SNF Report No. 13/00

Leadership and following in a public bad experiment

by

Eline van der Heijden
Erling Moxnes

SNF Project No. 4375
Spredning av energi- og miljøpolitikk

The Project has been funded by the Research Council of Norway
under the SAMRAM program

FOUNDATION FOR RESEARCH IN ECONOMICS AND BUSINESS ADMINISTRATION
BERGEN, MAY 2000

ISBN 82-491-0040-9
ISSN 0803-4036

CONTENTS

FOREWORD

This is a report in a series dealing with diffusion of energy and environmental policies and effects of international leadership. Eline van der Heijden works at CentER at Tilburg University. The project is funded by the SAMRAM program of the Research Council of Norway.

1. INTRODUCTION

Global and regional environmental problems can be described as commons problems. Prominent examples are climate change and ozone depletion. For each country, the marginal benefits of emitting pollutants typically exceed the extra environmental costs caused by relatively small own contributions to the total pollution. Still, for the entire region or globe, marginal environmental costs could be substantially higher than the marginal benefits. This situation calls for some sort of cooperation. In this report we investigate whether cooperation is encouraged by countries that take a leadership role and make unilateral emission reductions.

Our focus is on the competitive case. Will unilateral reductions by a leader, for instance a small country, lead to lower emissions by other nations? The competitive case can also be seen as representative of a situation with a small initial group of signatories (leaders), which compete with independent nations or groups of nations that have not committed themselves to any treaty. If unilateral reductions turn out to have a positive influence on the behavior of others in a competitive situation, it is not unlikely that unilateral reductions could also speed up negotiations and lead to better treaties. This question is, however, left for further research.

One aspect of the question we pose has been studied by the use of game theory, Hoel (1991). Hoel assumes that other nations than the leader choose policies dictated purely by their self-interest, and he seeks Nash equilibria of noncooperative games and the Nash bargaining solution in case of cooperation. In the noncooperative case he finds that unilateral actions lead to lower total emissions and increased total welfare. In the cooperative case, total emissions could go in either direction depending on the parameters of the problem. In case the leader is a small country, the effects become negligible.

Our focus is on other effects of unilateral reductions than those studied by Hoel. He explicitly states that he will not consider the possibility that unilateral actions by one nation "might lead to similar behavior from other countries", p.56. His choice is consistent with his assumption about individuals maximizing pure self-interest (in a narrow sense only considering direct economic effects of changes in emissions). However, both the assumption about narrow self-interest and the one about maximization are debatable. Laboratory experiments of public good and public bad games typically show deviations from Nash equilibria, Ledyard (1995). These could be caused by violations of one or both assumptions. First, consider utility functions[1]. Even within a broad set of utility functions based on self-interest, desires to be perceived as good, fair, innovative, clever etc. could stimulate leaders as well as followers to “unilateral reductions”.

Second, consider maximizing behavior. It is not obvious that actors have the appropriate information about themselves (e.g. costs, for which widely differing estimates exist), about others (e.g. their costs, utility functions, and beliefs), and that they are able to deduct maximizing strategies from such uncertain information. Discussing lessons from evolutionary game theory, Mailath (1998) points to the likely importance of learning by imitating observed successful behavior by others. He also points out that this perspective is relevant for the development of norms and conventions. Thus it seems pertinent to speculate whether a leading nation taking unilateral actions could come to influence directly the strategies of others. For instance, environmental interest groups are likely to point to leading nations as having appropriate strategies.

We will study the effect of leaders by the use of a laboratory experiment of a repeated public bad game, where the public bad represents pollution. The traditional design is altered in that in each round a leader makes his or her decision before the others, and then this decision is communicated to the followers before they make their decisions simultaneously. Cost functions are chosen such that the Nash solution becomes a corner solution, both when there is a leader and when there is not. Thus by design, we avoid potential confusion with the interior solutions found by Hoel for the competitive case[2]. Each group of participants get to play the game twice, once with a leader who has the same cost structure as the followers and once with a leader who face no direct costs of acting socially. The two treatments enable us to study the effect of the leader incentive structure on followers (indicating whether followers are involved in learning or are motivated by fairness issues), and to study leader behavior (do they act according to rational expectations regarding follower behavior).

As far as we know, no public good or public bad experiment has thus far been designed to investigate the impact of a leader. I.e. we have seen no other leader-follower experiment with a decision protocol that is mixed sequential-simultaneous. A major reason for choosing this design is that any nation in principle is free to pursue a leadership role

---

1 Recently, several theories have been advanced in which considerations of fairness, altruism, reciprocity, equity et cetera are incorporated (see e.g. respectively Fehr and Schmidt (1999), Levine (2000), Rabin (1993) and Dufwenberg and Kirchsteiger (1998), and Bolton and Ockenfels (2000)).

2 The internal solution is not a problem in itself. Keser (1996) finds that contributions in a public good experiment with an interior Nash equilibrium are similar to those in a corner solution public good experiment.

within a region or on a global scale. Other, purely sequential game designs[3], assume that actors make decisions in an exogenously determined order. While there seems to be interesting applications for such designs, they do not seem to fit regional or global pollution problems where numerous policy initiatives are needed over long time intervals. Neither is there a natural sequential order in which nations make such decisions, nor can such an order be forced on nations. A good reason for nations to object to ordering is that theory predicts the first movers to have an advantage and the last movers to carry the largest burdens. These predictions are confirmed in sequential step-level public good games[4] and in sequential common pool resource games[5]. In our design, theory holds that there is neither a first mover advantage nor a last mover disadvantage.

The next section outlines the hypotheses and describes the public bad game. Section 3 describes the experimental procedure. Section 4 presents the main results. We find that leaders on average have a statistically significant effect on the followers. Leader incentives are likely to matter for followers. Evidence is found of both a saturating effect of fairness and a tendency to mimic leader behavior. In the case with no direct leader costs of acting socially, the leaders behave as if leadership does not matter. Otherwise, the results are consistent with previous findings for public bad games without leaders. The last section contains a concluding discussion and gives some lines for further research.

---

3 See e.g. Budescu et al. (1995), Erev and Rapoport (1990), Morris et al. (1995), and Rapoport (1993).

4 See Erev and Rapoport (1990).

5 E.g. Budescu et al. (1995).

2. HYPOTHESIS AND EXPERIMENT

Our focus is on the effect of a leader in a public bad game. As mentioned in the introduction, we will use an experiment which has a corner solution for the Nash equilibrium. Assuming selfish (in a narrow sense including only direct costs and benefits) and maximizing subjects, we should expect that there is no effect of a leader. Hence our main null hypothesis reads:

H_0: *There is no effect of a leader in a public bad game with a Nash corner solution*

When we want to test this hypothesis it is because of the objections to standard assumptions raised in the introduction. We are also inspired by empirical findings in public good and public bad experiments of behavior that deviates from Nash predictions. The same mechanisms that cause these deviations could also be important for the effects of a leader.

