Analysing group contract design using a lab and a lab-in-the-field threshold public good experiment*

Bouma, J.A.[1], Nguyen, T.T.B.[2], Van der Heijden, E.[2], Dijk, J.J.[1,3],

[1] Netherlands Environmental Assessment Agency (PBL), the Netherlands

[2] Department of Economics, Tilburg University, the Netherlands

[3] Institute for Environmental Studies, VU University Amsterdam, the Netherlands

Abstract

This paper presents the results of a threshold public goods game experiment with heterogeneous players. The experiment is designed in close collaboration with the Dutch association of agri-environmental farmer collectives. Subjects are recruited at a university ("the lab") and a farm management training centre ("lab-in-the-field"). The treatments have two different distribution rules which are varied within treatment. After subjects have experienced both, they can vote for one of the two rules: either a differentiated bonus that results in equal payoff for all, or an undifferentiated, equal share of the group bonus. Between treatments, subjects can vote for a (minimum or average) threshold or are faced with an exogenous threshold. The results indicate that exogenous thresholds perform better, possibly because the focal point they provide facilitates coordination. With regard to the two distribution rules, the results are mixed: average contributions and payoffs are higher in the lab under the 'equal-payoff' rule, but there is no significant difference between the two in the lab-in-the-field, possibly because contributions in the lab-in-the-field are much less efficient. Overall, our results suggest that environmental payment schemes should not only consider farmer heterogeneity in the design of group contracts, but pay explicit attention to coordination problems as well.

JEL Codes:

H41, C92, C93, D70, Q57

Key words:

Threshold public goods games, endogenous choice, lab-in-the-field experiment, collective agri-environmental management, group contracts, distribution rules, heterogeneous subjects

* We thank Majelle Verbraak and Daan van Soest for helpful suggestions and discussions, and Bowie Goossens for programming the lab-in-the-field experiment.

1. Introduction

To reduce the spatial scattering of agri-environmental interventions and decrease administration costs, agri-environmental management in the Netherlands has recently been re-organized to target farmer collectives (Westerink et al. 2015). Individual farmers can no longer receive subsidies or payments for environmental services but instead have to organize themselves in collectives to become eligible. These collectives formulate a joint plan for agri-environmental management on the basis of which a group contract is negotiated with the regional government. The collective is responsible for meeting the targets of the group contract, and it pays individual farmers for the contributions made.

In defining the size and scope of the group contract, farmer collectives face the difficulty of reaching consensus in a heterogeneous group (Finus 2008). Collectives are heterogeneous in terms of farmer types (e.g. dairy vs. crop farmers) and scale of activities (e.g. small part-time vs. large commercial farmers), which results in significant differences in opportunity costs. As payments are currently uniform across farmers, in most collectives only a limited group of farmers participates in group contracts and group ambition levels remain low. The literature indicates that differentiated payments may increase both the number of farmers participating and the level of contributions (Kesternich et al. 2014). Collectives are reluctant to differentiate payments, however, as they argue that 'what's sauce for the goose is sauce for the gander'.

Experimenting is an effective tool to discover how different payments may affect contribution levels and group ambitions in different setting (Ledyard 1995). Hence, in close collaboration with the association of agri-environmental farmer collectives, we developed a threshold public goods game with heterogeneous players to examine if and how different payoff distribution rules and endogenous choices (of the distribution rule and the ambition level) affect group outcomes. We decided on a *threshold* public goods game as this most closely resembles the agri-environmental conservation scheme: only when the collective manages to provide a certain level (or threshold) of environmental services will the subsidy be paid. Our experiment builds on Kesternich et al. (2014) who find that, in a linear public good game, the distribution rule that equally distributes payoffs among heterogenous players ('equal-payoff-rule') increases contributions and improves group outcomes. In addition, we build on the finding of Gallier et al. (2017), that endogenous choice of the distribution rule improves group outcomes, and more generally on Dal Bó et al. (2010) who show that exogenously imposed policies yield lower cooperation levels than the same policy when it is endogenously (democratically) chosen.

