Sustainable Humanity, Sustainable Nature: Our Responsibility
Pontifical Academy of Sciences, Extra Series 41, Vatican City 2014
Pontifical Academy of Social Sciences, Acta 19, Vatican City 2014
www.pas.va/content/dam/accademia/pdf/es41/es41-barrett.pdf

Why Have Climate Negotiations Proved So Disappointing?

Scott Barrett

I'm grateful to the organizers for proposing this question for my title, because it's important. People often complain that the climate negotiations have been disappointing, only to wring their hands and say that we must do better. But unless we know the *reasons* why the negotiations have been disappointing, we won't know *how* to do better. Using a medical metaphor, if our diagnosis of the illness is wrong, our recommended treatment is unlikely to heal the patient. Indeed, the wrong treatment may only make the patient sicker.

One of the striking things about the climate negotiations is that the negotiators have admitted that they have failed to meet their own goal.

In the Framework Convention on Climate Change, adopted in 1992, parties agreed that atmospheric concentrations of greenhouse gases should be stabilized "at a level that would prevent dangerous anthropogenic interference with the climate system". Later, in the non-binding Copenhagen Accord adopted in 2009, countries recognized "the scientific view that the increase in global temperature should be below 2 degrees Celsius". Finally, in Cancun in 2010, the parties to the Framework Convention reaffirmed this goal, but added that it may need to be strengthened, limiting temperature rise to 1.5°C.

After Copenhagen, countries submitted pledges for reducing their emissions. However, analysis by Rogelj *et al.* (2010) shows that even an optimistic reading of these pledges implies that mean global temperature will surpass the 2°C temperature change target.

The negotiators agree with this assessment. In Durban in 2011, they noted "*with grave concern* the significant gap between the aggregate effect of Parties' mitigation pledges in terms of global annual emissions of greenhouse gases by 2020 and aggregate emission pathways consistent with having a likely chance of holding the increase in global average temperature below" the agreed threshold. Rogelj *et al.* (2010: 1128) describe this behaviour as being "equivalent to racing towards a cliff and hoping to stop just before it".

Note that the Copenhagen pledges are voluntary. It's possible that they'll be exceeded. As bad as things look now, they could turn out to be worse.

The problem isn't disagreement about what should be done. Support for the 2°C goal is universal. The problem is that this is a global goal. Everyone is responsible for meeting it, meaning that no country is responsible for meeting it. Limiting climate change requires very broad cooperation. It requires *collective action*.

The reason collective action has eluded us so far is that reducing emissions is a prisoners' dilemma game. Each country is better off when *all* countries reduce their emissions substantially. But each country has only a small incentive to reduce its *own* emissions.

The Kyoto Protocol asks some countries to reduce their emissions beyond "business as usual", and so confronts the prisoners' dilemma head on. But Kyoto has failed in its mission to limit emissions.[1] The Copenhagen Accord, by focusing on the need to avoid "dangerous" climate change, has tried to reframe the problem. Will Copenhagen succeed? Reframing is the right strategy. But we need a different framing. Although my main purpose is to explain why the negotiations have been disappointing, I also want to suggest how an understanding of this failure can provide insights into how we might do better. Countries are currently trying to negotiate a new kind of climate agreement for adoption in 2015. My paper ends by suggesting how a different framing of the climate collective action problem could turn the negotiations around. The climate negotiations needn't be as disappointing as they have been so far.

The "dangerous" climate change game

Here is a way to think about the "dangerous" climate change game. Let's say that there exists a red line for "danger", and that countries know what this red line is. For example, it might be the 2°C goal. Let's also say that the impact of crossing this threshold is expected to be so severe relative to the costs of staying clear of it that all countries, collectively, prefer to stay clear of it. Then it's obvious what countries should do if they act collectively: they should limit concentrations of greenhouse gases to avoid crossing the red line.

The problem, of course, is that countries don't act collectively. They're sovereign. They act independently. So, we should ask: What *incentives* do countries have to stay within the good side of the red line?

Under reasonable assumptions, the game I have just described is not a prisoners' dilemma. It's a "coordination game" with two Nash equilibria (Barrett 2013).[2] In one, countries stay just within the "safe" zone. In the

[1] See, especially, research by Aichele and Felbermayr (2011), discussed later in this paper, which takes into account the effect of Kyoto on trade "leakage".