Furthermore, we will investigate the importance of the incentive structure. Will it matter for the followers whether the leaders face the same costs as the followers or whether they have no direct costs of acting socially? If fairness is part of followers' utility functions, we should expect the leader costs to matter. Moreover, if fairness is the only reason why leaders have an effect, we should expect the effect of leadership to go away completely in the case leaders face no direct costs. If it does not vanish completely, an element of learning (imitation) could be present. We consider the test exploratory and do not formulate a null hypothesis.

Finally, we want to study the behavior of leaders. When facing the same costs of acting socially as the followers, leaders should behave like the followers, unless they expect the effect of leadership to be sufficiently large to justify an attempt to set the good example. We use the revealed effect of leadership to calculate the optimal action of leaders, and use this as a benchmark for leader behavior. In the case leaders face no direct costs of acting socially, any positive effect of leadership should imply a benchmark of zero investment in the public bad for leaders. For both treatments the null hypothesis is given by the rational response, given correct expectations about follower behavior:

H_0: *Leaders will behave according to self interest and correct expectations about follower behavior.*

2.1. The public bad game

The basic experimental design is similar to the public bad framework used by Andreoni (1995), with two notable exceptions: we introduce a leader and all subjects get full information about the investments of all the other players in the preceding round. In addition, parameter values differ.

The features of our public bad game are as follows. Subjects play in groups of five. In each round, subjects are endowed with 20 tokens, which they can allocate between two projects: project A (the public bad) and project B. Investing in project A gives a direct private return of 0.7 per token invested. Investing in project B gives a private return of 0.4 per token invested. However, investing in project A also has a negative external effect: each token invested in project A yields a negative return of 0.1 to all group members. So, payoff π_i to individual i (i=1, 2, ..,5) when she or he invests x_i^A in project A and x_i^B in project B reads

$$\pi_i = 0.7x_i^A + 0.4x_i^B - 0.1\sum_{j=1}^{5} x_j^A \qquad (1)$$

where x_j^A denotes the investment in project A (the public bad) by subject j (j=1, 2, ..,5), and $x_i^A + x_i^B = 20$. Using this last equality, we can rewrite Equation 1 as

$$\pi_i = 8.0 + 0.2x_i^A - 0.1\sum_{j \neq i} x_j^A \qquad (2)$$

From Equation 2 it follows that purely selfish, money-maximizing subjects have a dominant strategy to invest their total endowment in the public bad. That is, the unique Nash equilibrium is $x_i^A = 20$, which gives a total investment of 100 tokens in project A and a payoff $\pi_i = 4$. However, higher payoffs can be obtained if subjects invest in project B. If all members of a group decide to invest the total endowment in project B, i.e. $x_i^A = 0$, the payoff to each individual would be twice as much, namely 8, which is the socially optimal outcome. That is, the welfare maximizing solution is obtained by full cooperation[6].

---

6 The outcome $x_i^A = 0$, is also the unique symmetric Pareto efficient outcome. Note, however, that many other, asymmetric Pareto efficient outcomes exist. For instance, the outcome in which 4 players contribute everything to the public bad and one player contributes nothing to the public bad is Pareto efficient as in this situation no player can be made better off without making a player worse off.

One subject in each group is randomly selected to be a leader. The leader decides first on his or her investment. This decision is then communicated to the other four members in the group, after which they make their decisions simultaneously. I.e. the leader decides first and then the followers decide simultaneously (the decision-making protocol is mixed sequential-simultaneous). Independent of what the leader decides, there is still the same unique Nash corner solution for the followers.

With respect to the information given to the subjects, there is one difference between our design and the standard public bad game: usually individuals are only informed about the aggregate investments via their own payoff, whereas in our design subjects get feedback on the individual decisions of their group members. This has been done to avoid asymmetries between the leader, for which the investment has to be revealed, and the others. An experiment indicates that there is no significant effect of this extra information on average investments, Van der Heijden and Moxnes (1999).

The parameters of the game are summarized by the marginal per capita return, MPCR. Since all players (with one exception to be announced later) face the same MPCR, the leadership effects we find are for that particular set of parameters. In a public good game, Isaac and Walker (1988) find that contributions vary with MPCRs. This indicates that our choice of MPCR could influence the magnitude of the leader effect. Our game has a MPCR of 0.33[7], which is quite close to the MPCR used by Andreoni (1988) in his public bad game, 0.5.

In other words, the parameters (costs) have been set such that investments in project A are clearly profitable, for given investments by the others. From Equation 2 we see that individuals gain 4 tokens by going from fully social (x_i^A = 0) to fully selfish investments (x_i^A = 20). In relative terms this means that if all others have invested socially (x_j^A = 0), subject *i* reduces his or her profits by 33 percent by being cooperative instead of selfish (from 12 to 8). If all others have invested selfishly (x_j^A = 20), the loss for subject *i* is 100 percent by being fully social (from 4 to 0). These are rather large relative amounts which should stimulate money maximizing subjects to choose selfish investments and ignore leader decisions. Compared to actual gains from choosing polluting options as opposed to abatement options, the relative gains in the experiment seem high. This is particularly the case when comparing to the first abatement investments to move away from "business as usual". If "business as usual" means that

7 Using Equation 1 we find: $MPCR = \dfrac{d(0.1\sum_{j=1}^{5} x_j^A)/dx_i^A}{d(0.7x_i^A + 0.4(20 - x_i^A)/dx_i^A} = \dfrac{0.1}{0.3}$.

the economy is in an efficient state, marginal costs of the two investment options are equal. Hence, in reality, there should be virtually nothing to gain by acting selfishly and not making the first marginal reductions in pollution.

2.2. The experimental treatments

In the experiment we have employed two experimental treatments: the basic design just described and one in which the cost structure of the leader is changed such that the leader faces no direct costs of acting socially. We will refer to the two treatments by the terms "same-costs" and "no-leader-costs".

In the no-leader-costs treatment the direct costs of acting socially for the leader (*i*=L) is removed by changing the return on project B. Instead of having a return of 0.4 as in Equation 1, the leader faces a return of 0.6, such that the leader's profit function reads:

$$\pi_L = 0.7x_L^A + 0.6x_L^B - 0.1\sum_{j=1}^{5} x_j^A = 8.0 - 0.1\sum_{j \neq L} x_j^A \quad (3)$$

Now the profits of the leader only depend on investments in the public bad made by the others, $x_j^A, j \neq L$. Hence, the leader has a clear incentive to influence the others to invest as little as possible in the public bad. Followers still face the profit function of Equation 1. All subjects were informed about this asymmetry of the cost structure in the no-leader-costs treatment.

2.3. Benchmarks for optimal leader behavior

As mentioned earlier, according to standard game theory, leaders should behave like the followers, unless they expect the effect of leadership to be sufficiently large to justify an attempt to set the good example. Here we find benchmarks for how strong the effect must be for leaders to set the good example. In the no-leader-costs treatment, even the slightest positive effect of leadership implies that leaders should set the good example since it is costless for the leaders. In the same-costs treatment, the benchmark is the effect of leadership for which leaders become indifferent, i.e. where $d\pi_L / x_L^A = 0$. We rewrite Equation 1 by entering a reaction function, $x_j^A = \beta + \alpha x_L^A$, for the followers:

$$\pi_L = 0.7x_L^A + 0.4(20 - x_L^A) - 0.1\{x_L^A + 4(\beta + \alpha x_L^A)\} \quad (4)$$

To simplify, all reaction functions for the four followers are assumed equal and linear, where β is the constant and α denotes the slope, i.e. the sensitivity of followers' responses to the leader's decision. Differentiating with respect to the leader investment x_L^A and equating with zero, we find that the effect of leadership α must be equal to or greater than 0.5 to make it profitable for the leader to set the good example.

