Climate Treaties, Lock-In, and Increasing Returns[+]

Scott Barrett[*]

Columbia University and London School of Economics

This version
29 January 2025

Abstract. I present an approach to limiting climate change that differs from the main UN negotiations by focusing on sectors rather than economy-wide measures, on technology-fuel switching rather than emission reductions, and on global systemic change rather than national action. Though lock-in makes even small reductions in emissions costly and difficult for a treaty to enforce, increasing returns makes switching a tipping game with enforcement "built in." Switching treaties need only specify the new technology-fuel combination; incorporate trade measures and/or R&D funding to bolster increasing returns, where necessary; and provide the "big push" needed to assure selection of the superior equilibrium.

Key words. Climate treaties, technology switching, lock-in, increasing returns, tipping games, trade linkage, strategic R&D

JEL Codes. C72, F18, F53, F64, H87, O33, Q54

[+] Earlier versions of this paper were presented at the University of Durham, Universität Graz, Harvard, INRAE in Paris, the LSE, Oxford, the Université Paris Dauphine, Stanford, and the Abdus Salam International Centre for Theoretical Physics in Trieste. I am grateful to Jonas Oliver Piduhn for research assistance, and Hunt Allcott, Partha Dasgupta, Larry Goulder, Geoffrey Heal, Michael Oppenheimer, and Juan-Pablo Montero, and Eugene Tan for comments on an earlier draft. I received no funding for this work.

[*] Lenfest-Earth Institute Professor of Natural Resource Economics, Columbia University, and Centennial (visiting) Professor, London School of Economics; sb3116@columbia.edu.

1. Introduction

Carbon dioxide emissions are a byproduct of processes people employ to supply services like heating, lighting, and motive power. In every case, these processes combine a fuel with a technology: coal and the steam engine; gasoline and the internal combustion engine; electricity and electric lighting; coke and blast furnaces for steel-making; and so on. Often, these technology-fuel pairings are rooted in broader systems. For example, automobile transport requires not only cars and gasoline but roads and refueling stations. Most of the technology-fuel systems in place now were developed without a thought being given to their consequences for climate change. To reduce emissions, they need to change; but how?

To climate negotiators, the answer has always been obvious: just ask countries to reduce their emissions. Implicit in this "direct" approach is the belief that, once countries are compelled to reduce emissions, technologies and fuels will adjust efficiently. For the most part, economists have accepted this logic. However, though the direct approach is efficient when the emission reduction goal is modest, it may not be when the goal is radical change.[1]

Today's technology-fuel systems are entrenched. They embody a long history of investment, innovation, learning-by-doing, economies of scale, and, in some cases, network effects. Reinforcing these advantages, firms with a history of innovating in "dirty" technologies tend to continue to innovate in these technologies rather than in "clean" alternatives (Aghion et al

[1] The carbon pricing arrangements adopted thus far have had modest effects. British Columbia's celebrated carbon tax reduced emissions in transportation, but not aggregate emissions (Pretis 2022). Carbon taxes adopted by 15 Europe countries from as early as 1990 reduced cumulative emissions (mainly in sectors not covered by the Emissions Trading System) by only about five percent (Metcalf and Stock 2023). The European Union's ETS reduced emissions by less than four percent over the period 2008-2016 (Bayer and Aklin 2020) and by about 15 percent in manufacturing over the years 2005-2012 (Colmer et al. 2024). The Kyoto Protocol is estimated to have reduced emissions in Annex I countries by seven percent, but taking into account trade leakage, Aichele and Felbermayr (2012: 336) find that the treaty "has had at best no effect on world-wide emissions." Colmer et al. (2024) find no evidence of leakage, but it should be noted that trade-sensitive sectors have been protected under the ETS, either by being excluded or given free allocations of allowances. Most importantly of all, global emissions are expected to reach a record high in 2024.

2016).[2] There may even be political resistance to change.[3] All of these features give rise to, or are expressions of, "lock-in," conditions that make it difficult for a new technology-fuel-system to overturn the status quo (Unruh 2000; Jaffe et al. 2005; Seto et al. 2016; Fouquet 2016). Under conditions of lock-in, the policies economists normally recommend will reduce emissions, but they may fail to replace the existing technology-fuel system with a new one. That is, they may fail to achieve "transformational change."[4]

Another difficulty with the usual remedy is that, being a global public good, reductions in greenhouse gas emissions are vulnerable to free riding. The climate treaties negotiated so far specify targets and timetables (of varying degrees of specificity) for limiting each party's economy-wide emissions, a direct approach that succeeds only if the limits can be enforced. The Kyoto Protocol collapsed for lack of enforcement. The Paris Agreement avoids this fate by relying on voluntarism, but voluntary action is unlikely to support substantial emission reductions (Barrett and Dannenberg 2016). The record of these agreements speaks for itself: beginning with the Framework Convention in 1992, all the world's countries have declared that atmospheric concentrations of greenhouse gases must be stabilized, and yet, in the years since then, concentrations have steadily increased. Indeed, emissions are still increasing. Perhaps worst of all, a fossil-fuel-dependent capital infrastructure continues to accrete, making the reduction in emissions in the future even more difficult (see, for example, Hawkins-Pierot and Wagner 2022).

With "dangerous" climate change looming, negotiators are now pledging to achieve "net zero" emissions by a future date like 2050, a goal that can only be met by "transformational change." The contrast with the Kyoto Protocol is striking. Kyoto aimed to reduce the emissions of a subset of countries by just five percent, and still it failed. Because nothing was achieved before, more

[2] Arthur (2009) ascribes a psychological explanation for this behavior. "The old principle," he says, "lives on because practitioners are not comfortable with the vision—and promise—of the new" (Arthur 2009: 139).

[3] Meng and Rode (2019) find that, dollar for dollar, lobbying by firms that expect to lose from ambitious climate legislation is more effective than lobbying by firms that expect to gain from it.

[4] van der Ploeg and Venables (2022) make a similar point. Their analysis incorporates peer effects among households in addition to technological externalities, but ignores free riding. In their model, tipping is achieved by a temporary tax on a dirty good (in addition to a Pigouvian tax) sufficient to shift an economy to the "clean" equilibrium.

must be achieved now. But how can we achieve more now when we couldn't achieve even modest emission reductions before?

One suggestion is to augment the approach tried before by linking cooperation on trade to cooperation in reducing emissions. William Nordhaus (2015) has shown that an agreement to impose a carbon tax can be enforced by a "club" of countries that makes their cooperation on trade contingent on others' cooperation on climate change. However, his calculations show that this only works for a carbon tax less than $\$50/tCO_2$, a value below even today's social cost of carbon and certainly too low to bring about zero emissions (Geoffrey Heal 2022). Moreover, by assuming that countries outside the club won't retaliate, Nordhaus likely overstates the set of circumstances that allow linkage to be effective (Barrett and Astrid Dannenberg 2022). In a follow-up paper, Nordhaus (2021) shows that, with enough technical progress, trade linkage can achieve zero emissions. However, technical change in his model is exogenous. What will bring about the "rapid technological change" needed to make linkage effective?

In this paper I construct a framework for the design of a system of treaties aimed at "getting to zero" emissions in particular sectors. Rather than limit climate change directly, I propose addressing it strategically by changing the technology-fuel system that produces emissions as a waste product, supported, where appropriate, by trade and R&D policies targeted at particular technologies. The direct approach, I show, forces countries to play a prisoners' dilemma. By contrast, if increasing returns are sufficiently strong, technology switching is a coordination or tipping game. The advantage of the strategic approach is that it is able to sustain the cooperative outcome of the switching game as a Nash equilibrium. Enforcement is "built in." A treaty is needed only to coordinate switching.

I begin by modeling the classic emission reduction game, a prisoners' dilemma, and the technology-fuel switching game, a coordination or tipping game if increasing returns are sufficiently strong. In section 4 I describe conditions that allow switching by one or a few first movers to trigger a cascade of switching by other countries, and in section 5 I show that, where

this kind of "spontaneous" tipping is unreliable, a treaty can assure selection of the superior equilibrium by pushing participation past the tipping point. In sections 6-8 I show that, by linking cooperation on trade to switching, countries can amplify the returns to switching, but that a weaker form of linkage is preferable if linkage is to be attempted unilaterally rather than multilaterally. In section 9 I show that strategic R&D has a similar effect as trade linkage. Echoing Nordhaus (2021), in section 10 I show how both policies combined can be more powerful still. I conclude by discussing the implications for climate negotiations.

2. The emission reductions game

Assume a world of n symmetric countries, each of which must choose between two different technologies. The installed technology is A. The alternative technology is B. Both technologies are equally good at providing a valued service, such as heating, cooling, or transportation, but use of A results in greenhouse gas emissions and B is a low-emissions technology. In this section I consider emission reductions when countries use A, the *status quo*. In the next section I consider the decision to switch to B.

Using A, country i's emissions for a given sector are

$$E_i = E_0 e^{-\tau_i/\beta}, \tag{1}$$

where τ_i can be interpreted as i's carbon tax (equivalently, its emissions trading equilibrium price), and E_0 i's emissions in the absence of abatement. Setting $E_0 = 1$ and rewriting gives

$$\tau_i = -\beta \ln(E_i). \tag{2}$$

Facing this tax, A-users will reduce their emissions to the level at which $\tau_i = MC_i$, where MC_i stands for marginal abatement cost. Substituting and then differentiating gives $dMC_i/dE_i = -\beta/E_i$. Marginal cost is thus decreasing in emissions (increasing in abatement) for

$E_i \in (0, 1]$. Note that, as my model is static, it's best to think of MC_i as long run marginal cost, incorporating technical change induced by the policy. Calel and Dechezleprêtre (2016), for example, estimate that the European Union Emissions Trading System increased "low-carbon" innovation among regulated firms by up to 10 percent. Later I shall examine the role of policies that promote R&D directly and in a targeted way.

The total costs of limiting emissions to $\hat{E}_i$ are

$$TC_i(\hat{E}_i) = -\beta \int_{\hat{E}_i}^{1} \ln E_i \, dE_i \,. \tag{3}$$

Integrating by parts and evaluating the integral gives

$$TC_i(\hat{E}_i) = \beta\left[1 - \hat{E}_i(1 - \ln\hat{E}_i)\right]. \tag{4}$$

Obviously, $TC_i(1) = 0$. Applying L'Hôpital's rule, $\lim_{\hat{E}_i \to 0} \hat{E}_i \ln\hat{E}_i = 0$, and so $\lim_{\hat{E}_i \to 0} TC(\hat{E}_i) = \beta$. The parameter β thus represents the cost of eliminating country i's emissions using technology A.

Denote country i's social cost of carbon by γ. By the assumption of symmetry, the global social cost of carbon is, thus, γn. Country i's payoff is

$$\pi_i^A(E_i; E_{-i}) = \gamma[(1 - E_i) + (n - 1 - E_{-i})] - \beta[1 - E_i(1 - \ln E_i)], \tag{5}$$

where $E_{-i} = \sum_{j \neq i}^{n} E_j$. It is easy to confirm that $\pi_i^A(1; n-1) = 0$ and $\lim_{E_i \to 0} \pi_i^A(E_i; 0) = \gamma n - \beta$. Maximizing gives the symmetric Nash equilibrium, $E_{NE}^A = e^{-\gamma/\beta}$. Upon substituting,

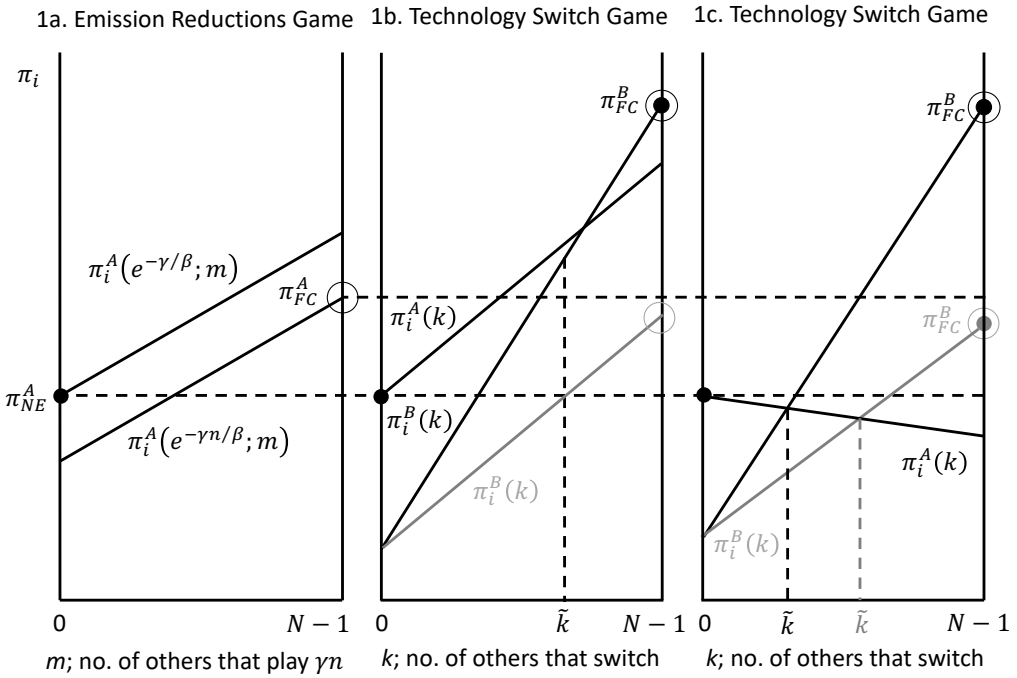
$$\pi_{NE}^A = \gamma n\left(1 - e^{-\gamma/\beta}\right) - \beta\left[1 - e^{-\gamma/\beta}(1 + \gamma/\beta)\right]. \tag{6a}$$

By cooperating fully, countries get

$$\pi_{FC}^{A} = \gamma n\left(1 - e^{-\gamma n/\beta}\right) - \beta\left[1 - e^{-\gamma n/\beta}(1 + \gamma n/\beta)\right]. \quad (6b)$$

Economists normally recommend that countries impose a carbon tax equal to γn—or, equivalently, that they limit their emissions to $E_{FC}^{A} = e^{-\gamma n/\beta}$. This approach, however, runs headlong into the prisoners' dilemma. Figure 1a, which is inspired by Schelling (1978), shows the payoffs to every country i as determined both by whether i emits E_{FC}^{A} or E_{NE}^{A} and by the number of other countries that emit E_{FC}^{A} rather than E_{NE}^{A}, represented by $m; m \in \{0,1,\ldots,n-1\}$. As shown in the figure, in the A game, free riding is a dominant strategy.[5]

Figure 1. Emission Reduction and Technology Switching Games



3. The switching game

[5] It is easy to confirm that this is true for $\gamma > 0$ and $n > 1$.

