UNILATERAL EMISSIONS ABATEMENT: AN EXPERIMENT

BODO STURM
JOACHIM WEIMANN

UNILATERAL EMISSIONS ABATEMENT: AN EXPERIMENT

Abstract

According to the model of HOEL (1991), a unilateral emissions abatement of a global pollutant leads to lower aggregated emissions in a game with a simultaneous decision protocol. Our experiment tests the Hoel model and examines the question of whether a leader can induce additional abatement of followers in a game with a mixed sequential-simultaneous decision protocol. Using an environment with a unique interior equilibrium, our experiments confirm the stylized outcomes of previous public good experiments. Changes in abatement and profits for the simultaneous decision protocol are in line with the theoretical predictions of the Hoel model albeit not significantly in every case. In the treatments with a mixed sequential-simultaneous decision protocol, during the first periods an abatement near social optimum is mostly chosen by the leader. However, in most cases the leader failed to induce cooperation, i.e. there are few followers who react cooperatively to the leader's signal. High efforts by the leader and the cooperative followers are exploited by the majority of defective followers.

JEL classification: Q50.

Bodo Sturm
Otto von GuerickeUniversity Magdeburg
Faculty of Economics and Management
Magdeburg Experimental Laboratory
Vilvredo Pareto Geb. C210
Universitätsplatz 2
39016 Magdeburg
Germany

Joachim Weimann
Otto von GuerickeUniversity Magdeburg
Faculty of Economics and Management
Magdeburg Experimental Laboratory
Vilvredo Pareto Geb. C210
Universitätsplatz 2
39016 Magdeburg
Germany
Joachim.Weimann@Wirtschafts-W.Uni-Magdeburg.de

1.   Introduction

Some of the most serious environmental problems can be characterized as international public good problems. The damage each country suffers depends on the aggregated emission of harmful material and not (only) on local emissions. The most prominent example is the climate change expected from global CO_2 emissions. It is not in the self interest of each individual country to abate the Pareto efficient amount of emissions, because parts of the total benefit generated by abatement cannot be internalised. Given this situation, environmental pressure groups often demand that local politicians take on a leading role and abate more emissions than is in the narrow self interest of the country. Obviously, they hope that "a good example" may encourage other countries to join the coalition of abating countries.

In a seminal paper, HOEL (1991) investigated the consequences of unilateral abatement, i.e. abatement of a single country (say country 1) above the equilibrium level, under the assumption that all countries decide simultaneously and that all other countries behave fully rationally and selfishly. It turns out that unilateral abatement leads to a greater total emission reduction and increases total welfare if country 1 has the lowest marginal benefit from abatement.[1] The assumption of selfish behavior of all countries except 1 rules out that "the good example" given by country 1 can infect other countries.[2] However, politicians and environmental pressure groups often claim a leader role for their own country regarding climate change in order to overcome the social dilemma situation all countries are confronted with.[3] Thus it remains an open question whether "leadership matters". In this paper, we try to answer that question by testing the Hoel model experimentally.

Whether abatement on the part of the leader has an influence on the behavior of the followers is analyzed in experiments by VAN DER HEIJDEN/MOXNES (2000). However, there are some important differences between our design and the design chosen by VAN DER HEIJDEN and MOXNES. In their experiments, subjects play a standard public bad game[4] where both the equilibrium solution for the simultaneous and mixed sequential-simultaneous decision protocol and the Pareto efficient outcome have a boundary solution. Therefore, they do not really test the Hoel model, which has interior solutions and allows deviations from the equilibrium in two directions. In VAN DER HEIJDEN/MOXNES (2000) the leader had a significant influence on the

---

[1] HOEL also shows that this positive result no longer holds if the unilateral abatement is followed by international negotiations on emission reductions. Unilateral reductions weaken the position of the leading country. Consequently, total emission reduction after international negotiations is lower in the case of ex ante reductions of the leading country compared to the case without unilateral abatement. We do not deal with negotiations on emission reduction in this paper but only look at decentralized decisions about abatement.

[2] HOEL explicitly points out that "I do *not* take up the question of whether such action from one country might lead to similar behavior from other countries" (p. 56).

[3] As an example we can quote the German Federal Minister for the Environment: "Germany remains the leader in international climate protection … And we have to maintain our leading position". Federal Minister for the Environment, Nature Conservation and Nuclear Safety (2002).

[4] See ANDREONI (1995) for a public bad game.

other subjects (followers). However, the influence of the leader on the followers is not strong enough to make leadership profitable for the leader.

In order to test the Hoel model experimentally, two methodological questions have to be answered. First, one has to decide how the experiment should be framed. Either subjects are confronted with the game as it is without any reference to environmental problems, or they are told the cover story which fits the story the Hoel model tells us. We decided on the latter because we wanted to test if unilateral *abatement of harmful material* induces further activities. A fair test of the model should take into account that it is designed to deal with *environmental* problems. Second, one has to decide how to inform subjects. Once again, we decided to design the experiment analogously to the model. This means that we gave subjects all the information the agents in the model are assumed to have. Game theoretic models implicitly make very strong assumptions about what people know. Normally, not only the rules of the game but also the equilibrium solution and the consequences of all kinds of deviations from rational behavior are assumed to be common knowledge. To test the model fairly, we had to ensure that these knowledge pre-conditions were fulfilled. Because the game has complicated interior equilibrium solutions, we could not assume that subjects would be able to compute the prediction of game theory and the Pareto efficient solution in the lab during the experiment. We therefore decided to teach the subjects before the experiment and to demonstrate the prediction of game theory as well as the Pareto solution. One may argue that this can influence subjects. But if this is the case, agents in the Hoel model should be influenced in the same way because they possess the same information we gave our subjects. Not informing subjects fully would mean not testing the model but testing the ability of students to find rational strategies in complicated games.

The paper is organized as follows. In section 2 the model is outlined and the parameters and functions used in the experiment are introduced.[5] Section 3 describes the experimental design. The results are presented in section 4 and discussed in section 5. Section 6 concludes.

2.   The laboratory version of the Hoel model

In the N-country version of the Hoel model, each country emits a global pollutant. Every country i, $i = 1,...,N$, chooses abatement X_i and possesses benefit and cost functions $B_i(X)$ and $C_i(X_i)$ with $B_i' > 0$, $B_i'' < 0$, $C_i' > 0$, and $C_i'' > 0$. The sum of individual abatements is $X = \sum_{i=1}^{N} X_i$.

In the Nash equilibrium (NE) of the game with a simultaneous decision protocol, country i abates to the amount at which *private* marginal benefit equals marginal costs, i.e.

$$B_i'(X) = C_i'(X_i) \tag{1}.$$

---

[5] A more complete version of the model can be found in appendix 1.

In the Pareto optimum (PO), aggregated abatement is chosen such that *social* marginal benefit equals marginal costs, i.e.

$$B^{'}(X) = C_i^{'}(X_i) \qquad \text{with} \qquad B^{'}(X) = \sum_{i=1}^{N} B_i^{'} \tag{2}.$$

Because social marginal benefit is above private marginal benefit, NE abatement is smaller than PO abatement, i.e. $X^{PO} > X^{NE}$.