[2] In a coordination game, people want to do what others are doing. A car may be driven on the left or right side of the road. So, on which side of the road should you

other, they breeze past the tipping point, making catastrophe inevitable. In general, game theory has trouble predicting how countries will behave in a coordination game. However, since the "bad" Nash equilibrium is so obviously bad, staying within the red line is focal (Schelling 1960). Moreover, with the help of a treaty, countries can virtually guarantee that they will coordinate around the "good" Nash equilibrium.

Here is how the treaty should be written. It should assign to every country an emission limit, with each country's limit chosen to ensure that, when all the limits are added up, concentrations stay within the "safe" zone. The limits should also be chosen to ensure that every country is better off staying within its assigned limit, given that all the other countries stay within *their* assigned limits. Finally, the agreement should only enter into force if ratified by every country.

The beauty of such a treaty is that it makes every country pivotal. If every other country behaves as required, each country has an incentive to behave as required. The reason: even the slightest slip up guarantees catastrophe.

This treaty, like all good treaties, transforms the game. In the treaty participation game, every country has a dominant strategy to participate. Why? Every country has nothing to lose by joining. If the agreement fails to enter into force, each country would be free to act as it pleased. If the agreement were to enter into force, however, then it would be binding on all parties, and catastrophe would be avoided – the outcome every country prefers. Every country is thus always better off participating.

How would the emission limits for individual countries be chosen? This is the bargaining problem. If countries were "symmetric", bargaining would be simple. Technically, a wide variety of allocations would satisfy the requirements I described above, but an equal allocation would be focal, and for that reason is to be expected. Asymmetry makes bargaining more complex, but so long as collective action promises all countries an aggregate gain, there will exist an allocation of responsibility that will be acceptable to every country (this allocation won't be unique and may require side payments, but these are relatively minor points compared with the imperative to avoid catastrophe).

drive? The answer depends on where everyone else is driving. In Italy, it's obvious that you should drive on the right. In the UK, it's obvious that you should drive on the left. These different outcomes (driving on the left and driving on the right) are each a "Nash equilibrium". Given that others are driving on the left (right), each driver chooses to drive on the left (right).

The threat of catastrophe simplifies the negotiation problem. It makes each county's promise to stay within its agreed limits *credible*.

The vulnerability in this game isn't the behavior of the countries. It is the credibility of the *science* – in particular, the science of locating the critical tipping point.

In the dangerous climate change game, it would be irrational for any country to exceed its assigned amount of emissions when doing so would cause atmospheric concentrations to cross the catastrophic tipping point. In this case, free rider deterrence depends on the credibility of *Nature's* threat to tip a critical geophysical system. Yes, in the dangerous climate change game, Nature is an important player.

The importance of scientific uncertainty

What would countries need to do collectively to prevent dangerous climate change?

The Copenhagen Accord says:

> To achieve the ultimate objective of the Convention to stabilize greenhouse gas concentration in the atmosphere at a level that would prevent dangerous anthropogenic interference with the climate system, we shall, recognizing *the scientific view that the increase in global temperature should be below 2 degrees Celsius* [emphasis added], on the basis of equity and in the context of sustainable development, enhance our long-term cooperative action to combat climate change.[3]

Why defer to the "scientific view"? The reason is that it simplifies the negotiations. It allows negotiators to bargain over individual country shares.

However, while scientists warned of "climate disaster" before the Framework Convention was adopted (see, in particular, Mercer 1978), I think the Framework Convention caused scientists to focus on this question at least as much as previous scientific research caused negotiators to focus on it.

Reference to "*the* scientific view" [emphasis added] implies that there is strong agreement among scientists about the threshold. There isn't. The only "scientific view" that I detect in the literature is that thresholds are likely to exist.[4]

[3] The Copenhagen Accord was written somewhat hastily, and this temperature target is identified without reference to a base level of temperature. In Cancun the following year, negotiators clarified that the temperature reference target was the pre-industrial level.

[4] Rapid changes in temperature have been observed in the paleoclimatic record, an example being the Younger Dryas; see Broecker (1997).

Although temperature thresholds are uncertain, the uncertainties involved in choosing a target are even greater than this. The Framework Convention specifies the target in terms of concentrations, not temperature, and converting temperature to concentrations introduces an additional layer of uncertainty – something known as "climate sensitivity" (Roe and Baker 2007). Moreover, we don't know the quantity of global emissions (expressed, perhaps, as a cumulative sum) needed to meet any particular concentration target, due to uncertainty in the carbon cycle. For example, there is uncertainty about how much of the CO_2 emitted will be taken up by soils and the oceans. The fifth assessment report by the Intergovernmental Panel on Climate Change calculates a global "budget" in terms of cumulative emissions that will keep temperature change within 2°C, but all of these values are probabilistic. Even if we knew for certain that 2°C were the true red line for climate change (and we don't know this), countries would have to decide how to balance the cost of reducing emissions with reductions in the risk of crossing the red line.