Another approach to reducing emissions is to orchestrate a switch from A to a new technology-fuel combination. There may be many alternatives to A. It may not be obvious which is best.[6] Here, for simplicity, I assume that there is just one alternative, technology B. Lock-in confers an advantage to A. However, there may be increasing returns to adopting B. In addition, B is a low-emission technology. For simplicity, I assume that B is a zero-emission technology. To bring about a switch, B's advantages need to be leveraged.

Numerous possibilities for switching have emerged in recent years. A simple switch would replace sulfur hexafluoride circuit breakers used in today's medium-voltage electrical switchgear with newly developed vacuum breakers that employ a mixture of nitrogen and oxygen (of these gases, only sulfur hexafluoride is a greenhouse gas). Similarly, carbon anodes (made from a mix of petroleum coke and coal tar) used in aluminum smelting today could be replaced with inert anodes, eliminating the emissions of both CO_2 and perfluorocarbons, another powerful greenhouse gas. Switching in these two cases is easy as the substitute technologies can be "dropped in." In other sectors, switching requires bigger changes. For example, in transportation, gasoline-powered automobiles could be replaced with battery-electric cars, diesel trucks with hydrogen-powered fuel cells, and heavy fuel oil container ships with ammonia-powered engines, but all of these changes require a new supporting infrastructure: for cars and small trucks, a network of recharging stations, supplied with zero-emission electricity; for long-haul trucks, a network of hydrogen refueling stations offering emission-free, "green" hydrogen; and, for shipping, new engines and fueling stations supplying "green" ammonia.

[6] For example, will the car of the future be all-electric, or will it run on hydrogen or synthetic fuels? If the answer is unclear, competition among the alternatives may mean that none displaces the *status quo* even if welfare would be improved were *any* of them to displace the status quo (Postrel 1990). Even in a centralized setting, if technologies are improving over time and their future prospects are uncertain, an inferior technology may be chosen (Cowan 1991). Jaffe, Newell, and Stavins (2005: 171) argue that, "to avoid accidentally helping to entrench the wrong technology, it is desirable for policy to be 'technology neutral,' encouraging all efforts that achieve specified objectives without focusing on a particular approach." I would only add the qualification that one of these objectives should be the need to transform the prisoners' dilemma into a tipping game.

Continue to assume that all countries are symmetric. Let k denote the number of other countries, from any country's perspective, that switch from A to B ($n - k$ is thus the number of other countries that don't switch). By switching i gets

$$\pi_i^B = \gamma(n-k-1)\left(1-e^{-\gamma/\beta}\right) + \gamma(k+1) - C\,(n-\alpha^B k)/n\,; \qquad (7a)$$

and by not switching i gets

$$\pi_i^A = \gamma(n-k)\left(1-e^{-\gamma/\beta}\right) + \gamma k - \beta\left[1-e^{-\gamma/\beta}\left(1-\ln e^{-\gamma/\beta}\right)\right](n+\alpha^A k)/n, \qquad (7b)$$

where the parameter C denotes the cost to i of replacing A with B if no other countries switch, and α^A and α^B reflect the returns globally to adoption of A and B, respectively. Initially, $k = 0$, making (7b) identical to (6a).

$\alpha^B > 0$ implies global increasing returns, $\alpha^B = 0$ constant returns, and $\alpha^B < 0$ decreasing returns. With constant returns, switching, like emission reductions, will be a prisoners' dilemma. There will be no strategic advantage to switching as opposed to imposing a carbon tax. However, as the slope of the payoff curves is higher for switching than for emission reductions (switching, by assumption, results in zero emissions), and switching entails a fixed cost (and, thus, increasing returns domestically, even with $\alpha^B = 0$), it is possible that the full cooperative outcome will still favor switching over imposing a tax equal to the global social cost of carbon.

With global increasing returns, the cost of switching by every country falls as more others switch.[7] The natural interpretation is that this reduction in cost is driven by economies of scale and learning (as the model lacks a time dimension, learning is captured by k, rather than cumulative production over time). Henceforth, I assume $n/(n-1) > \alpha^B > 0$. Adoption of B is costly but exhibits increasing returns.

[7] Heal (1994) was the first researcher to examine cooperation among countries due to interdependence in the costs, and not only the benefits, of reducing emissions.

Technology A may also exhibit increasing returns. However, because A is entrenched, α^A will likely be smaller than α^B. For example, the cost of deploying A today reflects a long history of innovation and learning by doing, whereas for B savings from such processes have yet to realized.

We see now that an implicit assumption in the previous section is that the tax $\tau = \gamma/\beta$ is too low to induce a switch to B. Indeed, to be sure we get the prisoners' dilemma in Figure 1a, I assume that even the tax level $\tau = \gamma n/\beta$ is too low to induce a switch.

The deterrent to switching by first-movers is the cost, C, which can be high. For example, the Norwegian Parliament has decreed that, by 2025, all new cars sold in the country must be zero emission. To promote the switch to electric cars, Norway has offered various sweeteners, estimated to cost about €1,370/tCO_2 (Fridstrøm 2021), many times most estimates of the *global* social cost of carbon, let alone the national one. According to a recent analysis (Allcott et al.: 2024), electric vehicle tax credits under the United States Inflation Reduction Act yield a return of $1.02 per dollar of public funds assuming the domestic social cost of carbon and $1.20 per dollar assuming the global social cost of carbon, which is more encouraging. However, the IRA does not have an end-goal. Indeed, at year-end 2024, electric vehicles comprised less than ten percent of new car sales in the US (in Norway, nearly 90 percent).[8] Also, the "industrial policy" requirements for vehicle assembly and (the sourcing of materials for and the manufacture of) batteries limit take up of the program, and could spur retaliation by other countries.

It is also important to remember that my analysis is sector-based, not economy-wide. Though many countries have adopted "carbon pricing," these policies typically exclude or impose weaker obligations on trade-sensitive industries.

[8] The subsidy induces additional sales in electric vehicles, but a higher subsidy, accompanied by an extensive recharging infrastructure, would be required to induce yet higher sales. Finally, the goal of zero emissions requires changing the generation of electricity. (Norway has abundant hydropower.)

Note that country i's cost of switching, $C\ (n - \alpha^B k)/n$, can be interpreted as depending either on the number of countries that switched prior to i's decision, as assumed by Arthur (1989), or the number of switchers at a given time. Under the first interpretation, i's switching costs are unaffected by whether others switch subsequently, as would be the case if switching entailed only an up-front cost that, once incurred, was sunk. In this case, the only strategic incentive to switch is to impel others to switch subsequently, thereby increasing supply of the global public good (a benefit that does not appear in Arthur's model). Under the second interpretation, switching entails a recurring cost, providing an additional incentive to switch—early switchers cause more others to switch, which, in turn, reduces the early movers' future costs of using A.[9]

From eqs. (7a) and (7b), i will switch iff $k > \tilde{k}$, where $\tilde{k}$ is the solution to $\pi_i^A(k) = \pi_i^B(k)$:

$$\tilde{k} = \frac{C - \beta\left(1 - e^{-\gamma/\beta}\right)}{C\alpha^B/n + \beta[1 - e^{-\gamma/\beta}(1 + \gamma/\beta)]\,\alpha^A/n}. \tag{8}$$

The numerator of (8) is i's net cost to switching if no others have switched (k = 0). The denominator is the reduction in this net cost for every increase in k. Country i should switch iff the reduction in cost due to increasing returns exceeds the net cost of switching. For example, if the RHS of (8) were equal to $0.5n$ it would pay i to switch iff at least a majority of countries switched (or were expected to switch).

Figure 1b shows the game of getting countries to switch, assuming that α^A is near zero. Two payoff curves are drawn for B, one black and one grey. If the grey curve applies, increasing returns are weak and switching is a prisoners' dilemma. If the black curve applies, increasing returns are strong and switching is a tipping game. In this case, universal switching sustains full cooperation as a Nash equilibrium. Notice, however, that, starting from the *status quo* in which all countries play A, switching by any country is risky. If few others switch, switchers could lose big-time.

[9] For example, more battery-electric vehicles would result in more recharging stations, lowering the cost to owning an EV. For US cities, Li et al. (2017) find that a one percent increase in the installed base of EVs increases investment in recharging stations by 0.6 percent.

Moreover, because the tipping point lies to the right of the mid-point of the horizontal axis, $(n-1)/2$, sticking with A is risk-dominant. For this reason, it can't be assumed that countries, acting independently, will risk going for the payoff-dominant equilibrium in which all switch. Relative to the efficient outcome, there may be "excess inertia" (Farrell and Klemperer 2007).

Figure 1c shows another interesting situation. Here, α^A is relatively large, causing the payoff to sticking with A to decrease as more countries switch. If the black payoff curve for switching applies, the tipping point is relatively low, making switching both payoff- and risk-dominant. If the grey payoff line applies, the switching equilibrium is welfare superior to the Nash equilibrium in the emission-reduction game, but the full cooperative payoff in the emission-reduction game is higher still ($\pi_{FC}^B < \pi_{FC}^A$). Switching is better, but only if countries see the strategic advantage in switching. In the experimental lab, Astrid Dannenberg and I have found that, under such conditions, players often fail to switch even though they would very likely have been better off if they had switched (Barrett and Dannenberg 2017).

Summarizing:

Proposition 1 (emission reductions versus switching). *If there are constant returns to switching globally, the switching game, like the emission reductions game, is a prisoners' dilemma. If increasing returns are sufficiently strong, switching is a tipping game. If, in addition, the payoff dominant Nash equilibrium of the tipping game is strictly preferred to the Nash equilibrium in the emission reduction game, countries will do better by switching, so long as they are sure to land on this superior equilibrium. Moreover, this is true even if imposition of a carbon tax equal to the global social cost of carbon supports an outcome that is mutually preferred to universal switching.*

4. Switching cascades

If countries are symmetric, universal switching is the unique Nash equilibrium if $\tilde{k} < 1$. If $\tilde{k} > 1$, universal switching requires coordination. As a general matter, universal switching becomes less of a sure thing the higher is $\tilde{k}$.

Asymmetries create more interesting possibilities. Suppose that countries are asymmetric in a very precise way. In particular, suppose that country i's social cost of carbon is given by

$$\gamma_i = \gamma i \; \forall i \in \{1, \ldots n\}. \tag{9}$$

In other words, countries are identified by an index that reflects their national social cost of carbon, with country 1 having the smallest value, country 2 the second smallest, and so on (for estimates, see Nordhaus 2017). By (9), the global social cost of carbon is now $\sum_{i=1}^{n} \gamma_i = \gamma n(n+1)/2$, where $\gamma = \gamma_1$. Assuming that countries remain symmetric as regards costs, country i has an incentive to switch *unilaterally* if $\pi_i^B(0) > \pi_i^A(0)$, or

$$\gamma_i > -\beta \cdot ln\left(1 - \frac{C}{\beta}\right). \tag{10}$$

Recall that A, the installed technology, is poorly suited to achieving zero emissions, whereas B is especially attractive for reaching this goal. As β is the cost of eliminating emissions (in the limit) using A, it seems reasonable to assume $\beta > C$, making the right-hand side of (10) positive.

