating parties to internalize, at least approximately, the global warming externality. Again, the basic challenge, as I see it, is to construct a relatively simple and relatively acceptable one-dimensional international quid-pro-quo mechanism, which automatically embodies the principle of "I will if you will."

Negotiating Prices versus Quantities

At first, for simplicity of exposition, I assume that a commitment to a uniform global price of carbon will be implemented as an internationally

Downloaded from http://direct.mit.edu/books/chapter-pdf/269406/9780262340380_ccs.pdf by guest on 23 May 2022

harmonized but nationally retained carbon tax. Later I indicate that the commitment is actually to an internationally equal *minimum* price on carbon emissions, which could be met in a variety of ways, including, in principle, imposing a uniform price floor in a cap-and-trade system. But for expositional purposes here, I pretend that the uniform price takes the form of an equal self-imposed tax on carbon emissions.

An internationally harmonized but nationally retained carbon tax (or price) has already been proposed as a potential solution to the global warming externality and has been examined on its merits.[4] In what follows, I briefly summarize some of the possible virtues of an internationally harmonized but nationally collected carbon tax (or price) that have already been noted in the literature. My foil here is an internationally harmonized cap-and-trade system (without a uniform price floor). This kind of global-design comparison is complicated and full of subjective judgments about what might or might not work better in practice and why or why not. Cap-and-trade systems are perhaps more widely used than taxes throughout the world to control pollution and, in that sense, are perhaps more visible or familiar market-like mechanisms than pollution taxes (although fossil-fuel taxes and subsidies are ubiquitous, if somewhat hidden, almost everywhere). My purpose here is merely to indicate that the perhaps less-familiar uniform carbon tax already has some significant arguments in its favor—as a prelude to some new arguments for negotiating a uniform price on carbon that I will later develop in this chapter.

Both quantity- and price-based controls are inherently uncertain for the period during which they apply (in between times of periodic review), but the uncertainty takes different forms. With cap-and-trade, total emissions are known, but the price or (marginal) cost is uncertain. With a carbon tax, the price or (marginal) cost of carbon emissions is known, but total emissions are uncertain. On the basis of economic models of climate change that include uncertainty, carbon taxes outperform tradable permits, both theoretically and in numerical simulations.[5] In the real world, above and beyond theory and numerical simulations, I think that energy price volatility is poorly tolerated by the general public. Swings in carbon prices, especially in extreme cases, could sour public opinion and discredit for some time thereafter (decades, generations) the entire idea of a market-based approach to the climate change problem. However, it is difficult for me to imagine the broad public getting quite so upset because total emissions fluctuate.

Downloaded from http://direct.mit.edu/books/chapter-pdf/269406/9780262340380_ccs.pdf by guest on 23 May 2022

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

The collected revenues from an internationally harmonized carbon tax remain within each country and could be used to offset other taxes or even be redistributed internally as lump-sum payments. This, I think, is a desirable property. By contrast, the revenues generated from an internationally harmonized cap-and-trade system flow as highly visible external transfer payments across national borders, which might be less easily tolerated by countries required to pay other countries large sums of taxpayer-financed money to buy permits.[6]

This extremely brief discussion of the advantages of an internationally harmonized carbon tax (compared with cap-and-trade) is not intended to be comprehensive. There are also legitimate arguments in favor of internationally harmonized tradable permits and against a carbon tax.[7]

A point in favor of tradable permits, frequently emphasized by its advocates, is the political appeal of giving free allowance permits to carbon-intensive industry groups (as contrasted with taxing them directly on their carbon emissions). As was pointed out, carbon taxes that are internally levied and collected by a national government could be used to reduce other, more distortionary, taxes—or they could even be distributed directly to the citizenry as lump-sum payments. But this redistribution aspect of a carbon tax is hidden behind the scenes as it were. Individual firms will prefer, and typically strongly prefer, what they perceive as the lesser burden to them of freely allocated permits over the greater perceived burden to them of pollution taxes. Indeed, studies show that the market value of the free allowances is typically significantly greater than the higher compliance costs of decarbonization that are incurred.[8] Firms and countries in a cap-and-trade regime will therefore struggle hard for a larger share of the total amount of freely distributed emissions allocations. The political appeal of freely distributed tradable permits is a double-edged sword. When negotiating emissions caps, a serious income distortion is introduced because a

Downloaded from http://direct.mit.edu/books/chapter-pdf/269406/9780262340380_ccs.pdf by guest on 23 May 2022

nation is much more concerned with the revenues from its own free quota allocations than it is concerned with overall international social optimality. Auctioning off the allowances would eliminate this income-effect distortion on the individually desired level of free permit allocations, but then we are effectively back in a tax-like system.

