

**A Meta-Architecture for Agreement:
How to Solve a Collective Action Problem by Creating a Collective Action Problem, and
How to Engineer an Effective and Ethical Climate Treaty***

Mark Bryant Budolfson

mbudolfs /AT/ princeton /DOT/ edu

Version 1.39

Abstract

The problem of how to secure global cooperation on a comprehensive emissions-reduction scheme is both the most important and the most intractable problem that stands in the way of an effective global response to climate change. This paper provides a straightforward, realistic, and non-coercive solution to this problem that is superior to all competitors along both ethical and practical dimensions, growing in superiority as assumptions about the dispositions of powerful nations are made increasingly pessimistic. A simple game theoretic result shows that even when our existent tools for solving collective action problems are useless, we can often escape a tragic dilemma by turning the force of such a dilemma against itself, by intentionally engineering and placing ourselves into a new dilemma, the predictable outcome of which is a solution to the initial, otherwise insoluble dilemma. A climate treaty can use this strategy by intentionally creating something like a multi-player prisoner's dilemma at the level of nations, the predictable outcome of which is universal ratification and long-run compliance to an effective climate treaty, even if compliance makes the current citizens of many nations worse off than the no-treaty status-quo, and even if nations are disposed not to comply whenever it is in the interest of their current citizens not to comply. The result is a 'meta-architecture' for securing agreement and compliance to a chosen global response to climate change that remains relatively agnostic about the details and architecture of the global response that ought to be chosen. More generally, the result is a framework for constructing economically and ethically optimal solutions to many previously intractable collective action problems.

* Thanks to John Devlin, Robert Keohane, Gideon Rosen, Peter Singer, and Jane Willenbring for helpful comments.

Anthropogenic climate change arises from a set of collective action problems that are impervious to our standard tools for achieving social cooperation and avoiding tragic outcomes. In particular, we are unable to rely on benevolent human dispositions, explicit bargaining, natural coordination, or governmental coercion to avoid a tragic outcome.¹ That is because a climactic tragedy of the commons arises not only at the level of individual people, but also at each higher level of collective abstraction, and there is no effective authority to provide the typical failsafe of governmental coercion. Furthermore, those who will suffer the most as a result of climate change are members of future generations who are incapable of influencing the current decisions that will determine their fate. Finally, ideology seems to preclude the possibility of any effective international agreement, because many negotiators refuse to consider any agreement that would make their nations worse off than the status quo, while others refuse to consider any agreement that isn't both fair and effective; this seems to make a stalemate inevitable, because it seems impossible for a single agreement to satisfy all of these constraints even in broad approximation.

¹ Derek Parfit provides a taxonomy of solutions to collective action problems and a classic discussion of solutions involving benevolent human dispositions in *Reasons and Persons*, Chapters 2-6; Ronald Coase provides a classic discussion of 'free-market' solutions in "The Problem of Social Cost"; Thomas Schelling provides a classic discussion of explicit bargaining in *The Strategy of Conflict*, and suggests such an approach to climate change in "What Makes Greenhouse Sense?"; Garrett Hardin provides a classic discussion of governmental coercion in "The Tragedy of the Commons"; Elinor Ostrom provides a taxonomy of solutions and a classic discussion of solutions involving natural coordination and bargaining, facilitated by appropriate background institutions in *Governing the Commons*. If the pessimistic assumptions outlined in this paper are correct, then Schelling's bargaining response to climate change would be unavailable, because according to the assumptions outlined below, an effective bargaining response would require the United States and other nations to act contrary to the interests of their current citizens, which would be impossible given those assumptions. In my paper "Why All Plausible Normative Theories are Sometimes Dramatically Collectively Self-Defeating", I show that Parfit's assumptions about morality and benevolence are mistaken by showing that morality and benevolence cannot be relied upon to solve collective action problems even in a world of normatively flawless agents.

If this pessimistic view of our situation is correct, are we doomed?