We know from (9) that if *any* country has an incentive to switch unilaterally, it will pay country n to switch unilaterally. Once any country or group of countries switches, the incentive for others to switch increases, provided there are increasing returns. Country i will switch if $\pi_i^B(n-i) > \pi_i^A(n-i)$. Upon substituting, this condition can be written as:

$$\begin{gathered} \gamma_i e^{-\gamma_i/\beta} > C - \beta\left[1 - e^{-\gamma_i/\beta}(1 + \gamma_i/\beta)\right] \\ -\left\{C\alpha^B\,(n-i)/n + \beta\left[1 - e^{-\gamma_i/\beta}(1 + \gamma_i/\beta)\right]\alpha^A\,(n-i)/n\right\}. \end{gathered} \tag{11}$$

The term on the top line of (11) is identical to (10). This is the condition that makes i want to switch unilaterally. The term on the bottom line is the effect of switching by $n - i$ others on i's incentive to switch, reflected in a reduction in the cost of switching.

It is easy to see how switching can be a Nash equilibrium. At a minimum we require that (10) hold for $i = n$ and that (11) hold for $i = 1, \ldots, n-1$. Dixit (2003) shows how switching (to use my terminology) can be the unique Nash equilibrium of either the simultaneous-move game or a sequential-move game. In the latter case, we can think of switching as triggering a cascade, with early switchers causing more countries to switch, with these switchers causing yet more to switch, and so on, until, finally, all countries have switched.[10] "Events" that could start a cascade include a discovery that γ is larger, or C smaller, than previously expected.

Proposition 2 (switching cascades): *With increasing returns to switching globally, asymmetries among countries can make universal switching the unique Nash equilibrium even if the vast majority of countries have no incentive to switch unilaterally.*

A striking example is the global switch from leaded to unleaded petrol combined with a catalytic converter. Beginning in the 1920s, lead was added to gasoline to raise octane and thus improve engine performance. Japan began regulating the lead content of petrol in the early 1970s with the aim of limiting public exposure to this toxic metal. Soon afterwards, the United States regulated lead in gasoline for a different reason. The US required that new cars be fitted with a catalytic converter to reduce emissions of carbon monoxide, hydrocarbons, and nitrogen oxides, and only required that new cars be run on unleaded gasoline because lead destroys the catalysts used in these devices. Though Japan and the US initially regulated lead for different reasons, over

[10] As also shown by Dixit (2003), the equilibrium in which every country adopts B may make some countries worse off relative to the *status quo ante* in which all adopt A. Indeed, a universal switch to B, supported as a Nash equilibrium, can lower the collective payoff of all n countries, even as it raises the payoff of some countries. In such a case, a treaty could potentially preserve the higher aggregate payoff if the countries that prefer A to B cooperated to pay the countries that prefer B to A not to switch. See also Heal and Kunreuther (2010) for an analysis of tipping sets in coordination games, inclusive of cascades. Heal and Kunreuther (2017) apply this logic to climate change.

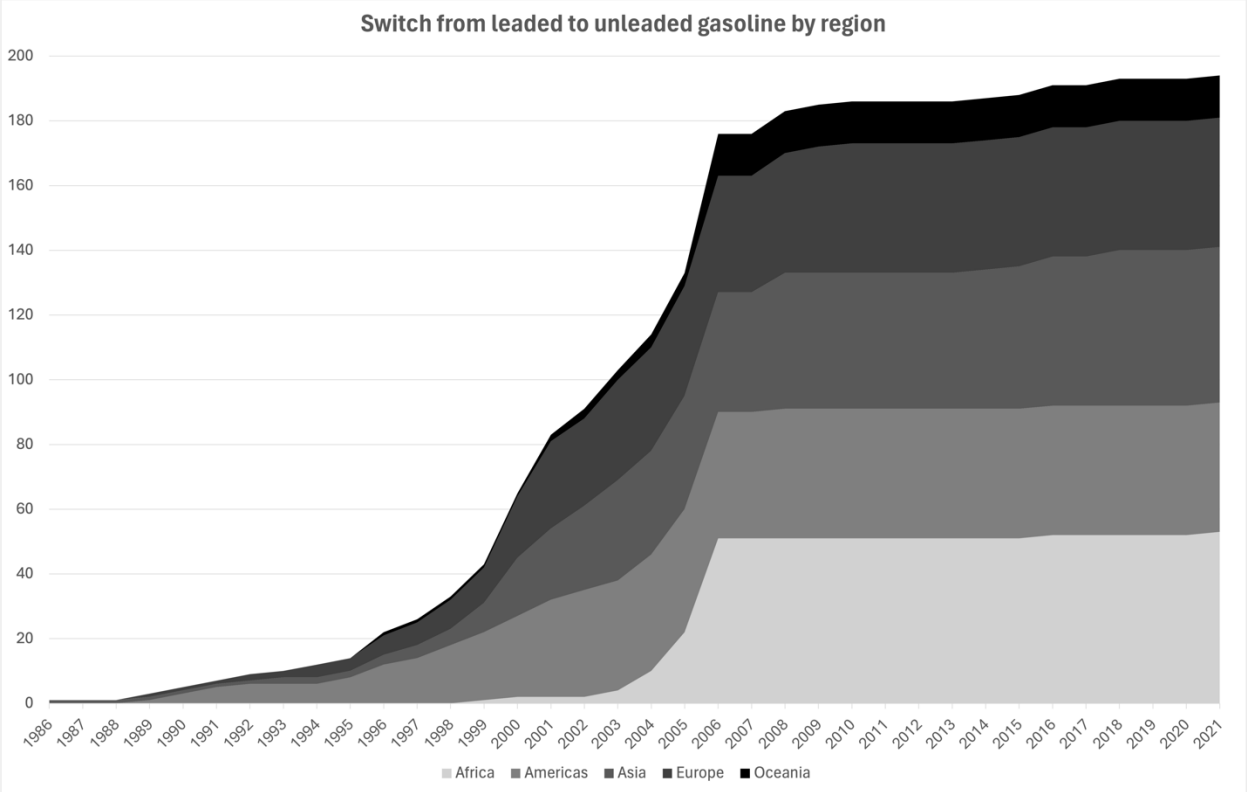
time their policies converged. Japan, a major exporter of automobiles to the US, adopted the US auto emissions standards, and the US, concerned about the public health consequences of continued use of leaded gasoline by older cars, regulated the lead content of gasoline directly.

Concurrent with these developments, other countries adopted their own regulations. Eventually, increasing returns set in, reinforcing the public health motivations for switching. Early movers paid the fixed costs of learning how to substitute for lead, how to modify engines to accommodate both the reformulated gasoline and the catalytic converter, and how to improve the converter itself. Though substituting for lead is costly (about one cent per liter), due to economies of scale in petroleum refining, spot prices of unleaded gasoline fell below those of leaded gasoline in some markets (Lovei 1996: 12). Gasoline exporters like Sudan switched when export markets for leaded petrol dried up. Gasoline importers switched when foreign suppliers stopped making leaded petrol. Due to economies of scale in automobile manufacturing, cars *not* fitted with catalytic converters became more expensive than the new kind (Lovei 1996: 29). Network effects were also important. For example, in Hungary, gasoline stations began selling unleaded petrol before being made to do so by law, "to supply the increasing number of Western European tourists visiting the country by cars that required unleaded fuel" (Lovei 1996: 30).

Start-to-finish, the process of banning lead in gasoline worldwide took 35 years (50 years, starting from Japan's first move to limit lead content). Japan got the ball rolling, banning lead in gasoline in 1986. Algeria came in last, switching to unleaded only in 2021; see Figure 2. The global switch was achieved without the help of a treaty.[11] A treaty wasn't needed, because the public policy aim of the ban was local environmental protection, not supply of a global public good. However, the forces powering this cascade are the same as would help supply a global public good: increasing returns (Heal 1994; Heal and Kunreuther 2017).

[11] Common standards can be a boon for trade; Schmidt and Steingress (2022) estimate that harmonized standards have increased global trade by as much as 13 percent. It is for this reason that two international agreements promote harmonization in cars, one adopted in 1958 and the other in 1998. However, vehicle emission standards still vary somewhat. For example, the EU sets different emission standards for diesel and petrol vehicles, and the United States does not. And, yet, many countries adopt the same standards as their main trading partners, just as many manufacturers adopt designs that meet the most stringent standards in the markets they supply.

Figure 2. The Global Switch to Unleaded Petrol



Source: Online appendix (to be prepared).

Another cascade may be underway now. The European Union phased out energy-inefficient incandescent bulbs in 2012 and halogen bulbs in 2018, and the United States is phasing out both kinds of bulbs in 2023. Compact fluorescent bulbs, though more energy efficient than incandescent and halogen bulbs, contain mercury, and are to be phased out by 2025 under a 2022 amendment to the Minamata Convention on Mercury, which currently has 141 parties, including major economies like Brazil, China, the European Union, India, Japan, and the United States. The effect of all these moves is to create a new standard: the light-emitting diode (LED) lightbulb. LEDs are much more energy efficient than ones being phased out. They also have a longer life and, being a solid-state technology, can function in new ways. Indeed, the economic

and technical advantages of solid-state lighting are so attractive that even the major producers of traditional lighting have made the transition to LEDs (Walsh Sanderson and Simons 2014).[12]

Might a cascade also drive a switch to electric vehicles? In just the last couple of years, countries like Canada, China, the European Union, Iceland, India, Japan, Singapore, South Korea, and the United Kingdom (in addition to states like California) have proposed or adopted laws either banning the sale of new cars powered by fossil fuels or requiring that new vehicles be zero-emission by some date, usually 2035. It's possible that increasing returns alone will bring about a cascade of switching worldwide, but as the impetus for banning fossil fuel vehicles is mainly to supply the global public good of limiting climate change, a treaty may be needed to effectuate a global switch.

5. Switching treaties

How could a treaty bring about a global switch? Consider the following game:

Stage 1: All countries decide whether to join a switching treaty.

Stage 2: Parties to the treaty decide collectively whether to switch.

Stage 3: Non-parties decide independently whether to switch.

In this model, parties have a slight edge over non-parties. In Stage 2, parties commit to collective action, whereas in Stage 3 non-parties act independently, taking the choices of parties as given. Cooperation in a treaty is a deliberate act, and the principle of customary international law known as *pacta sunt servanda* ("agreements must be kept") requires that countries take their

[12] The new bulbs won't only replace less efficient bulbs; they will also find new uses. Eventually, they will be integrated into the fabric of technology more generally.

treaty obligations seriously. However, the "commitment" in this model is modest, as international law also allows states to join or withdraw from a treaty as they please (Stage 1).[13]

The game is solved backwards. Assume again that countries are symmetric. For the situation depicted in Fig. 1b, non-parties switch for sure in Stage 3 if and only if $k \geq \tilde{k}$, where $\tilde{k}$ is the tipping point for switching, given by (8). Knowing this, parties switch in Stage 2 iff $k \geq \tilde{k}$. Finally, in Stage 1, every country cannot do better except by joining. The self-enforcing treaty sustains full cooperation by requiring that parties switch, with the treaty entering into force if and only if signed (and ratified) by at least $\hat{k}$ countries, where $\hat{k}$ represents the smallest integer greater than or equal to $\tilde{k}$.

It may make sense to require as well that parties be no worse off collectively for forming the treaty than for not forming it. Let k^+ be the solution to $\pi^B(k) = \pi^A_{NE}$:

$$k^+ = \frac{C - \beta\left(1 - e^{-\gamma/\beta}\right) + \gamma e^{-\gamma/\beta}}{C\,\alpha^B/n + \gamma e^{-\gamma/\beta}}. \qquad (12)$$

If we adopt this second condition, the minimum participation level for the treaty must be at least as large as max $\{\hat{k}, \check{k}\}$, where $\check{k}$ denotes the smallest integer greater than or equal to k^+. For the examples shown in Figure 1c, where $\pi^B(\tilde{k}) < \pi^A_{NE}$, choice of $\check{k}$ for the minimum participation level might seem more compelling than $\hat{k}$.

