Centre for Climate Change Economics and Policy

An ESRC Research Centre

Grantham Research Institute on Climate Change and the Environment

Luring others into climate action: Coalition formation games with threshold and spillover effects

Valentina Bosetti, Melanie Heugues, and Alessandro Tavoni

January 2015

Centre for Climate Change Economics and Policy Working Paper No. 199

Grantham Research Institute on Climate Change and the Environment

Working Paper No. 176

E·S·R·C Economic & Social Research Council

Global Green Growth Institute

The Grantham Foundation for the Protection of the Environment

University of Leeds

LSE The London School of Economics and Political Science

The Centre for Climate Change Economics and Policy (CCCEP) was established by the University of Leeds and the London School of Economics and Political Science in 2008 to advance public and private action on climate change through innovative, rigorous research. The Centre is funded by the UK Economic and Social Research Council. Its second phase started in 2013 and there are five integrated research themes:

1. Understanding green growth and climate-compatible development
2. Advancing climate finance and investment
3. Evaluating the performance of climate policies
4. Managing climate risks and uncertainties and strengthening climate services
5. Enabling rapid transitions in mitigation and adaptation

More information about the Centre for Climate Change Economics and Policy can be found at: http://www.cccep.ac.uk.

The Grantham Research Institute on Climate Change and the Environment was established by the London School of Economics and Political Science in 2008 to bring together international expertise on economics, finance, geography, the environment, international development and political economy to create a world-leading centre for policy-relevant research and training. The Institute is funded by the Grantham Foundation for the Protection of the Environment and the Global Green Growth Institute. It has nine research programmes:

1. Adaptation and development
2. Carbon trading and finance
3. Ecosystems, resources and the natural environment
4. Energy, technology and trade
5. Future generations and social justice
6. Growth and the economy
7. International environmental negotiations
8. Modelling and decision making
9. Private sector adaptation, risk and insurance

More information about the Grantham Research Institute on Climate Change and the Environment can be found at: http://www.lse.ac.uk/grantham.

Luring others into climate action: Coalition formation games with threshold and spillover effects

Valentina Bosetti[*,§], Melanie Heugues[*], Alessandro Tavoni[†]

This Draft December 2014

Abstract

We study the effect of leadership in an experimental threshold public 'bad' game, where we manipulate both the relative returns of two investments (the more productive of which causes a negative externality) and the extent to which the gains from leadership diffuse to the group. The game tradeoffs mimic those faced by countries choosing to what degree and when to transition from incumbent polluting technologies to cleaner alternatives, with the overall commitment dictating whether they manage to avert dangerous environmental thresholds. Leading countries, by agreeing on a shared effort, may be pivotal in triggering emission reductions in non-signatories countries. In addition, the leaders' coalition might also work as innovation and technology adoption catalyzer, thus producing a public good (knowledge) that benefits all countries. In our game, players can choose to tie their hands to a cooperative strategy by signing up to a coalition of first movers. The game is setup such that as long as the leading group reaches a pivotal size, its early investment in the externality-free project may catalyze cooperation by non-signatories. We find that the likelihood of reaching the pivotal size is higher when the benefits of early cooperation are completely appropriated by the coalition members, less so when these benefits spillover to the non-signatories. On the other hand, spillovers have the potential to entice second movers into adopting the 'clean' technology.

Key words

Climate change; international cooperation; R&D spillovers; threshold public goods game; coalition formation game; climate experiment.

§ Bocconi University, Milan, Italy. Email: valentina.bosetti@unibocconi.it
* Fondazione Eni Enrico Mattei, Milan, Italy
† Grantham Research Institute, London School of Economics, London WC2A 2AZ, England. Email:a.tavoni@lse.ac.uk.

The authors are grateful to Celse Jérémy for his invaluable help. We also thank Davide Rossi, Laura Dell'Acqua and Michele Peruzzi for assistance in running the experiments. In addition we are grateful to seminar participants at ASFEE, WCERE 2014, ISEE 2014 and at the FEEM workshop on "Climate Change and Public Goods", for their constructive comments. Bosetti gratefully acknowledges the financial support provided by the CASBS and the ERC-2013-StG 336703-RISICO. Heugues acknowledges funding on the Marie Curie fellowship INTCOP. Tavoni acknowledges the financial support of the Grantham Foundation for the Protection of the Environment, as well as the ESRC Centre for Climate Change Economics and Policy, which is funded by the UK's Economic and Social Research Council. Logistic support from the Bocconi Experimental Laboratory in the Social Sciences (BELSS) for hosting our experimental sessions is kindly acknowledged.

1. Introduction

Large scale cooperation on the provision of public goods is essential to overcoming many problems in modern and past societies, such as the spread of infectious diseases, resource overharvesting leading to either distributive inequalities (e.g. when countries share international waters) or stock collapse (e.g. in fishery exploitation), as well as ocean acidification, climate change and other global environmental problems. These problems have several features in common: (i) they are subject to sudden transitions from more benign states to harmful ones (tipping points); (ii) addressing them requires widespread cooperation in the face of individual incentives to refrain from it and 'free ride' on the effort of others; and (iii) the prospects of success hinge on the willingness of some to lead by example. The joint effect of these characteristics is appalling; if all actors wait for the others to show leadership, catastrophic and irreversible regime shifts may occur (Alley et al. 2003; Kriegler et al. 2009; Lade et al. 2013; Lenton et al. 2008). Providing the above public goods thus presents a challenge in terms of conciliating rational choice at the individual level with pro-social behavior. Here we investigate experimentally whether leadership and innovation diffusion can facilitate addressing such 'wicked problem'. Specifically, will the establishment of an institution where a coalition of agents restricts itself in the use of a polluting technology entice others to follow suit, given that the group collectively risks high losses if cooperation is insufficient?

While the experiment is framed neutrally[1], we will use avoidance of dangerous climate change as an illustration throughout the paper. To capture (i) and (ii), we consider a discrete public bad. The existence of a known threshold simplifies the challenge of reaching a meaningful agreement in negotiations, by transforming the underlying prisoner dilemma's game into one of coordination (Barrett and Dannenberg, 2012). Coordinating between two Pareto-ranked equilibria is an easier task than escaping the trap of a unique equilibrium where the dominant strategy is to defect and gamble on the effort of others or the clemency of Nature. However, even in the presence of a known threshold with the potential to trigger a catastrophe, coordination can be difficult, especially when the parties have different stakes in the game (Tavoni et al., 2011).

The rationale for focusing on (iii) is that in order to trigger breakthrough advancements in clean energy technologies, which are necessary for a transition to a low carbon economy that is compatible with economic

[1] The subjects were confronted with choices among two investment projects, labelled A and B. Compared to a frame that stressed the moral imperative for action (e.g. to reduce global pollution), this choice might induce less collaborative behaviour (Liberman, Samuels and Ross, 2004). What we are interested in is treatment effects rather than levels, so the framing effect on absolute levels of cooperation should wash away. Furthermore, unframed experiments have the advantage of being less prone to confounding effects originating from the frame.

growth, major efforts will be needed in both research and development (R&D) and large-scale deployment of new technologies, as well as in infrastructure development. One can think of electric vehicles as one obvious example of the magnitude of the required investments. Each country could invest independently in the required effort. However, this could still be insufficient to bring into reality some of the new technologies at a large enough scale. At best, this funding scheme will result in inefficient and redundant use of research funding. Countries (or companies innovating in those countries) might instead resort to common efforts, standardization, and development of gateway technologies that spark the formation of networks and allow large scale adoption of new technologies.

Innovation and technology cooperation has been frequently suggested as a possible way out of the negotiations deadlock (Carraro and Siniscalco, 1995; Barrett, 2003, 2006; Golombek and Hoel, 2004). This is the first objective of the present analysis, i.e. investigating the implications of linking coalition efforts with the ancillary benefits stemming from coordinated innovation.

The idea of a collective pursue of innovation, however, brings a new externality into the analysis. If cooperation on clean innovation hinges on partially sharing the associated collective burden (and benefiting from its yields), what about those that were not part of the agreement in the first place? Depending on the nature of the technologies, non-participants could, in principle, be excluded by such benefits, for example through a system of exclusive property rights. But would this be in the interest of the cooperating group? This is the second focus of our analysis, namely the role of spillovers.

Technology transfers within coalitions and between signatories and non-signatories have been documented to occur through climate policies linkages (see for example the work by Dechezleprêtre et al. (2008) and Seres et al. (2009) on technology transfers through the Clean Development Mechanism), but also simply because of trade flows, multinational enterprises, and skilled-labor mobility (Eaton and Kortum, 2001, 2006; Keller, 2010). Although empirical studies can hardly be definitive on the subject, technological transfers have been highlighted in the theoretical literature as one of the mechanisms that can in principle generate negative leakage (Golombek and Hoel (2004) and Van der Werf and Di Maria (2008)). Negative leakage occurs when countries that have not signed an environmental agreement reduce pollution in response to the efforts of an environmental coalition. This literature suggests that in principle it could be profitable for the coalition to let non-signatories benefit from the innovations brought about by it.