The dangerous climate change game with uncertain thresholds

Uncertainty about the threshold for "danger" changes the climate change game fundamentally.

Consider a very simple game – no treaty. In stage 1, countries choose their emission levels independently. In stage 2, Nature chooses the tipping point. When making their choices in stage 1, the players know the probability density function for the tipping point. What they don't know is which value under this function will be chosen by Nature – the "true" value for the tipping point.

An example will make this clear. Rockström *et al.* (2009) argue that atmospheric CO_2 concentrations should be constrained "to *ensure* [emphasis added] the continued existence of the large polar ice sheets". They note that the paleoclimatic record implies "that there is a critical threshold between 350 and 550 ppmv", and interpret this as saying that if concentrations are limited to 350 ppmv, then the ice sheets will be preserved, whereas if concentrations rise to 550 ppmv, then the ice sheets will be lost. In between these values there is a chance that the ice sheets will disappear, with the probability increasing with the concentration level. (For reference, last month's reading from Mona Loa was about 399.65 ppmv; at the start of the industrial revolution, concentrations were about 280 ppmv; when the Framework Convention was adopted they were 356 ppmv).

Assume that the probability density function is uniform over the range (350, 550).[5] Assume also that the expected aggregate benefit of reducing the threat of catastrophe exceeds the cost. What should countries do? Under reasonable assumptions (Barrett 2013), the collective-best outcome is to limit concentrations to 350 ppmv. This implements the "precautionary principle". Acting independently, however, countries have incentives to reduce emissions only up to the point where their expected *individual* marginal benefit equals marginal cost – under reasonable assumptions, a substantially smaller value. Indeed, very simple calculations show that it will probably pay individual countries to abate so little that will they blow right past the critical threshold. They will cross the 550 ppmv line, guaranteeing catastrophe.

You might think this is just theory and that country representatives wouldn't be this dumb. I think there are good reasons to take the prediction seriously.

Astrid Dannenberg and I have tested these predictions in the experimental lab with real people playing for real money (Barrett and Dannenberg 2012). Putting people into groups of 10, we find that when the threshold is certain, 18 out of 20 groups avoid catastrophe.[6] By contrast, when the threshold is uncertain, catastrophe occurs with probability 100% for 16 out of 20 groups and with probability no less than 80% for the rest.[7]

How sensitive are these results? Intuitively, there should exist a critical amount of uncertainty such that if uncertainty were greater than this amount, catastrophe would be bound to occur, whereas if uncertainty were less than this, catastrophe would be avoided. This is exactly what the theory predicts (Barrett 2013), and in further experiments (Barrett and Dannenberg 2014a), Astrid Dannenberg and I have shown that this result is also robust. To the left of a critical "dividing line" for threshold uncertainty, we find that catastrophe is avoided with high probability almost all the time. Just to the right of the dividing line, by contrast, catastrophe occurs with proba-

[5] This means that the probability that the threshold lies between 350 and 400 ppmv is the same as the probability that the threshold lies between 400 and 450 ppmv, between 450 and 500 ppmv, and between 500 and 550 ppmv. It also means that the probability that the threshold lies below 350 or above 550 ppmv is zero.

[6] In each of the two failing groups, just one individual, a bad apple, caused the trouble, pledging to contribute his or her fair share and then choosing to contribute nothing.

[7] Interestingly, theory predicts that uncertainty about the *impact* of crossing a critical threshold should make no difference to collective action (Barrett 2013), another prediction confirmed in the experimental lab (Barrett and Dannenberg 2012). It is only uncertainty about the tipping point that matters, and this is a purely scientific matter.

bility 100%. This research suggests that negotiators can't rely on science to solve their collective action problem. Even the new science of "early warning signals" won't be able to shrink uncertainty by enough to transform the behavior of nation states.

This research is helpful, I think, because it is completely consistent with the behavior we are observing in the real world. As noted before, countries have agreed to limit temperature change to 2°C, but they have pledged emission reductions that virtually guarantee overshooting of this target.