3. EXPERIMENTAL PROCEDURE

We ran three experimental sessions with three groups of five subjects in each session, one session in April 1997 and two in November 1999. In each session we ran both treatments, i.e. we employed a within-subject (group) design. To control for order effects we had two sessions with first 10 rounds of the same-costs treatment and then 10 rounds of no-leader-costs treatment, and one session in which this order was reversed. The subjects were students from the Norwegian School of Economics and Business Administration who were recruited from classes. No subject participated in more than one session. Subjects were told that they could earn between NOK 100 and 180 in about one hour (about 1.5 to 2.7 times a typical hourly wage to students) . They knew that rewards were contingent on performance.

Upon arrival subjects were seated behind computers such that groups were formed in a random way. The computers were separated by curtains, and the subjects could not identify the other members in their group (members of all groups were mixed). Groups remained the same during the session (the partner design by Andreoni (1988)). Instructions (in Norwegian) were divided and read aloud by the experimenter[8]. Subjects were encouraged to ask questions. Few questions were asked. After that the experiment started. In each treatment, all parameters of the experiment were common knowledge to all subjects. Subjects knew that they would play 10 rounds in the experiment, and then another 10 rounds with a different design. One exception is the session run in April 1997 which lasted for only 8 rounds for each of the treatments. In this session we used a rather time consuming manual procedure, receiving and passing out information on paper. The same room was used as in the computerized sessions.

Subjects decided how much of the endowment they wanted to invest in project A; the remaining amount was automatically invested in project B. Software was used to calculate private payoffs. The sequencing was as follows: In each round the leader first decided how much he or she wanted to invest in the public bad in that round. The followers were informed about this decision and about all follower decisions and own returns in the preceding round. Then followers made their decisions for that round. Information about follower decisions plus the leader return was then passed to the leader before he or she made his or her decision for the next round, and so on.

---

8 An English translation of the instructions is available upon request from the authors.

After the session, subjects were privately paid their earnings from all rounds. Each session lasted for about one hour, except the first one which lasted for one hour and a half. Average earnings were NOK 116.7 (when adjusting upwards the averages for those with only 8 rounds per treatment), including NOK 20 for showing up.

4. RESULTS

First we analyze averages over rounds. The tests performed are weak and conservative since we disregard the fact that the games are repeated over 10 rounds in each treatment. Next we perform regressions over all rounds. These are likely to overstate the accuracy of some findings since observations of groups over rounds are not totally independent observations[9]. Correct measures of accuracy are likely to lie in-between the two results.

4.1. Results based on average investments over rounds

Table 1 presents the average investments in the public bad (project A) for sessions A, B, and C and for groups 1, 2, and 3 within each session, i.e. averaged over all rounds. Sessions A and B were computerized, session C was carried out manually. In sessions A and C, the treatment same-costs came first, while in session B the treatment no-leader-costs came first. The table shows separate results for the two treatments, and it shows differences between investments in the second and the first treatment, irrespective of which treatment came first.

Table 1: Average investments in the public bad by followers and leaders in both treatments, and differences between investments in the second and the first treatment. Sessions A and C same-costs first, sessions B no-leader costs first.

	Same-costs		No-leader-costs		Second minus first treatment	
Group	Followers	Leader	Followers	Leader	Followers	Leader
A_1	13.18	13.30	15.35	11.90	2.17	-1.40
A_2	14.35	14.00	17.38	15.00	3.03	1.00
A_3	14.43	19.20	18.55	18.00	4.12	-1.20
B_1	19.15	19.90	14.48	12.50	4.67	7.40
B_2	16.88	14.00	14.38	8.80	2.50	5.20
B_3	18.88	17.00	15.40	9.00	3.48	8.00
C_1	15.47	15.63	17.81	13.13	2.34	-2.50
C_2	18.28	16.25	17.50	12.50	-0.78	-3.75
C_3	15.16	13.75	15.94	10.00	0.78	-3.75
Average	16.20	15.89	16.31	12.31	2.48	1.00

4.1.1. Some general findings

Observation 1: On average, subjects behave as weak free riders. Investment levels are closer to the Nash prediction of 20 than to the socially efficient level (investing 0).

9 It is more interesting that 10 groups follow the leader in one round than that one group follows the leader in 10 rounds.

This observation is in line with the findings of other public good and public bad studies (e.g. Weimann (1994), Andreoni (1995)).

It its sometimes argued that contact between subjects and experimenters can influence the results, the anonymity hypothesis. As our manual session implied more personal contact than the computerized sessions, one may expect to find differences in our experiment as well. However, a non-parametric Mann-Whitney U test with group averages as units of observation reveals no significant difference between the average investments in the computerized and manual sessions ($p=0.20$, $n_1=6$, $n_2=3$ for both treatments). So:

Observation 2: There is no significant difference between sessions conducted by computer and manually.

This result is in line with Weimann (1994), who concludes that personal contact has no significant effect on individual contributions in a public good game. Similarly, Bolton et al. (1998) find no evidence for the anonymity hypothesis. As we find no differences between the sessions, in what follows we treat all the data as if they came from the same type of experiment.

Observation 3: There is a tendency for investments in the public bad to increase over rounds.

Observation 3 can be seen from Figure 1 for sessions A, B, and C. The graphs show average investments in the public bad over groups for leaders and followers, round by round. The observed pattern is similar to the typical patterns found in standard public good and public bad experiments.

Observation 4: Leaders do not invest significantly more in the public bad in the second than in the first treatment.

Table 1 shows how average leader investments develop from the first to the second treatment, i.e. before and after the restart of the game. A Wilcoxon test with group averages as units of observation shows that the average investments are not significantly different between the first and the second treatment ($p=0.75$, $n_1=n_2=9$). This can also be seen from Figure 1.