Both approaches (an internationally harmonized but domestically collected carbon price and freely distributed marketable permits) are subject to immense—sometimes seemingly overwhelming—criticisms. In both cases, innumerable practical details must be attended to and worked out. In both cases, an effective international treaty needs to be binding, which raises uncomfortable issues of enforcement mechanisms and international sanctions. Additionally, there might be mixed hybrid systems, such as cap-and-trade with a uniform floor on the carbon price. I merely want to establish a level playing field where the idea of an internationally harmonized carbon tax already commands at least as much intellectual respect as an internationally harmonized cap-and-trade system (without a uniform price floor).

Throughout this chapter, I argue that it is difficult to resolve the global warming externality problem by directly assigning individual quantity targets. A meaningful, comprehensive, quantity-based treaty involves specifying as many different binding emissions quotas (whether in the form of tradable permits or not) as there are national entities. Each national entity has a self-interested incentive to negotiate for itself a high cap on carbon emissions—much higher than would be socially optimal. The resulting free-rider problem plagues a quantity-based approach. Even if there were a collective commitment to negotiate or vote on a second-stage worldwide total emissions cap, which I will later assume for the sake of argument, disagreements over the first-stage fractional subdivision formula (for disaggregating the negotiated or majority-voted aggregate worldwide quantity cap into individual quantity caps) would make it difficult to enact such a quantity-based approach.[9]

The inspiration for this chapter is the perception of a desperate need for some radical rethinking of international climate policy. As a possibly useful conceptual guide for what negotiations might accomplish, I sometimes ask the reader to temporarily suspend disbelief by considering what might happen in a "World Climate Assembly" (WCA) that votes on global carbon emissions via the basic principle of one-person-one-vote majority rule. In

Downloaded from http://direct.mit.edu/books/chapter-pdf/269406/9780262340380_ccs.pdf by guest on 23 May 2022

this conceptualization, nations would vote along a single dimension for their desired level of emissions stringency on behalf of their citizen constituents, but the votes are weighted by each nation's population.

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

It 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 warming externality problem cannot be resolved without a binding agreement on some overall formula for dividing emissions responsibilities among nations. Voluntary altruism alone will not solve

Downloaded from http://direct.mit.edu/books/chapter-pdf/269406/9780262340380_ccs.pdf by guest on 23 May 2022

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 noncompliance. The question then becomes: *Which* collective-commitment frameworks and formulas are more promising than which others?

Theory of Negotiating a Uniform Carbon Price

In this chapter, I examine the theoretical properties of a natural one-dimensional focus on negotiating a single binding price on carbon emissions, the proceeds from which are domestically retained. As previously mentioned, 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.

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. But it is not necessary for the conclusions of this chapter. Nations or regions could meet the obligation of a minimum price on carbon emissions by whatever internal mechanism they choose—a tax, a cap-and-trade system, a hybrid system, or whatever else results in an observable price of carbon above the internationally agreed minimum. I elaborate further on this issue in my concluding remarks.

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

1. Induce cost-effectiveness.
2. Be of one dimension based on a “natural” focal point to facilitate finding an agreement with relatively low transactions costs.
3. Embody “countervailing force”against narrow self-interest by automatically incentivizing all negotiating parties to internalize the externality via a simple, reciprocal, I will if you will, common-climate commitment formula.

Downloaded from http://direct.mit.edu/books/chapter-pdf/269406/9780262340380_ccs.pdf by guest on 23 May 2022

Using these theoretical properties as criteria, I now compare and contrast an idealized binding harmonized tax-like price with an idealized binding cap-and-trade system (without a uniform price floor).

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 articulate coherently. My argument here is necessarily intuitive or behavioral and relies on empirical counterexamples. In this situation, two important empirical counterexamples are the breakdown of the quantity-based Kyoto approach and the hugely underambitious "intended nationally determined contributions" actually volunteered by nations under the COP21 Paris Accord.