The prospects for scaling up climate cooperation nucleating at a small scale, although not necessarily hinging on the mechanisms of technological spillovers, has received increasing attention in related theoretical work (Ostrom, 2009; Dietz, et al. 2012; Sterner and Damon, 2011; Vasconcelos, et al., 2013; Tavoni, 2013). Network diffusion of behaviors and technology adoption may play an important part in catalyzing cooperation, since adoption by one agent often increases the likelihood that others will become aware of their existence and potential benefits relative to the status quo. Many studies have shown that that mutually reinforcing choices lead to accelerating diffusion of a behavior or to the adoption of a technology once a tipping point has been reached (Granovetter, 1978; Watts, 2002; Weir, 2004). Heal and Kunreuther (2012) focus instead on coordination in games with strategic complementarity, by resorting to the concept of 'tipping set', i.e. "a subset of agents who by changing from the inefficient to the efficient equilibrium can induce all others to do the same". They argue that international climate agreements have these characteristics, and motivate the theory with two often mentioned examples of strategic complementarity: the replacement of leaded gasoline with unleaded gasoline, and the phasing out chlorofluorocarbons through the Montreal Protocol on Substances that Deplete the Ozone Layer. Both examples show how unilateral action initiated by a subset of actors (in the United States) prompted others to follow suit immediately after[2]. This body of work suggests that unilateral action by a subset of agents might hold promise for promoting widespread cooperation notwithstanding the threat of free riding.

In the present paper, we investigate experimentally the role of increasing returns to coalition size (mimicking increasing returns to scale in innovation and adoption of clean technologies), as well as the implications that proprietary versus open knowledge policies might have. We employ a threshold public 'bad' game that is setup to test how these mechanisms play out in deterring or incentivizing players to be part of a coalition of early investors, or in responding to the coalition if they decide to stay out.

This experiment departs from standard public goods games in at least three ways: the presence of a threshold, which transforms it in a game of coordination with two Pareto-ranked equilibria (*tipping point avoidance* and *gamble*, as explained below); the possibility to form a coalition of Stackelberg leaders who invest in a

[2] Experimental work has also shed light on the role of leading by example in facilitating the provision of public goods (Moxnes and van der Heijden, 2003; Levati et al., 2007). Using a public bad experiment, Moxnes and van der Heijden (2003) ask themselves the following: "With regard to global or regional environmental problems, do countries that take unilateral actions inspire other countries to curtail emissions as well"? They find "a small but significant effect of a leader setting the good example", provided that the example is sufficiently 'good' (i.e. leader investments in the public bad are sufficiently low). Relatedly, İriş et al. (2014) find that contributions to a threshold public good drop when the investment decision is delegated to an appointed leader. This effect is attributable to the fact that delegates appear to focus on the lowest contribution level suggested by non-delegates (rather than the highest or average suggestions). Hence, negative examples can be detrimental to cooperation.

technology which is socially superior, but individually more costly; and the existence of technological spillovers that may be appropriated by the coalition or may diffuse to non-members.

We show that a narrow focus on targets is unlikely to be effective in catalyzing climate cooperation, since it exposes cooperators to the 'tyranny of free riders' refusing to take on sufficiently ambitious mitigation efforts. Such well-known negative result is alleviated when i) there exist increasing returns to entering in a coalition that are completely appropriated by the coalition and that are high enough to attract a pivotal group of participants; or ii) the fringe can partake in the benefits generated by the coalition, thus acting proactively even though from outside the coalition. This finding casts new light on the problem, by highlighting the game-changing potential of linking a climate agreement with technological agreements and the strategic implications of restricting access to the new technology.

Before detailing the experimental design in Section 3, we describe the main features of the game in the next section. Section 4 discusses main findings of our experiments and Section 5 draws some conclusive remarks.

2. The game

In this section we introduce the set-up of the game. We first present the dilemma with the main notation and constraints. Then we provide the stages of the game, to shed light on how coalition formation and technological cooperation can help coordination. Finally, we solve the game by backward induction.

2.1 The threshold public bad game

Consider N symmetric subjects playing a linear public bad game with a threshold. Each of them has an initial endowment e and decides how much to allocate between a high return but socially costly Project A (public bad) and a lower-return investment in an alternative project which does not cause negative externalities, Project B. The endowment is thus split between x_A and $x_B = e - x_A$. Investing in Project A (B) gives a private return of r_A (r_B). Returns on Project A are larger than returns on Project B, $r_A > r_B > 0$, but Project A has also a negative external effect: each unit invested in A yields a negative return of c_A to all subjects.

In addition to this traditional negative externality game, the group's aggregate investment determines whether a 'tipping' point has been reached. Namely, a threshold *T* determines the maximum safe collective investment in A. This threshold is common knowledge and can be interpreted as admissible global CO_2 concentrations that are compatible with full enjoyment of private earnings. To make the problem relevant, this safe level has to lie below the maximal public bad investment capacity (Ne). Players thus retain their earnings with certainty

(*tipping point avoidance*) if $Nx_A \leq T < Ne$; otherwise, with probability p they will be left with $q \in [0,1)$ of their private earnings (*gamble*).

Subjects' payoff function then takes the form:

$$\begin{cases} \pi(x_A, x_B) = r_A x_A + r_B x_B - c_A \sum x_A, & if \ \sum x_A = X_A \leq T \\ \pi(x_A, x_B) = (1-p)[r_A x_A + r_B x_B - c_A \sum x_A] + pq[r_A x_A + r_B x_B - c_A \sum x_A], & if \ \sum x_A = X_A > T \end{cases} \quad (1)$$

Where N, $e \in R^+$, x_A and $x_B \in [0, e]$ and $c_A < r_A - r_B < c_A N$. The first inequality means that the private net return of Project A is larger than the return of Project B: $r_A - c_A > r_B$; and the second inequality means that the individual opportunity cost of investing in the clean technology B is lower than the social marginal cost of pollution $c_A N$. The latter inequality is in line with the existing empirical evidence (Stern, 2007; IPCC, 2014[3]).

The social optimum entails that all players refrain from investing in A altogether.[4] In this case each subject gets $\pi(0, e) = r_B e$. But this is not an equilibrium, as each player has an incentive to deviate. By increasing x_A by one unit, any individual can get $\pi(1, e-1) = r_A + r_B(e-1) - c_A$ (while others get $\pi(0, e) = r_B e - c_A$). As long as the net return of Project A is larger than that of Project B the deviation pays off . Hence the dilemma arises as each individual strictly prefers invest everything in A, assuming all others refrain from investing. Obvioulsy, as more subjects follow this line of reasoning, the lower is everyone's expected payoff (because of the gradual negative externality term $c_A \sum x_A$ as well as of the stochastic implications of crossing the threshold). Risk-neutral players will either coordinate on threshold avoidance, or disregard the externality and make the most from investment in A. These two symmetric Nash equilibria correspond to $\underline{x}_A = T/N$ and $\overline{x}_A = e$ and are Pareto ordered. We denote with $\overline{\pi}$ the payoff associated with the tipping point avoidance ($\underline{x}_A = T/N$), and call $\underline{\pi}$ the payoff obtained when putting all eggs in Project A ($\overline{x}_A = e$).

2.2 Making coordination happen: coalition formation with technological cooperation

into capture the element of leadership we introduce a membership stage where players can opt to be part of a coalition. Being part of a coalition means signing up to a pre-specified investment strategy that is linked to the number of individuals who sign the agreement. In particular, the smaller the coalition, the more effort each coalition member is required to do in terms of constraining her/his investment in Project A. In what follows, we identify by s the number of members signing up to the coalition. For any $s \in [2, N-1]$, each coalition member

[3] More precisely see Working Group III, Chapter 10 – Mitigation: potential and costs, section "Social and environmental costs and benefits" pp. 851.

[4] This choice is made to mimic the nature of the climate change problem: in order to keep temperature below the agreed 2°C, global emissions will need to be nil by mid-century (IPCC 5th AR WGIII Summary for Policy Makers, 2014).

invests less than the equal share guaranteeing threshold avoidance: $x_A^s = X_A^s/s < T/N$. [5] Following the membership stage, those opting not to be in the coalition, non-members $n \in [N-s]$, are free to choose their investment given information on the size and aggregate investment of the coalition.

The resulting total investment in A can be then expressed as $X_A = X_A^s + X_A^n = \sum_{i=1}^{s} x_A^i + \sum_{i=s+1}^{N} x_A^i$. This determines the group's performance with respect to the threshold *T*, as well as the externality cost, $c_A X_A$, both affecting each individual's payoff.