In order to model unilateral abatement of country 1, HOEL assumes that this country derives some kind of extra marginal benefit from total abatements, measured by a parameter *h*. It is not important where this benefit comes from. It may be the case that the inhabitants of this country have altruistic preferences or just love to see a "cleaner" world. For country 1, the total benefit from abatement then is

$$B_1(X) + hX \qquad \text{with} \qquad h > 0 \tag{3}.$$

Given (3), in the new NE for the simultaneous decision protocol it holds that

$$B_1^{'}(X) + h = C_1^{'}(X_1) \quad \text{and} \qquad B_j^{'}(X) = C_j^{'}(X_j) \qquad \text{with} \qquad j = 2,...,N \tag{4}$$

The standard case (1) is contained in this formulation with $h = 0$. HOEL shows that, in the equilibrium described by (4), country 1 abates more and countries 2, …, *N* abate less than in the equilibrium given by (1) and that total abatements are greater in the case of unilateral abatement. He also shows that, under unilateral abatement, $B_1^{'} \leq B_j^{'}$ for all 2, …, *N* is a sufficient condition for an increase of total welfare.

We are seeking an answer to the question of whether "leadership", i.e. the pure timing of action, matters. Therefore, we have to analyze the implications of a change in the timing of action from a simultaneous decision protocol to a mixed sequential-simultaneous (in the following "sequential") decision protocol, i.e. country 1 decides first, the other countries are informed of this decision, and then decide simultaneously on their emission reduction. Although the cost and benefit functions remain unchanged, the variation from the simultaneous to the sequential decision protocol alters the game theoretical prediction for individual behavior. Country 1, the leader, now has a strategic advantage because it may choose the point on the best response function of the followers which maximizes its own profit. In the subgame perfect equilibrium (SPE) of the sequential game, country 1 abates less and country *j* abates more than in the NE of the simultaneous game.[6] Aggregated abatement in SPE is lower than in NE, i.e. leadership generates a negative environmental effect.

In order to test the model experimentally, the cost and benefit functions have to be specified. Table 1 shows the specification of the model, and in appendix 1 we derive the appropriate NE, SPE, and PO solution.

---

[6] See appendix 1.

Table 1: Cost and benefit functions

Cost function of country i, $i = 1, \ldots, N$	$C_i(X_i) = cX_i^2 + T$
Benefit function of country 1	$B_1(X) = b(AX - 0.5X^2) + hX$
Benefit function of country $j = 2, \ldots, N$	$B_j(X) = AX - 0.5X^2$

To test the Hoel model and the influence of leadership, we use a 2x2 factorial design varying the *sequence* of moves and the parameter h. In the simultaneous treatments (*sim*), all countries simultaneously decide on their abatement. In the sequential treatments (*seq*), country 1 decides first, the other countries are informed of this decision, and then simultaneously decide on their emission reduction. The variation of h in the simultaneous case is a direct test of the model. Table 2 summarizes the treatments.

Table 2: Experimental treatments

variable		sequence	
		sim	*seq*
parameter *h*	*h* = 0	treatment **T I** (*h* = 0, *sim*)	treatment **T III** (*h* = 0, *seq*)
	h > 0	treatment **T II** (*h* > 0, *sim*)	treatment **T IV** (*h* > 0, *seq*)

To implement the four treatments, we had to specify the parameters of the functions in table 1 in a way that ensures that the payoff functions are sufficiently steep. Table 3 summarizes the parameter values for T I - T IV and the abatements and payoffs subjects realize in NE, SPE, and PO. The parameters are chosen so that the NE has a solution in integers; numbers had to be rounded for the SPE and PO solution.[7]

---

[7] Subjects had to enter integers in the simultaneous treatments. We allowed for one decimal place in the sequential treatments.

Table 3: Summary of parameters, abatements, profits, and payments

T I and T III: $b = 30/47$, $h = 0$, $A = 500$, $c = 3$, $T = 40,000$, $N = 5$									
	Nash equilibrium[2]			Subgame perfect equilibrium[3]			Pareto optimum		
Country	1	2, …, 5	all	1	2, …, 5	all	1	2, …, 5	all
Abatement	30	47	218	18.44	48.16	211.07	79.45		397.23
Profits (Labdollar)	11,707	38,611	166,151	12,123	36,212	156,971	17,488	60,782	260,606
Payments[1] (EUR)	15.00	21.70		15.10	21.10		16.40	27.20	
T II and T IV: $b = 30/47$, $h = 3000/47$, $A = 500$, $c = 3$, $T = 40,000$, $N = 5$									
	Nash equilibrium[2]			Subgame perfect equilibrium[3]			Pareto optimum		
Country	1	2, …, 5	all	1	2, …, 5	all	1	2, …, 5	all
Abatement	40	46	224	24.59	47.54	214.75	81.63		408.16
Profits (Labdollar)	24,974	40,564	187,230	25,713	37,619	176,188	43,200	60,788	286,312
Payments[1] (EUR)	18.30	22.20		18.50	21.50		22.80	27.20	

[1] 40,000 Labdollars = 1 EUR; payments include 12 EUR show-up fee
[2] simultaneous decision protocol
[3] mixed sequential-simultaneous decision protocol

The efficiency loss in the NE and the SPE is considerable in all treatments. For $h = 0$, the profit for country 1 in PO is 49% (44%) above the profit in NE (SPE). The profit for country j in PO is 57% (68%) above the profit in NE (SPE). For $h > 0$, the profit for country 1 in PO is 73% (68%) above the profit in NE (SPE). The profit for country j in PO is 50% (62%) above the profit in NE (SPE). Free rider incentives are also considerable. If all other countries choose their PO abatement in T I, country 1 (j) can increase its profit by 72% (16%) compared to PO profit providing it chooses its best response. In T II and the same situation, country 1 (j) can increase its profit by 24% (16%). If all other countries choose their PO abatement, country j can increase its profit in T III (T IV) by 16% (19%) compared to PO profit providing it chooses its best response.[8]

Given this version of the Hoel model, we can now formulate the two central hypotheses which follow from standard game theory and which will be checked experimentally.

Hypothesis 1:

Parameter h has a significant influence on the abatement and profit of country i both in the simultaneous and sequential treatments. A higher abatement for country 1 and lower abatements for country $j = 2, …, 5$ should be observed in T II and IV ($h > 0$) than in T I and III ($h = 0$).[9] At the aggregated level, a positive environmental effect should be observed in T II and IV ($h > 0$) compared to T I and III ($h = 0$), i.e. aggregated abatement is higher for $h > 0$ than for $h = 0$. The change in abatement leads to higher individual profits for every country (and higher aggregated profit) for $h > 0$ compared to $h = 0$.[10]

---

[8] We assume that country 1 has no free rider option in the sequential treatments.

[9] However, the last part of the hypothesis is questionable due to the very small difference in NE and SPE abatement of country j between treatments with $h = 0$ and $h > 0$ (table 3).

[10] At this point, we have to discriminate between the increase in profits and the positive welfare effect which HOEL describes in his model for the simultaneous game. The increase in profits is caused by a change in prefer-

Hypothesis 2:

The variable *sequence* has an influence on abatement both at the individual and the aggregated level. A lower abatement for country 1 and higher abatements for country $j = 2, ..., 5$ should be observed in the sequential treatments (T III and IV) than in the simultaneous treatments (T I and II). At the aggregated level, a negative environmental effect should be observed in T III and IV compared to T I and II, i.e. aggregated abatement is lower for the game with leadership than for the game without leadership. Furthermore, the change in *sequence* leads to higher (lower) individual profits for country 1 (*j*) and lower aggregated profits for the game with leadership compared to the game without leadership.[11]

The most interesting point of hypothesis 2 is that the game theoretical prediction, that leadership results in less abatement for country 1 and a world with less aggregated abatement, is in direct contrast to the above-mentioned idea that leadership may be a solution to the social dilemma situation countries are confronted with.