With n different national entities, a quantity-based treaty involves assigning n different binding emissions quotas (whether tradable or not). Treaty making can be viewed as a coordination game with n different players. Such a game can have multiple solutions, often depending delicately on the setup, what is being assumed, and, most relevant here, the choice of negotiating instrument. In the case of Kyoto, the world had in practice arrived at a bad quantity-based 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 grossly underwhelming, even leaving aside the near impossibility of achieving the stated goal of keeping global warming below 2°C.

Thomas Schelling[10] introduced and popularized the notion of a focal point in game theory. Generally speaking, a focal point of an n-party coordination game is some salient feature that reduces the dimensionality of the problem and simplifies the negotiations by limiting bargaining 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.

Downloaded from http://direct.mit.edu/books/chapter-pdf/269406/9780262340380_ccs.pdf by guest on 23 May 2022

The concept of "transactions cost" is associated with the work of Ronald Coase.[11] The basic idea is that *n* parties to a negotiation can be prevented from attaining a socially desirable outcome by the costs of transacting the agreement among themselves. One could try to argue that, other things being equal, transactions costs increase proportionally with the number of parties *n*. Negotiating a one-dimensional price with single-peaked preferences has the important additional property of allowing a majority-rule voting equilibrium, which avoids the Arrow impossibility theorem.

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

The third desirable property is that the instrument or instruments of negotiation should embody a "countervailing force" against narrow free-riding self-interest by incorporating incentives that automatically internalize the externality. 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 my cap will be tradable as a permit). The self-interested part of me wants maximal leniency for myself. Other than altruism, there is no countervailing force on the other side encouraging me to lower my desired emissions cap because of the externality benefits I will be bestowing on others.

Downloaded from http://direct.mit.edu/books/chapter-pdf/269406/9780262340380_ccs.pdf by guest on 23 May 2022

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 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 (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.[12]

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.[13] I constructed a basic model indicating an exact sense in which each agent's extra cost from a higher international minimum emissions price is counterbalanced by that agent's extra benefit from inducing all other agents to simultaneously lower their emissions via the higher international minimum price (which might well take the form of a uniform price floor on a cap-and-trade system).

With further restrictions, the model showed that population-weighted majority rule for an internationally harmonized tax-like carbon price can come as close to an optimal price on emissions as the median per capita marginal benefit is close to the mean per capita marginal benefit. The key insight from this way of looking at things is that in voting 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.

Downloaded from http://direct.mit.edu/books/chapter-pdf/269406/9780262340380_ccs.pdf by guest on 23 May 2022

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

Might a Modified Cap-and-Trade Work as Well?

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

But maybe I am being unfair to tradable permits. Suppose we imagine trying to convert the *n*-dimensional problem of allocating carbon emissions permits into some one-dimensional aggregate-quantity analogue of a uniform tax-like price on carbon emissions. We might imagine a thought experiment where the cap-and-trade negotiators are sitting around a

Downloaded from http://direct.mit.edu/books/chapter-pdf/269406/9780262340380_ccs.pdf by guest on 23 May 2022

negotiating table limiting themselves to simple linear formulas for allocating individual emissions caps as a fraction of total world emissions.[14]

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

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

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

When a cap-and-trade system is used to control pollution *within* a nation, the government of that nation *assigns* the caps (or the fractions of

Downloaded from http://direct.mit.edu/books/chapter-pdf/269406/9780262340380_ccs.pdf by guest on 23 May 2022

emissions).[15] In this intranational case, there is a natural symmetry between a one-dimensional price p and a one-dimensional total quantity of emissions E. But there is no international government that has the unilateral power to assign caps or fractions. These caps or fractions must be *negotiated* among sovereign nations. This breaks the one-dimensional symmetry because now one tax-like price p is contrasted with the asymmetry of n vested sovereign interests jockeying for the n initial fractional distributions. There is thus a critical distinction between intranational and international cap-and-trade systems. In the international case, the initial distribution of caps is explicitly distributive, resulting in a war of words about who caused the global warming problem and who should bear the burden of remedying it, who is rich and who is poor, what is fair and what is unfair, and so on. There could also be a war of words about the green-fund transfers required to induce participation in a uniform-price treaty, but for reasons elaborated in footnotes 5 and 8 regarding the difference between first- and second-order transfers, I think that an internally retained price treaty takes a lot of pressure off the green-fund payments.