Our paper complements the literature by assessing whether the findings on benefit sharing and endogenous choice in standard public good games are transferable to a threshold public good setting. Compared to standard linear public good games, *threshold* public goods games are fundamentally different in incentive structure as the threshold turns the game into a coordination game (Ledyard 1995). Introducing endogenous choice may further exacerbate the coordination problem, while at the same time it may reduce the incentive to free-ride (Dal Bó et al. 2010). Although Rapoport and Suleiman (1993) added heterogeneity to the threshold public good game setting, the expected impact of adding benefit sharing and endogenous choice remains unclear. This is highly relevant for agri-environmental policy as environmental payment schemes are typically threshold based, hence our approach. To test whether the findings are applicable outside a lab setting, we take the lab to the field. Strictly speaking, our lab-in-the-field is more a lab than a field experiment, but by engaging students farm management, we aim to see whether subjects experienced in conducting experiments perform differently from subjects with no previous lab exposure, while controlling for differences in preferences, game understanding and other characteristics.

2. Background

Compared to standard linear public good games, *threshold* public goods games are fundamentally different in incentive structure as the threshold turns the game into a coordination game (Ledyard 1995). Whereas in standard public good games the Nash equilibrium in dominant strategies is to contribute nothing to the public good, threshold public goods games typically have multiple Nash equilibria: contributing zero is one of them, but all strategies in which total group contributions are equal to the threshold are equilibria too. The threshold acts as a reference or focal point and several studies have found a positive relation between a reference point and agents' effort provision and performance (Fiegenbaum et al. 1996, Koszegi and Rabin 2006, Abeler et al. 2009). Higher thresholds increase contributions but also increase the probability that the target is not met (Rapoport and Suleiman 1993, Finus 2008). Generally, the success rate of threshold public good games ranges from 40 to 60 percent (Alberti and Cartwright 2016), and higher *Step Returns*[1] lead to higher individual contributions and success rates (Croson and Marks 2000). Rapoport and Suleiman (1993) find that adding heterogeneity reduces success rates: with heterogeneity, the threshold is less often reached.

A number of studies have shown that burden or benefit sharing may help to improve game outcomes in games with heterogeneous players. In the public good games literature, there is ample attention for the impact of different burden or benefit sharing rules on contribution levels (Balafoutas et al. 2013; Gallier et al. 2017; Kesternich et al. 2018; Walker et al. 2000; Kesternich et al. 2014) and the impact of minimum contribution rules (*e.g.* Kesternich et al. 2014; Dannenberg et al. 2014; Kocher et al. 2016; Martinsson & Persson 2016; Sutter & Weck-Hannemann 2003, 2004). In the threshold public good game literature, attention for burden sharing and minimum contribution levels is limited. Tavoni et al. (2011) find that non-binding pledges can overcome barriers to provide threshold public goods, particularly in heterogeneous settings. Brick and Visser (2015) find that shared norms about equity determine the subject's choice for burden-sharing contribution schemes. Clearly, subject heterogeneity increases the coordination problem, and benefit or burden sharing may facilitate coordination by reducing heterogeneity in payoffs: Cadsby and Maynes (1999) show that in threshold public goods games with homogenous subjects, sufficiently high rewards elicit convergence of contributions to the threshold, rather than increase free-riding, and thus facilitate coordination. Corrazini et al. (2015) show, in the context of multiple threshold public goods, that salience helps to overcome coordination failure, and Palfrey et al. (2017) indicate that communication improves coordination, in line with the findings from the public goods literature (Ostrom et al. 1992).

Subject pool characteristics may influence game outcomes in several ways. Cooper et al. (1999, 2006) indicate that professionals may perform better than students because of experience, for example in overcoming coordination failure at the work floor. Similarly, List (2003) and Potters and Linden (2000) show that experience explains why senior managers or traders are more rational in their decisions than subjects without such experience. When comparing student subject pools, students of economics perform different from non-economic students. Here, the analysis has focused on economics students being less generous and more self-interested (Bauman and Rose 2011), which has turned out not be caused by the content of the study but rather by self-selection effects (Cipriani et al. 2009). Finally,

---

[1] Similar to the Marginal Per Capita Return (MPCR) in a standard public good game (see Isaac et al. 1984), the Step Return measures the value of the public good relative to that of the forgone private good for threshold public goods games.

Cadsy and Maynes (1998a,b) indicate that female students, and female professionals, are better at coordinating efforts, although they suggest that this does not necessarily improve game efficiency.

3. Methodology

3.1 Experimental design

In each experimental session, groups of four subjects play 12 rounds of a threshold public goods game. Group composition remained the same throughout the session, *i.e.* we used partner matching. In each round, subjects receive an endowment *e*, equal to 20 points, which they can use to make a contribution q_i to the public good. If total group contributions are at least as high as a threshold level *T*, a bonus is paid to the group members. This bonus is 1.6 times the threshold, corresponding to a Step Return of 0.4 (Croson and Marks, 2000). If the threshold is not met, no bonus is paid, and subjects' contributions are not returned. In the most basic setting, the bonus is divided equally among the group members, i.e. we apply the so-called equal-sharing-of-the-bonus rule, and the payoff π_i of group member *i* in a round is thus as follows:

$\pi_i = e - c_i q_i + 0.4T$ if $\sum_i q_i \geq T$;

$\pi_i = e - c_i q_i$ otherwise.