Table 1 also shows how follower investments develop from the first to the second treatment. A similar Wilcoxon test ($p=0.02$, $n_1=n_2=9$) indicates that followers invest on average significantly more (2.48) in the public bad in the second treatment. This observation suggests that the effect of leaders on followers decays over time, and perhaps more so than what has been observed for the willingness to cooperate in standard public bad games that are repeated. In a parallel experiment, Moxnes and Van der Heijden (2000), where a leader treatment is either preceded or followed by a standard public bad treatment, this decay is not observed. While we see no obvious reason why there should be a decay in the present experiment, we state the following observation:

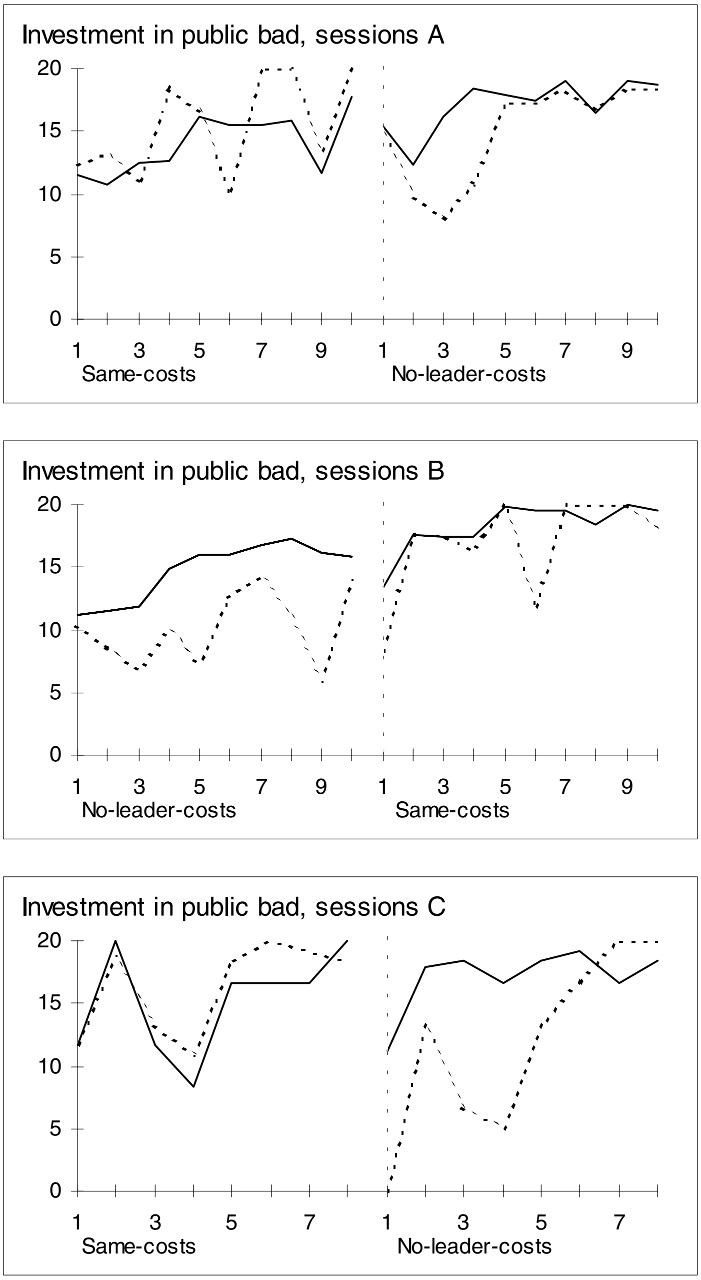

Figure 1: Average investments in the public bad: Dashed line: leaders; solid line: followers. Sessions A, B, and C.

Observation 5: Followers tend to invest more in the public bad in the second than in the first treatment.

4.1.2. Leader and follower behavior

What do the average investments say about leader and follower behavior?

Observation 6: In the treatment same-costs, leaders do not invest less in the public bad than the followers.

The result has been obtained by a Wilcoxon test with group averages as unit of observation ($p=0.44$, $n_1=n_2=9$). This observation could both be interpreted to mean that leaders have no effect on followers, and that leaders over time adjust their investments towards the mean of the followers. Further analysis to clarify this issue will follow since the causality is not clear when using average data over rounds.

Observation 7: In the treatment no-leader-costs, leaders invest significantly less in the public bad than the followers in the same treatment.

Observation 8: In the treatment no-leader-costs, leaders invest significantly less in the public bad than in the treatment same-costs, however, they invest substantially more than zero.

Thus in the treatment no-leader-costs, leaders tend to set a good example. However, they do not go all the way toward zero investments in the public bad even though this is not costly for them. The results in observations 7 and 8 are obtained by Wilcoxon tests with respective results ($p=0.01$, $n_1=n_2=9$) and ($p=0.01$, $n_1=n_2=9$). A 95 percent confidence interval for average leader investments extends from 10 to 14 tokens in the no-leader-costs treatment.

Figure 1 gives some more information about leader behavior. In the early rounds of the no-leader-costs treatment, it seems as if they try to set the good example. In later rounds it appears as if they resort to a certain degree of punishment, or they could be conforming to the type of strategy typically used in traditional public bad games.

A last observation concerns the decisions of the followers. A Wilcoxon test shows that the average investments in the public bad by the followers are not significantly different between the two treatments ($p=0.95$, $n_1=n_2=9$). So, *Observation 9*: There is no significant difference between the average follower investments in the treatments same-costs and no-leader-costs.

Finally, we summarize the relationship between leader and average follower investments by a regression (OLS) for each treatment still using average data over rounds. As we commented in connection with observation 6, the causality is not clear when using average investments over rounds. Hence we will not draw conclusions about behavior from this test in isolation. Assuming that causality runs from leaders to followers (based on ensuing results), this rather weak test based on averages indicates upper p-values for the effects of leadership.

Figure 2 shows the two regression lines together with the data points. In the same-costs treatment we find a constant equal to 8.7 (p=0.08) and a slope of 0.47 (p=0.15). In the no-leader-costs treatments the constant equals 11.4 (p<0.000) and the slope is 0.40 (p=0.02). If both leaders and followers follow each other perfectly, one should expect the data points to lie on a diagonal from the origin (slope equal to 1.0). A slope less than 1.0 shows that leaders make an effort to set the good example by choosing lower investments than the followers. These data do not by themselves prove that followers are influenced by the leaders. It could also be that leaders set an example relative to inflexible average investments of the followers.

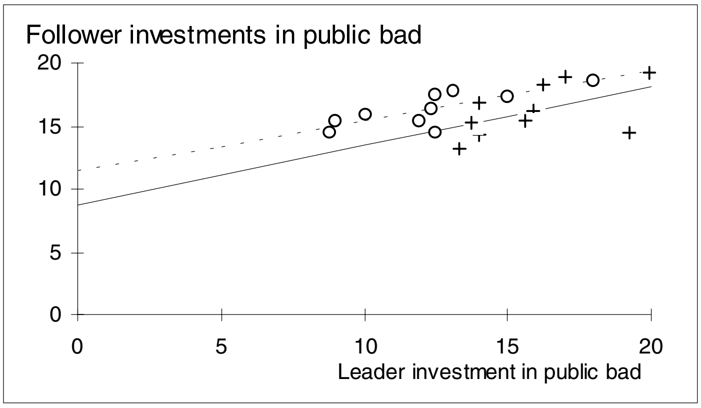

Figure 2: Regression lines between average follower and leader investments over all rounds. Treatment same-costs: solid line and + signs; no-leader-costs: dashed line and circles.

4.2. Results based on investments round by round

In this section we use OLS regressions, which include most of the factors analyzed separately above, however using data by round. Our main interest is in the effect of leader investments on follower investments. However, some new estimates of the effects of rounds and of the ordering of treatments will also follow. First we estimate a linear model. Next, we investigate alternative hypotheses about the effect of leadership.