Increasing returns are critical for switching to be a tipping game. Recall that, with constant returns to the adoption of both technologies ($\alpha^A = \alpha^B = 0$), switching is a prisoners' dilemma. A treaty may still help, but it probably won't be able to sustain the full cooperative outcome. To see this, note that, with constant returns, parties to the treaty are unable to influence how non-

[13] Canada became a party to the Kyoto Protocol but withdrew once it became clear that it would not comply with the agreement. If Canada had remained in the treaty, its non-compliance would have violated international law. By leaving, Canada avoided its legal responsibility to comply.

parties choose in Stage 3. In Stage 2, switching pays for each party if and only if at least $\check{k}$ countries participate in the treaty. In Stage 1, each country has nothing to lose by signing the treaty if fewer than $\check{k} - 1$ others have signed it, each gains by signing if exactly $\check{k} - 1$ others have signed it, and each loses by signing if $\check{k}$ or more others have signed it. In equilibrium, therefore, precisely $\check{k}$ countries sign the agreement, causing it to enter into force. The $\check{k}$ signatories switch to B; the $n - \check{k}$ non-signatories stick with A. This is a game of chicken. Signatories gain a little compared to the non-cooperative outcome; non-signatories free ride and gain more—a standard result in the literature on international environmental agreements.[14]

Proposition 3 (switching treaties): *If there are constant or weakly increasing returns, a treaty can sustain switching by only a subset of countries. If increasing returns are sufficiently strong, making switching a tipping game, a tipping treaty can assure selection of the payoff dominant Nash equilibrium, resulting in universal switching.*

6. Switching with trade linkage

Suppose increasing returns on their own are insufficient to make switching a tipping game.[15] Can linking trade cooperation to switching help?

Trade is bilateral. In its trade relations with country j, i chooses a policy, t_i^j, taking j's choice, t_j^i, as given. Assume that this choice is binary; every country i chooses $t_i^j \in \{0, t\}, t > 0$, where t can

[14] For surveys of this literature, see Finus (2001), Wagner (2001), Barrett (2003), and de Zeeuw (2015). Note that, if increasing returns are uncertain, it is also possible that, in equilibrium, only a fraction of countries will switch. In the online appendix I show that, if switching is a tipping game in expectation, one group of countries will move first, and the others "wait" until information about the increasing returns parameter has been revealed. If increasing returns turn out to be at least as strong as expected *ex ante*, the others will follow, and switching will be universal. However, if increasing returns turn out to be weaker than expected *ex ante*, the other countries may not follow. In this case, if switching entails a sunk cost, the first movers will be disappointed that their gamble didn't pay off, and yet they may still be better off than if they hadn't switched in the first place because they switched *en masse*, supplying a significant amount of the global public good.

[15] The arguments developed here also apply if it is uncertain *ex ante* whether switching is a prisoners' dilemma or a coordination game.

be interpreted either as a tariff or, since a very high tariff chokes off imports, as a ban on imports from the relevant sector (a ban is the more natural interpretation for technology switching).

In their trade relations, countries face a terms of trade prisoners' dilemma (WTO 2009; Bagwell and Staiger 2011): taking the trade policies of other countries as given, each has an incentive to put a tariff on imports, so as to shift the terms of trade in its favor; and, yet, if all countries do this, all will be worse off than if they cooperated and maintained free trade. Rather than model trade explicitly, let parameters represent the gains from trade. Assume $\pi_i^j(t;0) = a, \pi_i^j(0;0) = b, \pi_i^j(t;t) = c,$ and $\pi_i^j(0;t) = d$, with $a > b > c > d > 0$. Under the rules of the World Trade Organization, all WTO members refrain from imposing trade measures against other members, an outcome that is enforced by the threat of retaliation (reciprocity). Such an outcome can be shown to be an equilibrium of a simple game of trade cooperation (Barrett and Dannenberg 2022). Here I take the free trade equilibrium under an international trade agreement as given, and examine the incentive a subset of countries may have to make trade cooperation conditional on supply of the global public good. Nordhaus (2015) calls coalitions of this kind "climate clubs."

Let k denote the number of other countries, from any country's perspective, that both switch to B *and* restrict trade with non-switchers ($n - k$ is thus the number of countries that stick with A and retaliate against the switchers who restrict trade). To which group should country i belong? Letting $\tilde{\tilde{k}}$ denote the solution to $\pi_i^B(k) = \pi_i^A(k)$, it is rational for i to join the switchers/linkers iff $k > \tilde{\tilde{k}}$, where (see Appendix)

$$\tilde{\tilde{k}} = \frac{C - \beta\left(1 - e^{-\gamma/\beta}\right) + (n-1)(b-c)}{C\alpha^B/n + \beta[1 - e^{-\gamma/\beta}(1 + \gamma/\beta)]\,\alpha^A/n + 2(b-c)}. \qquad (13)$$

Linkage helps because it exhibits increasing returns. As more countries link, the incentive for the others to link increases. A limiting case is where there are constant returns to scale. Then (14) becomes

$$\lim_{\alpha^A,\alpha^B\to 0}\tilde{\tilde{k}} = \frac{C-\beta\left(1-e^{-\gamma/\beta}\right)}{2(b-c)} + \frac{(n-1)}{2}, \tag{14}$$

which is equivalent to the tipping point shown Barrett and Dannenberg (2022, eq. (1)). Eq. (14) shows that, depending on parameter values, linkage on its own may suffice to transform the prisoners' dilemma in switching into a tipping game.[16] However, the combination of increasing returns to switching and trade linkage will generally succeed better.

If there are increasing returns, trade linkage makes switching less attractive when relatively few countries switch, and more attractive when relatively many countries switch. From (8) and (13),

$$\tilde{\tilde{k}} \begin{matrix} > \\ = \\ < \end{matrix} \frac{(n-1)}{2} \; if \; \tilde{k} \begin{matrix} > \\ = \\ < \end{matrix} \frac{(n-1)}{2}, \tag{15a}$$

$$sgn\frac{\partial\tilde{\tilde{k}}}{\partial(b-c)} \begin{matrix} > \\ = \\ < \end{matrix} 0 \; if \; \tilde{k} \begin{matrix} < \\ = \\ > \end{matrix} \frac{(n-1)}{2}. \tag{15b}$$

From eq. (15a), if the tipping point without linkage is greater (less) than the mid-point of the Schelling figure, $(n-1)/2$, then the tipping point *with* linkage will also be greater (less) than the mid-point.[17] From eq. (15b), if, without linkage, the tipping point lies to the left (right) of the mid-point, an increase in the gains from trade cooperation will increase (decrease) the tipping point.

[16] Having a coordination game requires $\tilde{\tilde{k}} < (n-1)$. Substituting gives $C-\beta\left(1-e^{-\gamma/\beta}\right) < (n-1)(b-c)$. In words: when all other countries switch/link, the "last" country is better off switching/linking provided the gains to enjoying free trade relations with all the others exceeds the gains to free riding.

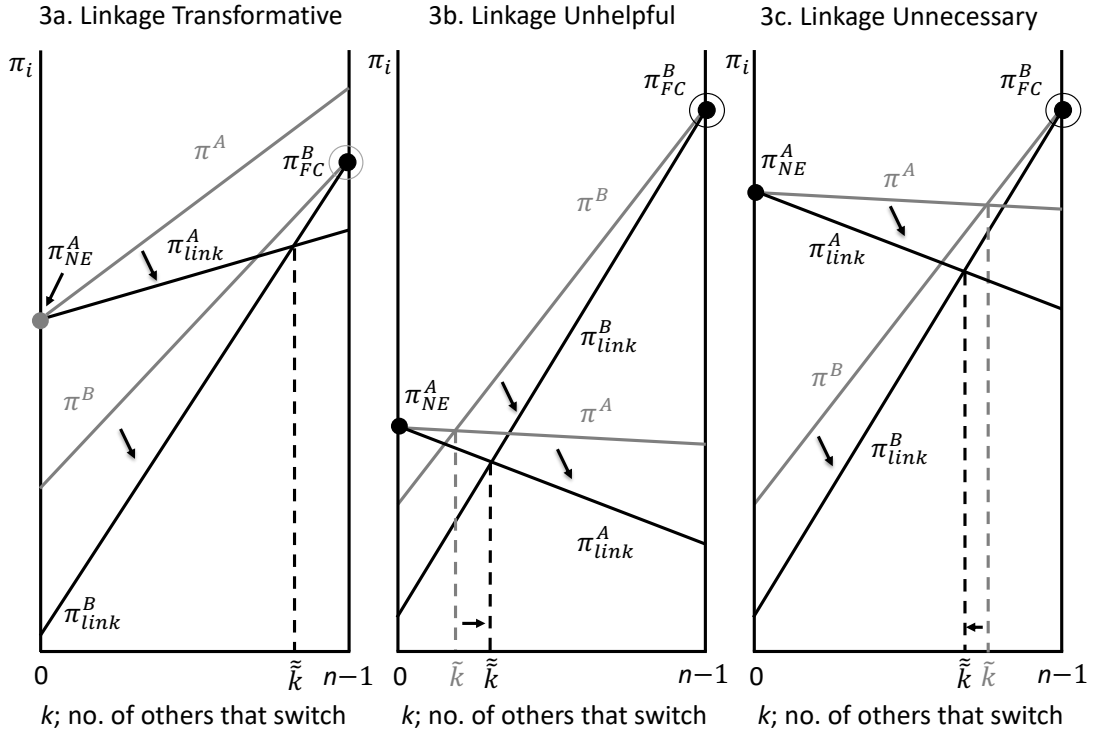
[17] To see why, suppose $\tilde{k} = (n-1)/2$. Then country i's decision of whether or not to switch will be unaffected by linkage. For if i switches, it will have free trade relations with half of the other countries (the switchers) and be in trade-conflict with the other half (the non-switchers); and if i doesn't switch, it will face the same situation, only with the identities of these two camps being reversed. Hence, $\tilde{k} = (n-1)/2 \iff \tilde{\tilde{k}} = (n-1)/2$. With this as a reference point, we can see that, to the left of the mid-point, linkage is more harmful to switchers than non-switchers. In this domain, switching only pays if *more* of the others switch relative to a regime without linkage. To the right of the mid-point, the opposite is true; linkage is more harmful to non-switchers than switchers. In this domain, switching pays even if *fewer* of the others switch relative to the regime without linkage.

This is because, with linkage, to the left of the mid-point, an increase in $(b-c)$ raises the relative cost of switching, whereas to the right it lowers the relative cost of switching. In short:

Lemma 1. *Relative to a regime of no linkage, linkage moves the tipping point closer to the mid-point,* $(n-1)/2$.

Figures 3a-c illustrate three interesting situations. In all of these figures, the grey payoff lines apply to a regime without linkage, and the black to a regime with linkage. The figures assume that payoffs in the no-link regime are inclusive of the gains from trade cooperation, $b(n-1)$. With linkage, if $k = 0$, a country pays a potentially huge penalty by switching. Relative to the no-link regime, linkage lowers its payoff by $(b-c)(n-1)$. If $k = n-1$, non-switchers pay the same penalty.

Figure 3. Switching with Trade Linkage



In Fig. 3a, without linkage, countries won't switch, even though they'd be better off collectively if they all switched (switching is a prisoners' dilemma), whereas with linkage, the game is transformed, making switching a tipping game. Importantly, for linkage to be transformational, we must have $\tilde{\tilde{k}} < n-1 < \tilde{k}$. That is, switching without linkage must be a prisoners' dilemma. Moreover, under these circumstances (see below), $\tilde{\tilde{k}} > (n-1)/2$. That is, the payoff and risk dominant equilibria differ, making coordination without the support of a treaty unreliable.

Lemma 2. *If switching without linking is a prisoners' dilemma (that is, if $\tilde{k} > n-1$), then, at best, switching with linking is a coordination game in which the payoff and risk dominant equilibria differ (that is, $\tilde{\tilde{k}} > (n-1)/2$), and vice versa.*

Proof. See Appendix.

In Fig. 3b, switching is a tipping game with or without linkage. Moreover, as the tipping point is relatively low, the payoff and risk dominant equilibria coincide under both regimes. However, linkage raises the tipping point and increases the penalty to switching if few others switch, making switching riskier and coordination harder.[18] Linkage is unnecessary and might thwart switching.[19] Finally, in Fig. 3c, linkage lowers the tipping point, but doesn't cause the payoff and risk dominant equilibria to align. Linkage also increases the penalty to switching when few others switch. In this situation, linkage is unnecessary and could be either helpful or unhelpful.

Proposition 4 (unilateral trade linkage). *Trade linkage is unambiguously of potential help only if increasing returns are weak and linkage is able to transform the switching game from a prisoners' dilemma into a tipping game. Under these circumstances, the payoff and risk dominant Nash equilibria in the tipping game differ, making unilateral switching and linking risky.*

[18] Of course, linkage also increases the penalty to not switching when nearly all others switch, and so makes the payoff dominant equilibrium even more stable. However, given that the default is to stick with A, the tried-and-true technology, the more important effect of linkage is to discourage switching if k is low.

[19] This possibility is hinted at by the lab experiment in Barrett and Dannenberg (2017), particularly the comparison between the B-8 and B-10 treatments, though there are also some differences between the situation depicted in Fig. 3b and the one studied in this experiment.

Once again, a treaty can help. Consider the following game:

> Stage 1: Every country chooses independently whether to be a signatory or a non-signatory to a switching/linking treaty.
>
> Stage 2: Signatories decide collectively whether to switch and impose trade measures against countries that behave differently.
>
> Stage 3: Non-signatories decide independently whether to follow the lead of signatories or to stick with A and retaliate.