3. Experimental design

Each of the four treatments was played with six groups of five subjects. Thus, we have six independent observations for each cell of table 2 and 120 subjects participating in the experiment. All subjects are economics undergraduate students familiar with fundamental game theoretic concepts. Each of the, in total, 24 groups played the game 10 times and subjects were informed about the number of repetitions. The experiment was fully computerized[12] and anonymous. The subjects were seated in soundproof booths and had no contact before, during, and after the experiment. We had eight sessions with three groups (15 subjects) playing the game in parallel. Each session lasted about one hour.

The information to the subjects was organized as follows. During the experiment subjects were informed about their individual and aggregated abatement, the aggregated abatement of all other countries, the individual profits of all countries, and the aggregated profit for all expired periods. The subjects received written instructions about the rules of the game, their role (country 1 or country 2, ..., 5), the parameters, and the functional forms.[13] Furthermore, their computers were equipped with a payoff simulator which had two elements.[14] The first was the profit maximizing function.[15] In the simultaneous treatments, each country could put in the expected abatement of the other countries and then its profit maximizing response was com-

---

ences due to parameter *h,* which is an element of the real payoff function of country 1. The positive welfare effect assumes that country 1 voluntarily abates more than in the equilibrium and the resulting welfare loss is compensated for by the gain of all other countries.

[11] Here the same problem as in footnote 10 appears. The difference in NE and SPE profits of country 1 between the simultaneous and sequential treatments is too small to get significant results (see table 3).

[12] We used Z-tree for programming. See FISCHBACHER (1999).

[13] See appendix 3 for the instructions.

[14] See appendix 2 for a screenshot of the payoff simulator.

[15] Instead of the term "best response" we use the term "profit maximizing response" because the latter is more neutral in our view.

puted. In the sequential treatments, country j used the same program as in the simultaneous treatment and country 1, the leader, could put in its own abatement and then the profit maximizing response of the other countries was computed. Additionally, the expected profit and total abatement were computed for both cases. The second element was the simulator, which subjects could use to evaluate the consequences of non-profit-maximizing actions. The subjects here could put in the expected abatement of the others and their own arbitrary abatement.[16] The payoff simulator was identical for all subjects in a treatment and was visible both on the input and the output screen. Before the experiment started, questions were answered and all subjects played two test rounds against the computer. The subjects knew that in the test periods they were playing against four automated systems whose behavior would not change.

As already mentioned in the introduction, we decided to use the frame also employed in the Hoel model and to inform subjects comprehensively. For this purpose, we invited the subjects to attend a separate lesson held before the experiment on the same day or a day before the experiment was carried out. We conducted five identical lessons with 24 subjects each (one from each of the 24 groups). This procedure guaranteed the anonymity within the groups. At the beginning of the lessons, subjects were told that they should imagine that they were the head of a delegation from their country at an international conference on emissions abatement of a global pollutant.[17] Given all necessary information for their country (costs and benefits), they had to decide on the level of domestic abatement. Then the most important features of the decision situation were explained. We demonstrated the Nash equilibrium for the simultaneous game and the subgame perfect equilibrium for the sequential game. We showed that everybody was better off in the Pareto efficient solution but that there were strong incentives to deviate from the efficient solution. At the end of the lecture, the input screen, the output screen, and the payoff simulator were shown and explained using an example. The lesson lasted about one hour. Subjects were informed that, in the experiment, countries would have different roles (country 1 and the other countries). However, the information on which role each subject would play and the timing of action was first given in the experiment. At the beginning of the experiment, subjects received a show-up fee of 12 EUR and they were told that possible negative payoffs had to be settled using this fee.

---

[16] Subjects who tried to maximize the collective profit given the expected abatement of the others could compute their "best response" with the help of the simulator.

[17] From the viewpoint of game theory, the frame of the decision problem is irrelevant for individual behavior. On the other hand, if we assume that people have "green preferences", i.e. that they behave more cooperatively in a social dilemma with an environmental frame, our frame provides a "worst case" scenario for game theory.

4.   Results

The payoffs per subject range from 14.40 EUR to 27.30 EUR including the show-up fee. The average payoff over all subjects is 22.10 EUR. Figures 1, 2, and 3 display the average abatement over all ten rounds for country 1, the other countries j, and all countries. The PO, NE, and SPE values are marked.

Figure 1: Abatement country 1

Figure 2: Abatement country j

Figure 3: Total abatement
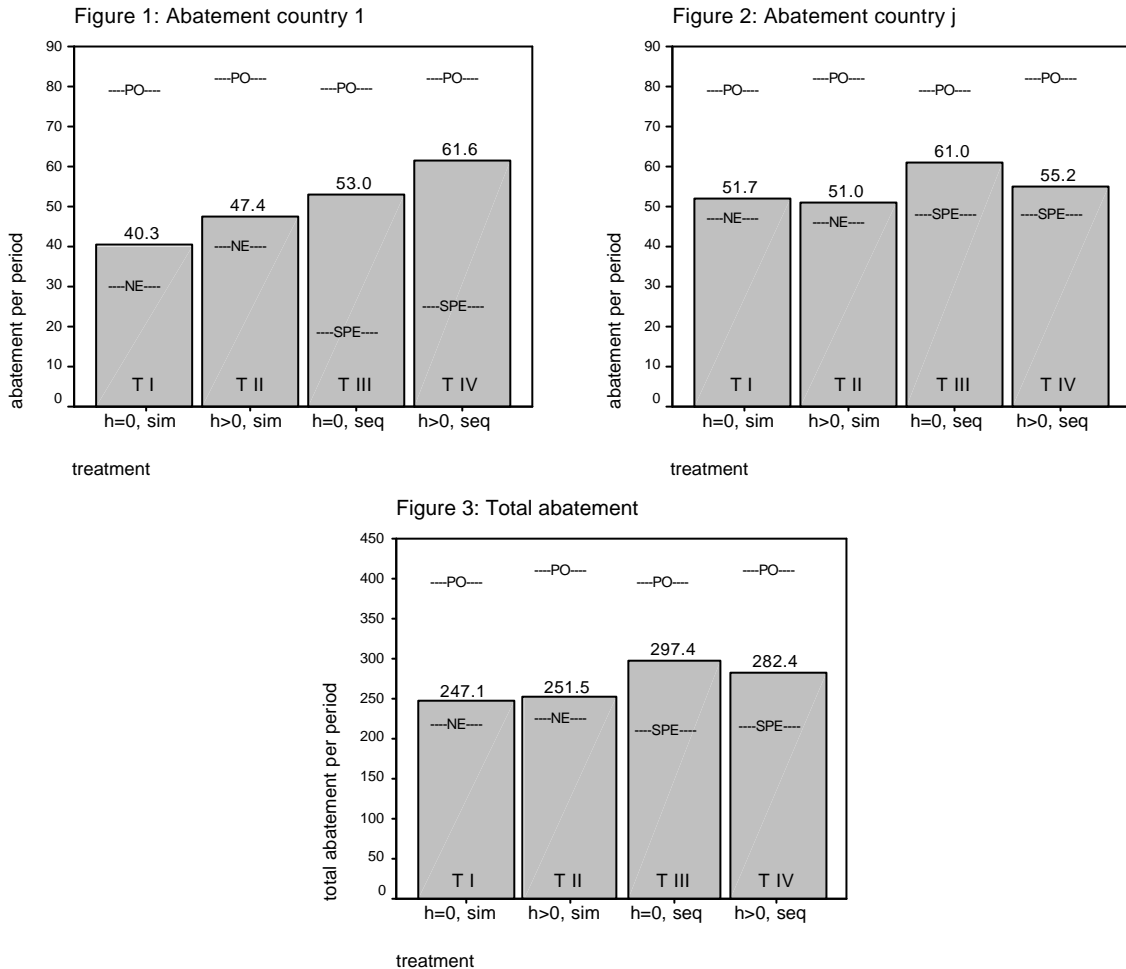
We summarize our findings with respect to the mean abatement (for country 1, j, and total) as follows:

Observation 1:

   a. *The mean abatement of country 1 is ceteris paribus higher in the $h > 0$ treatments than in the $h = 0$ treatments. Abatement is higher in the sequential than in the simultaneous treatments. However, none of these differences between treatments is significant (exact two-sided MW-U test, 5% level).*

   b. *The mean abatement for country j (j = 2, …, 5; in what follows abbreviated by country j) is higher in the sequential treatments than in the simultaneous treatments. As for*

country 1, no significant differences between treatments can be found (exact two-sided MW-U test, 5% level).

c. *The remarkable difference between the mean abatement and the SPE value in the sequential treatments for both countries is significant for country 1 but not for country j (two-sided t-test, 3% level).*

d. *Total abatement is higher in the sequential treatments than in the simultaneous treatments but the difference is not significant. The difference between total abatement and the SPE value in the sequential treatments is not significant (two-sided t-test, 5% level).*

The most striking observation is that country 1 does not reduce its abatement in the change from a simultaneous to a sequential decision protocol – as theory predicts – but raises its abatement. Due to the high variance of individual behavior, the difference between treatments is not significant. However, we still must reject the hypothesis that country 1 abates the SPE values in the sequential treatments. All in all, observations *1a-1c* confirm the visual impression from figures 1 and 2 that, on the one hand, the behavior of country *j* does not vary much between the treatments but, on the other hand, the treatment variables *h* and *sequence* display a strong influence on the behavior of country 1.

Although we have only 6 independent observations per cell in table 2, it is interesting to look at the correlation of the mean abatement of country 1 and *j*. Whereas the Spearman rank correlation yields $r_S = .200$ (p = .704) and .429 (p = .397) for the simultaneous treatments T I and II respectively, values of $r_S = .986$ (p = .000) and .783 (p = .066) are obtained for the sequential treatments T III and IV respectively, i.e. there is a strong positive correlation of leader's and follower's abatement in both sequential treatments. The simple linear regression in table 4 supports this result and yields a significantly positive coefficient for the sequential treatments. These results are an indication that (at least some) followers follow the leader's example and that abatements of followers and leaders are positively correlated.

Table 4: Regression analysis

Model	$a_j = b_0 + b_1 a_1$				
	b_0	$p(b_0)$	b_1	$p(b_1)$	R^2
T I ($h = 0$, *sim*)	51.2	.001	.010	.928	.002
T II ($h > 0$, *sim*)	24.9	.168	.552	.147	.446
T III ($h = 0$, *seq*)	30.4	.001	.579	.000	.965
T IV ($h > 0$, *seq*)	9.3	.527	.745	.024	.759

$a_j (a_1)$ is the mean abatement of country j (1) over rounds, N = 6 ind. obs.

Figures 4-7 show the abatements in the four treatments round by round.

Figure 4: Abatement in T I (h=0, sim)

Figure 5: Abatement in T III (h=0, seq)

Figure 6: Abatement in T II (h>0, sim)

Figure 7: Abatement in T IV (h>0, seq)
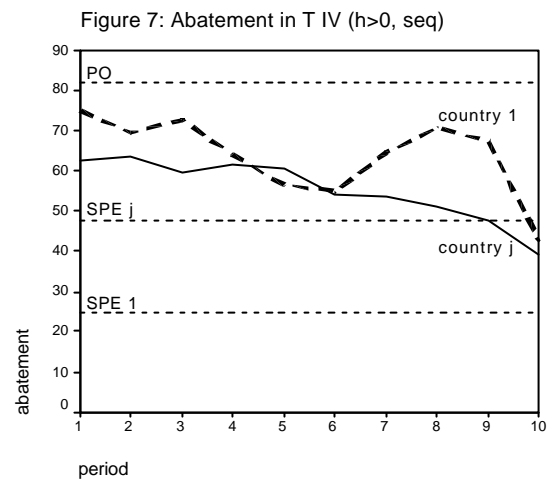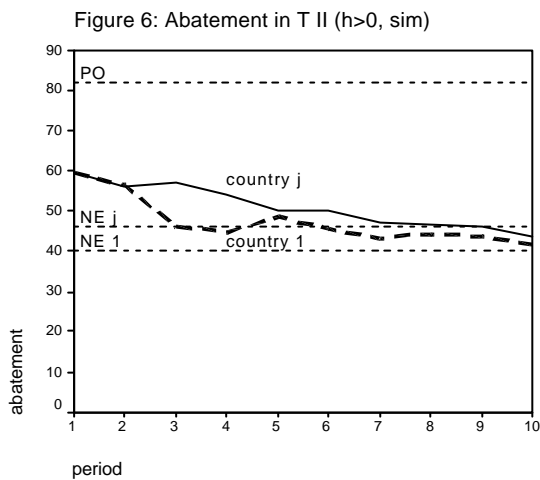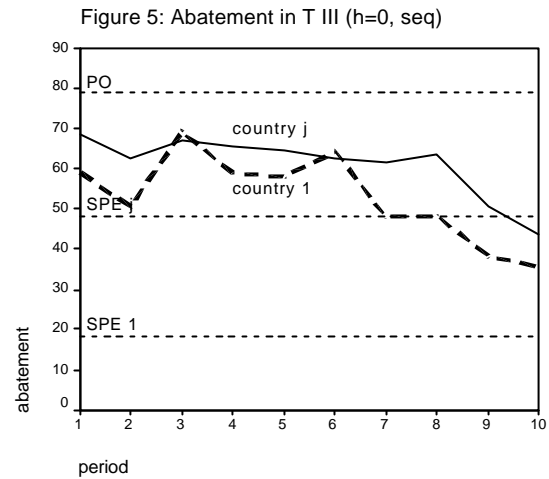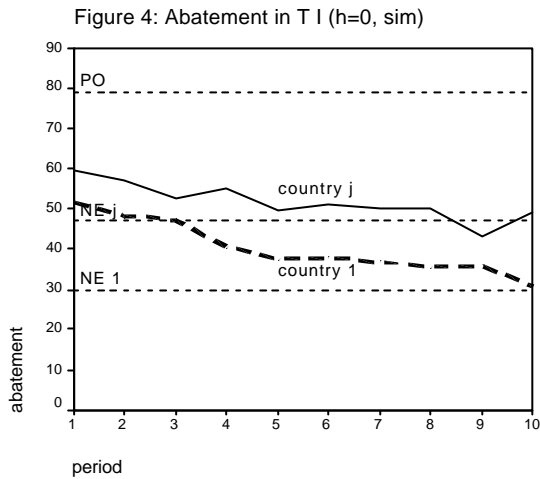
The round-by-round analysis of the abatements confirms the findings of observation 1:

Observation 2:

   a. *The hypothesis that the abatement of country 1 is equal to the NE/SPE value must be rejected in 13 of the 20 rounds of the sequential treatments (two-sided t-test, 5% level). The hypothesis cannot be rejected for any round of the simultaneous treatments.*

   b. *The same hypothesis must be rejected in only eight of the 40 rounds of all treatments for country j. With one exception, all cases appear in the first four rounds (two-sided t-test, 5% level).*

   c. *The hypothesis that abatements for country j are equal to the PO level must be rejected in all 20 rounds of the simultaneous treatments and in 11 of the 20 rounds of the sequential treatments. The same hypothesis must be rejected for country 1 in 19 of the 20 rounds of the simultaneous treatments and in only 3 of the 20 rounds of the sequential treatments (two-sided t-test, 1% level).*

The overall impression from figures 4-7 is that subjects show a clear tendency towards the equilibrium strategy during the course of the game and that this tendency is more pronounced

in the simultaneous treatments compared to the sequential treatments. In the sequential treatments, the downward trend to the SPE is interrupted by phases of constant and even increasing abatement, especially for country 1. In order to gain more insight into the way subjects adjust their behavior, we use the coefficient

$$\boldsymbol{a}_t = \frac{\frac{1}{N}\sum_{i=1}^{N}\left|X_{t,i} - X_{t,i}^{Eq}\right|}{X_{t,i}^{PO} - X_{t,i}^{Eq}}$$

as a simple measure of the deviation of individual behavior from individual rationality. $X_{t,i}$ is the individual abatement of country i in period t, $X_{t,i}^{Eq}$ is i's individual equilibrium (NE or SPE) abatement in period t and $X_{t,i}^{PO}$ is i's individual PO abatement in period t. Coefficient $\boldsymbol{a}_t$ measures the mean absolute value of the deviation of individual abatement from equilibrium abatement as a fraction of the difference between equilibrium and PO abatement, i.e. $\boldsymbol{a}_t$ summarizes the information about individual abatement behavior in figure 4-7. We observe $\boldsymbol{a}_t = 0$ (1) for country 1 or j if all countries of this type play their equilibrium (PO) abatement.

Figure 8: Alpha for country 1

Figure 9: Alpha for country j

T I (h = 0, sim)   T II (h > 0, sim)   T III (h = 0, seq)   T IV (h > 0, seq)
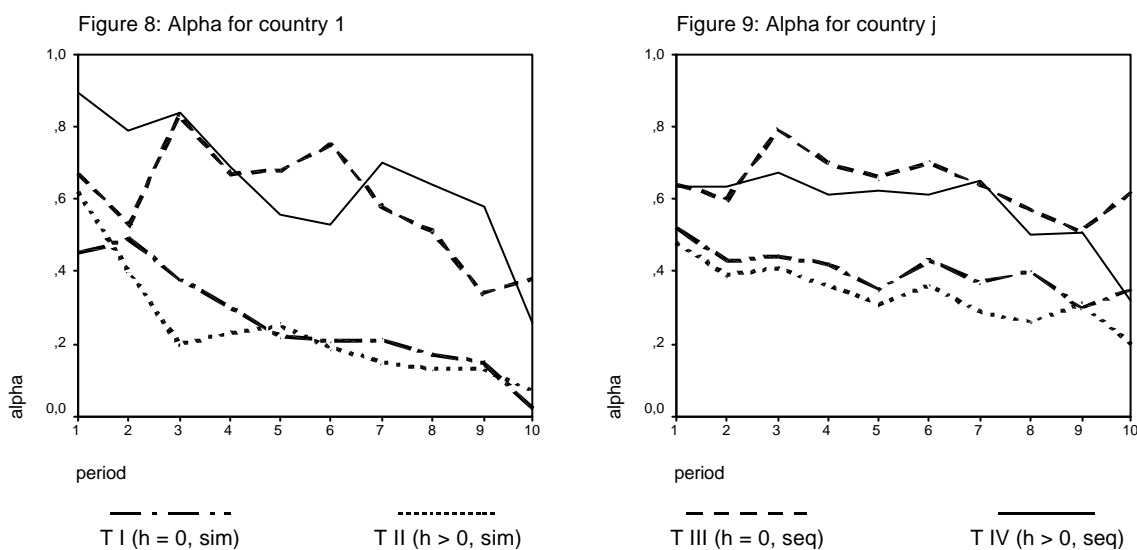
As expected, the $\boldsymbol{a}_t$ values in figure 8 and 9 decrease, although not monotonically. At first glance we observe that the $\boldsymbol{a}_t$ values are higher for the sequential treatments than for the simultaneous treatments for both countries over almost all rounds. The similarity between the $\boldsymbol{a}_t$ values for the simultaneous treatments, on the one hand, and for the sequential treatments, on the other hand, is striking.[18] Furthermore, the decrease in the $\boldsymbol{a}_t$ values seems to be sharper for country 1 than for country j. Because the NE and SPE are not at a boundary, $\boldsymbol{a}_t$ measures departures from NE and SPE behavior in both directions. Abatements below the NE or SPE level may therefore serve as a means of punishing subjects who behave selfishly. In order to analyze the structure of the deviations from NE and SPE behavior, we classified the individual deci-

---

[18] However, significant differences can only be observed for few periods.

sions of country j into three groups: cooperative behavior, best response or NE/SPE behavior, and abatements that are below the best response or NE/SPE.[19] The corresponding intervals are described at the bottom of figures 10-13, which show the fractions of the three behavioral patterns for each treatment.

Figure 10: Individual behavior of country j

T I (h=0, sim)
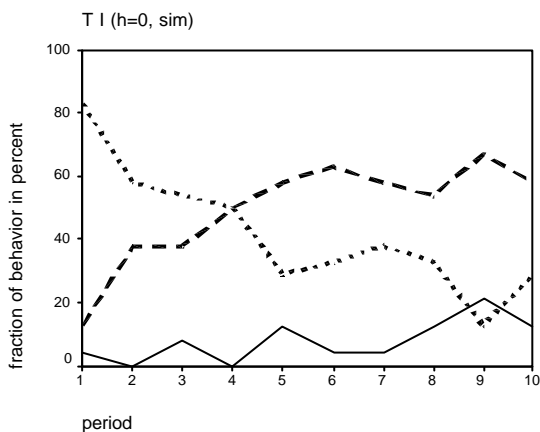
Figure 11: Individual behavior of country j

T III (h=0, seq)
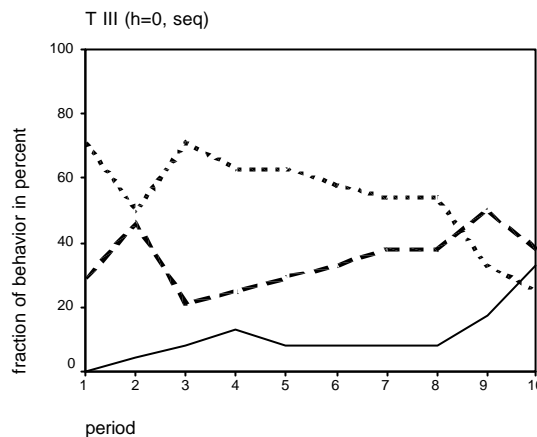
Figure 12: Individual behavior of country j