The parameter c_i represents the cost of contributing, and it is used to introduce heterogeneity among the subjects. In most experiments, heterogeneity is introduced via endowments (rich and poor players, e.g. Tavoni et al. 2015, Kesternich et al. 2018, Levati et al. 2007), and sometimes via the MPCR (Kesternich et al. 2014), but we introduce heterogeneity in costs since this best mimics actual group decision-making. Hence, in each group two group members are of the low-cost type, with $c_i = c_L = 0.6$, and two group members are of the high-cost type, with $c_i = c_H = 1.2$. Types are randomly assigned, and subjects do not change their cost-type throughout the game. Note that group contributions above the threshold do not affect the bonus and are not compensated.[2] Note also that the above setting represents a social dilemma as $c_H > c_L > 0.4$. Not contributing is a Nash equilibrium, as well as most combinations of contributions with $\sum_i q_i = T$.

We run several variations of this basic setup in a between-subjects setting (see also Table 1), varying with the two distribution rules: the 'equal-sharing-of-the-bonus' rule and the 'equal-payoff' rule (Kesternich et al. 2014). The equal-sharing-of-the-bonus rule implies that if the threshold is met, each group member receives an undifferentiated, equal share of the group bonus, namely 1.6*T*/4. Under the equal-payoff rule, there is differentiation in bonus payments. Group members receive different shares of the group bonus in such a way that every group member receives in principle the same payoff, namely $\pi_i = (4e - 2c_l q_l - 2c_h q_h + 1.6T)/4$.[3]

In line with Putterman et al. (2011), we let subjects choose their distribution rule after several rounds. More specifically, in phase 1 (rounds 1-4) and 2 (rounds 5-8), subjects play with alternating, exogenously determined distribution rules. Before phase 3 (rounds 9-12) commences, the group chooses the distribution rule through majority voting (with the computer randomly deciding the outcome in case of a tied vote). Besides the distribution rule, we vary two other factors: the threshold origin (whether the threshold is determined exogenously or endogenously), and – within the

[2] In our view, this best mimics the situation with the farmers in the Netherlands, in which the bonus is based on the level as specified in the contract (and not on total contributions).
[3] In some rare cases with a very extreme contribution distribution, equal payoffs may not be feasible. In those cases, the equal-sharing-of-the-bonus rule is automatically implemented. Subjects are informed about this in the instructions.

endogenous threshold treatments – whether the threshold is based on the average or lowest proposed threshold in the group.

In the endogenous threshold treatments, subjects (simultaneously) propose a group threshold before making a contribution. Depending on the treatment, the group threshold is then based on the average or the lowest proposed threshold, and subjects are informed about the group threshold *T* before making their contribution choice. To ensure comparability between the exogenous and endogenous threshold treatments, we first conducted the endogenous treatments, and based the exogenous thresholds in each round on the median of the thresholds proposed (per round and per group). This is a common approach in the literature (see for instance Sutter and Weck-Hahneman 2003, 2004). Finally, inspired by Kallbekken et al. (2011), we pay explicit attention to learning in game design: we ask subjects at the start of the experiment which of the two distribution rules they prefer, and measure their preference again after they have experienced both distribution rules. After the experiment, subjects fill in a short questionnaire about their game perceptions and background characteristics.

3.2. Experimental procedures

We conducted the lab experiment with a total of 204 participants at Tilburg University's CentERLab in March 2017. The lab-in-the-field experiment was conducted with a total of 148 participants[4] at Van Hall Larenstein University of Applied Sciences in Leeuwarden in November 2017. The lab experiment in Tilburg was programmed and conducted with z-Tree software (Fischbacher 2007). For the lab-in-the-field experiment in Leeuwarden, we implemented the oTree open source platform (Chen et al. 2016) using the Internet Explorer browser in the students' computer labs (Goossens 2017).

Table 1 summarizes the experimental design, the treatments and treatment names, and the number of groups per treatment. To account for potential ordering effects, we alternated the order of the distribution rules within a treatment. For instance, treatment ESPM implements the equal-sharing-of-the-bonus rule in rounds 1-4 and the equal-payoff rule in rounds 5-8, whereas treatment EPSM implements the rules in the reversed order.