We are not. Even when our standard tools for avoiding tragic outcomes are useless, we can often escape a tragic dilemma by turning the force of such a dilemma against itself. The trick is to intentionally create a new tragic dilemma, the predictable outcome of which is a solution to the initial, otherwise insoluble dilemma. In more detail and applied to the problem of climate change, the idea is that we can intentionally create tragic dilemmas, which are situations in which we can predict that everyone will 'defect', thereby making everyone worse off than they would have been if everyone had chosen to 'cooperate' instead; furthermore, we can engineer and introduce international treaties that create such tragic dilemmas, the predictable outcome of which is that all nations will agree to the treaty, thereby making themselves worse off than they would have been if they had all chosen not to agree to the treaty instead; finally, we can engineer a climate treaty that has this 'tragic' flavor, but where the treaty makes the current generation of only a few high-emitter nations worse off as a mere side-effect of a mechanism within the treaty that effectively combats climate change. Such a treaty might be our only effective option if nations are disposed to reject or pull out of treaties whenever they make their current citizens worse off, and if it is impossible to effectively combat climate change without making the current citizens of a few influential nations worse off than the no-treaty status quo.

The motivation for such a treaty emerges most clearly from reflection on the following two claims, which constitute a worst-case description of our situation with respect to a climate treaty:

No nation will agree to a treaty if agreeing would make their current citizens worse off than not agreeing. (Call this the *Realist Assumption*.)

Any effective climate treaty requires the agreement of the United States, China, Brazil, and other high-emitter nations, and any effective treaty would make the current citizens of those nations worse off than the no-treaty status quo – i.e., worse off than they would be without the existence of a climate treaty. (Call this the *Current Generation Assumption*.)

From these two assumptions, it might seem to follow that the United States, China, Brazil, and other high-emitter nations will never agree to an effective treaty, and thus that there is no hope for an effective treaty. However, those conclusions do not really follow from these assumptions. That is because the introduction of a treaty can change what it is in the interest of nations to do, thereby making it the case that nations are better off by agreeing to the treaty than not agreeing, even if agreeing to the treaty makes nations worse off than the no-treaty status quo – i.e., worse off than they would have been if the treaty had never been introduced. Although this might seem paradoxical at first glance, denying it would imply that tragic dilemmas can never arise – but such dilemmas can arise, even at the level of nations, and the threat of such dilemmas provides part of the intellectual justification for global authorities like the World Trade Organization (WTO). For example, imagine that global trade rules did not exist

except for the rule that duties could only be introduced by treaties open to all nation; if so, then it would be possible to create a tragic dilemma at the level of nations by introducing a treaty that imposed duties on non-signatory nations: given plausible assumptions, all nations would ultimately sign on to such a treaty in order to avoid having non-reciprocated duties imposed on their exports, even though universal ratification of such a treaty would make all nations worse off than they would have been if the treaty never existed, because imposing symmetrical duties on all nations would have no real positive effect for any nation, but would have the effect of imposing additional economic inefficiency on all nations. This illustrates the way in which the introduction of a treaty can change what it is in the interest of nations to do, thereby making it the case that nations are better off by agreeing to the treaty than not agreeing, even though agreeing to the treaty makes nations worse off than the no-treaty status quo.

At first glance, a treaty that creates a tragic dilemma at the level of nations might seem necessarily irrational and undesirable. However, a treaty with this tragic flavor is exactly what we need if we face an intergenerational tragedy of the commons that is impervious to our standard tools for solving collective action problems, and that can only be solved by finding a novel way of changing our situation to make it the case that some influential nations would best serve the interests of their current citizens by acting in a way that makes those citizens worse off than the status quo. The only practical way of solving such a problem seems to be by creating something like a tragic dilemma at the level of nations, the predictable outcome of which is collective action that makes some nations worse off than the status quo.