If the science of climate change were much more certain, the prospect of catastrophe would give countries the discipline they needed to act in their collective-best interest. It would make the dangerous climate change game a coordination game. Scientific uncertainty makes the emission reductions game a prisoners' dilemma. The most important thing about a prisoners' dilemma is that the collective-best outcome cannot be sustained by non-cooperative behavior. It requires enforcement, something that the international system is very bad at doing.

A climate Doomsday Machine

It is tempting to consider an analogy to nuclear arms control. Herman Kahn (1961: 107) proposed construction of a Doomsday Machine,

> a device whose function is to destroy the world. This device is protected from enemy action (perhaps by being situated thousands of feet underground) and then connected to a computer, in turn connected to thousands of sensory devices all over the United States. The computer would be programmed so that if, say, five nuclear bombs exploded over the United States, the device would be triggered and the world destroyed. Barring such problems as coding errors (an important technical consideration), this machine would seem to be the 'ideal' [deterrent]. If Khrushchev ordered an attack, both Khrushchev and the Soviet population would be automatically and efficiently annihilated.

A Climate Doomsday Machine would connect all the world's nuclear bombs to a computer, which in turn would be linked to a sensor at the top of Mona Loa in Hawaii. This is where readings are taken of atmospheric concentrations of greenhouse gases. Today, as noted before, the concentration level is about 400 ppmv. The computer could be programmed to destroy the world should this level top, say, 500 ppmv. With the trigger for catastrophe being certain, theory and experimental evidence strongly suggest that this device would give the world all the encouragement needed to stay within 500.

Of course, I'm not seriously proposing this. The proposal is unacceptable.[8] However, the idea behind it is worth thinking about. The Doomsday Machine is a purely strategic device. Its sole purpose is to change the incentives countries have to rein in their emissions and save the world from dangerous climate change. It works by transforming the prisoners' dilemma into a coordination game. Thinking about it begs the question: Are there *acceptable* strategic approaches that could have a similar effect? I shall return to this point later in the paper.

Framing reconsidered

Has a focus on "dangerous" climate change really made no difference? Under certain conditions, theory suggests that uncertainty about the threshold could mean that behavior won't change at all, even though the consequences of failing to act will be much worse because of the threat of catastrophic climate change (Barrett 2012).

But is this result to be believed? Many predictions of analytical game theory are disproved in the experimental lab. For example, cooperation in a prisoners' dilemma typically exceeds the Nash equilibrium prediction (though the level of cooperation declines rapidly as the players learn that their efforts to cooperate are not reciprocated). In a one-shot test of the theory, Astrid Dannenberg and I found that cooperation was higher for the prisoners' dilemma with an uncertain threshold for "catastrophe" compared to a prisoners' dilemma without any risk of "catastrophe" (Barrett and Dannenberg 2014b). This suggests that, given the risk from "dangerous" climate change, the wording of Article II of the Framework Convention has probably helped (though it is the real risk rather than the wording that would affect behavior). Unfortunately, our experiment also showed that the additional cooperation wasn't enough to prevent catastrophe from occurring.

Strategies of reciprocity

My description of the dangerous climate change game left out the role of an international agreement. I explained before that a treaty could change the incentives in the game with a certain threshold, ensuring coordination. Could a treaty help overcome the incentive to free ride when the threshold

[8] Nor did Kahn recommend the Doomsday Machine: "If one were presenting a military briefing advocating some special weapon system as a deterrent … the Doomsday Machine might seem better than any alternative system; nevertheless, it is unacceptable". (Kahn 1961: 104–105).

is very uncertain? Theory suggests that an agreement would help very little (Barrett 2013). The reason is the difficulty of enforcing an agreement to limit emissions.

Atmospheric concentrations of CO_2 are determined by the aggregate behavior of *all* countries (as mediated by the carbon cycle). Strategies of reciprocity work very well in two-player games. They work less well when the number of players is large.

In the climate change game, how many players really matter? Some countries are bigger emitters than others. However, the top ten emitters account for only about two-thirds of total emissions, and stabilizing concentrations requires driving global net emissions to zero, necessitating the engagement of nearly *all* countries.[9]

The temptation to free ride is further aggravated by the high marginal cost of reducing emissions substantially. It is sometimes argued that reducing emissions is cheap. If this were true, however, collective action would be easy.