We now explore the case where returns to Project B increase with the size of the coalition, thus reducing the returns gap between the two investment alternatives. Inspired by the literature on multi-issue bargaining (Schelling, 1960), we assume that members of the coalition, by curtailing investments in Project A, also increase the productivity of Project B. Adoption of new technologies typically entails several externalities and the rational of our set up is that, by getting together, players are leveraging on the coalition size to reduce those externalities, thus reducing the return wedge between the two technologies. The larger s, the greater this positive externality on Project B is. Hence $r_B(1+sI)$ is the increased return to B resulting from a coalition of size *s*, where $I \in [0,1)$ is the percentage rate of technological improvement. [6]

In order to account for this positive externality, the payoff to its beneficiaries takes now the form:

$$\begin{cases} \hat{\pi}(x_A, x_B) = r_A x_A + r_B(1+\boldsymbol{sI})x_B - c_A X_A, & if\ X_A \le T \\ \hat{\pi}(x_A, x_B) = (1-p+pq)[\, r_A x_A + r_B(1+\boldsymbol{sI})x_B - c_A X_A], & if\ X_A > T \end{cases} \quad (2)$$

We investigate two alternative setups for what concerns the implications for the fringe of technological cooperation among signatories. In the first case, the positive externality is appropriated by coalition members' only, with non-signatories payoffs given by (1). In a second setup we assume that this positive externality diffuses to the fringe as well, whose payoff then also follows equation (2). We refer to these as the *no spillover* and the *spillover* cases, respectively.

We now move on to the discussion of the actual parameters and treatments utilized in the laboratory, while we refer to Appendix 1 for the equilibrium solution of the two stage game.

[5] Under this assumption, average investment by non-members above T/N can still be compatible with avoiding the probabilistic loss triggered when exceeding T.

[6] To keep the social dilemma set up, we impose that $r_B(1+NI) < r_A - c_A$, i.e. even when all subjects cooperate, the net return of Project A remains larger than the increased return of Project B.

3. Experimental design and hypotheses tested

Groups of $N = 7$ subjects face an unframed game, in neutral language. Each player is endowed with $e = 50$ experimental currency units (1 ECU corresponding to 0.05 Euros), which are to be entirely allocated between two projects, A and B. Each unit invested in A yields an individual return $r_A = 10$ and causes a cost $c_A = 1$ to each group member; investment in B yields a lower return $r_B = 6$, but carries no external cost, $c_B = 0$. The threshold is set at $T = 105$ECU$= 30\%\ N * e$, meaning that for a group to avoid the probabilistic losses, it has to limit collective investment in A to at most 30% of total endowment (or equivalently invest at least 70% in externality-free Project B). Otherwise, all subjects in a group face a 50% probability of losing their earnings: $p = 0.5$ and $q = 0$.

With regard to the increased competitiveness of Project B resulting from coalitional investments in it, we test the following cases: $I = \{0\%;\ 2\%;\ 7\%\}$. We will refer to *I*=2% as the condition with Low Innovation returns, and to *I*=7% as the case with High Innovation returns.[7] We also manipulate whether the returns to innovation are appropriated by coalition members only, or benefit the fringe as well (Spillover condition).

We test for the effect of four conditions, yielding the five treatments (and the control one) which are sketched out in Table 1. The threshold public bad game without coalition formation stage serves as benchmark (T0). It consists solely of the investment decision stage, where players simultaneously and independently choose their investment in Project A (and which determines the residual, if any, to be invested in B). T1 captures the implication of the addition of a membership stage with coalition formation, while the remaining four treatments differ in the returns to innovation (T2 and T4) as well as in who its beneficiaries are(T3 and T5).

	COALITION	$I = 2\%$	$I = 7\%$	**SPILLOVER**
T0 (Threshold Public Bad Game)				
T1 (Coalition)	✓			
T2 (Coalition & Low Innovation)	✓	✓		
T3 (Coalition & Low Innovation with Spillover)	✓	✓		✓
T4 (Coalition & High Innovation)	✓		✓	
T5 (Coalition & High Innovation with Spillover)	✓		✓	✓

Table 1. Features of the different treatments

[7] Recalling (2), for positive *I* and coalition size *s*, the return to B increases to $r_B(1 + sI)$.

In particular, subsequent treatments explore the implications of increasing returns to coalition participation. Returns from Project B are increased proportionally to the coalition size, hence reducing the return wedge between the two investments (Appendix 2 reports the full parameterization for different treatments). However, while in T2 and T4 only coalition members benefit from this increase in Project B returns, in T3 and T5 every player can benefit.

As an illustration, Table 2 reports the information given to players at the membership stage for treatment T1 (additional information on implications for Project B returns and whether they were available to coalition members only or to all players were provided under other treatments and are summarized in Appendix 2). This includes levels of investment in Project A each member is going to be tied to and how they vary depending on the resulting coalition size.[8] The table also reports, for each coalition size, the remaining allowed investment in Project A for non-members which is consistent with investments not exceeding the threshold, both at the group (penultimate row) and individual level (assuming symmetric behavior, bottom row).[9] This information is also provided again to non-signatories at the investment strategy stage, together with information concerning the actual size of the coalition that has formed.

	Number of participants joining a coalition (s)					
	7 (all)	**6**	**5**	**4**	**3**	**2**
Investment in Project A for each member (ECU)	15	13	11	9	7	5
Aggregate investment in Project A by the coalition (ECU)	105	78	55	36	21	10
Amount left to be invested before reaching 105 ECU	0	27	50	69	84	95
Corresponding symmetric individual investment not to exceed 105 ECU for non-members	0	27	25	23	21	19

Table 2 – T1 (Coalition Only): Information provided to players during the game

Once information about investments by non-signatories are collected, each player is informed on the resulting aggregate investment in A, whether the threshold has been crossed or not, and her/his final payoff (conditional on the 50% probability for instances where the threshold has been crossed). Finally, for instances where the threshold is crossed a virtual coin is tossed and the effective payoff is communicated to the group. The full Table containing information for other treatments, including the potential improvement to Project B returns induced by coalition size, are reported in Appendix 2.

[8] For the limit case of the grand coalition, the parameterization replicates the symmetric cautious equilibrium.

[9] In the treatments with coalition (T1 to T5), subjects were informed that a coalition only forms if at least 2 participants in the group choose to join a coalition and that members of a coalition cannot alone guarantee that the sum of all investments in Project A stays below 105 ECU, except when all 7 join the coalition.

The experiment was conducted between May 2013 and February 2014 in the BELSS Lab at the Bocconi University (Italy). We recruited 434 subjects (respectively 70, 84, 63, 70, 77 and 70 subjects in Treatment 0, 1, 2, 3, 4 and 5), corresponding to between 9 and 12 independent group observations per treatment. No subject participated in more than one session. Sessions lasted between 60 and 90 minutes and were run on visually isolated computer terminals. For programming the interactive games, we used the software z-tree (Fischbacher 2007). Subject earned 13.20 Euros on average.

At the beginning of a session, written instructions with a neutral frame (context and language of the experiment abstracted from interpretations of any sort) were provided to the subjects. Before starting the experiment, subjects completed a comprehension questionnaire to ensure that they fully understood all the procedures.

Each session consisted of 2 practice rounds and 8 independent rounds, i.e. the subjects played the game described above 10 times. The subjects were informed that only the 8 independent rounds would be considered to determine the final payout. At the beginning of each round subjects were randomly assigned to groups of seven and were given an endowment $e = 50$ to be used in the investment decision. They were not aware of whom they were grouped with and each subject was not matched up with the same other 6 participants for more than a single round. At the end of each round, the participants were informed of their (potential) earnings for the round, given their choices and the choices made by the other 6 group members.

At the end of the experiment, subjects were paid according to one randomly selected round (out of the 8 rounds). Payments were settled at the end of the experiment in cash. Since subjects were informed that the round to be selected for payment was determined randomly, and could be any of the non-practice rounds, it is reasonable to expect that they played as if each round was payoff-consequential. Before moving to the results, let us discuss the set of hypotheses that we set out to test.

The first conjecture refers to the potentially positive effect that leadership signaling per se may have in catalyzing cooperation. By comparing Treatments 0 and 1, we are able to assess whether the opportunity to form a coalition changes the aggregate investment behavior. The experimental literature seems to confirm the theoretical prediction of small coalitions that only partially internalize the externality by investing slightly more in the public good than in voluntary contribution mechanisms (Dannenberg et al., 2014). Certain design features, such as imposing a minimum participation rule or introducing an endogenous rule for determining coalitional contributions, increase cooperation. In our game, where the latter features are absent, we expect that Treatment 1 will induce the formation of small coalitions and little switch away from the public bad.