If signatories switch and link in Stage 2, non-signatories will either switch and link in Stage 3, so as to avoid being the target of trade measures, or stick with A and retaliate in Stage 3. They will stick with A and retaliate if $k < \tilde{\tilde{k}}$. They will switch and link if $k \geq \tilde{\tilde{k}}$. Reasoning backwards, signatories have a collective incentive to switch and link in Stage 2 iff $k \geq \tilde{\tilde{k}}$. Finally, in Stage 1, no country loses by signing the treaty, irrespective of how many others sign it, and each gains by signing it provided enough others sign it. Hence, in equilibrium, participation in the treaty is full; all countries switch to B; trade remains free. The treaty solves the coordination problem.

Proposition 5 (linked treaties): *If switching/linking is a tipping game, a tipping treaty assures selection of the payoff dominant Nash equilibrium, resulting in universal switching and free trade.*

Notice that this result applies even if linkage is unhelpful or unnecessary if undertaken unilaterally.

How might this work in practice? Consider the aluminum sector. Aluminum is made using a century's-old process. A carbon anode is dipped in a "pot" containing alumina and cryolite, and electricity is then run through the anode to initiate the smelting process. The reaction releases

oxygen, which reacts with the carbon in the anode to produce both CO_2 and perfluorinated carbon compounds or PFCs, another greenhouse gas, as byproducts. These process emissions can be eliminated by replacing carbon anodes with "inert" anodes. The US and EU have been holding discussions on whether and how to impose carbon tariffs on aluminum imports from countries that fail to meet certain emissions-intensity standards, with a focus on the energy input rather than the process. The approach suggested here is different.[20] It would be to negotiate a treaty requiring that parties manufacture aluminum using inert anodes, and that they trade in aluminum products only with other parties.

7. "Strong" vs. "weak" trade linkage

To this point, like Nordhaus (2015), I have assumed that players are either in or out of the club. Here I consider the possibility that a country can avoid being the target of trade measures by switching, without having to join parties to the agreement in imposing trade measures against non-switchers. We might call this "weak linkage," and the approach considered previously, "strong linkage."

Under weak linkage, countries play one of three roles. They can be "leaders," switching and imposing trade measures against non-switchers; "non-switchers," sticking with A and retaliating against the leaders; and "follower-switchers," complying with the environmental aims of the agreement, but maintaining free trade relations with all countries. Continue to assume $\tilde{k} > n - 1$, and suppose that there are k leaders. Let $\acute{k}$ be the solution to $\pi_i^{FS}(k) = \pi_i^{NS}(k)$. Country i does better being a follower-switcher if $k \geq \acute{k}$, where (see Appendix)

$$\acute{k} = \frac{C - \beta\left(1 - e^{-\gamma/\beta}\right)}{C\alpha^B/n + \beta[1 - e^{-\gamma/\beta}(1 + \gamma/\beta)]\,\alpha^A/n + (b - c)}. \tag{16}$$

[20] One problem with basing carbon tariffs on the electricity input is that it creates an incentive for aluminum manufacturers to switch to hydropower, forcing other electricity users to buy power produced from fossil fuels.

For linkage of this type to make switching a tipping game, we require $\acute{k} < n - 1$. (If $\acute{k} > n - 1$, switching remains a prisoners' dilemma under weak linkage.)

The following lemmas compare and contrast strong and weak linkage:

Lemma 3. $\lim_{(b-c)\to 0} \acute{k} = \lim_{(b-c)\to 0} \tilde{\tilde{k}} = \tilde{k}$. *In words, as the gains from trade cooperation shrink, all else equal, the tipping points under both weak and strong linkage converge to the tipping point without trade linkage.*

Lemma 4. $\tilde{\tilde{k}} \lesseqgtr n - 1 \Leftrightarrow \acute{k} \lesseqgtr n - 1$. *If switching is a tipping game under strong (weak) linkage game, then switching is also a tipping game under weak (strong) linkage. If switching is a prisoners' dilemma under strong (weak) linkage, then switching is also a prisoners' dilemma under weak (strong) linkage.*

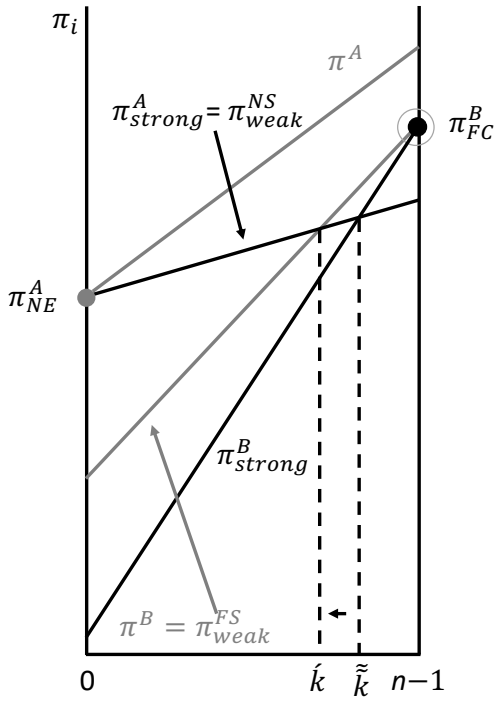
Proof. See Appendix.

Lemma 5. $\acute{k} < \tilde{\tilde{k}} \Leftrightarrow \acute{k} < n - 1, \acute{k} < \tilde{\tilde{k}} \Leftrightarrow \tilde{\tilde{k}} < n - 1$. *If switching is a tipping game under either weak or strong linkage, then the tipping point is lower under weak linkage than under strong linkage.*

Proof. See Appendix.

Lemma 5 is illustrated in Figure 4.

Figure 4. Weak vs. Strong Linkage



Lemma 6. $\lim_{(b-c)\to\infty} \acute{k} = 0$ *and* $\lim_{(b-c)\to\infty} \tilde{\tilde{k}} = (n-1)/2$. *As the gains from trade cooperation increase, the tipping points under weak and strong linkage diverge. Under weak linkage, switching and linking by even a single country may suffice to impel all other countries to switch. By contrast, under strong linkage, a majority of countries must switch and link for others to be impelled to do likewise.*

Since we know that a treaty can coordinate selection of the payoff-dominant Nash equilibrium in a tipping game, the above lemmas imply:

Proposition 6 ("weak" v. "strong" trade linkage): *Weak linkage has no particular advantage over strong linkage in a treaty. However, if the gains from trade cooperation are sufficiently large, the payoff and risk dominant Nash equilibria under weak linkage will coincide, possibly helping coordination to succeed even without a treaty. Indeed, under weak linkage, a small coalition, possibly even a coalition of one, may suffice to tip all countries to switch.*

Proposition 6 helps to interpret the scant evidence that exists in this area. Though the Montreal Protocol on Substances that Deplete the Ozone Layer applies strong linkage as a rule, it allows weak linkage as an exception; under Article 4, parties may refrain from imposing trade measures on "any State not party to this Protocol, if that State is determined, by a meeting of the Parties, to be in full compliance with [the emission reduction obligations]." The European Union's Aviation Directive, which extended the EU's emissions trading system to air travel beyond EU airspace for flights departing from or landing at EU airports, applies weak linkage unilaterally, but was soon suspended after other countries, including the US, threatened to retaliate. This suggests that a coalition of one may be too small for aviation. Recently, the US invited the EU to join it in establishing a Global Arrangement on Sustainable Steel and Aluminum. Under this proposal, countries that fail to meet the standard chosen by the US-EU partnership would be subject to tariffs—another example of weak linkage. We shall have to see whether this invitation is accepted and, if it is, whether a US-EU coalition is big enough to tip behavior globally.

8. Strategic R&D

Can R&D also help transform what would have been a prisoners' dilemma into a tipping game? Again, an example may help frame our thinking. International shipping is characterized by strong network externalities—the reason the International Maritime Organization establishes common standards. A technology-centered agreement for this sector would focus not on reducing emissions but on changing the technology-fuel standard. Today's standard is the diesel engine powered by heavy fuel oil. As a first step, R&D is needed to identify the best alternative—a primary aim of Mission Innovation, a coalition of the willing supported by 22 countries.

For purposes of discussion, suppose that the best alternative to bunker fuel turns out to be ammonia. How to proceed? Ammonia is expected to cost several times as much as heavy fuel oil, making switching a prisoners' dilemma. A switch to ammonia also presents technical challenges. It requires new engines, new on-board storage tanks (necessitating new ship designs), and new port facilities: in short, a new technology-fuel system. Clearly, vessel owners won't want their ships to run on ammonia unless a network of refueling infrastructure is available, just as no country will want to build an ammonia fuel network unless assured that lots of ships will run on ammonia. Lock-in is a barrier to switching if one or a small number of countries switch. But as more ports switch to ammonia, more ship owners will want their vessels to run on ammonia, and as more ships run on ammonia, more ports will want to switch. It is by a similar process that an international agreement known as MARPOL, negotiated under the IMO, orchestrated adoption of new standards for oil tankers (Mitchell 1994; Barrett 2003).

Another aim of Mission Innovation is to establish a "global port infrastructure to support vessels operating on zero-emission fuels so that by 2030, 10 large trade ports covering at least three continents supply zero-emission fuels."[21] Whether ten large ports is enough to tip the switch to ammonia is unclear, but the logic behind this program plainly reflects the need to transform the prisoners' dilemma into a tipping game.[22] Mission Innovation isn't only about innovation; it's about linking innovation to future deployment (see also Barrett 2006). This approach to addressing climate change is strategic.

Suppose that switching costs depend on total R&D spending, denoted by R.[23] Specifically, assume $C(R) = Ce^{-\eta R}$, where $R = \sum_{i=1}^{n} r_i$, and r_i represents country i's contribution to R&D. Substituting $C(R)$ for C in (7a), the gross cost of switching becomes $Ce^{-\eta R}(n - \alpha^B k)/n$. An increase in R thus has two effects on the payoff to switching. It raises the left-hand intercept of

[21] https://explore.mission-innovation.net/mission/zero-emissions-shipping/

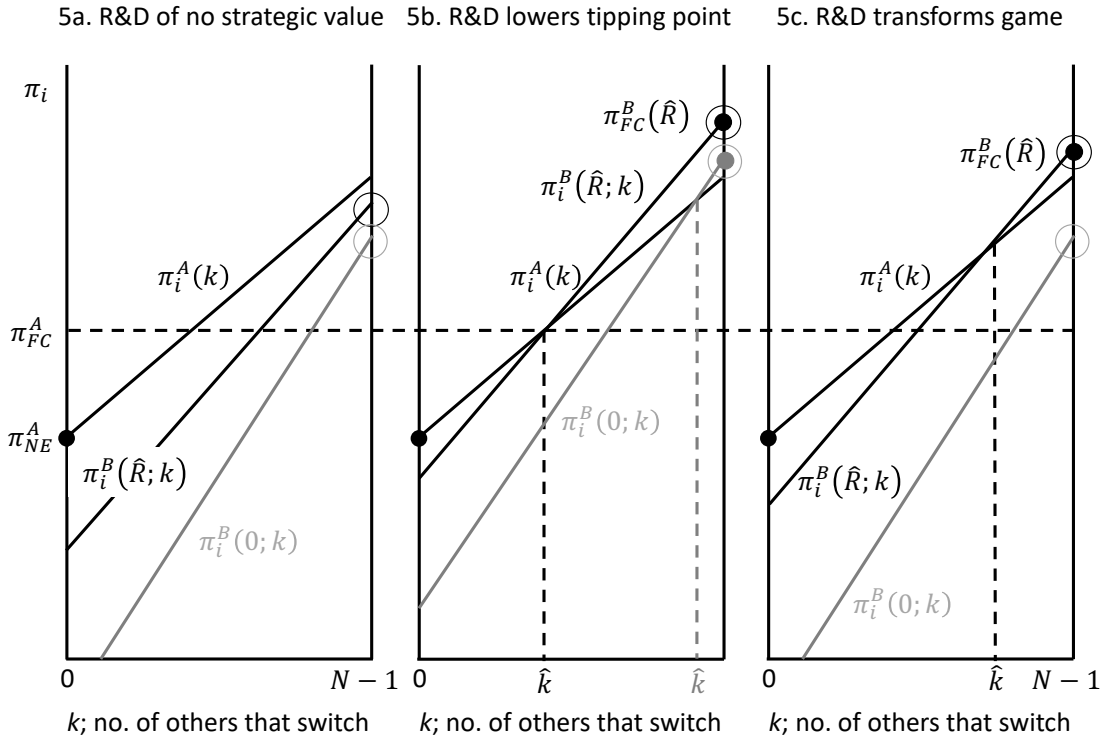
[22] Of the world's ten biggest ports by volume, seven are in China. China's participation in a strategy to change shipping is essential.

[23] My approach broadly follows Barrett (2006); see also Hoel and de Zeeuw (2010). My purpose here is to cast R&D in a more general framework.

the π^B curve and, for $\alpha^B > 0$, it flattens the slope. Because R&D lowers the cost of switching, it lowers the reduction in costs achievable by increasing returns.