Table 1. Experimental design

Threshold origin	Threshold based on	Distribution rule order (phase 1 - phase 2)	Treatment name	Lab no. of subjects (groups)	Lab-in-the-field no. of subjects (groups)
Exogenous	Minimum*	Equal sharing - equal payoff	XSPM	48 (12)	
Exogenous	Minimum*	Equal payoff - equal sharing	XPSM	40 (10)	
Endogenous	Minimum	Equal sharing - equal payoff	ESPM	60 (15)	28 (7)
Endogenous	Minimum	Equal payoff - equal sharing	EPSM	56 (14)	20 (5)
Endogenous	Average	Equal sharing - equal payoff	ESPA		48 (12)
Endogenous	Average	Equal payoff - equal sharing	EPSA		52 (13)
Total number of subjects (groups)				204 (51)	148 (37)

* The exogenous thresholds were based on the median of the thresholds proposed under the ESPM and EPSM treatments.

As Table 1 shows, we did not conduct all treatments in both locations. The exogenous threshold treatments were only implemented in the lab, and the treatments with thresholds based on average proposals were conducted only in the lab-in-the-field. In fact, the average threshold treatments were only included after the lab experiment, at the request of the association of farmer collectives, as they argued that although most collectives indeed take the lowest ambition level, other collectives use the

[4] In total, 160 students participated, but we had to exclude 3 groups from the analysis as we did not fully succeed in preventing communication between them.

average. Given that the lab experiment indicated no significant differences between the endogenous and exogenous threshold treatments, we dropped the exogenous treatments in the lab-in-the-field.

Apart from the difference in treatments, the lab and the lab-in-the-field experiment were identical except that a minimum level of framing was added to the instructions of the lab-in-the-field experiment, in order to clarify game design. Framing was added at the request of the association of farmer collectives, for whom the lab-in-the-field was a pre-test for conducting the experiment with farmers at a later stage. We minimized framing to ensure that results remained comparable: The experimental instructions, with the framing highlighted, are provided in Annex A. Note that whereas the lab experiment was in English, the lab-in-the-field experiment was in Dutch.

Finally, we did not pay a show up fee to subjects in the lab-in-the field as their participation was mandatory: the experiments were scheduled by the school administration requiring mandatory attendance. Students at the lab participated voluntarily and were recruited through a mailing list. They were paid a 5 euro show up fee. Both in the lab and the lab-in-the-field, students were paid earnings from three rounds, the computer randomly selecting one round per phase to be paid out. In all sessions, the value of each point was 0.10 eurocent. In both locations, the experimental sessions were run by the same experimenters. Each experimental session lasted approximately one hour, and on average subjects earned about 11 euro (including the show up fee) in the lab and 6 euro in the lab-in-the-field. After the experiment subjects filled in a short questionnaire to collect information about their game understanding, social preferences and other characteristics. We used this information to analyse game behavior, and the differences in game behavior between the lab and the lab-in-the-field.

3.3. Expected results and statistical methods

First, in line with Kesternich et al. (2014), we expect the equal-payoff rule to result in higher average contributions and payoffs than the equal-share-of-the-bonus rule. We do not expect the change in game incentives resulting from having adapted the experiment to a threshold public goods game to affect this result.

Second, we are ambiguous about the effect of endogenous choice. With regard to the endogenous choice of the threshold, Dal Bó et al. (2010) find that an institution or policy that is exogenously imposed, such as a threshold, yields lower cooperation levels than an endogenously (democratically) chosen one. However, their finding reflects a public goods setting, whereas in a threshold public goods setting an exogenous threshold may improve coordination by providing a clear focal point. With regard to the endogenous choice of the distribution rule, we have no reason to expect a different result from Gallier et al. (2017): they show that any distribution rule that was endogenously selected is an improvement upon an exogenously imposed rule. Still, in our case the coordination problem of having to decide on the preferred distribution rule may again reduce the advantage of endogenous choice.

Third, given the differences in exposure to experiments between the two subject pools, and the general notion that conducting lab-in-the-field experiments introduces heterogeneity and potential noise (Schram 2005, Levitt and List 2007), we expect subjects in the lab-in-the-field to encounter more coordination problems than students in the lab.