4.2.1. A linear effect of leadership?

The dependent variable is a vector consisting of the average follower investment in the public bad $x_{i,t}^{F}$ for group *i* in round *t*. We estimate the following model:

$$x_{i,t}^{F} = \alpha_0 + \alpha_1(1-N_i)x_{i,t}^{L} + \alpha_2 N_i x_{i,t}^{L} + \alpha_3 N_i + \alpha_4 t + \alpha_5 R_i + \varepsilon_{i,t} \qquad (5)$$

where $x_{i,t}^{L}$ is the investment by the leader in the same round. The dummy variable N_i takes the value 1 for the no-leader-costs treatment (and zero otherwise). The dummy variable R_i takes the value 1 if it concerns rounds after the restart, i.e. the second treatment in a session (and 0 otherwise). The regression results are depicted in Table 2[10].

Table 2: Estimation results for follower investments (group averages) in the public bad.

Variable	symbol	Coefficient	p-value
Constant	α_0	8.98	0.0000
Inv. leader, same-costs	α_1	0.28	0.0000
Inv. leader, no-leader-costs	α_2	0.15	0.0024
Effect of no-leader-costs	α_3	1.99	0.071
Rounds	α_4	0.37	0.0000
Second treatment	α_5	2.33	0.0000

Note: R^2=0.41, F=23.0, number of observations *n*=168.

The results confirm our previous findings. There is a significant upward trend in investments over rounds. The constant in front of the second treatment dummy is significantly positive, implying that the average investments in the public bad are higher in the second treatment of a session than in the first one.

The most interesting and new finding is that there are highly significant effects of leadership in both treatments. In the sample, the treatment same-costs has a stronger leadership effect than the treatment no-leader-costs, 0.28 versus 0.15. The difference is not significant at the 5 percent level. The effect of the treatment no-leader-costs α_3=1.99 indicates that follower investments in the public bad are higher when followers

10 The regression results seem robust to the specification chosen. For instance, they do not change when we include fixed effects for the groups or if we use a forward or stepwise selection procedure.

know that leaders face no direct costs of acting socially. While such an effect seems likely if followers value fairness, it is not quite significant at the 5 percent level.

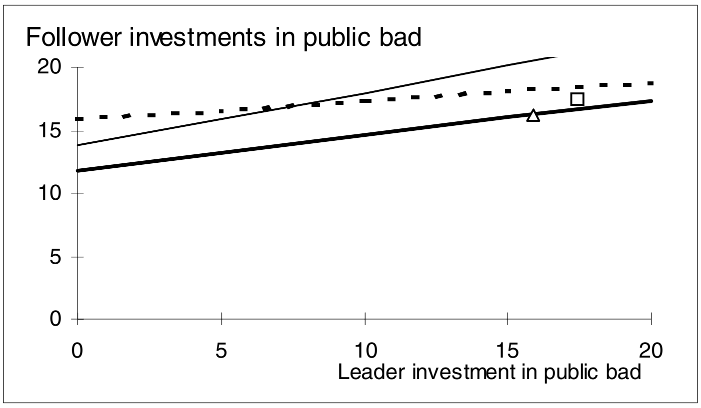

Figure 3: Reaction curves, treatment same-costs: thick solid line; no-leader-costs: thick dashed line.
From Moxnes and Van der Heijden (2000), reaction curve: thin solid line; square: average investments in their no-leader treatment.

The effects of leadership are illustrated in Figure 3. The reaction curve for followers in the treatment same-costs (thick solid line) is constructed by letting the curve go through the coordinates for the average leader (15.89) and follower (16.20) investments for that treatment (the triangle). The slope is given by α_1. The reaction curve for the treatment no-leader-costs (thick dashed line) takes the same starting point, however, the effect of no-leader-costs α_3 is added, and the slope α_2 is used.

The square (coordinates 17.4 and 17.4) in Figure 3 denotes average investments in the no-leader treatment in the parallel public bad experiment by Moxnes and Van der Heijden (2000). It is interesting to note that the same-costs treatment in the present experiment produces average follower investments close to this point for leader investments equal to 17.4. The main difference between these two experiments is that the same-costs treatment uses a mixed sequential-simultaneous procedure (i.e. a leader) while the no-leader treatment in the parallel experiment uses a standard, completely simultaneous procedure. The proximity indicates that this difference in the experimental design is of little importance, as long as leaders invest approximately the same as the average subject would do with no leader.

The figure also shows the reaction curve for the leader treatment of the parallel experiment (thin solid line). The parallel experiment used the identical same-costs treatment. In that experiment leaders were secretly instructed to invest from zero to 6 tokens in the public bad. No observations were made outside of that range. Comparing reaction curves from the same-costs treatments for the two experiments, we see that they differ. This puzzle calls for further analysis.

4.2.2. A non-linear effect of leadership?

Are the results of the two same-costs treatments in the two studies inconsistent, or could they be aligned somehow? A natural first idea is to propose a non-linearity. However, if we assume that the reaction curve for the parallel experiment saturates as leader investments move above the upper limit of 6 tokens in that experiment, the slope of that curve would be approaching zero at high leader investments. This is inconsistent with the present study, showing a slope significantly larger than zero for large leader investments. The same lack of consistency occurs if we assume that the reaction curve from the present study saturates at low leader investments. Hence, the two results cannot be aligned by introducing non-linearities. To test this claim we estimate the following non-linear model for the present experiment:

$$x_{i,t}^{F} = \theta_0 + \theta_1(1-N_i)x_{i,t}^{L} + \theta_2(1-N_i){x_{i,t}^{L}}^2 + \theta_3 N_i x_{i,t}^{L} + \theta_4 N_i {x_{i,t}^{L}}^2 + \theta_5 N_i + \theta_6 t + \theta_7 R_i + \varepsilon_{i,t} \qquad (6)$$

All variables are defined in connection with Equation 5. Table 3 shows that both second order terms come out as clearly insignificant. Hence we discard the hypothesis about a simple non-linear relationship for the reaction curves.

Table 3: Estimation results for a non-linear leadership model.

Variable	symbol	Coefficient	p-value
Constant	θ_0	8.76	0.0000
Leader inv., same-costs	θ_1	0.34	0.13
Leader inv.squared, same-costs	θ_2	-0.0026	0.78
Leader inv., no-leader-costs	θ_3	0.09	0.59
Leader inv.squared, no-leader-costs	θ_4	0.0027	0.72
Effect of no-leader-costs	θ_5	2.38	0.010
Rounds	θ_6	0.38	0.0000
Second treatment effect	θ_7	2.30	0.0000

Note: R^2=0.42, F=16.3, number of observations *n*=168.

4.2.3. A two-component model of leadership?

In another attempt to solve the puzzle, we advance an alternative theory to explain the apparent lack of consistency between the two studies. We propose that the leadership effect consists of two components. One component is a willingness to sacrifice something when the leader is expected to do so. The other component is a tendency to mimic leader deviations around the expected investment level.