But perhaps a formulation of this generality is biased against cap-and-trade. We might try to imbue the distribution coefficients with dimensionality-reducing salient qualities by imagining “naturally symmetric” focal allocations of the fractional coefficients. One such seemingly symmetric formula might be that each country is assigned the same fractional reduction of emissions from some agreed-on baseline year. The Kyoto Protocol of 1997 adopted just a little of the spirit of this idea for developed countries alone, with the hope that some variant of it might later be extended to developing countries. The high-income industrialized countries (Annex I) agreed to “binding” commitments (but without any enforcement mechanism) to reduce greenhouse gas emissions in 2010 by an average of 5% relative to 1990 levels (although allowing some individually negotiated variations around that 5% average). Developing countries were exempt from any “binding” commitments. Overall, the Kyoto Protocol did not come close to fulfilling its initial aspirations. The United States and Australia did not ratify; Canada, Japan, and Russia eventually dropped out; and individual compliance was at best spotty.[16] Furthermore, and perhaps most distressingly, non-Annex I countries did not formally agree to any actual future “binding” commitments going forward from 2012. The Kyoto experience is subject to multiple interpretations. For me, it largely

Downloaded from http://direct.mit.edu/books/chapter-pdf/269406/9780262340380_ccs.pdf by guest on 23 May 2022

testifies to the great difficulty of negotiating binding international quantity caps on the major emitters. In the language employed here, it has been overwhelmingly problematic to assign binding quantity-like distributional coefficients on a worldwide basis. The Paris COP21 agreement of December 2015 "solved" this problem only by making all targets completely voluntary as "intended nationally determined contributions." In COP21, there was no pressure for nations to cut back emissions by 5% or any other uniform amount.

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

I abstain from further speculation. My point here is that no matter what quantity-like initial allocation mechanism I can imagine, an attempt to modify an international cap-and-trade system by making it one dimensional seems likely to founder for essentially the same reasons that an unmodified international cap-and-trade system founders. In a quantity-based system with n different sovereign nations, I fear there will be intractable negotiations for n different distributional assignments (a_i, b_i), with no force countervailing each nation's free-riding desire to secure for itself a selfishly lenient emissions fraction of the total emissions E.[17]

Here is what I think is the essence of the one-price versus n-quantities negotiation problem as elaborated in this section. A quantity-type system based on a formula like $E_i(E) = a_i + b_iE$ involves *two* layers of negotiations. First, the n parties must agree on the quantity-like distributional coefficients (a_i, b_i).Second, the parties must agree on the single worldwide aggregate level of emissions E. By contrast, a price-based system involves only *one* layer of negotiation, focused on agreeing to a single one-dimensional uniform price p. The latter is not an easy task, but it would seem generally

Downloaded from http://direct.mit.edu/books/chapter-pdf/269406/9780262340380_ccs.pdf by guest on 23 May 2022

easier to negotiate one price layer than two quantity layers (whose first layer involves assigning *n* quantity-like distributional coefficients). Admittedly, this argument depends on a particular way of framing the issue, but it seems to me that, in international negotiations among *n* sovereign nations, there may be an irreducible asymmetry between one price instrument versus *n* quantity instruments.

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

Concluding Remarks

At the end of the day, there is no air-tight logic in favor of a negotiated price over negotiated quantities, only a series of partial arguments. One argument is that the revenues from a tax-like carbon price are nationally collected, so that the contentious distributional side is somewhat hidden, and there is at least the appearance of fairness as measured by equality of marginal effort. A second desirable feature, I have argued, is the natural salience and relatively low transaction costs of negotiating one price as against negotiating *n* quantities, which, although somewhat imprecise, is in my opinion

Downloaded from http://direct.mit.edu/books/chapter-pdf/269406/9780262340380_ccs.pdf by guest on 23 May 2022

an important distinction. A third argument is the self-enforcement mechanism that constitutes the main theme of this chapter, namely, the built-in countervailing force of an imposed uniform price of carbon, which tends to internalize the externality and gives national negotiators an incentive to offset their natural impulse to otherwise bargain for a low price.