The second conjecture is that increasing the returns of being in the coalition (by increasing returns to Project B to coalition members only) increases the willingness to be part of the coalition. Treatment 2 offers coalition members an increased return to Project B that is proportional to the number of participants and this effect is stronger in Treatment 4. We expect the coalition to be largest in T4, followed by T2 and T1.

As far as the fringe is concerned, two conflicting mechanisms are at work. On the one hand, the theory on leadership suggests that larger coalition sizes may induce pro-social behavior in non-signatories as well by making the target within reach (Kosfeld et al., 2009). Therefore, in response to larger coalitions in T4 and T2, a reduction in average Project A investment by the fringe may occur. On the other hand, as the constraint on investment in A by non-coalition members becomes less binding as the coalition size increases (see Table A2), free riding incentives will pull in the opposite direction. The third conjecture is that these two opposing effects will (partly) cancel out and the overall group success in avoiding the tipping point will be mostly determined by the number of signatories. Combining the second and third conjecture, we hypothesize that success will be higher in T4, followed by T2 and T1.

A different set of incentives comes into play in treatments where innovation benefits spillover to non-coalition members as well (T3 and T5). First, fewer subjects will sign up early on to the coalition, compared to treatments where the benefits are appropriated by the coalition only (due to the larger return gap between A and B). This forms the basis for our fourth conjecture, namely that individuals will response to the spillover condition by less likely enrolling in the coalition (coalition size in T3 should be smaller than in T2, and also in T5 compared to T4).

Ceteris paribus, reduction in coalition size should in turn reduce both the pivotal and the free riding effects, again with inconclusive effects. However, under spillovers the fringe benefits from a smaller opportunity cost of investing in Project B due to the reduced wedge on returns. The fifth conjecture is thus that when the increased returns to B spillover to non-members, the fringe will invest less in Project A (comparing T3 to T2, and T5 to T4). The fourth and fifth conjectures point in opposing direction with regards to overall group success in tipping point avoidance, so we resort to empirics to establish which effect (shrinking coalition size versus less uptake of the polluting technology by the fringe) dominates.

4. Results

In our analysis, three are the crucial indicators of group performance: the size of the coalition (s), the frequency of threshold crossing (T^+) and the total investment in the public bad

(X_A), which determines the group's distance from the threshold (by how much investments are below or above *T*). The last metric is relevant as it captures the gradual component of the external costs of investing in A. In Table 3 we report the summary statistics for these key indicators across all treatments. The first four rows of the table recall the basic assumptions for each of the treatments, while the bottom six report the statistics concerning the basic indicators of performance. All numbers reported are averages over all periods.

Table 3. Descriptive statistics of results

ASSUMPTIONS	**T0 (Threshold Public Bad Game)**	**T1 (Coalition)**	**T2 (Coalition & Low Innovation)**	**T3 (Coalition & Low Innovation with Spillover)**	**T4 (Coalition & High Innovation)**	**T5 (Coalition & High Innovation with Spillover)**
Number of stages	1	2	2	2	2	2
Coalition effect on costs	-	None	Internal only	Internal and External	Internal only	Internal and External
Repetition of game with random clustering	8	8	8	8	8	8
***I*: Increased return to B per member ($s \geq 2$)**	-	0%	2%	2%	7%	7%
DESCRIPTIVE STATISTICS	**T0**	**T1**	**T2**	**T3**	**T4**	**T5**
s ; $N - s$ (coalition size; fringe size)	-	1.7 ; 5.3	2.5 ; 4.5	1.8 ; 5.2	4.3 ; 2.7	2.5 ; 4.5
X_A (Total Investment in Project A)	191	166	154	136	117	139
X_A^s (Total Investment in Project A by the Coalition)	-	6.5	6.8	6.3	9.7	7.4
X_A^n (Total Investment in Project A by the Fringe)	27.3	28.8	29.4	23.6	27.1	25.6
T^+ (Groups that exceeded T)	89%	83%	81%	68%	56%	59%
By how much above *T*	82%	58%	47%	30%	11%	32%

Let's begin with the most conservative criterion to assess treatment effects in this game, failure to avoid crossing the threshold. Figure 1 Shows that while in T0, T1 and T2 the vast majority of groups fails to stay below the target (failure rates above 80%), in T3, T4 and T5 failure rates drop to around 60%. Similarly, total investments in Project A in this second group of treatments is significantly lower, as summarized in Table 3. We will devote the remaining of this section to the analysis of the mechanisms underlying these results.

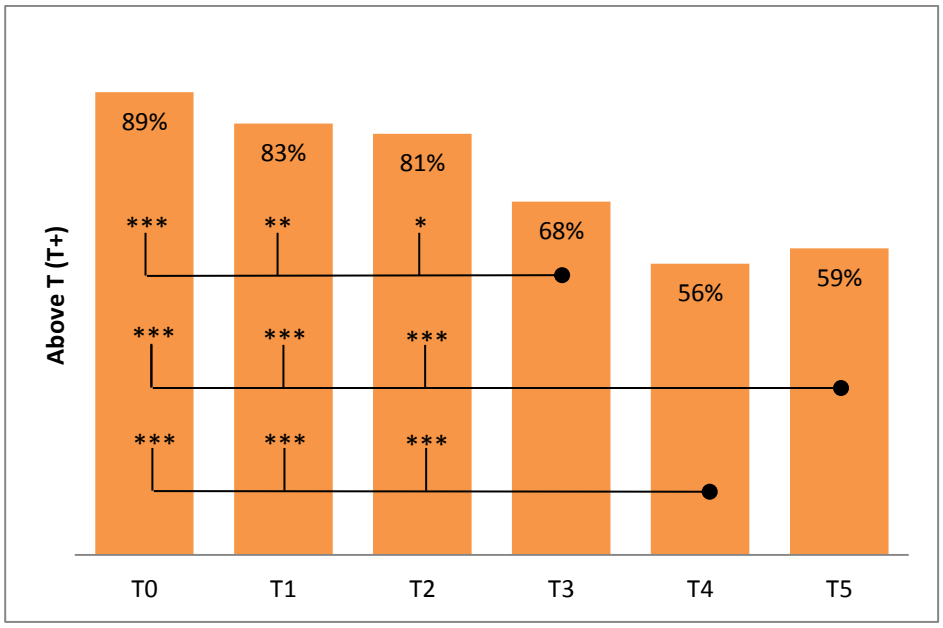

Figure 1: Percentage of groups that exceeded the threshold *T*. Lines emphasize statistical differences across treatments (*: p<0.01, **: p<0.05, *: p<0.1)**

To summarize the treatment effects along many metrics, in Table 4 we map the differences across conditions in five indicators discussed thus far. Namely, in addition to total investment in A (X_A) and coalition size (s), we compare total and average fringe investment in A (X_A^N, x_A^N, respectively), to get a sense of the relative implications that different incentives have on non-signatories' behavior.

	T1 (Coalition)	T2 (Coalition & Low Innovation)	T3 (Coalition & Low Innovation with Spillover)	T4 (Coalition & High Innovation)	T5 (Coalition & High Innovation with Spillover)
T0 (Threshold Public Bad Game)	X_A < **	X_A < ***	X_A < *** T^+ < ***	X_A < *** T^+ < ***	X_A < *** T^+ < ***
T1 (Coalition)		s > ***	X_A < *** T^+ < ** X_A^N < *** x_A^n < ***	s > *** X_A < *** T^+ < *** X_A^N < ***	s > *** X_A < *** T^+ < *** X_A^N < *** x_A^n < ***
T2 (Coalition & Low Innovation)			s < *** X_A < ** T^+ < * x_A^n < ***	s > *** X_A < *** T^+ < *** X_A^N < *** x_A^n < **	X_A < *** T^+ < *** X_A^N < ** x_A^n < ***
T3 (Coalition & Low Innovation with Spillover)				s > *** X_A < ** X_A^N < *** x_A^n > ***	s > ***
T4 (Coalition & High Innovation)					s < *** X_A^N > *** x_A^n < **

Table 4. Statistical differences in treatments (column versus row), *** p<0.01, ** p<0.05, * p<0.1. Reported entries are at least significant at the 10% level according to the two-sample Wilcoxon rank-sum test. The table is read starting from the column treatment (e.g. total investment in A, X_A,is significantly lower in T1 than in T0, and coalition size s is significantly larger in T2 than in T1).

Comparing T0 with T1 (first row and first column in Table 4), we find that the option of signaling leadership is helpful as total investment in A is significantly reduced in T1, but not sufficiently to significantly reduce the probability of crossing the threshold (which happens 83% of times in T1). This finding confirms, in a discrete public bad setting, the theoretical result advanced by Barrett (1994) and Carraro and Siniscalco (1993) that the option to form coalitions with voluntary participation leads only to modest improvements. This pessimistic result has been confirmed experimentally for linear public goods[10], but to our knowledge the present experiment is the first to show the limited gains brought about by voluntary coalition participation in a setting where lack of restraint in the use of a technology (the public bad) causes negative and potentially catastrophic

[10] See Dannenberg et al., 2014 for a recent experiment, and references therein for earlier experimental work on coalition formation.

externalities. As we will see below, coalition formation becomes much more consequential (as measured by the metrics in Table 4), once it is interacted with the other conditions.