The first component is motivated by frequent findings of fairness (and reciprocity) in experiments, see references in the introduction. The nearly significant difference between our two treatments signals that fairness is important also for the effect of leadership: in the same-costs treatment leaders make (irreversible) sacrifices, and fairness implies that followers should make a sacrifice as well. Since there is more than one follower in each group, and the Nash equilibrium predicts no following at all, it seems likely that following due to fairness should be limited. We postulate that the follower sacrifices saturate as expected leader investments in the public bad decline.

Before we move to the second component we need to define what we mean by an expected investment level for leaders. One alternative is that followers form adaptive expectations about leader investments, i.e. that they start out with an expectation given by the first round investment and adjust this gradually as new leader investments are observed in later rounds. A special case of adaptive expectations is that the expected investment level for leaders is based on only the first round leader investment. One reason to expect a heavy weight on the first round, is that during the first round, followers have to make up their minds about an appropriate or fair response to the first round leader investments. This response is likely to set a standard for later rounds. A somewhat similar and unwarrantedly enduring effect of initial conditions is found in an experiment by Brekke and Moxnes (2000).

The second component is a variation in follower investments with variations in leader investments around the expected investment level for leaders. We can think of several motivations for such a behavior. First, there is learning, where followers imitate leader actions, presumably because they suspect leaders to do what they do for some good reason. Second, followers could perceive leader actions to be norm setting and follow suit for no other reason. Third, followers could make adjustments to maintain an appropriate degree of fairness when leaders deviate from their expected investment levels. Fourth, it could be that followers try to stimulate leaders to set the good example, by reacting positively to leader initiatives. Equation 4 showed that if (average) follower reactions are sufficiently strong (α>0.5), leaders have an incentive to reduce investments in the public bad. Having a leader with a presumably social inclination, even lower values of α should be positively motivating for leaders. Whether leaders perceive and act on estimates of α is not important, it is sufficient that followers believe they do.

No matter what the underlying motivations are, behavior could be described by an anchoring and adjustment heuristic, Tversky and Kahneman (1974). Follower investments are anchored to the expected leader investments and are adjusted according to deviations from the expected by the leaders. According to the anchoring and adjustment heuristic, adjustments will typically be insufficient, i.e. with a slope less than one. We will not be able to distinguish between the four motivations. A slope less than one seems likely no matter what the motivation is.

In the following we test the proposed two component model of leadership by the following regression:

$$x_{i,t}^F = \beta_0 + \beta_1(1-N_i)Ex_{i,t}^{L\,\gamma} + \beta_2(1-N_i)(x_{i,t}^L - Ex_{i,t}^L) +$$

$$\beta_3 N_i Ex_{i,t}^{L\,\gamma} + \beta_4 N_i(x_{i,t}^L - Ex_{i,t}^L) + \beta_5 N_i + \beta_6 R_i + \beta_7 t + \varepsilon_{i,t} \qquad (7)$$

Variables are defined as in Equation 5. In addition we introduce $Ex_{i,t}^L$ denoting the expected leader investment. The first component of the leadership effect is captured by $Ex_{i,t}^L$ raised to the power of γ, by the constant β_0, and the effects of treatments N_i and second rounds R_i. The second component is captured by expressions with the term $x_{i,t}^L - Ex_{i,t}^L$, denoting leader variations around the expected leader investment.

The model is kept as simple as possible. Since we postulate that the first component effect of fairness saturates, we primarily want to test a model with an exponent γ greater than 1. The higher γ is, the more quickly saturation takes place when leader investments drop. The second component is assumed linear even though this is not realistic for some large variations in leader investments. However, linearity suffices in light of the puzzle we want to solve: the difference between the two linear reaction curves in Figure 3. The formulation of the expected leader investment is also kept as simple as possible. Adaptive expectations are captured by the expression

$$Ex_{i,t}^L = \phi_k Ex_{i,t-1}^L + (1-\phi_k)x_{i,t}^L \qquad (8)$$

with the initial condition

$$Ex_{i,0}^L = x_{i,0}^L \qquad (9)$$

Primarily we want to test the special case with $\phi_k = 1$, i.e. where $Ex_{i,t}^L$ is constant and equals $x_{i,0}^L$. The subscript k denotes the two treatments: k=same-cost or k=no-leader-costs. Including k allows for different forms of expectation formation in the two treatments.

Table 4: Estimation results for two component leadership model, assumptions: γ=3 and $\phi_k = 1$.

Variable	symbol	Coefficient	p-value
Constant	β_0	9.90	0.0000
Expected leader inv., same-costs	β_1	0.00080	0.0000
Leader deviations, same-costs	β_2	0.27	0.0000
Expected leader inv., no-leader-costs	β_3	0.00034	0.0215
Leader deviations, no-leader-costs	β_4	0.16	0.0005
Effect of no-leader-costs	β_5	2.06	0.0099
Second treatment effect	β_6	1.59	0.0017
Round	β_7	0.40	0.0000

Note: R^2=0.47, F=20.1, number of observations n=168.

Table 4 shows the results for the case with γ=3 and $\phi_k = 1$. All parameters are significant at the 5 percent level, most of them with very low p-values. The effects of leader deviations from expected investment levels (second component) are close to the effects of leader investments in Table 2, as one should expect. The effect of the no-leader-costs treatment is slightly greater, and it is now significant at the 1 percent level. The effect of the second treatment in a session is 32 percent lower than in Table 2, however still significant. The effect of round numbers, 0.40, is slightly larger than in Table 2. The effect is also close to the round effects found in the parallel experiment using instructed leaders, 0.44. The new and interesting finding is that both terms for the expected leader investments (first component) are significant and with expected signs.

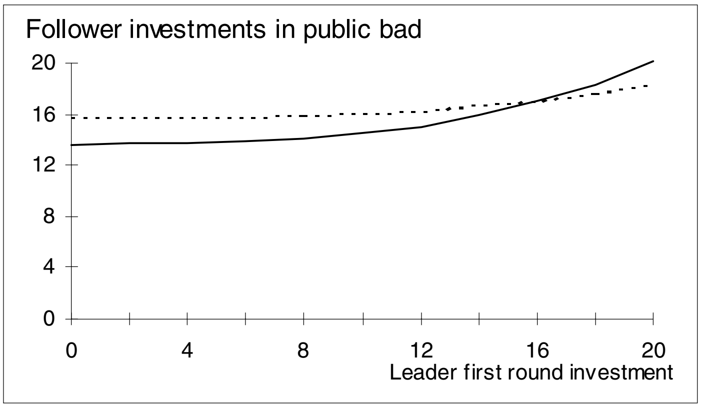

Figure 4: Follower reaction curves based on only the first component of leadership. Solid line: same-costs treatment; dashed line: no-leader-costs treatment. (R_i=1 and the average round number, t=5.5).