T II (h>0, sim)
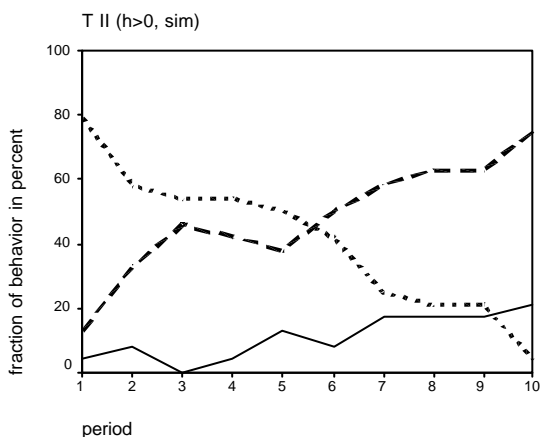
Figure 13: Individual behavior of country j

T IV (h>0, seq)
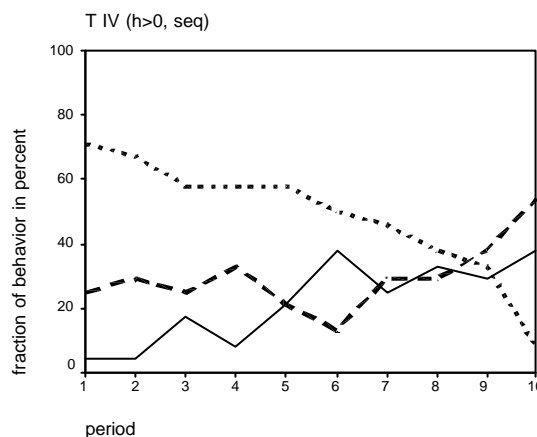
·········· Cooperative behavior, i.e. abatement between max (bR, Eq) and PO.

– – – Abatement between min (bR, Eq) and max (bR, Eq) with "bR" as individual best response to the aggregated abatement of the others in the current period and "Eq" as individual abatement in NE or SPE (with +/- 20%).

——— Abatement below min (bR, Eq).

A significant fraction of abatement decisions of country j are *below* the best response level, and in all treatments the proportion of this kind of deviation from individual rational behavior increases during the game. This may be interpreted as a sign of the existence of some punishment behavior or negative reciprocity shown by the subjects in the role of country j. Note that particularly in the sequential treatments there are at least some subjects in the role of country j

---

[19] As punishment behavior of country 1 is extremely rare (we observe it in 5 of 240 cases) we only consider country j here.

who show an increasing tendency to abate below the best response level. This may be a reaction of these subjects to the frustrating experience made in the early rounds that the high abatement of country 1 does not motivate the other countries j to follow suit. Furthermore, figures 10-13 confirm the finding that cooperative behavior (of country j) is stable more frequently and for longer in the sequential treatments than in the simultaneous treatments.

Figures 14-16 show the average profits earned by the subjects. These profits are not identical to the welfare measure used in the Hoel model because there the "extra profit" country 1 derives from abatements above the Nash level (measured by h) is not part of the welfare of country 1. The positive welfare effect of "over-abating" of country 1 results because the welfare loss suffered by this country is overcompensated for by the welfare gains of the other countries.

Figure 14: Profit for country 1

Figure 15: Profit for country j

Figure 16: Total profit
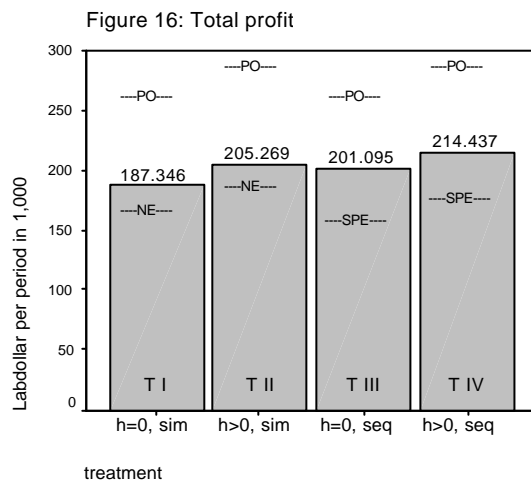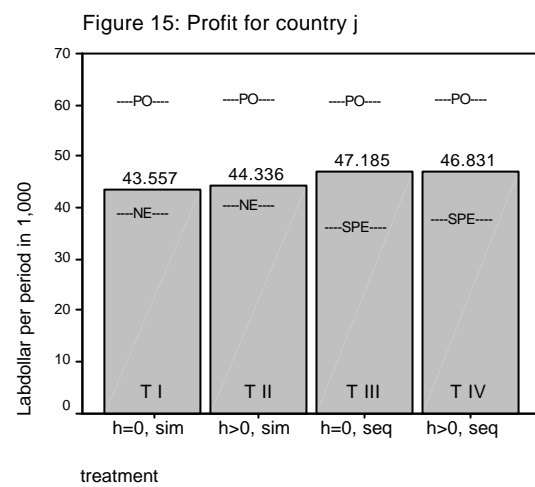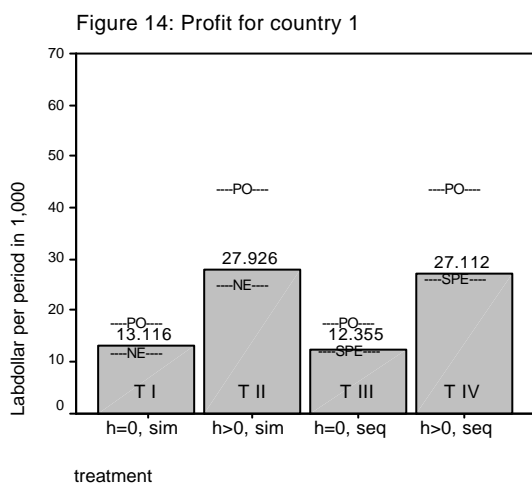
We summarize our findings concerning the profits in the following.

Observation 3:

  a. *Country 1 earns a significantly higher profit in the $h > 0$ treatments than in the $h = 0$ treatments (MW-U test, two-sided, 1% level). But, comparing the sequential with the simultaneous treatments, it turns out that country 1 is unable to increase its profits by using leadership. Although abatement behavior with leadership differs a lot from that*

without leadership, there is no significant difference between the respective profits (MW-U test, two-sided, 5% level). The hypothesis that the profit of country 1 is equal to the NE or SPE value cannot be rejected for all treatments (two-sided t-test, 5% level).

b. *Country j is able to increase its profits above the NE or SPE value in all treatments. The difference is particularly high although not significant in the sequential treatments (two-sided t-test, 5% level).*

c. *Total profit is slightly above the NE or SPE value in all treatments. The difference is particularly high although not significant in the sequential treatments (two-sided t-test, 5% level).*

5.   Discussion

Our data support the Hoel model for the treatments with a simultaneous decision protocol at least for the second half of the 10 rounds. Therefore, it seems fair to state that the Hoel model describes actual behavior surprisingly well, i.e. we may confirm hypothesis 1, in an environment where subjects act simultaneously. On the other hand, our findings for the simultaneous treatments are in line with the stylized facts of many public good experiments: abatement starts between the NE and PO level, decays during the course of the game, and then displays a final-round effect. BROSIG ET AL. (2003) have shown that, in standard public good games, subjects try to coordinate their behavior in order to realize the efficient outcome but that this coordination is only successful if *all* subjects stick to their promise to cooperate. Normally, this is not the case and cooperation breaks down after a few periods. This line of reasoning seems to be in line with our observations.