To see how such a strategy might work in practice, imagine a treaty that has two components: a duty imposed on imports from nations that are not signatories to the treaty, where this duty is determined by the normalized emissions differential between signatory nations and non-signatory nations,² and a cap and trade scheme among signatories to the treaty, where the cap decreases each year along a path that is insensitive to the number of signatories to the treaty.³ Many nations would initially join such a treaty, in part because it would be to the advantage of many nations to impose an unreciprocated duty on imports from nations that are not initially signatories to the treaty, especially in light of the fact that the treaty would not require deep emissions reductions early on, but would only require significant reductions later as the cap decreased and the largest emitters joined the treaty, at which point the cost to early signatories would tend to be offset by payments from later-joining nations in exchange for emissions permits; furthermore, many other nations would initially join such a treaty out of a long-run concern for their own citizens and perhaps even the world, even if such a treaty was not to their short-run economic advantage. As nations joined the treaty, incentives would increase for other nations to join as well, because additional signatories would mean

² For ease of exposition, I use ‘signatory’ to refer to nations that have *ratified* the treaty. There are many ways of normalizing the emissions differential into units that allow ‘apples-to-apples’ comparison of the emissions differential between signatory and non-signatory nations: the simplest are either *by population*, or *by economic output* – i.e., tying duties to the difference in emissions *per person* in signatory nations versus non-signatory nations, or to the difference in emissions *per unit of GDP* in signatory nations versus non-signatory nations. Perhaps the most attractive approach would be a complex metric that takes both population and economic output into account, as well as other factors such as economic development and growth.

³ Such a treaty would also include an allocation scheme for emissions permits, any number of which would be consistent with the suggestions in this paper. Such a treaty might also include a mechanism to redistribute to developing nations some of the duties collected, in order to secure the universal compliance of developing nations and to offset the more ethically significant costs of the treaty to developing nations.

additional nations imposing duties on imports from non-signatory nations; in addition, as more nations joined the treaty and the cap decreased over time, the cap and trade scheme would become increasingly effective, thereby increasing the emissions differential between signatories and non-signatories, thereby increasing the magnitude of each individual duty imposed. As a result, even if the initial set of signatories was comparable to the limited set of nations that initially ratified the Kyoto Protocol, the treaty would set in motion a chain reaction that would make it in the interest of an increasing number of nations to join over time, eventually driving the costs of not joining so high that even the United States and China would ultimately find it in their interest to join rather than continue to hold out.

Of even greater importance, once nations joined such a treaty, it would never be in their interest to pull out, because the costs of pulling out always outweigh the costs of staying in after the point at which it is initially in a nation's interest to join, if the treaty has the 'self-enforcing' structure described above.⁴ This solves the most serious problem for climate treaties, which is the problem of securing not merely initial agreement to the treaty, but more importantly *long-run participation and compliance*.⁵ This problem is insurmountable for more familiar treaty proposals. For example, even on the unrealistic assumption that the United

⁴ A treaty is *self-enforcing* in the relevant sense if and only if it is both individually and collectively rational to maintain agreement to the treaty from the point of view of all nations. (For a similar definition, see Scott Barrett, *Environment and Statecraft*, pg. xiv.)

⁵ The self-enforcing structure is also consistent with other complementary incentive schemes to encourage compliance. For example, one promising addition would be for duties from all nations to be held by a single global administrator until the end of each year, at which point each nation's proceeds would be disbursed only if that nation complied with the treaty's provisions in the previous year; duties could then be subtracted from the accounts of non-compliant signatory nations based on their degree of non-compliance, with the proceeds distributed to compliant signatories.

States and China might initially agree to a more familiar treaty that is costly enough to be effective, once the negative effects of such a treaty became salient to the populations and leaders of those nations, those nations would certainly pull out, resulting in yet another failed climate treaty and a decade or more of additional wasted time.⁶ More importantly, the failure of a treaty that is costly enough to be effective under such circumstances would create deep hostility and resistance to future treaties, especially among nations that incurred significant costs under the failed treaty. For such reasons, it is likely that the world has only one shot at a climate treaty that is costly enough to be effective – and it is almost certain that more a familiar treaty proposal would waste that single opportunity, whereas a treaty with the self-enforcing structure described above would provide a realistic means to lock in universal participation and compliance over the long run.

The main hurdle for the particular treaty proposal described above is compliance with WTO regulations. Fortunately, there is clear precedent for its permissibility under WTO rules.⁷ Perhaps even more importantly, it is clear that the WTO *should* permit such a treaty upon its

⁶ In particular, a Republican president would almost certainly pull out of such a treaty if pulling out was much less costly to the US in the short run than staying in. (Although the US Constitution requires a two-thirds majority in the Senate and a Presidential signature for ratification, it allows a President to pull out of any treaty unilaterally without consulting either the Senate or the House of Representatives.)