In order to test our second conjecture, we need to compare T1, T2 and T4. We find that individuals' propensity to join a coalition responded to the innovation incentives as expected. The larger the innovation benefits, the larger the coalition size (as evident from the differences in T2 and T4 relative to T1 and T2, respectively).

How did the fringe respond to the increasingly larger leading group in T2 and T4? Is the leading by example effect prevailing, or is it free riding? Alternatively, it may be that the two effects largely balance each other out, as laid out in our third conjecture. In our experimental sample, when the innovation benefits are low, the two effects cancel each other out: the coalition is larger in T2 with respect to T1, but there is no statistical difference in either the overall investment in Project A, nor in the failure rate in avoiding the tipping point. However, conjecture three is not confirmed when looking at larger benefits from innovation (T4). Under T4, the incentives to participate in a voluntary coalition are highest, and the subjects responded by signing up to it more frequently than in other treatments: the average number of participants increases to four out of seven. The implication of a larger coalition is that the leadership effect prevails over the free riding effect. Both the total investment in Project A and, more interestingly, the average fringe investment are significantly lower in T4 than in T2. This suggests that for leadership to be effective a critical mass is necessary. The resulting implication is that, overall, threshold crossing is significantly lower in T4 compared to any of the other treatments investigated so far (T0, T1 and T2).

Conjecture four is confirmed: treatments where innovation benefits spillover to the fringe (T3 and T5) imply a significant reduction in the coalition size, with respect to the equivalent treatment with no spillovers (in T3 s is smaller than in T2 and, similarly, s is smaller in T5 than in T4). Note however that while this difference is statistically significant in both cases, the drop in coalition size that is witnessed when comparing T5 to T4 is much larger in magnitude (with average coalition dropping from 4.3 in T4 to 2.5 in T5, a value that is comparable to the small coalition size observed in T1 and T2). This bears important implications for the overall reaction of the fringe, which is a key determinant of the threshold crossing indicator (T^+).

Lastly, spillovers reduce the average fringe investment as suggested in conjecture five. However, what this implies for the whole group is influenced by the actual size of the fringe. Let us first discuss the behavior of non-signatories comparing T3 with T2. Spillovers reduce the coalition size, but given that this effect is only marginal, the negative implications of this reduction are more than compensated by the proactive behavior that spillovers induce in the fringe investment strategy. The average Project A investment by each fringe member, x_A^n, is lower in T3 than in T2 (and in T1), and the probability of crossing the threshold is reduced. A

different balance of these effects is at play when comparing T5 with T4: although x_A^n is again smaller in T5, this is not enough to compensate for the large drop in coalition size relative to T4 (from 4.3 to 2.5, on average).

The results from pairwise comparisons are confirmed by the linear regressions reported in Table 5, which also allows us to control for potential learning effects, as well as other individual fixed effects.

	s	T+	T+	X_A^n	x_A^n
T2	0.587***	-0.0278	0.0675	-18.88**	0.500
	(0.195)	(0.0690)	(0.0620)	(8.861)	(1.136)
T3	0.0146	-0.158**	-0.156***	-29.79***	-5.182***
	(0.189)	(0.0670)	(0.0596)	(8.604)	(1.098)
T4	2.285***	-0.277***	0.0945	-81.91***	-1.745
	(0.184)	(0.0653)	(0.0682)	(8.388)	(1.081)
T5	0.677***	-0.246***	-0.136**	-35.83***	-3.228***
	(0.189)	(0.0670)	(0.0605)	(8.604)	(1.098)
s			-0.162***		
			(0.0157)		
Constant	2.494***	0.622***	1.027***	125.2***	25.34***
	(0.206)	(0.0730)	(0.0758)	(9.381)	(1.211)
Round Dummy	Yes	Yes	Yes	Yes	Yes
Observations	416	416	416	416	412
R-squared	0.349	0.108	0.296	0.262	0.150

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5. OLS regression on Group level data. Reference treatment is T1. T0 is excluded from the analysis

The first column looks at the impacts on coalition size, whereas the second and third models predict threshold crossing, while the last two columns report model results for total and average fringe investments. The coalition size is significantly larger in treatments T2, T4 and T5 than in T1 (1% significance) but it is not in T3 (T0 is excluded as it does not feature coalition formation. However, the probability of crossing the threshold is significantly lower in T3, due to the behavior of the fringe. Looking at total and average investments one can clearly notice how self-restraining behavior by the average fringe member is maximized in T3 (last column in Table 5). T4, on the other hand, implies the largest coalition size, and hence the minimum total investment in Project A by the fringe. This, as noted above, follows from the increased participation in the coalition rather than from the fringe behavior (whose average investment in A remains largely unchanged).

We further search for individual features that might influence either the individual amounts invested in A or the choice to be in the coalition or the level of investments in Project A for those participants that are not part of the coalition (using a random effect robust regression).

	x_A	In coalition	x_A^n
T2	-7.63***	0.21***	-3.86
	(2.18)	(0.05)	(2.45)
T3	-9.16***	0.11**	-8.50***
	(2.02)	(0.04)	(2.23)
T4	-11.56***	0.42***	-5.46***
	(1.76)	(0.04)	(1.94)
T5	-8.80***	0.21***	-5.87***
	(1.98)	(0.04)	(2.07)
Gender (0-Male; 1-Female)	2.45**	-0.05**	2.39*
	(1.08)	(0.03)	(1.23)
Understand	0.10**	-0.01**	0.09*
	(0.04)	(0.00)	(0.05)
Risk aversion (life metric)	0.48**	-0.00	0.64***
	(0.21)	(0.00)	(0.246)
Risk aversion (finance metric)	0.37	3.91e-07	0.411
	(0.279)	(0.00)	(0.3)
Prominent Player (0-same; 1-less)	2.78***	-0.06***	2.06***
	(0.63)	(0.01)	(0.75)
Nationality (0-foreigner; 1-italian)	-2.07	0.09***	-0.94
	(1.27)	(0.03)	(1.37)
Controlling for Rounds	Yes	yes	yes
Constant	6.61	0.59***	9.41*
	(4.73)	(0.12)	(5.43)
Observations	2800	2800	1761
Number of subject number	350	350	342

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6. Results of the random effect robust regression on individual data. Reference treatment is T1.

A few words on the non-self-explaining additional regressor variables used in the individual is deemed (Appendix 3 reports the descriptive statistics). “Understanding” is a continuous variables coded between 0 and

100 that maps the numbers of correct answers to the preliminary questionnaire we run to make sure individuals have correctly understood the basic structure of the game. Risk aversion was self-reported both framed in the context of life-threatening risk and in the context of financial risks. Finally, the "Prominent player" variable encodes the answer to the question: "Suppose in the game you just played you were the representative of your group. So, the remaining 6 participants in your group will do the same choices as you. If you were to repeat the same experiment as the one you just took part in, what would be your choice as a leader?" and was only asked in treatments T1 to T5.

When looking at individual investment in A by the subjects, unconditional on their choice at the coalition stage, and controlling for treatments and round effects, gender, risk aversion measured through the life threats question and understanding seem to have a very mild implications. The strongest and most significant effect seems to be associated with the Prominent Player variable, suggesting that individuals who have invested more in Project A might have constrained themselves more were they able to trust or, better, enforce reciprocity of the group. The decision to enter the coalition and the investments for those in the fringe only follows a very similar pattern.

5. Discussion

We have explored empirically the prospects of cooperation in a threshold public bad game designed to capture the tradeoffs faced by countries choosing to what degree to transition from incumbent polluting technologies to cleaner alternatives, with the overall commitment dictating whether they manage to avert dangerous climate change.

Our analysis suggests two possible situations. The first is one where the potential benefits of innovation generated within agreements fostering early investments in a clean technology are deemed very large. The expected returns to cooperation are sufficiently large that a pivotal number of participants is lured into action. The fringe also reacts proactively to the diminished burden they have to shoulder. Ex-ante it would make sense to promote this process by committing to some form of appropriation of the knowledge created within the coalition.

The second scenario is one where the expectations from the new technology are more modest. In this case, our experiment suggests that the negotiation is likely to result in a coalition which is not large enough to be pivotal. Leveraging on the effort of second movers by fostering clean technology uptake by the fringe would be recommendable here.