The most important question we sought to answer with this experiment was whether or not "leadership matters". Having a leader may open a way to solving the coordination problem just mentioned. If the leader starts each round with the PO abatement, this could serve as a kind of focal point for the followers. Our data show that this does not happen in the sequential treatments, although the variance of individual behavior is large and this results in only a few significant results. In most groups (10 of 12) of the sequential treatments, we observe a similar pattern of leader behavior.[20] Country 1 tries to be a real leader and to present a "good example", i.e. its abatement is equal to or near the social optimum. The reaction of the followers is mixed. There are some groups (3 of 12) which succeed in abating at the PO level from the first round to the second-last. One of these groups even plays PO in the last round. However, in most cases (7 of 12) the leader fails to induce cooperation, i.e. in these groups there are few followers who react cooperatively to the leader's signal. High efforts by the leader and the cooperative followers are exploited by the majority of defective followers. This behavioral pattern is the explanation for the fact that the mean profit of country 1 does not exceed the SPE values in the sequential treatments although the mean abatement of country 1 is significantly

---

[20] In 2 of 12 groups we observe SPE behavior in both the leader and the followers.

higher than the SPE values and that the mean profit of country j exceeds the SPE values in the sequential treatments. Based on these observations, we must reject hypothesis 2, but otherwise our data do not support the idea that leadership is an effective means to create stable cooperation.

6.   Conclusion

The primary objective of our experiment was to test the Hoel model and to analyze the influence of leadership. Since the external validity of results gained in singular laboratory experiments is restricted to the specific laboratory environment, we have to admit that we are not able to give any recommendations for environmental policy purposes. However, we may come to the conclusion based on our results that, firstly, the Hoel model describes the individual behavior surprisingly well in an environment with a simultaneous decision protocol and, secondly, leadership matters a lot but is not able to increase the profit of the leader and to overcome the social dilemma situation all countries are confronted with. Only the followers who free ride at the expense of the leader and the cooperative followers in cooperative groups can increase their profits. In particular, the experiments show that countries that want to increase their own profit and the total profit of the group with leadership should not put too much hope in the effect of their good example. Even if some follow this example, the probability that other followers free ride and cooperation breaks down very soon is high. All in all, leadership does not seem to be an appropriate tool to overcome social dilemma problems, even if leaders have a strong incentive to induce cooperative behavior.

References

ANDREONI, J. (1995): Warm Glow versus Cold-Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments, *The Quarterly Journal of Economics*, 110, p. 1-21.

BARRETT, S. (1994): Self-Enforcing International Environmental Agreements, *Oxford Economic Papers*, 46, p. 878-894.

BROSIG, J./OCKENFELS A./WEIMANN, J. (2003): The Effect of Communication Media on Cooperation, *German Economic Review*, 4, p. 217-242.

Federal Minister for the Environment, Nature Conservation and Nuclear Safety (2002): Press Release from 31. 07. 2002, 177/02.

FISCHBACHER, U. (1999): z-Tree: A Toolbox for Readymade Economic Experiments, Working Paper No. 21, University of Zurich, 1999.

HOEL, M. (1991): Global Environmental Problems: The Effects of Unilateral Actions Taken by One Country, *Journal of Environmental Economics and Management*, 20, p. 55-70.

VAN DER HEIJDEN, E./MOXNES, E. (2000): Leadership and Following in a Public Bad Experiment, working paper, SNF Report No. 13/00, Bergen.

Appendix 1

In this appendix we derive the solutions for the Nash equilibrium for the simultaneous decision protocol, the subgame perfect equilibrium for the mixed sequential-simultaneous decision protocol, and the Pareto optimal solution based on the specification used in the experiment.

Nash equilibrium (NE) for the simultaneous decision protocol

Country 1 maximizes the difference between benefit and costs of abatement given the abatement of all other countries, i.e.

$$\max_{X_1} \boldsymbol{p}_1 = b\left[A(X_1 + X_{-1}) - 0.5(X_1 + X_{-1})^2\right] + h(X_1 + X_{-1}) - \left(cX_1^2 + T\right)$$

with $\quad X_{-1} = \sum_{j=2}^{N} X_j$ (A1).

The reaction function of country 1 is

$$R_1(X_{-1}) = \frac{1}{2c+b}(bA + h - bX_{-1})$$ (A2).

Since all other countries ($j = 2,...,N$) are identical in their benefit and cost functions, we can substitute abatement X_{-1} by $X_{-1} = (N-1)X_j$. The reaction function

$$R_1(X_j) = \frac{1}{2c+b}(bA + h - (N-1)bX_j)$$ (A3)

describes the best response of country 1 to the abatement chosen by country j, X_j.

Country j, $j = 2,...,N$, maximizes the difference between benefit and costs of abatement given the abatement of all other countries, i.e.

$$\max_{X_j} \boldsymbol{p}_j = A(X_j + X_{-j}) - 0.5(X_j + X_{-j})^2 - (cX_j^2 + T) \qquad \text{with} \quad X_{-j} = \sum_{i \neq j}^{N} X_i$$ (A4).

The reaction function of j is

$$R_j(X_{-j}) = \frac{1}{2c+1}(A - X_{-j})$$ (A5).

Since country j and all other $N-2$ countries k, $k \neq j \wedge k \neq 1$, have identical benefit and cost functions, we can substitute abatement X_{-j} by $X_{-j} = (N-2)X_j + X_1$. The reaction function of j,

$$R_j(X_1) = \frac{1}{N-1+2c}(A - X_1)$$ (A6),

describes the best response of country j to the abatement of country 1, X_1. The NE for the simultaneous decision protocol results as the intersection of the reaction functions (A3) and (A6), and the interior solution for the NE is

$$X_1^{NE} = \frac{2cbA + h(N-1+2c)}{2cb + 2(N-1)c + 4c^2}, \qquad X_j^{NE} = \frac{2cA - h}{2cb + 2(N-1)c + 4c^2},$$

and $\qquad X^{NE} = X_1^{NE} + (N-1)X_j^{NE} = \dfrac{(b+N-1)A + h}{N-1+2c+b}$ \hfill (A7).

For $b=1$ and $h=0$, we have the perfectly symmetric NE[21] for the simultaneous decision protocol with

$$X_1^{NE} = X_j^{NE} = \frac{A}{N+2c} \qquad \text{and} \qquad X^{NE} = \frac{NA}{N+2c}$$ \hfill (A8).

For $b<1$ and a sufficiently small h, we have $X_1^{NE} < X_j^{NE}$, i.e. the country with the smaller marginal benefit from abatement (here country 1) abates less than all other countries in NE.

We are interested in analyzing the effects of a marginal increase in h. It is easy to show that the results of the general model hold for our specification, i.e. $\partial X_1^{NE}/\partial h > 0$, $\partial X_j^{NE}/\partial h < 0$, and $\partial X^{NE}/\partial h > 0$. For profits, we can show that $\partial p_1^{NE}/\partial h > 0$, $\partial p_j^{NE}/\partial h > 0$, and $\partial p^{NE}/\partial h > 0$, i.e. payoffs increase with a marginal increase in h, which is an element of the real payoff function of country 1 here. This effect is independent of the relative size of the marginal benefit of country 1.