Figure 5 illustrates three possibilities, assuming that $R \in \{0, \hat{R}\}$, where $\hat{R} > 0$. In Figure 5a, an investment of $\hat{R}$ lifts the π^B curve, but switching remains a prisoners' dilemma. This R&D doesn't stimulate widescale deployment. In Fig. 5b, switching is a coordination game, even without R&D. However, by reducing the tipping point to switching, an investment of $\hat{R}$ makes the full cooperative outcome both payoff- and risk-dominant, and lab experiments suggest that this should improve the chances that coordination will succeed without a treaty (Barrett and Dannenberg 2017). Finally, in Fig. 5c, investment of $\hat{R}$ transforms the switching game more fundamentally, turning it from a prisoners' dilemma into a coordination game. In the remainder of this section, I focus on this last situation.

Figure 5. Switching with R&D



I have taken R to be binary. How to interpret $\hat{R}$? As the purpose of R&D is to convert the prisoners' dilemma into a coordination game, we can calculate $\hat{R}$ as the minimum expenditure on R&D necessary to bring about this transformation. If R is continuous, $\hat{R}$ is thus the *threshold* for transformation. For simplicity, I shall continue to assume that public good of R&D is *discrete*.

Assume that, for $R = 0$, switching is a prisoners' dilemma; that is, $\pi^A(k) > \pi^B(k;0) \forall k$, and $\pi^B(n-1;0) > \pi^A(0)$. For $R = \hat{R}$ to make switching a tipping game we require: $\pi^A(0) > \pi^B(0;\hat{R})$ and $\pi^B(n-1;\hat{R}) > \pi^A(n-1)$. For switching to be efficient, we require $\pi^B(n-1;\hat{R}) > \pi^A(0)$. (See Appendix.) Of course, if $\eta\hat{R}$ is sufficiently large, we will have $\pi^A(0) < \pi^B(0;\hat{R})$, making switching a dominant strategy. In this case, a treaty may be needed to supply $\hat{R}$, but not to tip switching. Here, I shall assume that $\pi^A(0) > \pi^B(0;\hat{R})$, meaning that a treaty is needed both to promote supply of $\hat{R}$ *and* the switch to B.

If the above conditions apply, collective R&D and technology switching are strong complements. R&D pays only if switching occurs, and switching occurs only if the investment in R&D is made. The supply of both thus requires a system of two treaties, one to supply the R&D and one to coordinate the switch to B. We already know that a switching treaty can support the efficient outcome, provided the above conditions hold and the expenditure $\hat{R}$ is made. Aggregate willingness to pay for $\hat{R}$ is $n\left[\pi^B_{FC}(\hat{R}) - \pi^A_{NE}\right]$. *Collectively,* states will gain by paying this cost iff

$$\hat{R} \leq n\left\{\gamma n e^{-\gamma/\beta} - \left[Ce^{-\eta\hat{R}}[n - \alpha^B(n-1)]/n - \beta\left[1 - e^{-\gamma/\beta}(1 + \gamma/\beta)\right]/n\right]\right\}. \qquad (17a)$$

It will not pay any country to supply $\hat{R}$ *unilaterally* if

$$\hat{R} > \gamma n e^{-\gamma/\beta} - \left[Ce^{-\eta\hat{R}}[n - \alpha^B(n-1)]/n - \beta\left[1 - e^{-\gamma/\beta}(1 + \gamma/\beta)\right]/n\right]. \qquad (17b)$$

If (17a) and (17b) both hold, international cooperation is needed to fund the R&D, which is a discrete public good. There are two symmetric Nash equilibria (in pure strategies), zero funding and funding by an amount $r_i = \hat{R}/n \; \forall \; i$. Full funding can also be supported by many asymmetric

Nash equilibria. In particular, funding of R&D could be shared by a subset of countries, such as the members of Mission Innovation. Adoption of a new standard, by contrast, will require a new agreement, applicable to all members of the IMO.

Solving the coordination problem now requires a two-part treaty:

Part I Stage 1: Every country chooses independently whether to be a signatory to a B-specific R&D treaty;

Stage 2: Signatories decide collectively whether to fund the R&D, knowing that each signatory will pay an equal share of the cost;

Part II Stage 1: Every country chooses independently whether to be a signatory to a switching treaty;

Stage 2: Signatories decide collectively whether to switch.

Stage 3: Non-signatories decide individually whether to switch.

Begin by solving Part II. The tipping point for the switching game will be $\tilde{k}(\hat{R})$, where $\tilde{k}(\hat{R})$ is given by (8) after substituting $Ce^{-\eta\hat{R}}$ for C. Hence, in Stage 3, every non-signatory will switch provided at least $\hat{k}(\hat{R})$ others switch, where $\hat{k}(\hat{R})$ is the smallest integer greater than or equal to $\tilde{k}(\hat{R})$. In Stage 2, signatories will switch collectively iff the number of signatories is equal to or greater than $\hat{k}(\hat{R})$. Finally, in Stage 1, no country loses by joining the treaty, and all gain by joining it if at least $\hat{k}(\hat{R}) - 1$ others join it. In equilibrium, therefore, all countries participate in the switching treaty; all countries switch to B.

Play in Part II is predicated on a coalition of states having invested $\hat{R}$ previously. Let $\tilde{k}^{\hat{R}} = \left[\pi_{FC}^{B}(\hat{R}) - \pi_{NE}^{A}\right]/\hat{R}$. Then the minimum coalition willing to fund $\hat{R}$ is the smallest integer greater

than or equal to $\tilde{k}^{\hat{R}}$. Denote this value $\hat{k}^{\hat{R}}$. In Stage 2, signatories will decline to invest in R&D if participation is less than $\hat{k}^{\hat{R}}$, and pay for the R&D fully if participation is at least $\hat{k}^{\hat{R}}$. Finally, in Stage 1, precisely $\hat{k}^{\hat{R}}$ countries will sign the treaty. Should another country join, it would have to contribute its share to the R&D budget, without changing the final outcome. This is another chicken game.[24]

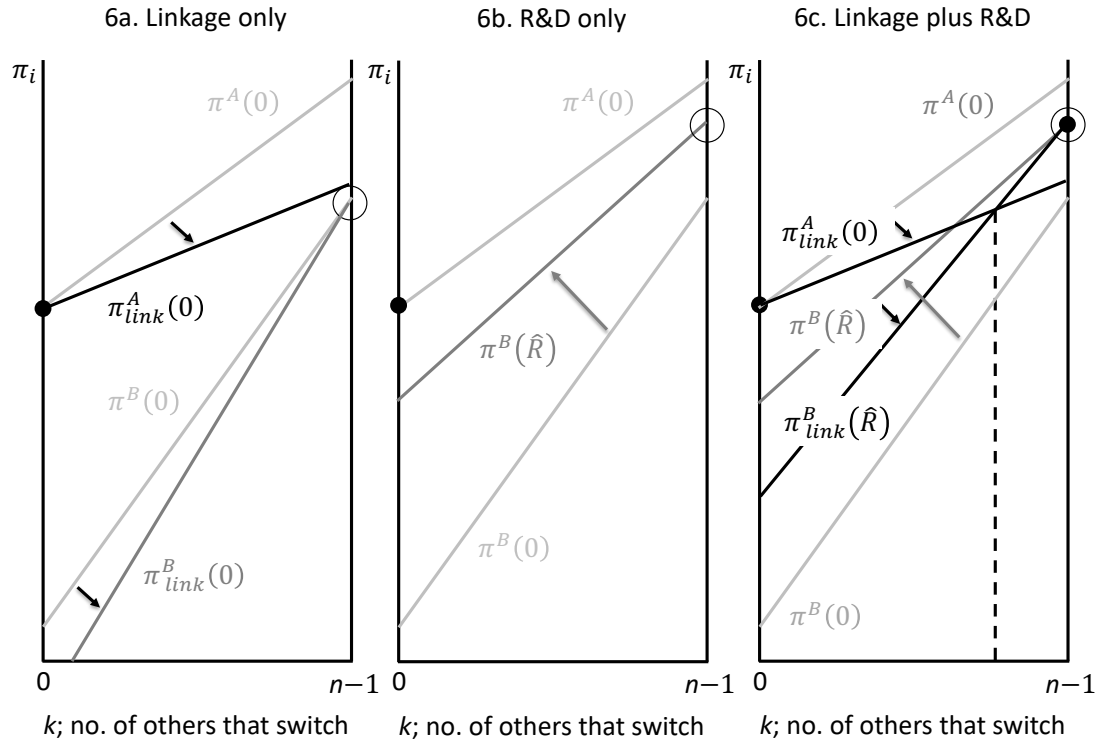
Proposition 7 (two-part R&D-switching treaty): *If R&D can transform switching from a prisoners' dilemma into a tipping game, a two-part treaty consisting of an R&D treaty, supported by a subset of countries, and a multilateral switching treaty can assure universal switching.*

9. R&D coupled with trade linkage

Where the two-way combinations of increasing returns and trade linkage or increasing returns and strategic R&D each fall short of transforming the switching game, the three-way combination of increasing returns, trade linkage, and R&D may succeed better. This possibility is illustrated in Figure 6. The combinations of increasing returns and (strong) linkage (Fig. 6a) and increasing returns and R&D (Fig. 6b) are unable to transform the switching prisoners' dilemma, but the three-way combination is able to do so (Fig. 6c). R&D improves the economics of switching for any level of take up, and linkage raises the payoff to switching as more countries switch. Selection of the preferred equilibrium, and full funding of the R&D, is again achieved by the two-part treaty described in Section 8, except for the Part II treaty being replaced by the linked game presented in Section 6.

[24] If R&D is a joint product, meaning that it provides private benefits to contributors (training of their scientists, for example) in addition to the public good of knowledge, participation would increase.

Figure 6. The Combined Effect of Linkage and R&D on Switching



10. Conclusions and implications for negotiations

From the start, climate negotiators saw their task as getting countries to limit their emissions. The Framework Convention framed collective action this way in 1992, and negotiators have stayed on this track ever since. In Kyoto, in 1997, emission limits were negotiated for a subset for countries in the expectation that an enforcement mechanism could be added later. No such mechanism could be agreed, and support for Kyoto eroded even as negotiators gathered in Copenhagen in 2009 to negotiate tougher "commitments" shared by a greater number of countries. Asking an approach that was already faltering to achieve even more was fanciful, and the negotiations collapsed. To rescue collective action, negotiators pivoted on the spot, asking countries to pledge "contributions" towards meeting a common temperature change goal, with the understanding that the pledges were to be nationally determined and strictly voluntarily. Like the previous approach, this new one, formalized in Paris in 2015, asks countries to play the prisoners' dilemma. It is unlikely to be effective.

A different approach, akin to the one studied in this paper, is starting to emerge. In Paris, separate from the United Nations process, a group of 20 countries, including the United States, China, the European Union, India, Japan, and Brazil, announced Mission Innovation, an "all-in, all-sector approach to transform global energy markets."[25] Seeing that, even as emissions were increasing, prices of LED lightbulbs, batteries, and solar and wind energy were falling, Mission Innovation aims to fund R&D so as to lower the prices of other critical technologies. I noted previously Mission Innovation's goal of zero-emission shipping, but there are others. Its clean hydrogen mission, for example, aims to reduce "end-to-end costs to a tipping point of 2 USD/kg by 2030."[26] As with shipping, innovation in hydrogen is tied to deployment—in this case, "at least 100 large-scale integrated clean hydrogen valleys worldwide by 2030… [so as to] generate economies of scale, pushing down clean hydrogen costs to catalyse the development of a global clean hydrogen economy."[27] By targeting R&D and linking it to deployment, Mission Innovation aims to tip the switch to new technology-fuel systems.

Ten months after Paris, countries adopted the Kigali Amendment to the Montreal Protocol, which phases down the production and consumption of hydrofluorocarbons, a powerful greenhouse gas. The Kyoto Protocol failed to reduce HFC emissions, but this new agreement makes cooperation on trade in HFCs and products containing HFCs contingent on participation in the collective effort to phase down these chemicals (strong linkage). That is, Kigali makes cooperation a tipping game. The United States Senate never voted on the Kyoto Protocol, and the Paris Agreement had to be constructed in such a way that the United States could participate without Senate approval. By contrast, in 2022, the Senate ratified the Kigali Amendment in the usual way, by a two-thirds majority, another indication of the power of tipping agreements.

[25] https://obamawhitehouse.archives.gov/blog/2015/11/29/announcing-mission-innovation
[26] https://explore.mission-innovation.net/mission/clean-hydrogen/
[27] http://mission-innovation.net/missions/hydrogen/. A hydrogen "valley" is defined as an integrated system comprising production, storage, and distribution in a defined geographic area.