Specifically, in order to disentangle the push and pull factors behind the incentives to join environmental agreements (and more broadly behind technology adoption), we have introduced several modifications to the threshold public bad game employed in the baseline treatment. These modifications capture some realistic features of current negotiation platforms and may ease the problem of equilibrium selection. Namely, we incrementally add: (i) a membership stage where motivated investors in the 'green' technology can lead by example and (partially) correct the externality; (ii) a first mover advantage of differing magnitudes, which increases the competitiveness of the 'green' technology; (iii) the presence of spillovers benefitting second movers with the same increased return to the green technology as the one enjoyed by early investors.

The temporal dimension introduced with the above conditions leads to nontrivial strategic effects. Effectively, non-signatories play a game of their own, where the maximal safe investment in the 'dirty' technology is determined by the number of those that showed leadership by restricting themselves in its use. In particular, the interplay of (i)-(iii) can either catalyze or deter investments in the clean technology, by affecting participation to the treaty and consequently the incentives for the fringe. From the point of view of the latter, the presence of leaders (i) has potentially conflicting effects. This is due to the coexistence of increased free riding incentives (the target is within reach and second movers may optimistically assume that others will take it upon themselves to restrict their use of the polluting technology) and opposite incentives to cooperate (early commitments to the common good may entice the fringe to follow suit). Increased competitiveness of the socially preferable technology (ii) will affect both groups differently, depending on whether the third condition, spillovers to the fringe, is active.

Perhaps unsurprisingly, the two treatments where the subjects cooperated most are those in which it is less costly to do so, i.e. the gap in the cost between the clean and the polluting investments is smallest. The distribution of burdens between the two groups, however, is rather different. While most of it is taken on by the coalition when its members retain the benefits of R&D, the reverse is true when R&D benefits trickle through to the fringe: coalition size drops by about 40%, but the fringe, lured by the spillovers, embraces the new technology. This effect is even more marked when the magnitude of the benefits from R&D in clean technology is smaller: here the drop in coalition size is more modest under positive spillovers, while the effect on fringe behavior remains strong, leading to a significantly higher chance of avoidance of the threshold for dangerous climate change.

These findings point to the importance of adding R&D to the bargaining table in climate negotiations. Reducing the cost-effectiveness gap with respect to the incumbent technology (e.g. fossil fuels) by means of investments by a set of motivated innovators, may suffice to lure more reluctant players towards an environmentally

superior, but individually more costly alternative. Our results suggest that especially when investments in R&D can only provide limited returns, all the parties are better off when followers can also profit from these investments.

Of course, caution must be used when extrapolating to the climate negotiations. Unfortunately, the problem we face has many more layers of complexity, including asymmetry of payoffs for different countries and uncertainty about the location of the threshold for dangerous climate change. These will make the matter of coordination more difficult, as agreement is inevitably harder to reach when objectives differ and the target is fuzzy. In terms of asymmetries, even in our simple setup we note that the timing element introduces differences in incentives and expected payoffs between the leaders and the followers. In fact, we find that in groups that successfully avoided the threshold the average investment in the clean technology is about the same in the fringe and in the coalition. Conversely, in unsuccessful groups the average fringe investment in the public good was only half than the corresponding investment by a signatory, further evidence of the importance of coordinating on equitable burdens. Uncertainty about the threshold works in a similar direction, by hindering its role as a coordination mechanism and pulling parties towards widespread defection. We maintain that it is therefore all the more important to induce participation by the more reluctant players, and our experimental findings suggest that the diffusion of innovation may be an important lever for climate action.

References:

Alley, R.B., Marotzke, J., Nordhaus, W.D., Overpeck, J.T., Peteet, D.M., Pielke Jr., R.A., Pierrehumbert, R.T., Rhines, P.B., Stocker, T.F., Talley, L.D, Wallace J.M. (2003). Abrupt Climate Change. *Science,* 299 (5615), 2005-2010.

d'Aspremont, C., Jacquemin, A., Gabszewicz, J.J., Weymark, J. (1983), On the Stability of Collusive Price Leadership. *Canadian Journal of Economics*, 16 (1), 17-25.

Barrett, S. (1994). Self-enforcing international environmental agreements. *Oxford Economic Papers*, *46*, 878–894.

Barrett, S. (2003). *Environment and statecraft: the strategy of environmental treaty-making*. Oxford: Oxford University Press.

Barrett, S. (2006). Climate treaties and 'breakthrough' technologies. *American Economic Review,* 96, 22-25.

Barrett, S., Dannenberg, A. (2012). Climate negotiations under scientific uncertainty. *Proceedings of the National Academy of Sciences*, 109, 17372–17376.

Bohm, P. (1993). Incomplete international cooperation to reduce CO2 emissions: alternative policies. *Journal of Environmental Economics and Management*, 24, 258-71.

Carraro, C., Siniscalco, D. (1992). The International Protection of the Environment: Voluntary Agreements among Sovereign Countries. In P. Dasgupta and K.G. Mäler (eds.), *The Economics of Transnational Commons*, Clarendom Press, Oxford.

Carraro, C., Siniscalco, D. (1993). Strategy for the international protection of the environment. *Journal of Public Economics*, 52, 309-328.

Carraro, C., Siniscalco, D. (1995). R&D Cooperation and the Stability of International Environmental Agreements. *CEPR Discussion Papers* 1154.

Dannenberg, A., Lange, A. and Sturm, B. (2014), Participation and Commitment in Voluntary Coalitions to Provide Public Goods. Economica, 81: 257–275

Dechezleprêtre, A., Glachant, M., Ménière, Y. (2008). The Clean Development Mechanism and the International Diffusion of Technologies: An Empirical Study. *Energy Policy*, 36, 1273-1283.

Dietz, S., Marchiori, C., Tavoni, A., (2012). Domestic politics and the formation of international environmental agreements. Working Paper 87, *Grantham Research Institute*.

Eaton, J., Kortum, S. (2001). Trade in Capital Goods. *NBER Working Papers* 8070.

Eaton, J. Kortum, S. (2006). Innovation, Diffusion and Trade. *NBER Working Papers* 12385.

P. N. Edwards, "Infrastructure and Modernity: Scales of Force, Time, and Social Organization in the History of Sociotechnical Systems, "in Modernity and Technology, ed. T. J. Misa et al. (MIT Press, 2002).

IPCC (2014). Climate change 2014: Mitigation of climate change. In Edenhofer, O., Pichs-Madruga, R., Sokona, Y., Farahani, E., Kadner, S., Seyboth, K., Adler, A., Baum, I., Brunner, S., Eickemeier, P., Kriemann, B., Savolainen, J., Schlömer, S., von Stechow, C., Zwickel, T., Minx, J.C. (eds.), *Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, UK and New-York, NY, USA.

İriş, D., Lee, J., Tavoni, A. (2014). Delegation and Public Pressure in a Threshold Public Goods Game: Theory and Experimental Evidence. Mimeo.

Golombek, R., Hoel, M. (2004). Unilateral Emission Reductions and Cross-Country Technology Spillovers. *The B.E. Journal of Economic Analysis & Policy, Berkeley Electronic Press,* 2, 1-27.

Granovetter, M. (1978). Threshold models of collective behavior. *The American Journal of Sociology*, 83 (6), 1420-1443.

Hare B., Meinshausen, M. (2006). How much warming are we committed to and how much can be avoided? *Climatic Change*, 75, 111-149.

Heal, G., Kunreuther, H. (2012). Tipping climate negociations. In R. Hahn and A. Ulph (eds.), *Climate Change and Common Sense: Essays in Honour of Tom Schelling,* Oxford University Press.

Hoel, M. (1992). International Environment Conventions: the Case of Uniform Reductions of Emissions. *Environmental and Resource Economics*, 2, 141-159.

Keller, W. (2010). International Trade, Foreign Direct Investment and Technology Spillovers. In B. Hall and N. Rosenberg (eds.), *Handbook of the Economics of Innovation*, Volume 2, Elsevier Publishers.

Kosfeld, M., Okada, A. and Arno Riedl (2009). Institution Formation in Public Goods Games. *The American Economic Review,* 99 (4), 1335-1355.

Kriegler, E., Hall, J.W., Hermann, H., Dawson, R., Schellnhuber, H.J. (2009). Imprecise probability assessment of tipping points in the climate system. *Proceedings of the National Academy of Sciences*, 106 (13), 5041-5046.

Liberman, V., Samuels, S. M., Ross, L. (2004). The name of the game: predictive power of reputations versus situational labels in determining prisoner's dilemma game moves. Personality and Social Psychology Bulletin, 30(9), 1175-1185

Lade, S., Tavoni, A., Levin, S., Schlueter, M. (2013). Regime shifts in a social-ecological system. *Theoretical Ecology,* 6, 359-372.