Another problem is the lack of correlation between a country's contribution to emissions and its vulnerability to climate change. To illustrate, William Nordhaus (2011) has calculated that the "social cost of carbon" is more than twice as large for Africa, a continent of more than 50 states, as it is for the United States. Moreover, this gap is growing. Yet, Africa's emissions are tiny when compared to those of the United States. Africa is both more vulnerable to climate change and less able to prevent it from occurring.

Finally, globalization amplifies the incentives to free ride. Abatement by a single country or coalition of countries will tend to shift emissions towards the countries that fail to act – a phenomenon known as "leakage".

It is well known that infinitely repeated play of the prisoners' dilemma can allow the full cooperative outcome to be sustained as a (subgame perfect) Nash equilibrium, provided discount rates are sufficiently low. The reason is that, should a country "cheat" on an agreement to limit emissions, the others can reciprocate. This suggests that cooperation should be easy.

The flaw in this perspective is that it considers only the interests of individual countries. It ignores these countries' collective interests.

Imagine that all the world's countries come together and negotiate an agreement that maximizes their collective interests. Later, one country announces that it will withdraw. This withdrawal would harm the other states, and they would like to punish this country (or, better yet, threaten to punish it, hoping to deter its withdrawal). In the context of a treaty, they would

[9] Unless, that is, substantial amounts of carbon are removed from the atmosphere.

naturally *cooperate* to punish the deviant country. To deter a deviation, their punishment must be big enough that the deviant would be better off remaining in the agreement than withdrawing and facing the punishment. But the punishment must also be credible. Given that this country has withdrawn, the remaining $N - 1$ countries must be better off when they impose the punishment than when they do not impose it (or when they impose a weaker punishment). Because so many countries remain in the agreement, it will only pay these countries to cut their abatement a little. A larger punishment wouldn't be credible. A small punishment, however, would be too little to deter a defection. Continuing in this way, it is easy to see that an agreement to limit emissions is only self-enforcing if the number of countries participating is very small. For then, should one country withdraw, the remaining countries would have an incentive to drop their abatement significantly. However, once participation shrinks to such a low level, the treaty achieves very little.

It may be possible for countries to sustain a high level of participation, but this would only be true if the gains to cooperation were small. It may also be possible to sustain a high level of participation if the ambition of the treaty were set very low. What isn't possible is for countries to sustain a high degree of cooperation when the gains from cooperation are very large.[10]

I have so far focused on participation. What about compliance? A flaw in the approaches to enforcement taken previously is that they either ignore participation (as in Chayes and Chayes 1995) or fail to distinguish between participation and compliance (as in Downs, Rocke, and Barsoom 1996). Under the rules of international law, countries are free to choose whether or not to participate in a treaty. However, the countries that choose to participate are legally obligated to comply with it (*pacta sunt servanda*, meaning "agreements must be kept"). The easiest way to avoid needing to comply is therefore not to participate in the first place – or to withdraw after becoming a party. From the perspective of game theory, the problem is coming up with a credible punishment that is large enough to deter non-participation. Once this is done, deterring non-compliance is easy. Remember, larger deviations can only be deterred by larger punishments, and larger punishments are less credible. What's the biggest harm a party could ever do? Behaving as it would were it not a party to the agreement (any bigger harm would not be credible). So, if the parties can deter non-participation, they can easily deter a smaller deviation of non-compliance. From both perspec-

[10] All of these points are developed in detail in Barrett (2003).

tives, then, deterring non-participation is the binding constraint on enforcement (Barrett 2003).

This is theory. Is the reasoning compelling? I know of no example from international cooperation that challenges this perspective. The World Trade Organization might appear to be an exception, but trade isn't a global public good. Trade is a bilateral activity, and strategies of reciprocity are very effective in sustaining cooperation amongst pairs of players. The Montreal Protocol on protecting the ozone layer might appear to be another exception, but as I shall explain later, this treaty works very differently.

Enforcement of the Kyoto Protocol

The Kyoto Protocol looks at climate change as a prisoners' dilemma, demands that certain countries cooperate, and then does nothing about enforcement.

How do we know this? The United States participated in the Kyoto negotiations. President Clinton signed the treaty. However, the United States never ratified Kyoto. One reason for this is that there were no consequences to the United States for not ratifying the agreement. Non-participation by the United States was not deterred.

Canada ratified Kyoto, but failed to adopt the domestic legislation needed to implement its obligations. As a consequence, Canada's emissions exceeded the limit set by Kyoto limit. Once in this situation, Canada had three options. It could buy permits or offsets to stay in compliance; it could stay in the agreement and be in non-compliance; or it could withdraw from the agreement. In contrast to the first option, withdrawal would be costless. In contrast to the second option, withdrawal would not violate international law. Not surprisingly, Canada decided to withdraw. The Kyoto Protocol could not deter Canada from withdrawing.