Subgame perfect equilibrium (SPE) for the mixed sequential-simultaneous decision protocol

Country 1 maximizes the difference between benefit and costs of abatement given the knowledge that the N-1 other countries will behave according to their best response function (A6)

$$\max_{X_1} p_1 = b\left[A\left(X_1 + (N-1)R_j(X_1)\right) - 0.5\left(X_1 + (N-1)R_j(X_1)\right)^2\right] + h\left[X_1 + (N-1)R_j(X_1)\right] - \left(cX_1^2 + T\right)$$
\hfill (A9),

i.e. country 1 can use its first mover advantage by choosing a point on the follower's best response function which maximizes its profit by backward induction. The solution for the subgame perfect equilibrium is

$$X_1^{SPE} = \frac{2cbA + h(N-1+2c)}{2cb + (N-1+2c)^2}, \qquad X_j^{SPE} = \frac{A(N-1+2c) - h}{2cb + (N-1+2c)^2},$$

and $\quad X^{SPE} = X_1^{SPE} + (N-1)X_j^{SPE} = \dfrac{2cbA + A(N-1)(N-1+2c) + 2ch}{2cb + (N-1+2c)^2}$ \hfill (A10).

---

[21] See BARRETT (1994).

We can show that $X_1^{SPE} < X_1^{NE}$, $X_j^{SPE} > X_j^{NE}$, and $X^{SPE} < X^{NE}$, i.e. the change from the simultaneous decision protocol to the mixed sequential-simultaneous decision protocol leads to a lower aggregated abatement. Regarding profits, the change from the simultaneous decision protocol to the mixed sequential-simultaneous decision protocol leads to a higher profit for country 1, a lower profit for country j, and (for the chosen parameters) a lower aggregated profit. The influence of parameter h on abatements and profits is the same as in the simultaneous case.

Pareto optimum (PO)

In the Pareto optimal allocation, total profit from abatement is maximized, i.e. the global planner has the following optimization problem:

$$\max_{X} \boldsymbol{p} = \boldsymbol{p}_1 + (N-1)\boldsymbol{p}_j = (b+N-1)(AX - 0.5X^2) + hX - \left(\frac{c}{N}X^2 + NT\right) \tag{A11}.$$

In the PO all countries adjust their marginal abatement costs to the marginal social benefit from abatement. Since all countries have the same marginal abatement costs, $C_i' = 2cX_i$, in the PO all countries choose an equal abatement independent of the protocol of play given by

$$X_i^{PO} = \frac{A(b+N-1)+h}{N\left(b+2\frac{c}{N}+N-1\right)} \quad \text{and} \quad X^{PO} = \frac{A(b+N-1)+h}{b+2\frac{c}{N}+N-1} \tag{A12}.$$

It is easy to show that $X^{PO} > X^{NE}$ holds. For abatement, it follows that $\partial X_i^{PO}/\partial h > 0$ and $\partial X^{PO}/\partial h > 0$. For profits, we can show that $\partial \boldsymbol{p}_1^{PO}/\partial h > 0$, $\partial \boldsymbol{p}_j^{PO}/\partial h > 0$, and therefore $\partial \boldsymbol{p}^{PO}/\partial h > 0$.

Appendix 2

Screenshot: Payoff simulator (T I and II)

payoff simulator				
profit maximizing response for country 1		**simulation for country 1**		
expected abatement of the others	?	expected abatement of the others	?	
profit maximizing response for country 1	#value!	abatement for country 1	?	
total abatement	#value!	total abatement	0	
profit for country 1	#value!	profit for country 1	#value!	
profit maximizing response for country j, j=2, .., 5		**simulation for country j, j=2, .., 5**		
expected abatement of country 1	?	expected abatement of country 1	?	
expected abatement of the other j	?	expected abatement of the other j	?	
profit maximizing response for country j	#value!	abatement for country j	?	
total abatement	#value!	total abatement	0	
profit for country j	#value!	profit for country j	#value!	
profit for country 1	#value!	profit for country 1	#value!	

Appendix 3

Instructions for T I ($h = 0$, *sim*).

Instructions (Translation from German)

Welcome to the Magdeburg Experimental Laboratory MAXLAB!

Please read the instructions carefully in the next 15 minutes. If you have any questions, please open the door behind you or give a sign.

In this laboratory experiment, you can earn money depending on your decisions. Your earnings are computed in Labdollars (LD). The exchange rate between EUR and LD is 1:40,000, i.e. 40,000 LD are equal to 1 EUR.

After the experiment, your credit in LD will be converted into EUR and paid in cash. In addition to your earnings from the experiment, you have received a 12 EUR show-up fee before the experiment.

You will make your decisions during the experiment anonymously. Only the experimenter knows your identity. He will keep your information in confidence.

In the experiment, you are the representative of a country at an international environmental conference. You negotiate with representatives of other countries on the emissions abatement of a global pollutant. Five countries take part in the conference. Your are the representative of country 1.

In each period, you have to decide on the abatement of your country 1, V_1.

The profit function of country 1 is

$$G_1(V, V_1) = \underbrace{\frac{30}{47}\left(500V - \frac{V^2}{2} - 62.667\right)}_{\text{benefit for country 1}} - \underbrace{\left(3V_1^2\right)}_{\text{costs for country 1}}.$$

The profit maximizing response function (GMA) of country 1 is

$$V_1^{GMA}(V_{-1}) = \frac{1.875}{39} - \frac{15}{156}V_{-1}, \qquad \text{with} \quad V_{-1} = \sum_{i=2}^{5} V_i.$$

The simulator in the lower part of the screen will help you to make your decision. The simulator assists in both computing the profit maximizing response for country i and simulating the effects of different abatement decisions on the profit of country i. The simulator is identical for all countries.

The procedure: The experiment will be replicated ten times. A repetition is denoted as a period. In each period, you will make your decision under the same conditions. All countries decide simultaneously on their abatement V_i, i.e. each country i decides at the same time without knowing the decision of all other countries.

The cumulated profit G_{1kum} shows your current earnings from the experiment. After the last period, you will be paid your earnings in cash (G_{1kum} divided by 40,000). If you make a loss ($G_{1kum} < 0$), a message will appear that your show-up fee will be used to cover the loss. If your earnings are still negative after the last period, you will have to cover your loss with your show-up fee. The greatest possible loss cannot exceed your show-up fee of 12 EUR.

Table 1 depicts abatement and profit for countries 1 to 5 in the Nash equilibrium and the social optimum (the row for your country is marked).

Table 1: Nash equilibrium and social optimum

	Nash equilibrium		Social optimum	
	V_i	G_i	V_i	G_i
country 1	**30**	**11,707**	**79**	**17,546**
country 2	47	38,611	79	60,765
country 3	47	38,611	79	60,765
country 4	47	38,611	79	60,765
country 5	47	38,611	79	60,765
sum	218	166,151	395	260,606

Before the experiment starts, you will take part in a test phase which last 2 periods. In this test phase, you can familiarize yourself with the decision environment and the simulator. The only difference between the test phase and the experiment is that the representatives of the other countries are substituted by four computer agents in the test phase. The computer agents will make the same decision in both periods. The profits of the test phase are not relevant for your earnings in the experiment. After the test phase, the actual experiment starts.

When you are familiar with the instructions and if you do not have any questions, please click on the "continue" button. The test phase will start in a few minutes.

Good luck!

CESifo Working Paper Series

(for full list see www.cesifo.de)

_____