Lenton, T. M., Held, H., Kriegler, E., Hall, J.W., Lucht, W., Rahmstorf, S., Schellnhuber, H.J. (2008). Tipping elements in the Earth's climate system. *Proceedings of the National Academy of Sciences,* 105 (6), 1786-1793.

Levati, M. V., Sutter, M., van der Heijden, E. (2007). Leading by Example in a Public Goods Experiment with

Heterogeneity and Incomplete Information. *Journal of Conflict Resolution*, 51 (5), 793-818.

Moxnes, E., van der Heijden, E., (2003). The effect of leadership in a public bad experiment. *Journal of Conflict Resolution*, 47 (6), 773-795.

Ostrom, E. (2009). A Polycentric Approach for Coping with Climate Change. *World Bank Policy Research* Working Paper 5095. Washington, D.C., World Bank.

Rogelj, J., Hare, W., Lowe, J., van Vuuren, D., Riahi, K., Matthews, B., Hanaoka, T., Jiang, K., Meinshausen, M. (2011). Emission pathways consistent with a 2 °C global temperature limit. *Nature Climate Change*, 1, 413–418.

Rockström, J., Steffen, W., Noone, K., Persson, A., Chapin III, F.S., Lambin, E., Lenton, T.M., Scheffer, M., Folke, C., Schellnhuber, H., Nykvist, B., De Wit, C.A., Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P.K., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R.W., Fabry, V.J., Hansen, J., Walker, B., Liverman, D., Richardson, K., Crutzen, P., Foley, J. (2009). Planetary boundaries: exploring the safe operating space for humanity. *Ecology and Society,* 14 (2): 32.

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

Seres, S., Haites, E., Murphy, K. (2009). Analysis of Technology Transfer in CDM Projects: An Update. *Energy Policy*, **37**, 4919-4926.

Schelling, T. (1960). *Strategy of Conflict*, Harvard University Press, Cambridge.

Sonnemans, J., Schram, A., Offerman, T. (1998). Public good provision and public bad prevention: the effect of framing. *Journal of Economic Behavior and Organization*, 34, 143- 161.

Stern, N. (2007). *The economics of climate change. The Stern review*. Cambridge: Cambridge University Press.

Sterner, T., Damon, M. (2011). Green growth in the post-Copenhagen climate. *Energy Policy,* 39, 7165–7173.

Tavoni, A., Dannenberg, A., Kallis, G., Löschel, A. (2011). Inequality, communication and the avoidance of disastrous climate change in a public goods game. *Proceedings of the National Academy of Sciences*, 108, 11825–11829.

Tavoni, A. (2013). Game theory: Building up cooperation. *Nature Climate Change*, 3, 782–783.

van der Werf, E., Di Maria, C. (2008). Carbon leakage revisited: unilateral climate policy with directed technical change, *Environmental Resource Economics,* 39, 55–74.

Vasconcelos, V, Santos, F. C., Pacheco, J. M. (2013). A bottom-up institutional approach to cooperative governance of risky commons. *Nature Climate Change,* 3, 797–801.

Vasconcelos, V., Santos, F., Pacheco, J., Levin, S. (2014). Climate policies under wealth inequality. *Proceedings of the National Academy of Sciences*, 111, 2212–2216.

Watts, D. J. (2002). A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99, 5766-5771.

Weir, S., Knight, J. (2004). Externality effects of education: Dynamics of the adoption and diffusion of an innovation in rural Ethiopia. *Economic Development and Cultural Change*, 53, 93-113.

Appendix 1: Equilibrium solutions of the 2-stage game

To provide the equilibrium solutions, we solve the game using backward induction, beginning with the fringe decision.

Fringe investment stage:[11]

The threshold public bad game is played by the non-members. The reasoning holds for both *no spillover* and *spillover* cases.

- When $s < 2$, no coalition forms in the membership stage. The game is characterized by the two equilibria described above.
- For any $s \in [2, N-1)$, non-members play the threshold public bad game with a different effective threshold than the one contemplated by members in the prior stage, as the latter have already invested part of their endowment in A. The threshold for the group of non-members becomes $X_A^n = T - X_A^s$. Again, risk-neural non-members will either coordinate on tipping point avoidance, or gamble and invest their total endowment in A. For each possible coalition size, there are thus two symmetric equilibria: non-members invest respectively $x_A^n = (T - X_A^s)/(N - s)$ or $x_A^n = e$.
- When $s = N - 1$, the best-response of the sole pivotal non-member is unique and it is to coordinate with the coalition, i.e. to choose $x_A^n = T - (N-1)x_A^s$.
- When $s = N$, there is no fringe and each subject contributes to reach the Pareto superior equilibrium investing the pre-determined amount $x_A^s = T/N$, thus guaranteeing loss avoidance.

Membership stage:

Using the concept of internal and external stability (d'Aspremont et al. 1983), for a coalition to be stable two conditions must hold: a member has no incentive to leave (internal stability: $\pi^s(s) > \pi^n(s-1)$), and a non-member has no incentive to join (external stability: $\pi^n(s) > \pi^s(s+1)$). Call s^* the stable coalition size. As we just established the fringe investment stage has several equilibria. Below we provide the stability conditions when members and non-members coordinate on threshold avoidance.[12] We provide the reasoning for both *no spillover* and *spillover* respectively.

No spillover:

[11] In this paper we consider only the symmetric equilibria. Hence, we restrict attention to symmetric investments by fringe individuals and coalition members, x_A^n and x_A^s respectively. Nonetheless note that in addition to this equilibria the game between non-members has also multiple asymmetric equilibria ensuring avoidance of the tipping point, as any investment profile such that $X_A = T$ is a Nash equilibrium.

[12] Note that the reasoning for the inefficient equilibrium is the same except that payoffs are multiplied by $(1 - p + pq)$ as non-members invest their full endowment e in Project A such that $X_A > T$.

Payoffs at the safe equilibrium when being member and when being non-member are respectively:

$$\overline{\pi}^{s}(s) = r_A \underline{x}_A^s(s) + r_B(1 + sI)(e - \underline{x}_A^s(s)) - c_A T$$

$$\overline{\pi}^{n}(s-1) = r_A \underline{x}_A^n(s-1) + r_B(e - \underline{x}_A^n(s-1)) - c_A T$$

The stability function (Carraro and Siniscalco, 1992) is:

$$\overline{\pi}^{s}(s) - \overline{\pi}^{n}(s-1) = (r_A - r_B)[\underline{x}_A^s(s) - \underline{x}_A^n(s-1)] + r_B sI\left(e - \underline{x}_A^s(s)\right) \qquad (3)$$

Let's start with the case $I = 0$ (i.e. no technological improvement): no coalition is stable when $r_A > r_B$ (i.e. $\overline{\pi}^{s}(s) - \overline{\pi}^{n}(s-1) < 0, \forall s$). When $I > 0$, $\overline{\pi}^{s}(s) - \overline{\pi}^{n}(s-1)$ is negative (positive) if the gain of leaving the coalition $(r_A - r_B)[\underline{x}_A^s(s) - \underline{x}_A^n(s-1)]$ is higher (lower) than the additional revenue thanks to the technological cooperation $r_B sI\left(e - \underline{x}_A^s(s)\right)$. In other words, a stable coalition (bringing together at least two subjects) is achievable if technological cooperation brings enough additional revenue to members such that it compensates the loss of not leaving the coalition.[13]

Spillover:

Similarly, payoffs at the cautious equilibrium are:

$$\overline{\pi}^{s}(s) = r_A \underline{x}_A^s(s) + r_B(1 + sI)(e - \underline{x}_A^s(s)) - c_A T$$

$$\overline{\pi}^{n}(s-1) = r_A \underline{x}_A^n(s-1) + r_B(1 + (s-1)I)(e - \underline{x}_A^n(s-1)) - c_A T$$

The stability function becomes:

$$\overline{\pi}^{s}(s) - \overline{\pi}^{n}(s-1) = [r_A - r_B(s-1)I]\left(\underline{x}_A^s(s) - \underline{x}_A^n(s-1)\right) + r_B I\left(e - \underline{x}_A^s(s)\right) \qquad (4)$$

With $I > 0$, the gain of leaving the coalition $(r_A - r_B(s-1)I)[\underline{x}_A^s(s) - \underline{x}_A^n(s-1)]$ is smaller than when there is no spillover and has to be compared with the additional revenue originating from the presence of one additional member: $r_B I\left(e - \underline{x}_A^s(s)\right)$. Given the assumption on $x_A^s(s)$ increasing in s, the latter decreases in s. As a consequence, for a stable coalition to be achievable, the first term has to decrease more in s than the second term. Comparing equations (3) and (4), it becomes apparent that it is more difficult to achieve a stable coalition when technological improvement spills to non-members.

Appendix 2: Returns to Project A and Project B

[13] Stable coalition sizes under both *no spillover* and *spillover* are further detailed in the next section for the parameterization considered in the experiment and subgame perfect equilibria of the game are provided. The latter will be denoted $\{s^*; X_A^*\}$ defining the stable coalition size and the corresponding total investment in A.