Compliance with the Kyoto Protocol by other parties is uneven (Haita 2012). However, there are ways to get around compliance. For example, Japan is maneuvering to achieve compliance partly by purchasing "assigned amount units" from Ukraine, when Ukraine's emissions are well below its "assigned amounts". In other words, in buying these units from Ukraine, Japan can comply without emissions being reduced anywhere. This may seem crazy but the treaty was written to allow this trading in "hot air".

Finally, countries like China and India are not subject to limits on their emissions. They participate in Kyoto. They comply with it. But Kyoto does not require that these countries do anything.

Overall, did Kyoto contribute to meeting the objectives of the Framework Convention? Did it reduce global emissions? Econometric analysis of the Kyoto Protocol by Rahel Aichele and Gabriel Felbermayr (2012: 351) shows that Kyoto did reduce the emissions of participating countries. However, its effect on *global* carbon emissions "has been statistically indistinguishable from zero".

The most important indicator of whether the objectives of the Framework Convention are being met is whether the growth in atmospheric concentrations is slowing. It isn't. If anything, the rate of increase has gone up.[11] We are no closer now to addressing this great problem than we were more than twenty years ago when the Framework Convention was adopted.

Pivot: From Kyoto to Paris

The Copenhagen talks were supposed to provide a successor agreement to the Kyoto Protocol. They failed.

A new agreement is now being negotiated under the "Durban Platform". It is supposed to be ready for adoption by 2015, when the parties to the Framework Convention meet in Paris. It is supposed to be implemented by 2020.

Since Kyoto's emission limits ended in 2012, this leaves a gap of eight years. To fill the gap, Kyoto was given an extension in the form of the Doha Amendment (which has yet to enter into force). However, it is a further sign of Kyoto's failings that Japan, New Zealand, and Russia declared their intention not to participate this time around.[12]

The new agreement being negotiated now won't repeat all of Kyoto's mistakes, but there is no indication yet that it will improve much on what countries would have done in the absence of cooperation. Kyoto referred to its emission limits as "commitments". However, countries were never truly committed to meeting these limits; they couldn't be committed to meeting these limits so long as Kyoto lacked the means to compel parties to do more than they were willing to do unilaterally. It is a sign of where the current round of negotiations are going that in Durban countries agreed to negotiate a new "protocol, legal instrument or agreed outcome with legal force", and that in the recent Warsaw talks they agreed to negotiate "contributions" rather than "commitments".

The world clearly needs a new model for cooperation on climate change.

[11] The data can be found at http://www.esrl.noaa.gov/gmd/ccgg/trends/

[12] The European Union will participate, because this agreement only requires that Europe meet the target it declared it would meet unilaterally. The new government in Australia has introduced legislation to repeal the previous government's climate legislation, a sign that Australia may not ratify the Doha Amendment.

Why the Montreal Protocol succeeded

Kyoto lacks a strategic design. The emissions targets and timetables were chosen in the expectation that they would be met. No consideration was given to whether the treaty created *incentives* for them to be met.

The Montreal Protocol, negotiated to protect the stratospheric ozone layer, was designed very differently. Remarkably, while the Montreal Protocol was not intended to reduce greenhouse gases, it has been much more successful at doing this than the Kyoto Protocol. It turns out that ozone in the stratosphere is a greenhouse gas (protecting the ozone layer will thus add to climate change), as are the chemicals that deplete stratospheric ozone (reducing these emissions will thus help mitigate climate change) and many of their substitutes (use of these will thus add to climate change). Calculations by Velders *et al*. (2007) show that, by phasing out the ozone-depleting substances that double as greenhouse gases, the Montreal Protocol has done four times as much to limit atmospheric concentrations as the Kyoto Protocol aimed to do.

Why did Montreal succeed where Kyoto failed? A key reason for its success is the threat to restrict trade – in particular, a ban on trade in controlled substances between parties and non-parties (Barrett 2003). The most important motive for the trade restriction was to enforce participation in the agreement (Benedick 1998: 91). If participation could be enforced, then trade leakage would be eliminated; moreover, compliance could also be enforced (Barrett 2003). Crucially, the trade restrictions in the Montreal Protocol are a strategic device. Their purpose was not to be used; their purpose was to change behavior.