Figure 4 illustrates the first component of leadership for the two treatments. The two reaction curves are produced by Equation 7 using the estimated coefficients and using $R_i = 1$ for the illustration. To isolate the effect of the first component, the effect of variation in leader investments is nullified by setting $\beta_2 = \beta_4 = 0$. We see that both treatments show tendencies towards saturation at low expected leader investments (here given by first round investments). The willingness to follow the leader is less pronounced for the no-leader-costs treatment. With positive coefficients for β_1 and β_3 saturation comes as no surprise since we have assumed $\gamma = 3$. However, estimating the model with different values of γ, it turns out that a value of 3 produces the highest F-value and the highest R^2. The change in fit going from a model with $\gamma = 3$ to a linear model with $\gamma = 1$ is characterized by the F-value dropping from 20.1 to 16.4 and the R^2 dropping from 0.47 to 0.42. Correspondingly the p-values for both β_1 and β_3 increase by assuming a linear effect. Hence a curved and saturating relationship produce a better fit to the data. A more advanced test shows that for the treatment same-costs the curved relationship is significantly better than the linear one[11].

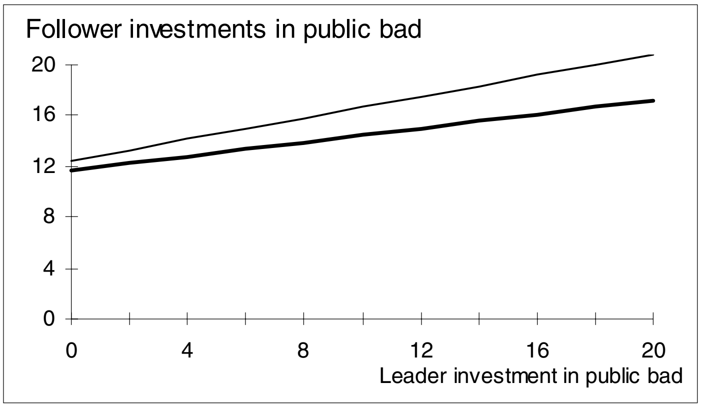

Figure 5: Two follower reaction curves for the same-costs treatment.
Thick solid line: high first round leader investments (10.7) and $\beta_2 = 0.27$.
Thin solid line: low first round leader investments (3.0) and $\beta_2 = 0.42$.
($R_i = 1$ and $t = 5.5$).

Next we what to see if the estimated two component model can align the apparently contradictory results of the present and the parallel experiment, see the discrepancy in the corresponding Figure 3. We concentrate on the same-costs treatment which is common for the two experiments. Figure 5 shows the reachtion curves. The thick solid line

---

[11] A more elaborate test is carried out by including both a linear term and the third order term for expected leader investments. In the treatment same-costs, the third order terms comes out as highly significant (p=0.0001) while the first order term is not (p=0.43). Similar results are obtained if a second or fourth order term is used instead of the third order one, i.e. we cannot distinguish between the higher order terms. In the treatment no-leader-costs, a similar tendency is observed, however neither the third nor the first order terms come out as significant (p=0.26 and p=0.92 respectively). This is not surprising since the test is rather weak due to the fact that first and higher order terms are highly correlated (with data in the first quadrant).

is the reaction curve for the present experiment using Equation 7. We equate the expected leader investment with the average first round investment of 10.7 for this experiment, and we use the estimated value of β_2=0.27 for the second component effect of leadership. The thin line is the corresponding reaction curve for the case when the expected leader investment is 3.0, which was the average investment level for leaders in the parallel experiment. Instead of using the estimated value of β_2 for this case, we use the effect of leader investments found in the parallel experiment, i.e. β_2=0.42. The saturating first component effect of leadership, see the solid line in Figure 4, implies that the reaction curve for the case with high expected leader investment (10.7) lies below the curve for the case with low expected leader investment (3.0). The higher value for β_2 in the latter case serves to further widen the gap between the two curves. Clearly Figure 5 and Figure 3 show similar patterns. Thus the estimated two component model seems able to align the results from the two experiments. By itself, this comparison of the two experiments tend to strengthen our belief in the two component model of leadership, i.e. a saturating effect of fairness and a partial tendency to imitate leaders.

Finally, we want to see if expected leader investments are formed as adaptive expectations, see Equation 8. We search for a pair of values of ϕ_k that maximize the F-value for the model. We find $\phi_{same-\cos ts}$=1.0 and $\phi_{no-leader-\cos ts}$=0.5 yielding a F-value of 21.7. This is only a moderate improvement over the earlier F-value of 20.1. For the treatment same-costs, we find that the F-value is maximized by our earlier assumption that it is the initial investments that matter for expectations and serve as an anchor.

The entire improvement in the F-value is due to the reduced value of ϕ for the no-leader-costs treatment. The coefficient of 0.5 implies that the expected leader investment is a moving average lagging approximately one and a half round behind the actual leader investments. With this coefficient, the first component effect is still positive but no longer significant, i.e. β_3=0 (the other p-values are reduced). The observed effect of leadership is given exclusively by the second component, i.e. by the difference between leader investments and the moving average of the same variable. This effect is strengthened with β_4=0.38 versus the earlier value of 0.16. This implies that there will be no effect of leadership for constant leader investments. According to this model, leaders should start out with high investments and reduce these over the rounds to provoke a positive effect of leadership. Since the improvement in F-value is small, it seems risky to disregard entirely a first component effect.

4.2.4. Leader behavior

What are the implications of the reaction curves for optimal leader behavior? In the no-leader-costs treatment, we cannot distinguish clearly between models that show a lasting effect of leadership and one that predicts only a transitory effect. If there is a lasting effect, leaders should invest zero in the public bad to ensure a maximum effect on followers. If there is only a transitory effect, leaders should start out with high investments and reduce these over rounds. While leaders do invest less than the followers in the no-leader-costs treatment, their investments are much above zero and the tendency over rounds is upwards, not downwards. Hence, it seems that leaders underestimate the potential effects they could have on followers.

In the same-costs treatment, the slope of the reaction curve in the linear model, 0.28, does not reach the benchmark suggesting fully social behavior for the leaders, α=0.5. The same conclusion can be reached by using the two component model. According to the two component model, the optimal response of leaders is to start out with an investment of 11 and increase it to 20 in the remaining rounds. If remaining investments are set equal to 0, the loss in payoffs is 29 percent. Hence, in the same-costs treatment, it is not profitable for leaders to set the good example when using correct expectations. As noted earlier, in this treatment, leaders did not deviate from the followers.

5. CONCLUSIONS

With regard to global or regional environmental problems, do nations that take unilateral actions inspire other countries to curtail emissions as well? Using game theory, Hoel (1991) has shown that in a competitive situation, unilateral actions will both improve global welfare and lead to lower total emissions. The effect depends on the relative size of the leading country. Hoel explicitly leaves out the question whether unilateral actions "might lead to similar behavior from other countries." We have investigated this question by a public bad experiment with a leader. The experiment is designed with a Nash corner solution, such that we do not have to deal with the effects of a leader on the interior solution investigated by Hoel. Each of the nine groups of five subjects play the game twice, with two treatments, one in which all subjects are confronted with the same costs and one in which leaders face no costs of acting socially. The order of the two treatments is varied over groups.

The leader experiments lead to the same type of behavior found in earlier studies of public bad games: subjects behave as weak free riders, closer to the Nash equilibrium than to the social optimum; and investments in the public bad increase over rounds.