This paper exposes the mechanisms behind these developments. It also offers a blueprint for developing more such efforts. By exploiting increasing returns in a new technology-fuel system, supported, where appropriate, by trade linkages and targeted R&D cooperation, the strategic approach is able to overcome lock-in and make universal switching self-enforcing. Of course, the approach is piecemeal. It ignores general equilibrium effects. It won't bring global emissions to zero. But it will improve on the default arrangement, the main UN negotiations; and as elements of the approach are already emerging, it is important that economists understand the conditions and strategies needed to make it work efficiently and effectively.

References

Aghion, Philippe, Antoine Dechezleprêtre, David Hémous, Ralf Martin, and John Van Reenen (2016), "Carbon Taxes, Path Dependency, and Directed Technical Change: Evidence from the Auto Industry," *Journal of Political Economy* 124(1): 1-51.

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

Allcott, Hunt, Reigner Kane, Maximilian S. Maydanchik, Joseph S. Shapiro, and Felix Tintelnot (2024), "The Effects of 'Buy American': Electric Vehicles and the Inflation Reduction Act," NBER Working Paper 33032.

Arthur, Brian W. (1989), "Competing Technologies, Increasing Returns, and Lock-In by Historical Events," *Economic Journal* 99(394): 116-131.

Arthur, Brian W. (2009), *The Nature of Technology: What It Is and How It Evolves*, New York: Free Press.

Bagwell, Kyle and Robert W. Staiger. 2011. What do trade negotiators negotiate about? Empirical evidence from the World Trade Organization. *American Economic Review* 101(4): 1238-1273.

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

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

Barrett, Scott and Astrid Dannenberg (2016), "An Experimental Investigation into 'Pledge and Review' in Climate Negotiations," *Climatic Change* 138(1): 339-352.

Barrett, Scott and Astrid Dannenberg (2017), "Tipping Versus Cooperating to Supply a Global Public Good," *Journal of the European Economics Association* 15(4): 910-941.

Barrett, Scott and Astrid Dannenberg (2022), "The Decision to Link Trade Agreements to the Supply of Global Public Goods," *Journal of the Association of Environmental and Resource Economists* 9(2): 273-305.

Bayer, Patrick and Michaël Aklin (2020), "The European Union Emissions Trading System Reduced CO2 Emissions Despite Low Prices," *Proceedings of the National Academy of Sciences* 117(16): doi/10.1073/pnas.1918128117.

Calel, Raphael and Antoine Dechezleprêtre (2016), "Environmental Policy and Directed Technological Change: Evidence from the European Carbon Market," *Review of Economics and Statatistics* 98(1): 173-191.

Colmer, Jonathan, Ralf Martin, Mirabelle Muûls, and Ulrich J. Wagner (2024), "Does Pricing Carbon Mitigate Climate Change? Firm-Level Evidence from the European Union Emissions Trading System," *Review of Economic Studies*, forthcoming.

Cowan, Robin (1991), "Tortoises and Hares: Choice Among Technologies of Unknown Merit," *Economic Journal* 101(407): 801-814.

Dixit, Avinash (2003), "Clubs with Entrapment," *American Economic Review* 93(5): 1824-1829.

Farrell, Joseph and Paul Klemperer (2007), "Coordination and Lock-In: Competition with Switching Costs and Network Effects," in Mark Armstrong and Robert Porter (eds.), *Handbook of Industrial Organization*, volume 3, Amsterdam: Elsevier, 1967-2072.

Finus, Michael (2001), *Game Theory and International Environmental Cooperation*, Cheltenham, UK: Edward Elgar.

Fouquet, Roger (2016), "Path Dependence in Energy Systems and Economic Development," *Nature Energy* 1: 1-5.

Fridstrøm, Lasse (2021), "The Norwegian Vehicle Electrification Policy and Its Implicit Price of Carbon," *Sustainability* 13, 1346. https://doi.org/10.3390/su13031346.

Hawkins-Pierot, Jonathan T. and Katherine R.H. Wagner (2022), "Technology Lock-In and Optimal Carbon Pricing," CESifo Working Paper NO. 9762.

Heal, Geoffrey (1994), "Formation of International Environmental Agreements," in Carlo Carraro (ed.), *Trade, Innovation, Environment*. Dordrecht: Kluwer, 301-322.

Heal, Geoffrey (2022), "Economic Aspects of the Energy Transition," *Environmental and Resource Economics* https://doi.org/10.1007/s10640-022-00647-4.

Heal, Geoffrey and Howard Kunreuther (2010), "Social Reinforcement: Cascades, Entrapment, and Tipping," *American Economic Journal: Microeconomics* 2(1): 86-99.

Heal, Geoffrey and Howard Kunreuther (2017), "An Alternative Framework for Negotiating Climate Policies," *Climatic Change* 144: 29-39.

Hoel, Michael and Aart de Zeeuw (2010), "Can a Focus on Breakthrough Technologies Improve the Performance of International Environmental Agreements?" *Environmental and Resource Economics* 47: 395-406.

Jaffe, Adam B., Richard G. Newell, and Robert N. Stavins (2005), "A Tale of Two Market Failures: Technology and Environmental Policy," *Ecological Economics* 54: 164-174.

Li, Shanjun, Lang Tong, Jianwei Xing, and Yiyi Zhou (2017), "The Market for Electric Vehicles: Indirect Network Effects and Policy Design," *Journal of the Association of Environmental and Resource Economists* 4(1): 89-133.

Lovei, Magda (1996), "Phasing Out Lead from Gasoline: World-Wide Experience and Policy Implications," World Bank Environment Department Paper No. 40.

Meng, Kyle C. and Ashwin Rode (2019), "The Social Cost of Lobbying Over Climate Policy," *Nature Climate Change* 9: 472-476.

Metcalf, Gilbert E. and James H. Stock (2023), "The Macroeconomic Impact of Europe's Carbon Taxes," *American Economic Journal: Macroeconomics* 15(3): 265-286.

Mitchell, Ronald B. (1994). 'Regime Design Matters; Intentional Oil Pollution and Treaty Compliance,' *International Organizat*ion, vol. 48, pp. 425-458.

Nordhaus, William (2015), “Climate Clubs: Overcoming Free-Riding in International Climate Policy,” *American Economic Review,* 105(4): 1339-1370.

Nordhaus, William (2017), “Revisiting the Social Cost of Carbon,” *Proceedings of the National Academy of Sciences* 114(7): 1518-1523.

Nordhaus, William (2021), “Dynamic Climate Clubs: On the Effectiveness of Incentives in Global Climate Agreements,” *Proceedings of the National Academy of Sciences* 118(45): 1339-1370.

van der Ploeg, Frederick and Anthony J. Venables (2022), “Radical Climate Policies,” Policy Research Paper 10212, World Bank.

Postrel, Steven R. (1990), “Competing Networks and Proprietary Standards: The Case of Quadraphonic Sound,” *Journal of Industrial Economics* 39(2): 169-185.

Pretis, Felix (2022), “Does a Carbon Tax Reduce CO_2 Emissions? Evidence from British Columbia,” *Environmental and Resource Economics* 83: 115-144.

Schelling, Thomas C. (1978), *Micromotives and Macrobehavior*, New York: WW Norton & Company.

Schmidt, Julia and Walter Steingress (2022), “No Double Standards: Quantifying the Impact of Standard Harmonization on Trade,” *Journal of International Economics* 137: 103619.

Seto, Karen C., Steven J. Davis, Ronald B. Mitchell, Eleanor C. Stokes, Gregory Unruh, and Diana Ürge-Vorsatz (2016), “Carbon Lock-In: Types, Causes, and Policy Implications,” *Annual Review of Environment and Resources* 41: 425-452.

Unruh, Gregory C. (2000), “Understanding Carbon Lock-In,” *Energy Policy* 28: 817-830.

Wagner, Ulrich J. (2001), "The Design of Stable International Environmental Agreements: Economic Theory and Political Economy," *Journal of Economic Surveys* 15(3): 377-411.

Walsh Sanderson, Susan and Kenneth L. Simons (2014), "Light Emitting Diodes and the Lighting Revolution: The Emergence of a Solid-State Lighting Industry," *Research Policy* 43: 1730-1746.

World Trade Organization. 2009. *World Trade Report 2009: Trade Policy Commitments and Contingency Measures*. Geneva: World Trade Organization.

de Zeeuw, Aart (2015), "International Environmental Agreements," *Annual Review of Resource Economics* 7: 151-168.

Appendix

Solution to $\tilde{\tilde{k}}$. If i joins the switchers/linkers it gets

$$\pi_i^B = \gamma(n-k-1)\left(1-e^{-\gamma/\beta}\right) + \gamma(k+1) - C\,(n-\alpha^B k)/n + bk + c(n-k-1). \quad (A.1a)$$

If i sticks with A and retaliates against the switchers/linkers it gets

$$\pi_i^A = \gamma(n-k)\left(1-e^{-\gamma/\beta}\right) + \gamma k - \beta\left[1-e^{-\gamma/\beta}(1+\gamma/\beta)\right](n+\alpha^A k)/n + b(n-k-1) + ck. \quad (A.1b)$$

$\tilde{\tilde{k}}$ in eq. (13) is found by (A.1a) equal to (A.1b).

Proof of Lemma 2. Suppose the lemma is false. Then $\tilde{\tilde{k}} \leq (n-1)/2$, or

$$\frac{C-\beta\left(1-e^{-\gamma/\beta}\right)+(n-1)(b-c)}{C\alpha^B/n+\beta[1-e^{-\gamma/\beta}(1+\gamma/\beta)]\,\alpha^A/n+2(b-c)}\leq\frac{n-1}{2},$$

which implies $\tilde{k}\leq(n-1)/2$, a violation of the lemma. ■

Solution to $\acute{k}$. Country i gets

$$\pi_i^{FS}=\gamma(n-k-1)\left(1-e^{-\gamma/\beta}\right)+\gamma(k+1)-C\,(n-\alpha^B k)/n+b(n-1) \qquad (A.2a)$$

by being a follower-switcher and

$$\pi_i^{NS}=\gamma(n-k)\left(1-e^{-\gamma/\beta}\right)+\gamma k-\beta\left[1-e^{-\gamma/\beta}(1+\gamma/\beta)\right](n+\alpha^A k)/n \\ +b(n-k-1)+ck \qquad (A.2b)$$

by being a non-switcher. $\acute{k}$ in eq. (16) is found by setting (A.2a) equal to (A.2b).

Proof of Lemma 4. Rewrite $\tilde{\tilde{k}}$ as $\tilde{\tilde{k}}=[\Omega+(n-1)(b-c)]/[\Psi+2(b-c)]$, where $\Omega=C-\beta\left(1-e^{-\gamma/\beta}\right)$ and $\Psi=C\alpha^B/n+\beta\left[1-e^{-\gamma/\beta}(1+\gamma/\beta)\right]\alpha^A/n$. Similarly, rewrite $\acute{k}$ as $\acute{k}=\Omega/[\Psi+(b-c)]$. Substituting, we get $\tilde{\tilde{k}}<n-1\Leftrightarrow\Omega<\Psi(n-1)+(n-1)(b-c)$. Also, $\acute{k}<n-1\Leftrightarrow\Omega<\Psi(n-1)+(n-1)(b-c)$. Seeing that these inequalities are identical proves the lemma. ■

Proof of Lemma 5. $\acute{k}<\tilde{\tilde{k}}$ implies $\Omega/[\Psi+(b-c)]<[\Omega+(n-1)(b-c)]/[\Psi+2(b-c)]$. Rearranging gives $\Omega<\Psi(n-1)+(n-1)(b-c)$. Dividing both sides by $\Psi+(b-c)$ gives $\acute{k}<n-1$. The other part of the lemma is obtained by reversing the inequality. ■

Conditions under which R&D transforms switching from a prisoner's dilemma into a tipping game. For switching to be a prisoners' dilemma for $R=0$, we require $\pi^A(k)>\pi^B(k;0)\forall k$, or

$$C\,(n-\alpha^B k)/n-\beta\left[1-e^{-\gamma/\beta}(1+\gamma/\beta)\right](n+\alpha^A k)/n>\gamma e^{-\gamma/\beta}, \qquad (A.3a)$$

and $\pi^B(n-1;0)>\pi^A(0)$, or

$$\gamma n e^{-\gamma/\beta}>C[n-\alpha^B(n-1)]/n-\beta\left[1-e^{-\gamma/\beta}(1+\gamma/\beta)\right]. \qquad (A.3b)$$

For $R=\hat{R}$ to make switching a tipping game we require: $\pi^A(0)>\pi^B\left(0;\hat{R}\right)$, or

$$Ce^{-\eta\hat{R}}-\beta\left[1-e^{-\gamma/\beta}(1+\gamma/\beta)\right]>\gamma e^{-\gamma/\beta}; \qquad (A.4a)$$

$\pi^B\left(n-1;\hat{R}\right)>\pi^A(n-1)$, or

$$\gamma e^{-\gamma/\beta}>Ce^{-\eta\hat{R}}[n-\alpha^B(n-1)]/n-\beta\left[1-e^{-\gamma/\beta}(1+\gamma/\beta)\right][n+\alpha^A(n-1)]/n\,. \qquad (A.4b)$$

Finally, for switching to be efficient we require $\pi^B\left(n-1;\hat{R}\right)>\pi^A(0)$, or

$$\gamma n e^{-\gamma/\beta}>Ce^{-\eta\hat{R}}[n-\alpha^B(n-1)]/n-\beta\left[1-e^{-\gamma/\beta}(1+\gamma/\beta)\right]. \qquad (A.5)$$

Online Appendix

OA.1 Uncertain switching costs

Consider again Figure 1b. If the gray payoff line applies, the parameter α^B is "low" and switching is a prisoners' dilemma; if the black line applies, $\alpha^B$is "high" $(\alpha_H^B > \alpha_L^B \geq 0)$ and switching is a coordination game. The more realistic situation is one in which, before deployment of a novel technology, the value of α^B is unknown. What kind of game is switching when the returns to switching come into focus after switching has occurred?