How do the trade restrictions work? Imagine that very few countries are parties to an agreement to limit emissions, and your country is contemplating whether or not to join. If you join, you will have to reduce your emissions. Your country will pay the cost, and the benefits will be diffused; the incentives to free ride will not be blunted. If, in addition, you are now also prohibited from trading with non-parties – the vast majority of countries – than you will be doubly harmed. By joining, you not only forfeit the benefits of free riding; you also lose the gains from trade.

Now imagine that almost every other country is a party to the same agreement. If you join, you still lose the benefits of free riding. But now you are able to trade with the vast majority of countries. If the gains from trade exceed the loss from free riding, your country will be better off joining. Put differently, if every country is a party to such an agreement, none will wish to withdraw. The agreement will sustain full participation by means of a self-enforcing mechanism – the trade restriction. Most remarkably, in equilibrium, trade will never be restricted, just as the Doomsday Device would never be detonated. It is the credible threat to restrict trade (detonate the device) that disciplines behavior.

Notice that a key feature of this strategy is that "enough" countries participate in the agreement. If too few participate, none will want to participate. If enough participate, everyone will want to participate. Somewhere in between there exists a "tipping point" for participation. Once this tipping point has been identified, the treaty only needs to coordinate participation – something treaties can do very easily. As I said before, treaties need to ensure that countries are steered towards the desired outcome. In the Montreal Protocol, this was achieved by the minimum participation clause (Barrett 2003).

The trade restrictions are an acceptable alternative to the Doomsday Machine. Like the Doomsday Machine, trade restrictions transform the prisoners' dilemma into a coordination game.

Coordination in climate treaties

The Montreal Protocol could be amended to achieve more for the climate. Hydrofluorocarbons (HFCs) do not deplete the ozone layer, and so are not currently regulated by the Montreal Protocol. However, HFCs are a very potent greenhouse gas – one of the six gases controlled by the Kyoto Protocol. Kyoto has done very little to limit HFCs. In May 2011, the United States, Canada, and Mexico proposed amending the Montreal Protocol to control HFCs. If adopted, such an amendment would represent a significant departure from the approach taken so far to address climate change. It would mean addressing one piece of the problem, rather than all of it in a comprehensive way. And it would likely involve using trade restrictions for purposes of enforcement. I want to underline that the application of trade restrictions is purely strategic. Other proposals for trade restrictions in climate policy are very different; their purpose is to be *used*, not to alter behavior strategically (Barrett 2011).

There are other opportunities for transforming the climate change game from a prisoners' dilemma into a coordination game, especially the application of technology standards (Barrett 2003, 2006).

Let me give one example. Another greenhouse gas known as perfluorocarbons or PFCs is emitted in the process of manufacturing aluminum. Apparently, these emissions can be eliminated if the anodes being used now are replaced with inert anodes. According to the United States Environmental Protection Agency's webpage, "This technology is being pursued aggressively through a joint R&D program that has been established between the aluminum industry and the U.S. Department of Energy in its

Industrial Technology Program".[13] The new anode is expected to be available in 10-15 years.

Here, then, is a suggestion. A new agreement should be negotiated requiring that producers adopt the new technology, and that all parties agree to import aluminum only from countries that participate in the agreement. This approach creates a "tipping" phenomenon. Provided enough countries join the agreement, all will want to join it. Why? To be outside the agreement when most countries are inside means not being able to trade in aluminum with most of the world.

This proposal for aluminum would be different from Montreal. The trade restriction would be based on a process standard, not a product standard. However, I believe it would still be effective. I also believe it would be compatible with the WTO, partly because of the exemptions allowed under Article XX, but also because it would be adopted by a multilateral agreement.

A final point. Another reason Montreal works is that it includes side payments to address related equity issues. Side payments could also be included in an agreement establishing a aluminum production standard. As in Montreal, any side payments should be based on the "incremental costs" of adopting the standard. Transfers should be small, and the countries giving the money should know what they are getting for their money.

Again, this is just one example of how the negotiations could be made more effective.[14] My aim here is not to develop a comprehensive approach to future climate negotiations but to suggest a new direction. If the negotiators understood their job as needing to achieve coordination, and to think strategically, they would achieve more – and the climate negotiations wouldn't prove so disappointing.

References

Aichele, Rahel and Gabriel Felbermayr (2012). "Kyoto and the Carbon Footprint of Nations", *Journal of Environmental Economics and Management* 63: 336-354.