⁷ For example, see the discussion of ‘border tax adjustments’ and other instruments in Joost H. B. Pauwelyn, “U.S. Federal Climate Policy and Competitiveness Concerns: The Limits and Options of International Trade Law”, available at <http://www.nicholas.duke.edu/institute/internationaltradelaw.pdf>, Frank Biermann and Rainer Brohm, “Implementing the Kyoto Protocol without the USA: the strategic role of energy tax adjustments at the border”, and Roland Ismer and Karsten Neuhoff, “Border tax adjustment: a feasible way to support stringent emission trading”. Precedent also establishes that initial non-signatories like the United States and China would not be permitted to levy reciprocal duties on signatory countries to cancel the effect of the signatories’ duties, because the duties imposed by signatories would have protected status as a result of their function to internalize externalities in a way that is permitted by existing WTO rules.

inevitable review, because if the Realist Assumption and the Current Generation Assumption above are true, then allowing such a treaty is the only way to correct the most catastrophic case of market failure in history⁸ – which means that allowing such a treaty is then the only way that the WTO can act in accord with its mission and treaty obligations of promoting ethical and economically efficient free trade.

In any event, the main point is not to defend the details of the particular treaty proposal described above, but rather to illustrate the general way in which a treaty can secure universal agreement and long-run compliance by incorporating the self-enforcing structure described above, which creates something like a tragic dilemma at the level of nations, the predictable outcome of which is universal ratification and adherence to the treaty over the long run, despite the fact that agreeing to the treaty makes some nations worse off than the no-treaty status quo. If the Realist Assumption and the Current Generation Assumption are roughly true, then a treaty with this structure is the only hope for an effective climate treaty, regardless of which of the existing architectures for a global response are chosen to fill in the details of the treaty. In this way, the current proposal solves the seemingly intractable problem of identifying a ‘meta-architecture’ that ensures cooperation and compliance to a chosen global response, while remaining relatively agnostic about the details and architecture of the global response that ought to be chosen.⁹

⁸ “Climate change is the greatest market failure the world has ever seen...”, *The Economics of Climate Change: The Stern Review*, pg. xviii.

⁹ For the distinction between ‘architecture’ and ‘meta-architecture’, see Jonathan Wiener “Incentives and meta-architecture”, in Joseph Aldy and Robert Stavins (eds.) *Architectures for Agreement*, pg. 67. For discussion of

The novel self-enforcing structure described above is also the key to constructing the most *ethical* climate treaty possible, because it is the key to avoiding the ethical and practical failings of existing approaches to climate treaties. For example, the *idealist approach* identifies the climate treaty that it would be best to propose with the treaty that would be best from an ethical perspective if adopted.¹⁰ The problem with this approach is that treaties that would be best from an ethical perspective if adopted seem to have no chance of being ratified and followed over the long run, given the actual dispositions of powerful nations. As a result, simple insistence on such treaties would ensure that the problem of climate change is never successfully addressed, which cannot be the best course of action even from an ethical perspective, because of the disastrous consequences that would ensue. As a result, the idealist approach fails because it mistakenly equates the question of what we should do in our actual, imperfect circumstances with the question of what it would be best to do if we found ourselves in more ideal circumstances.

alternative architectures for a global response, see *The Economics of Climate Change: The Stern Review*, Part IV and Part VI, the Intergovernmental Panel on Climate Change (IPCC) report *Climate Change 2007: Mitigation of Climate Change*, Chapter 13, Aldy and Stavins (eds.) *Post-Kyoto International Climate Policy*, and Aldy and Stavins (eds.) *Architectures for Agreement*.