The new and interesting finding is that there are significant effects of leader decisions on follower investments in the public bad. Hence we reject the null hypothesis about no effect of leaders. Comparing the two treatments, we find that there is a stronger effect of leadership in the same-costs treatment than in the no-leader-costs treatment. This suggests that fairness is an important motivating factor for the effect of leadership.

For the same-costs treatment we find that follower behavior is consistent with a two component model of the effect of leadership. The first component is a saturating effect of "expected leader investments", where expectations are built on first round leader investments. As expected leader investments decline, followers reduce their investments less and less. We interpret this effect to say that considerations of fairness have a limited potential with respect to reductions in follower investments. The second component is a tendency to partially imitate leader deviations from the expected leader investments. We cannot say if this tendency reflects some learning effort or an attempt to influence leaders. The two component model succeeds to align the present results with results from a parallel experiment where leaders were secretly instructed to invest very little in the public bad, Moxnes and Van der Heijden (2000).

In the no-leader-costs treatment, the two component model is also found to be consistent with the data. However, an alternative model, where expected leader investments are formed as adaptive expectations, gives a slightly better fit. In the latter model it is only the tendency to imitate leader deviations around the expected leader investments that matter. Thus if this model is the correct one, there is no permanent effect of leadership, only a transitory one.

Leader behavior is also observed. In the no-leader-costs treatment, it would be optimal for the leaders to invest nothing in the public bad or to start out with high investments and reduce them over time. However, leaders only make limited reductions in investments and the investments tend to increase over time. Often they seem to engage in punishment of the followers by investing heavily in the public bad. This is particularly the case in the last rounds of the experiment. Thus, leaders behave as if they have wrong (too pessimistic) expectations about the reactions of followers. This is a reason to reject the null hypothesis about leaders acting rationally with correct expectations. In the treatment same-costs, leaders have no incentive to act socially, even when considering the revealed effects on followers. Accordingly, we found no difference between leaders and followers in this treatment with respect to average investments in the public bad. For this treatment the null hypothesis is not rejected.

While the experiment demonstrates that an effect of leadership should be expected, one should be careful in applying the exact numbers to given environmental problems. The only point we will make here is that the experiment does not indicate that the effect is negligible. Using numbers from the two component model for the same-costs treatment, a leader that cuts back its emissions by 40 percent (from an initial (or expected) level of 17.4 to 10.4), would cause an average reduction of 11.4 percent in follower emissions. The costs to the leader is an 11.5 percent reduction in payoffs.

While the effect we have measured is likely to reflect some common trait of human decision making, other factors are also important. In laboratory experiments, contributions to public goods are found to increase with identification of, and eye contact with, other players, Bohnnet and Frey (1995), and there are positive effects of free and costly communication, Isaac et al. (1985) and Isaac and Walker (1991) respectively. In reality, one should also expect effects of NGOs (environmental groups and polluter organizations), environmental costs, abatement costs, income levels, history of pollution etc. The way in which these factors influence the effect of leaders in regimes with competition or with negotiations, are questions for further research.

REFERENCES

Andreoni, J. (1988). "Why Free Ride? Strategies and Learning in Public Good Experiments." *Journal of Public Economics* **37**:291-304.

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**(1):1-21.

Bohnnet, I. and Frey, B.S. (1995). "Ist Reden Silber und Schweigen Gold? Eine ökonomische Analyse." *Zeitschrift für Wirtschafts- und Sozialwissenschaften* **115**:169-209.

Bolton, G.E., Katok, E., and Zwick, R. (1998). "Dictator Game Giving: Rules of Fairness versus Acts of Kindness." *International Journal of Game Theory* **27**(2):269-99.

Bolton, G.E. and Ockenfels, A. (2000). "A theory of equity, reciprocity and competition." *American Economic Review* **90**(1):166-193.

Brekke, K.A. and Moxnes, E. (2000). "Do Numerical Simulation and Optimization Model Results improve Management. Experimental Evidence." Bergen: SNF.

Budescu, D., Suleiman, R., and Rapoport, A. (1995). "Positional order and group size effects in resource dilemmas with uncertain resources." *Organizational Behavior and Human Decision Processes* **61**:225-238.

Dufwenberg, M. and Kirchsteiger, G. (1998). "A Theory of sequential reciprocity." Discussion paper 9837: CentER, Tilburg University.

Erev, I. and Rapoport, A. (1990). "Provision of Step-Level Public Goods, the Sequential Contribution Mechanism." *Journal of Conflict Resolution* **34**(3):401-425.

Fehr, E. and Schmidt, K.M. (1999). "A theory of fairness, competition and cooperation." *Quarterly Journal of Economics* **114**(3):817-868.

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

Isaac, R.M., McCue, K.F., and Plott, C.R. (1985). "Public Goods Provision in an Experimental Environment." *Journal of Public Economics* **26**:51-74.

Isaac, R.M. and Walker, J.M. (1988). "Group size effects in public goods provision: The voluntary contributions mechanism." *Quarterly Journal of Economics* **103**:179-199.

Isaac, R.M. and Walker, J.M. (1991). "Costly Communication: An Experiment in a Nested Public Goods Problem." in *Contemporary Laboratory Research in Political Economy*, edited by T. Palfrey. Ann Arbor: University of Michigan Press.

Keser, C. (1996). "Voluntary contributions to a public good when partial contribution is a dominant strategy." *Economics Letters* **50**(3):359-366.

Ledyard, J. (1995). "Public goods: A survey of experimental research." in *The handbook of experimental economics*, edited by A.E. Roth and J. Kagel. Princeton, NJ: Princeton University Press.

Levine, D.K. (2000). "Modeling altruism and spitefulness in experiments." *Review of Economics Dynamics* (forthcoming).

Mailath, G.J. (1998). "Do People Play Nash Equilibrium? Lessons From Evolutionary Game Theory." *Journal of Economic Literature* **XXXVI**(September):1347-1374.

Morris, M.W., Sim, D.H., and Girotto, V. (1995). "Time of decision, ethical obligation, and causal illusion." Pp. 209-239 in *Negotiation as a social process*, edited by R.M. Kramer and D.M. Messick. Thousand Oaks: Sage.

Moxnes, E. and Van der Heijden, E. (2000). "The effect of leadership in a public bad experiment." R 15/00. Bergen: SNF.

Rabin, M. (1993). "Incorporating fairness into game theory and economics." *American Economic Review* **83**:1281-1302.

Rapoport, A. (1993). "Order of play in strategically equivalent games in extensive form." in *Workshop in experimental game theory*. Stony Brook.

Tversky, A. and Kahneman, D. (1974). "Judgment under Uncertainty: Heuristics and Biases." *Science* **185**(Sept.):1124-1131.

Van der Heijden, E. and Moxnes, E. (1999). “Information feedback in public-bad games: A cross-country experiment.” WP No.42/1999. Bergen: SNF.

Weimann, J. (1994). “Individual behaviour in a free riding experiment.” *Journal of Public Economics* **54**(2):185-200.