Barrett, Scott (2003), *Environment and Statecraft: The Strategy of Environmental Treaty-Making*, Oxford: Oxford University Press.

Barrett, S. (2006). "Climate Treaties and 'Breakthrough' Technologies". *American Economic Review* (*Papers and Proceedings*) 96(2): 22-25.

Barrett, Scott (2011). "Rethinking Climate Change Governance and Its Relationship to the World Trading System", *The World Economy* 34(11): 1863-1882.

Barrett, Scott (2013). "Climate Treaties and

[13] http://www.epa.gov/aluminum-pfc/resources.html

[14] For other examples of sectoral approaches to reducing emissions, see Barrett (2011).

Approaching Catastrophes", *Journal of Environmental Economics and Management*, 66(2): 235-250.

Barrett, Scott and Astrid Dannenberg (2012). "Climate Negotiations Under Scientific Uncertainty", *Proceedings of the National Academy of Sciences*, 109(43): 17372-17376.

Barrett, Scott and Astrid Dannenberg (2014a), "Sensitivity of Collective Action to Uncertainty about Climate Tipping Points", *Nature Climate Change* 4: 36-39.

Barrett, Scott and Astrid Dannenberg (2014b), "Negotiating to Avoid 'Gradual' versus 'Dangerous' Climate Change: An Experimental Test of Two Prisoners' Dilemmas", in Todd L. Cherry, Jon Hovi, and Dave McEvoy (eds.), *Toward a New Climate Agreement: Conflict, Resolution, and Governance*, London: Routledge.

Benedick, R.E. (1998). *Ozone Diplomacy: New Directions in Safeguarding the Planet*, Enlarged Edition, Cambridge, MA: Harvard University Press.

Broecker, W.S. (1997). "Thermohaline Circulation, the Achilles Heel of Our Climate System: Will Man-Made CO_2 Upset the Current Balance?" *Science* 278:1582-1588.

Chayes, A. and A.H. Chayes (1995), *The New Sovereignty*. Cambridge, MA: Harvard University Press.

Downs, G.W, D.M. Rocke, and P.N. Barsoon (1996), "Is the Good News About Compliance Good News About Cooperation?" *International Organization* 50: 379-406.

Environmental Protection Agency (2006). *Global Mitigation of Non-CO_2 Greenhouse Gases*, Washington, DC: EPA.

Haita, Corina (2012). "The State of Compliance in the Kyoto Protocol", International Center for Climate Governance; http://www.iccgov.org/FilePagineStatiche/Files/Publications/Reflections/12_Reflection_December_2012.pdf

Kahn, Herman (1961), "The Arms Race and Some of Its Hazards", in Donald G. Brennan (ed.), *Arms Control, Disarmament, and National Security*, New York: George Braziller, pp. 89-121.

Mercer, J.H. (1978). "West Antarctic Ice Sheet and CO2 Greenhouse Effect: A Threat of Disaster", *Nature* 271: 321-325.

Nordhaus, William (2011). "Estimates of the Social Cost of Carbon: Background and Results from the RICE-2011 Model". Cowles Foundation Discussion Paper No. 1826, Yale University.

Rockström, J., W. Steffen, K. Noone, Å. Persson, S. Chapin III, E.F. Lambin, T.M. Lenton, M. Scheffer, C. Folke, H.J. Schellnhuber, B. Nykvist, C.A. de Wit, T. Hughes, S. van der Leeuw, H. Rodhe, S. Sörlin, P.K. Snyder, R. Costanza, U. Svedin, M. Falkenmark, L. Karlberg, R.W. Corell, V.J. Fabry, J. Hansen, B. Walker, D. Liverman, K. Richardson, P. Crutzen, and J.A. Foley (2009). "A Safe Operating Safe for Humanity", *Nature* 461: 472-475.

Roe, G.H. and M.B. Baker (2007). "Why is Climate Sensitivity So Unpredictable?" *Science* 318: 629-632.

Rogelj, J., J. Nabel, C. Chen, W. Hare, K. Markmann, M. Meinshausen, M. Schaeffer, K. Macey, and N. Höhne (2010). "Copenhagen Accord Pledges are Paltry", *Nature* 464: 1126-1128.

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

Velders, G.J.M., S.O. Anderson, J.S. Daniel, D.W. Fahey, and M. McFarland (2007), 'The Importance of the Montreal Protocol in Protecting Climate', *Proceedings of the National Academy of Sciences* 104 (12): 4814-4819.