¹⁰ For example, see the various “allocation methods” for national emission targets and emissions trading surveyed in the IPCC report *Climate Change 2007: Mitigation of Climate Change*, pg. 770; in addition, almost all work by ethicists related to global responses to climate change falls under the rubric of the idealist approach. A possible exception is Peter Singer, who explicitly notes the problem with the idealist approach in “One Atmosphere”, in Stephen Gardiner et. al. eds. *Climate Ethics*. However, because Singer’s proposed cap and trade scheme, which allocates permits on an equal per capita basis, would likely cost the United States more than one hundred billion dollars per year and would be prohibitively costly to other large-emitter nations as well (see Eric Posner and David Weisbach, *Climate Change Justice*, pp. 123 and 212-213), a simple-minded insistence on Singer’s proposal would be misguided for the same reason that idealist approaches are misguided in general: such an approach is not advisable even from an ethical perspective, because it has no chance of actual success, given the actual dispositions of powerful nations. The only promising way around such a problem would be to make an equal per capita allocation of permits part of a treaty that has the self-enforcing structure described above.

Another approach, the *hard-headed realist approach*, identifies the climate treaty that it would be best to propose with the treaty that would leave all nations at least as well off as the no-treaty status quo, and, subject to that constraint, would be best from an ethical perspective if adopted.¹¹ The problem with this approach is that it has unacceptable consequences of its own, because it implies that the treaty that it would be best to propose is a treaty on which poor nations that are vulnerable to climate change must make large transfer payments to the rich nations that are causing climate change in order to ensure that the rich nations are made no worse off by emissions reductions. Defenders of the hard-headed realist approach insist that this regrettable feature is necessary for an effective treaty, because they claim that there is no way that nations would agree to treaties that make themselves worse off than the no-treaty status quo. However, as the discussion above reveals, this idea is confused even from a rational choice perspective that takes the Realist Assumption for granted, because the introduction of a treaty can change what it is in the interest of nations to do, thereby making it the case that nations are better off by agreeing to the treaty than not agreeing, even if agreeing to the treaty makes them worse off than the no-treaty status quo.¹² As a result, the hard-headed realist approach fails because it implies that we should favor a treaty that is dramatically worse from

¹¹ For example, see Eric Posner and Cass Sunstein, "Climate Change Justice", especially pp. 1569-1570, Eric Posner and David Weisbach, *Climate Change Justice*, pp. 6, 86, and 143, Richard Stewart and Jonathan Wiener, *Reconstructing Climate Policy: Beyond Kyoto*, pp. 102-103, Jonathan Wiener, "Incentives and meta-architecture", in Aldy and Stavins (eds.) *Architectures for Agreement*, pp. 75-76.

¹² More generally, the reasoning of hard-headed realists overlooks the fact that acts of coercion and influence can change what is in the interest of nations to do relative to the status quo. For further discussion, see Thomas Schelling, "An Essay on Bargaining" in *The Strategy of Conflict*.

an ethical perspective than other treaties that would be equally effective, even given the Realist Assumption – namely, well-crafted treaties with the self-enforcing structure described above.

The other main approach, the *incrementalist approach*, does not identify the treaty that it would be best to propose with a treaty that would actually be effective, but rather with a treaty that would serve merely as a starting point for future agreements to reduce emissions, with the hope that nations would, over time, accept incremental additions to the initial treaty, thereby gradually deepening the emissions-reduction effect of the initial agreement.¹³ One problem with this approach is that it has already been tested over several decades, with little or no success. More importantly, the world now has only a few decades left to prevent the accumulation of potentially catastrophic levels of greenhouse gasses in the atmosphere, and it is implausible that the incrementalist approach could lead to sufficient emissions reductions within anything like that timeframe – in particular, it is implausible that such an approach could generate a reduction in global emissions of more than 50% from current levels, as is necessary merely to halt the accumulation of greenhouse gasses in the atmosphere.¹⁴ Perhaps most

¹³ For example, see Axel Michaelowa, “Graduation and Deepening”, and Jeffrey Frankel, “Formulas for quantitative emissions targets”, both in Aldy and Stavins (eds.) *Architectures for Agreement*, and Robert Keohane and David Victor, “The Regime Complex for Climate Change”, *Perspectives on Politics*.