For treatments where the implications of innovation were tested and the return wedge between Project A and B was reduced proportionally to coalition size, the information on the change in the returns to Project B was provided at the membership stage, in addition to the information presented in Table 2 in the main text (for the sake of simplicity, in Table A2 we lump together different information that was provided in different treatments).

Depending on the treatment, participants were informed either told that the returns would apply to coalition members only, or to all players.

	Number of coalition members						***No Coalition***
	7 (all)	**6**	**5**	**4**	**3**	**2**	**0**
Gross return from Project A (T0 –T5)	10	10	10	10	10	10	10
Return from Project B to all (T1)	6	6	6	6	6	6	6
Return from Project B to members only (T2) [to everybody in T3]	6.8	6.7	6.6	6.5	6.4	6.2	6
Return from Project B to members only (T4), [to everybody in T5]	8.9	8.6	8.1	7.7	7.3	6.8	6

Table A2 – Returns to Project A and B under different treatments and coalition sizes

Appendix 3: Descriptive statistics and additional regressors in the individual model.

Variable	Obs	Mean	Std. Dev.	Min	Max
Gender	364	.5961538	.4913427	0	1
Understand	364	88.47604	10.49185	46	100
Risk av. (life)	364	5.123626	2.454791	1	11
Risk av. (finance)	364	5.167582	2.136381	1	11
Prominent Player	350	1.948571	1.000106	0	4
Nationality	364	.7912088	.4070043	0	1

The Prominent Player question was asked to participants to treatments T1 to T5 and was phrased as follows:

"Suppose that you are the representative of your group. Therefore, the remaining 6 participants in your group will be bound to your decision. If you were to repeat the same experiment as the one you just took part in, what would be your choice as a leader?"

The Understanding Questionnaire is available from the authors upon request.

Appendix 4: ISTRUCTIONS TO STUDENTS FOR T0

(Instructions for other treatments as well as Ztree codes are available from the authors upon request)

Welcome and thank you for participating in this experiment.

This experiment is about decision-making. Please read carefully the whole instructions. The instructions will help you to understand correctly the experiment. Once all the participants to the experiment have read the instructions, an assistant will read them aloud and the experiment will begin.

Your earnings in this experiment will depend upon your decisions and the decisions made by other participants. All your decisions will be anonymous. In the experiment all amounts are stated in ECU (Experimental Currency Units) and at the end of the experiment, your earnings will be converted into Euros. The exact procedure is detailed at the end of the instructions.

From now on and until the end of the experiment, we ask you to remain silent. If you have any questions, raise your hand and an assistant will come to answer your questions privately.

RULES OF THE GAME

The experiment consists of 2 practice rounds and 8 independent rounds (which will be used to determine your final payout, as explained at the end).

At the beginning of each round, you will be randomly assigned to a group of 7 participants (including yourself). In each round, each of you is given 50 ECU to be used in the investment decision detailed below. You will not know who the other 6 participants in your group are.

The group assignment will change after every round. You will not be matched up with the same other 6 participants for more than a single round.

THE INVESTMENT DECISION

All of you will face the same decision-making problem: to decide on how you will use the 50 ECU. You can invest it in two different projects: **Project A** and **Project B**. You decide how much you want to invest in Project A; the remaining part of your 50 ECU is then automatically invested in Project B.

Project A: Each ECU invested in Project A yields a direct payoff of **10 ECU**. In addition to the payoff to yourself, each ECU invested in Project A yields a cost of **1 ECU** to you and to each of the other 6 participants in your group.

Similarly, investments in Project A by any other participant yield a cost to you and to each of the other participants.

Project B: Each ECU invested in Project B yields a direct payoff of **6 ECU**. **No additional cost** is charged for investing in Project B.

Once all the participants in your group have decided on how to invest the 50 ECU between Project A and Project B, net earnings will be calculated.

- If the total sum of ECU invested by the participants in your group in Project A is **equal or below 105 ECU**, each of you will be paid according to the investment decisions in your group (see next section for details).
- **However,** if the total sum of ECU invested by the participants in your group in Project A is **greater than 105 ECU**, you and the rest of your group will lose all your earnings with 50% probability.

COMPUTING EARNINGS

Your earnings at the end of a round are calculated as follows:

Earnings = ECU invested in Project A x 10

+ ECU invested in Project B x 6

– Sum of all investments in Project A

If the sum of all investments in Project A in your group is equal or below 105 ECU, you will keep the earnings for sure; if the sum of all investments in Project A in your group is greater than 105 ECU, your earnings will be 0 with 50% probability (and as above with 50% probability).

Example 1: *assume that you invest 30 ECU in Project A: you receive a direct payoff of 300 ECU from Project A (=30x10). The remaining 20 ECU (=50-30) are automatically invested in Project B and yield a payoff of 120 ECU (=20x6). Together this generates a direct payoff of 420 ECU (=300+120).*

Assume furthermore that the other 6 participants in your group invest on average 30 ECU in Project A. That gives a total investment of 180 ECU (=30x6) in Project A for these 6 persons. Together with your own investment of 30 ECU, this gives a total investment in Project A of 210 ECU (=180+30). This yields a cost of 210 ECU for you (and for each of the other participants). A direct payoff of 420 ECU minus a cost of 210 ECU gives you final earnings of 210 ECU.

However, as the sum of all investments in Project A exceeds 105 ECU, the related earnings for you and the rest of your group will be 0 with 50% probability.

Direct payoff from Project A	**Direct payoff from Project B**	**Sum of all investments in Project A**	**Net Earnings**
30ECU x 10 = 300ECU	20ECU x 6 = 120ECU	30ECU + 30ECU x 6 = 210ECU	50% probability: 300 + 120 - 210 =210ECU 50% probability: 0

Table 1 – Earnings in Example 1

Table 2 below summarizes the earnings for several combinations of investment decisions in Project A by you and the other 6 participants.

To limit the size of the table, we only mention investments in steps of 5 ECU. However, you can use all integers from 0 up to and including 50 when you choose your investment in Project A. To know your corresponding earnings when investments differ from those reported in the table, you can then use the general formula provided above.

The first column provides your investment decision in Project A, whereas the first row is the mean investment decision in Project A by the other participants in your group. Inside the table you can find your earnings in ECU associated with such choices.

Note that the shaded part of the table provides the combinations of choices for which the sum of all investments in Project A exceeds 105 ECU. The related earnings will be the value in the shaded area with 50% probability and 0 with 50% probability.

Mean investment by the others / Investment by you	**0**	**5**	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**45**	**50**
0	300	270	240	210	180	150	120	90	60	30	0
5	315	285	255	225	195	165	135	105	75	45	15
10	330	300	270	240	210	180	150	120	90	60	30
15	345	315	285	255	225	195	165	135	105	75	45
20	360	330	300	270	240	210	180	150	120	90	60
25	375	345	315	285	255	225	195	165	135	105	75
30	390	360	330	300	270	240	210	180	150	120	90
35	405	375	345	315	285	255	225	195	165	135	105
40	420	390	360	330	300	270	240	210	180	150	120
45	435	405	375	345	315	285	255	225	195	165	135
50	450	420	390	360	330	300	270	240	210	180	150

Table 2 – Earnings arising from the investment decisions made by you and the others.

PROCEDURE FOR MAKING YOUR DECISION

In this experiment, you are asked to choose your investment level in Project A. You can choose any integer from 0 up to and including 50 when you choose your investment in Project A. The remaining part of your 50 ECU is then automatically invested in Project B.

Remind that if the total sum of ECU invested by the participants in your group in Project A is **equal or below 105 ECU**, each of you will get his earnings according to the investment decisions in your group. However, if the total sum of ECU invested by the participants in your group in Project A is **greater than 105 ECU**, you and the rest of your group will lose all your earnings with 50% probability.

The corresponding individual investment to not exceed 105 ECU if all participants invest the same amount (i.e. 105 ECU divided by 7 participants) is 15 ECU.

At the end of each round, you will be informed of your earnings given your choice and the choice made by the other 6 participants in your group.

Once the experiment is completed you will receive **the payout corresponding to the earnings in one randomly selected round** (out of the 8 rounds; note that the 2 practice rounds are not evaluated for payment purposes). **Payments are settled at the end of the experiment,** in cash, according to the following exchange rate: 1 ECU = 0.05€. Since the round to be selected for payment will be determined randomly, and could be any of the non-practice rounds, you should behave in each round as if it was the relevant one for payout.

Before we start the experiment, we would like to give you some review questions to ensure that you fully understand all the procedures. Once all the participants have answered the questions, the experiment will begin. Should you have any questions, feel free to raise your hand to ask for assistance.