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

My argument here is sufficiently abstract that it is open to enormous amounts of criticism on many different levels. There are so many potential complaints that it would be incongruous to list them all and attempt to address them one by one. These 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 previously noted, the important thing is acquiescence by each nation to a *binding minimum price* on carbon emissions, not the particular internal mechanism by which this obligation is met. A system of national carbon taxes with revenues kept in the taxing country is a relatively simple and transparent way to achieve internationally harmonized carbon prices. But it is not absolutely necessary for the conclusions of this chapter. In principle, nations or regions could meet the obligation of a minimum price on carbon emissions by whatever internal mechanism they choose—a tax, a cap-and-trade system with a tax floor, some other hybrid system, or whatever else results in an observable price of carbon above the uniform minimum.[18]

Of course any nation or region could choose to impose a carbon tax or price above the international minimum. 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 chapter is primarily expository and exploratory. *Any* proposal to resolve the global warming externality will face a seemingly overwhelming array of practical administrative obstacles and will need to overcome powerful vested interests. That is the nature of the global warming externality problem. The theory of this chapter seems to suggest that negotiating a uniform minimum price on carbon can have several desirable properties, including, especially, helping to internalize the global warming externality. To fully defend the relative "practicality" of what I am proposing would probably require a book not a chapter. In any event, *this* article is not primarily about practical considerations of international negotiations. I leave that important task mostly to others.[19] However, I do want to mention just a few real-world considerations that have been left out of my mental model yet seem especially pertinent.

An example of a relatively small practical issue that I am waving aside is just where in the production chain a carbon price should be collected. I think the presumption would be that the carbon price should be collected by the country in which the carbon dioxide is actually released into the atmosphere. One might try to argue that a carbon price should be collected downstream as close as possible to the point where the carbon is burned. But this would involve an impractically large number of collection points. It is much easier to collect the price upstream at various chokepoints where the carbon is first introduced into the carbon-burning economy.[20]

A truly critical issue is that a binding international agreement on a uniform minimum carbon tax or price requires some serious compliance mechanism. To begin with, the carbon price must be observable. For enforcement, perhaps there is no practical alternative to using the international trading system for applying tariff-based penalties on imports from noncomplying nations. Nordhaus (2015) advocates such an approach with uniform border tariffs on imports from nonmember countries imposed by a "Climate Club" of member nations who agree to impose a harmonized carbon price on themselves. Cooper (2010) has argued for an expansive interpretation, whereby the internationally agreed charge on carbon emissions would be considered a cost of doing business, such that failure to pay the charge would be treated as a subsidy that is subject to countervailing duties under existing provisions of the World Trade Organization.[21]

An efficient carbon price naturally produces more winners than losers (by the metric of the modified Pareto criterion). In the case of the global warming externality, which has been characterized as the greatest public

Downloaded from http://direct.mit.edu/books/chapter-pdf/269406/9780262340380_ccs.pdf by guest on 23 May 2022

goods problem of all time, it seems reasonable to suppose that there might be many times more winners than losers from imposing a uniform carbon price. Because countries here get to keep their own carbon price-generated revenues, welfare-compensating transfers, to the extent that they are made at all, should, at least for small changes, be relatively modest second-order, deadweight-loss triangles instead of the relatively immodest first-order rectangle transfers associated with tradable permits from, say, an initial assignment of caps that are equal per capita.[22]

I close by noting again that global warming is an extremely serious as-yet-unresolved international public goods problem. With the failure of a Kyoto-style quantity-based approach, the world has seemingly given up on a comprehensive global design, settling instead in the 2015 Paris COP21 agreement for completely voluntary and sporadic national, subnational, and regional "contributions." These partial measures seem far from constituting a socially efficient response to the global warming externality. Perhaps, as previously suggested, a quantity-based focus on negotiating emissions caps embodies a bad design flaw. The arguments of this chapter suggest a way in which negotiating 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.

Notes

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

2. For more about the coherence of this quid-pro-quo mechanism, see chapters 2 and 4.

3. *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 references cited there. For more information on the Paris COP21 Accords, see the Wikipedia entry for "Paris Agreement" and the many other references cited there.

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

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

Downloaded from http://direct.mit.edu/books/chapter-pdf/269406/9780262340380_ccs.pdf by guest on 23 May 2022

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

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

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

9. One could try to argue that binding green-fund equity payments are required to get *n* countries to agree in the first place to negotiate a uniform carbon price, also representing an *n*-dimensional problem. However, footnotes 5 and 7 suggest that the required green-fund payments may be smaller than the absolute value of the (positive or negative) transfers involved in a cap-and-trade regime that starts off, say, with equal per capita permit assignments. Cramton, Ockenfels, and Stoft (chapter 12, this volume) argue additionally that choosing a green-fund equity-payment formula for a uniform price can be reduced to a one-dimensional focal problem.