¹⁴ From the authoritative Intergovernmental Panel on Climate Change (IPCC) report *Climate Change 2007: The Physical Science Basis*: “The concentration of greenhouse gas in the atmosphere depends on the competition between the rates of emission of the gas into the atmosphere and the rates of processes that remove it from the atmosphere. For example, carbon dioxide (CO₂) is exchanged between the atmosphere, the ocean, and the land through processes such as atmosphere-ocean gas transfer and chemical (e.g. weathering) and biological (e.g. photosynthesis) processes. While more than half of the CO₂ emitted is currently removed from the atmosphere within a century, some fraction (about 20%) of emitted CO₂ remains in the atmosphere for many millennia. Because of slow removal processes, atmospheric CO₂ will continue to increase in the long term even if its emission is substantially reduced from present levels. ... More specifically, the rate of emission of CO₂ currently greatly exceeds its rate of removal, and the slow and incomplete removal implies that small to moderate reductions in its

importantly, the incrementalist approach is by its nature piecemeal, and would therefore fail to create a harmonized global price for greenhouse gas emissions, which economists tend to agree is a necessary condition for any effective strategy, because a single global price for emissions is necessary not only to reduce emissions efficiently, but also to create sufficient incentives for the kind of large-scale global investments in low-carbon energy and green technology that are necessary to reduce global emissions by more than 50% from current levels over the coming decades.¹⁵ Because the incremental approach has no chance of creating such an unprecedented Archimedean lever of incentives, it has no real chance of being effective.

In light of the problems with these existing approaches to climate treaties, the possibility of a treaty with the self-enforcing structure described above holds the key to the most ethical climate treaty possible in our imperfect circumstances, and transforms the project of crafting such a treaty into a well-defined engineering problem that can be readily solved by experts on the relevant economic mechanisms and dispositions of nations. The primary constraint on such a treaty is that it must predictably lead to universal ratification and compliance in the long run via the cascade of economic incentives described above. Subject to

emissions would not result in stabilization of CO₂ concentrations, but rather would only reduce the rate of its growth in coming decades. A 10% reduction in CO₂ emissions would be expected to reduce the growth rate by 10%, while a 30% reduction in emissions would similarly reduce the growth rate of atmospheric CO₂ concentrations by 30%. A 50% reduction would stabilize atmospheric CO₂, but only for less than a decade. After that, atmospheric CO₂ would be expected to rise again as the land and ocean sinks decline owing to well-known chemical and biological adjustments” (pp. 824-825). For further ramifications for greenhouse gas stabilization levels, see *The Economics of Climate Change: The Stern Review*, Chapter 8; for a clear summary, see Hal Harvey and Sonia Aggarwal, “The Costs of Delay”, ClimateWorks Foundation.

¹⁵ *The Economics of Climate Change: The Stern Review*, Part IV and Chapter 22; William Nordhaus, *A Question of Balance*, pg. 29: “...placing a near-universal and harmonized price or tax on carbon is a necessary and perhaps even a sufficient condition for reducing the future threat of global warming”.

that constraint, the treaty should be designed to distribute the costs of emissions reductions as fairly as possible.

At first glance, it might seem that the devil is in the details regarding fairness. However, on further reflection, the details are relatively uncontroversial insofar as they are relevant to the engineering problem of creating the basic structural features of the treaty, because the basic structural features can only ensure fairness within a range, letting the ultimate details fall where they may in the inevitable round of political adjustment before the treaty is formally introduced for ratification – and it is easy to make accurate comparative judgments of the fairness of treaty proposals whenever one proposal is within the general range of fairness while the other falls outside that range. As a result, the most important consideration is to ensure that the basic structure of the treaty is crafted by experts with good ethical and practical judgment, and that such experts design the treaty to ensure fairness subject only to the constraint of feasibility. Fortunately, the self-enforcing structure described above is a significant advantage even here, because it ensures that an effective treaty can be developed by a team of independent experts motivated only by a concern for fairness and efficiency, with no need to accommodate myriad demands from national representatives and others in order to ensure universal participation and compliance in the long run once the treaty is introduced.