Assume that, before choosing whether to switch, each symmetric country holds the same prior probability, p, $p \in [0,1]$, of α^B being high ($1-p$ is thus the prior probability that $\alpha^B$is low). Then the expected payoff to switching lies somewhere between the black and gray π^B lines in Fig. 1b. If countries choose simultaneously and are risk neutral, none will switch if each believes switching is a prisoners' dilemma $\left(\pi^A(n-1;p) > \pi^B(n-1;p)\right)$, whereas each *might* switch if each believes switching is a tipping game $\left(\pi^B(n-1;p) > \pi^A(n-1;p)\right)$. Upon substituting, the latter condition can be rewritten as

$$\bar{\alpha}^B > \frac{C - \gamma e^{-\gamma/\beta} - \beta\left[1 - e^{-\gamma/\beta}(1+\gamma/\beta)\right]\left[1 + \alpha^A\,(n-1)/n\right]}{C\,(n-1)/n}, \qquad \text{(OA. 1)}$$

where $\bar{\alpha}^B = p\alpha_H^B + (1-p)\alpha_L^B$.

If countries had to choose simultaneously whether to switch, we know that a treaty can assure coordination provided inequality (OA.1) holds. Under certainty, simultaneous choice makes sense because there is no benefit to waiting. However, under uncertainty, when k signatories switch, they not only reduce emissions and the expected switching costs of signatories due to increasing returns; if the costs of switching are public knowledge, they also reduce uncertainty about these costs—another positive externality—creating an incentive for non-signatories to

"wait" before deciding whether to switch so that they may learn whether the posterior probability that α^B is high differs from the prior, p. By waiting, non-switchers "learn-by-observing."

If increasing returns are uncertain, the treaty game becomes:

> Stage 1: All countries decide simultaneously whether to join a switching treaty.
>
> Stage 2: Parties to the treaty decide collectively whether to switch, and then, upon doing so, reveal information about the costs of switching;
>
> Stage 3: Non-parties decide independently whether to switch (equivalently, whether to accede to the treaty, should it have entered into force).

Suppose that k countries switch, producing a signal, θ_k, about α^B. This signal is "stronger" (packs in more information) the higher is k. Upon observing θ_k, Bayesian non-signatories will calculate the posterior probability that α^B is high as

$$p_k = \frac{P(\theta_k|\alpha_H^B) \cdot p}{P(\theta_k|\alpha_H^B) \cdot p + P(\theta_k|\alpha_L^B) \cdot (1-p)}, \qquad \text{(OA. 2)}$$

where, $P(\theta_k|\alpha_H^B)$ is the likelihood that signal θ_k would be observed given that α^B is high and $P(\theta_k|\alpha_L^B)$ the likelihood that θ_k would be observed given that α^B is low. (The denominator is thus the prior probability of observing signal θ_k.)

Given that k other countries switch, revealing the posterior probability p_k, each non-signatory will switch in Stage 3 iff $\pi^B(k, p_k) \geq \pi^A(k, p_k)$. To simplify, assume $\alpha^A = 0$. Then, early action by the k signatories will impel the others to switch provided the number of signatories is at least

$$\tilde{k}(p_k) = \frac{Cn - \beta\left(1 - e^{-\gamma/\beta}\right)}{C\alpha_k^B}, \qquad \text{(OA. 3)}$$

where $\alpha_k^B = p_k\alpha_H^B + (1 - p_k)\alpha_L^B$; otherwise, each of the $n - k$ non-signatories will not switch.[28]

In Stage 2, signatories anticipate this response by non-signatories. However, signatories have to choose in Stage 2 without knowing p_k. Their best guess is that p_k will equal p. Hence, they can do no better collectively than to switch iff the number of signatories is at least $\tilde{k}(p)$, where $\tilde{k}(p)$ is calculated by substituting $\bar{\alpha}^B$ for α^Bin (8). As the number of switchers must be an integer, signatories will switch in Stage 2 iff the number of signatories is at least as large as the smallest integer greater than or equal to $\tilde{k}(p)$. Denote this integer $\hat{k}(p)$.

How will countries choose in Stage 1? If, from country i's perspective, fewer than $\hat{k} - 1$ others participate, i is no worse off for participating, for the agreement will not enter into force whichever way i decides; if precisely $\hat{k} - 1$ other countries participate, i will gain (in expectation) by participating, as its participation will cause the agreement to enter into force, obligating i *and* all the other $\hat{k} - 1$ parties to switch and supply the public good; and if $\hat{k}$ or more countries participate, i will be worse off for participating, since, by "waiting," i can decide whether to participate in Stage 3, after observing $\alpha_{\hat{k}}^B$. As the game is one-shot, there is no cost to "waiting," only a benefit.

The model assumes that decisions are irreversible (an assumption I relax in the next section). Choices by signatories are made before $\hat{p}_k$ is revealed. Choices by non-signatories are made after $\hat{p}_k$ is revealed, but before $p_n$is revealed, where p_n denotes the posterior probability that α^B is high after all n countries have switched. The combination of irreversibility and uncertainty opens the door to two kinds of *ex post* error.

[28] The numerator of (OA.3) represents $\pi^A(0) - \pi^B(0)$, which is positive by assumption, and the denominator is the expected reduction in cost due to one more country switching. The ratio thus represents the expected number of countries that must switch in order to close the initial gap in payoffs, making switching attractive for every other country.

First, if $\tilde{k}(p_{\hat{k}}) > \hat{k}(p) \geq \tilde{k}(p_n)$, non-signatories won't switch, but they (and the original signatories) would all have been better off if they had switched. Of course, because non-signatories don't switch, they won't observe p_n and so will have no reason to feel regret. However, signatories will regret having switched previously. After all, they switched to induce non-signatories to switch, and this didn't happen. And, yet, if $\pi^B\left(\hat{k}-1, p_{\hat{k}}\right) > \pi^A(0, p_{\hat{k}})$, every signatory will be better off than if none had switched, because they switched *en masse*, supplying a significant quantity of the public good.

Second, if $\tilde{k}(p_{\hat{k}}) \leq \hat{k}(p) < \tilde{k}(p_n)$, all countries will switch and, after doing so, discover that their beliefs about α^B were optimistic. In this case, non-signatories will regret switching if $\pi^A\left(\hat{k}, p_{\hat{k}}\right) > \pi^B(n-1, p_n)$ and be glad that they switched if this inequality is reversed. (Of course, given their prior decision to switch, the original signatories will only be pleased that the non-signatories switched subsequently.) Again, though a non-signatory may regret its own prior decision to switch, because *all* non-signatories switch as a group, and because each country gains from mass switching, each country may be better off than if no country had switched.[29] Moreover, this second error is less of a concern than the first one for the simple reason that $p_k \to p_n$ as $k \to n$. The risk of regret for second movers decreases in the treaty's minimum participation level.

Proposition OA.1 (uncertain switching costs). *If switching costs are uncertain, and this uncertainty can only be reduced by making the switch, countries will join a switching treaty in tranches. A first tranche of "leaders" will participate, much as in the certainty case, provided they expect switching to be a tipping game. Whether the others follow depends on whether the leaders' costs turn out to be higher or lower than expected ex ante. Though countries may find ex post that their prior decision to switch was unwise, switchers are still likely to be better off compared to the status quo ante because switching is done en masse.*

[29] An obvious solution to the prisoners' dilemma is for all countries to commit to switching. The situation described here is similar, except that no commitment is needed. Ex ante, switching is individually rational.

OA.2 Switching, reversibility, and sunk costs

Let us now allow countries to reverse their decision to switch, but also recognize that switching may involve a sunk cost. In particular, suppose that, after switching, a country earns

$$\pi_i^B = \gamma(n-k-1)\left(1-e^{-\gamma/\beta}\right) + \gamma(k+1) - (1-s)\,C(n-\alpha^B k)/n\,, \qquad \text{(OA. 4)}$$

where $s \in [0,1]$ represents the fraction of the cost of switching that is sunk.

If $s = 0$, there are no sunk costs, and countries can only gain by switching. By switching, countries learn about the costs of switching without having to pay for this information. For example, if countries switch in the belief that switching is a tipping game only to discover later that switching is a prisoners' dilemma, each will withdraw from the agreement and revert to A. Switching ultimately fails, but the exercise in discovering whether switching pays is costless, and switchers have no reason to feel regret.[30] If s is positive but small, countries will again switch back to A, but now they will regret their prior decision to switch because of the sunk cost they incurred in switching. Finally, if s is sufficiently large, it won't pay any party to switch back to A. In this case, countries will behave in the same way as described in the previous section, where switching was assumed to be irreversible.

Incorporating the possibility of ex post reversion to A, the treaty game becomes:

> Stage 1: All countries decide whether to join a switching treaty.
>
> Stage 2: Parties decide collectively whether to switch, and then, upon doing so, reveal information about the costs of switching.

[30] Indeed, because course reversal is costless, countries have an incentive to switch even if the chances are slim that switching is a tipping game. If there are no sunk costs, learning-by-switching is costless. Sunk costs play a similar role here as in the theory of "contestable markets."

Stage 3: Non-parties decide independently whether to switch/accede to the treaty, and then, upon doing so, reveal additional information about the costs of switching.

Stage 4: Parties choose independently whether to remain in the agreement or to withdraw and switch back to A.

The novelty here is Stage 4. In this stage, every party, acting independently, will stick with B (not withdraw) iff $\pi_i^B(k) \geq \pi_i^A(k)$ or

$$\gamma(n-k-1)\left(1-e^{-\gamma/\beta}\right)+\gamma(k+1)-(1-s)\,C(n-\alpha_k^B k)/n \geq \gamma(n-k)\left(1-e^{-\gamma/\beta}\right)+\gamma k-\beta\left[1-e^{-\gamma/\beta}(1+\gamma/\beta)\right]. \quad \text{(OA. 5)}$$

Rearranging, sticking with B pays if

$$s \geq 1-\frac{\beta\left(1-e^{-\gamma/\beta}\right)}{C(n-\alpha_k^B k)/n}. \quad \text{(OA. 6)}$$

The fraction on the far-right of (OA.6) represents, from country i's perspective, the cost of reversing course (the cost of A plus the damage to i caused by i's emissions being higher when i adopts A) relative to the cost of staying the course. If the fraction exceeds one, it is rational for i to stay the course, irrespective of any sunk costs. If the fraction is less than one, it is rational for i to stay the course only if a sufficient proportion of the cost of switching is sunk.

There are again two situations of interest. First, $\hat{k}$ countries may switch in Stage 2, and no more countries switch in Stage 3. In this case, the critical value for s is found by substituting $\hat{k}-1$ for k in (OA.6). Second, all $n-\hat{k}$ non-signatories may switch in stage 3, in which case the critical value for s is found by substituting $n-1$ for k in (OA.6). In the second situation, p_n would have to be significantly lower than $p_{\hat{k}}$ for switching back to A to pay, a prospect that becomes increasingly remote as $\hat{k}$ approaches n.

As in the previous section, signatories may discover that they failed to induce others to switch, and yet be better off for having switched themselves. This will be the case for the first situation described above if $\pi^B\left(\hat{k}-1; s=0\right) \geq \pi_{NE}^A$ and for the second situation if $\pi^B(n-1; s=0) \geq \pi_{NE}^A$. Sunk costs bind countries to their previous choice, but if these conditions apply, this is all for the good.

Proposition OA.2 (reversibility and sunk costs): *When the extent of increasing returns is uncertain, and switching is reversible but entails sunk costs, switching is risky, especially for first movers. However, because signatories to a switching treaty switch en masse, supplying a significant amount of the global public good, first movers may still be better off for having made the switch.*

The prospects of tipping coupled with the sunk costs of switching play a similar role as commitment. The prospects of tipping are a lure for switching *en masse*, and sunk costs a barrier to these same countries later reversing course. Though these features may be unable to support full cooperation, they may sustain a superior outcome compared to the emission reductions game.