10. See Schelling (1960). Also see the 2006 special 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.

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

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

13. See Weitzman (2014).

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

Downloaded from http://direct.mit.edu/books/chapter-pdf/269406/9780262340380_ccs.pdf by guest on 23 May 2022

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

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

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

18. A minimum carbon price could theoretically be attained in a 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 *n*-dimensional quantity-based binding agreement among *n* different nations.

19. See Bodansky (2010) or Barrett (2005).

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

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

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

References

Asheim, G. B. 2012, August 11. A distributional argument for supply-side climate policies. *Environmental and Resource Economics*. Published online.

Downloaded from http://direct.mit.edu/books/chapter-pdf/269406/9780262340380_ccs.pdf by guest on 23 May 2022

Barrett, S. 2005. *Environment and Statecraft: The Strategy of Environmental Treaty Making*. Oxford: Oxford University Press.

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

Bosetti, V., and J. Frankel. 2012, August. Sustainable cooperation in global climate policy: Specific formulas and emissions targets. Mimeo.

Coase, R. H. 1960. The Problem of Social Cost. *Journal of Law & Economics* 3:1–44.

Colman, A. M. 2006. Thomas C. Shelling's psychological decision theory: Introduction to a special issue. *Journal of Economic Psychology* 27:603–608.

Cooper, R. N. 2010. The case for charges on greenhouse gas emissions. In *Post-Kyoto International Climate Policy: Architectures for Agreement*, ed. J. Aldy and R. Stavins. Cambridge, England: Cambridge University Press.

Cramton, P., A. Ockenfels, and S. Stoft. 2015. An international carbon-price commitment promotes cooperation. *Economics of Energy & Environmental Policy* 4 (2): 51–64.

Cramton, P., and S. Stoft. 2012. Global climate games: How pricing and a green fund foster cooperation. *Economics of Energy & Environmental Policy* 1 (2): 125–136.

Ellerman, A. D. 2010. The EU's emissions trading scheme: A prototype global system? In *Post-Kyoto International Climate Policy: Architectures for Agreement*, ed. J. Aldy and R. Stavins. Cambridge, England: Cambridge University Press.

Goulder, L. H., M. A. C. Hafstead, and M. Dworsky. 2010. Impacts of alternative emissions allowance allocation methods under a federal cap-and-trade program. *Journal of Environmental Economics and Management* 60 (3): 161–181.

Goulder, L. H., and A. R. Schein. 2013. Carbon taxes vs. cap and trade: A critical review. *Climate Change Economics* 4 (3): 1–28.

Hoel, M., and L. Karp. 2002. Taxes vs. quotas for a stock pollutant. *Resource and Energy Economics* 24:367–384.

Libecap, G. D. 2013. Addressing global environmental externalities: Transaction costs considerations. *Journal of Economic Literature* 52 (2): 424–479.

Metcalf, G. E., and D. Weisbach. 2009. The design of a carbon tax. *Harvard Environmental Law Review* 33 (2): 499–556.

Nordhaus, W. D. 2007. To tax or not to tax: Alternative approaches to slowing global warming. *Review of Environmental Economics and Policy* 1 (1): 26–44.

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

Downloaded from http://direct.mit.edu/books/chapter-pdf/269406/9780262340380_ccs.pdf by guest on 23 May 2022

Nordhaus, W. D. 2015. Climate Clubs: Designing a mechanism to overcome free-riding in international climate policy. *American Economic Review* 105 (4): 1339–1370.

Pizer, W. 1999. Optimal choice of policy instrument and stringency under uncertainty: The case of climate change. *Resource and Energy Economics* 12:255–287.

Schelling, T. C. 1960. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

Victor, D. 2011. *Global Warming Gridlock*. Cambridge, England: Cambridge University Press.

Weitzman, M. L. 1974. Prices vs. quantities. *Review of Economic Studies* 41 (4): 477–491.

Weitzman, M. 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.

Downloaded from http://direct.mit.edu/books/chapter-pdf/269406/9780262340380_ccs.pdf by guest on 23 May 2022