In sum, when the novel approach to solving collective action problems outlined in this paper is available, it allows us to bring about solutions to collective action problems that cannot be achieved by any other means. This is especially important when we face an intergenerational tragedy of the commons, or some other situation in which we have decisive

reason to do something that we are unable to do using any of the familiar approaches to collective action. In such cases, the approach outlined above is often the best approach possible in our imperfect world.¹⁶

Sidebar / Appendix:

Strategy for Introducing a Climate Treaty

Any climate treaty should be introduced under only the most favorable conditions that can be realistically expected, especially if the world has only one shot at a climate treaty that is costly enough to be effective. As a result, any climate treaty should be introduced only if the following two conditions are satisfied:

¹⁶ For example, on a relatively small scale, the collective action problems that stand in the way of a coordinated global aviation fuel tax and a coordinated global financial instruments trading tax are examples of otherwise intractable problems that are amenable to a similar treatment. On a much larger scale, runaway deficit spending by nations that benefits current citizens at an unacceptable cost to future generations and at an unacceptable risk to the current generation of other nations is an example of another serious intergenerational tragedy of the commons that is likely amenable to a similar treatment – although it is less clear how to work out the details of such a treatment in an optimal way. A promising starting point for thinking about the space of options is a duty levied on imports from nations that do not ratify a treaty that imposes a set of flexible but fiscally conservative restrictions on debt, where such duties are justified on the ‘anti-externality’ ground that the fiscal policy of non-signatories imposes an unacceptable risk of a serious or even catastrophic harm on other nations by imposing an unacceptable risk of sparking global financial panic and collapse.

(1) The treaty should be introduced early in the term of a US President who supports the treaty, enabling him or her to sign the treaty and publicly endorse its permissibility under all international laws and treaties, and to use his or her power and influence to ensure that international court decisions establish as *precedent* that the treaty is permissible. This will *institutionalize* the permissibility of the treaty in a way that cannot be reversed by anything short of the most dramatic (and hence unlikely) power politics. The US President can do all of this unilaterally even if the US Senate is initially disposed to reject the treaty unanimously, and even if powerful nations such as China oppose the treaty.

(2) Upon introduction, the treaty should be ratified by a substantial proportion of developed nations, as well as a substantial proportion of developing nations.

The possibility of a treaty with the self-enforcing structure described above is the key to satisfying both conditions (1) and (2), because the structure of such a treaty ensures that it would actually be effective, which means that nations will recognize that the costs of the treaty would be non-futile, including the political costs to leaders who support the treaty. This increases the likelihood that (1) will be satisfied relative to any particular US President, and the design of such a treaty ensures satisfaction of condition (2), partly for economic reasons, and partly because many nations are already disposed to ratify such a treaty out of a long-run

concern for their own citizens and the world, as demonstrated by the nations that ratified the Kyoto Protocol.¹⁷

Furthermore, if both (1) and (2) are satisfied, then other strategic challenges do not arise. For example, with respect to the detailed treaty proposal discussed above, on the realistic assumption that (1) is satisfied by a supportive US President and that (2) is satisfied by an initial coalition that includes most European nations, Japan, Russia, and other nations that ratified the Kyoto Protocol, it is unrealistic to think that China would have the power and influence to convince international courts to rule against both the US administration and that coalition regarding the permissibility of the treaty under WTO rules, and it is thus unrealistic to think that China could retaliate against signatory nations by imposing reciprocal duties on imports, because a favorable WTO ruling would mean the impermissibility of such reciprocal duties.

As a result, it is realistic to think that a treaty with the self-enforcing structure described above would succeed if introduced under favorable conditions, even given maximally pessimistic assumptions about opposition to the treaty from China and the US Senate.¹⁸

[END]

¹⁷ It is worth noting that even the flawed and ineffective Kyoto Protocol was ratified by an impressive set of nations, and was signed by US President Bill Clinton. As a result, it is realistic to think that a superior treaty with the self-enforcing structure described above would ensure satisfaction of (2), and that a US President might well be willing to expend the political capital necessary to satisfy condition (1) relative to such a treaty. (If that is not immediately clear, it might help to imagine a US President such as Al Gore.)

¹⁸ Thanks to Robert Keohane for encouraging me to discuss these issues.