To analyse the experimental data, following Gallier et al. (2017), Kesternich et al. (2014), and Croson and Marks (2000) we perform non-parametric tests and random-effects regressions (adjusted for group clusters). We use Mann–Whitney U (MW-U) tests to compare results between different treatments (between subjects), and Wilcoxon matched-pairs signed-ranks (W S-R) tests to compare results across phases in a treatment (within subjects).

Game efficiency indicators are the group threshold, the group contribution, the 'frequency of threshold met' (or success rate), the 'over-contribution rate', and the group payoff. The 'Frequency of threshold met' is the ratio between the number of times that groups meet their threshold and the total number of rounds played, and the 'over-contribution rate' is an indicator that measures the efficiency of group contributions in a particular round, determined by $(\sum_i q_i - T)/\sum_i q_i$. The higher the ratio, the more inefficient the group has played, while a negative ratio means that the group has not met the threshold. For groups with group contributions equal to zero, this ratio has a zero value. Unless indicated otherwise, we run the tests using group averages as units of observations. In addition to the non-parametric tests, we run ordinary least square random-effects regressions on individual contributions and individual payoffs. We focus mostly on phase 1 as subjects' behaviour is less affected by previous rounds' results, and hence, the potential impacts of game factors can be estimated in a cleaner way.

4. Results

4.1 Summary statistics

Table 2 presents summary statistics for the group threshold, the group contribution, the success rate, the over-contribution rate, and the group payoff as the average over all rounds of each treatment. The table indicates that average group contributions range between about 26 and 44 points, and that realized group thresholds vary between approximately 22 and 42 points across treatments. The exogenous threshold treatment (XSPM) in the lab has the highest average group payoff (80 points), while the endogenous average threshold treatment (ESPA) lab-in-the-field treatment has the lowest average group payoff (68 points). On average, group contributions seem to be lower in the lab than in the lab-in-the-field, and the standard deviations of group thresholds show that variation between groups is higher in the lab than in the lab-in-the-field. By definition, the realized group thresholds are higher for the average threshold treatments (ESPA and EPSA) than for the minimum threshold treatments, but differences in group contributions between these treatments are much smaller.

Table 2. Summary statistics for each treatment

Treatment	Average group threshold	Average group contribution	Average success rate (%)	Average over-contribution rate	Average group payoff
Lab results					
XSPM	22.28 (15.62)	25.78 (16.97)	76 (26)	0.03 (0.30)	80.29 (11.41)
XPSM	26.63 (12.48)	29.13 (11.24)	76 (16)	0.06 (0.14)	77.39 (8.04)
ESPM	24.65 (13.57)	28.70 (14.92)	71 (23)	0.07 (0.30)	72.47 (7.72)
EPSM	29.02 (13.76)	32.07 (13.94)	67 (13)	-0.00 (0.24)	74.97 (7.79)
Lab-in-the field results					
ESPM	23.62 (10.42)	34.35 (13.64)	80 (12)	0.28 (0.13)	70.33 (5.95)
EPSM	23.85 (4.66)	39.16 (7.08)	90 (11)	0.36 (0.19)	70.70 (10.95)
ESPA	41.90 (10.09)	39.84 (10.43)	48 (22)	-0.20 (0.27)	68.06 (10.21)
EPSA	42.21 (11.72)	43.85 (12.15)	63 (28)	-0.03 (0.30)	76.11 (11.66)

Standard deviations of groups' average values over 12 rounds are given in parentheses.

Success rates (based on the 'frequency of threshold met') range from slightly below 50 to about 90 percent. Compared to the literature, the upper bound of this range is relatively high: Croson and Marks

(2000) and Alberti and Cartwright (2016) report success rates of 40 to 60 percent, which corresponds to the success rates observed in the treatments with average proposed threshold.

Table 2 shows that a high success rate does not always translate into high payoffs. Specifically, payoffs in endogenous minimum threshold treatments ESPM and EPSM in the lab-in-the-field are lower than in the corresponding lab treatments. This is mainly due to high over-contribution rates. Indeed, over-contribution rates are especially high in the lab-in-the-field treatments (about 20 to 36 percent), while in the lab this rate is maximally about 7 percent (we will elaborate on this in more detail below). So even though most groups in the endogenous minimum threshold ESPM and EPSM lab-in-the-field treatments reach the threshold, the inefficiency of their choices results in relatively low payoffs. The negative sign of over-contribution rates in some other treatments explains the low average success rate in these treatments: on average groups 'under-contribute' and thus do not meet the threshold.

Figure 1. Average threshold and group contribution over time in all endogenous threshold treatments

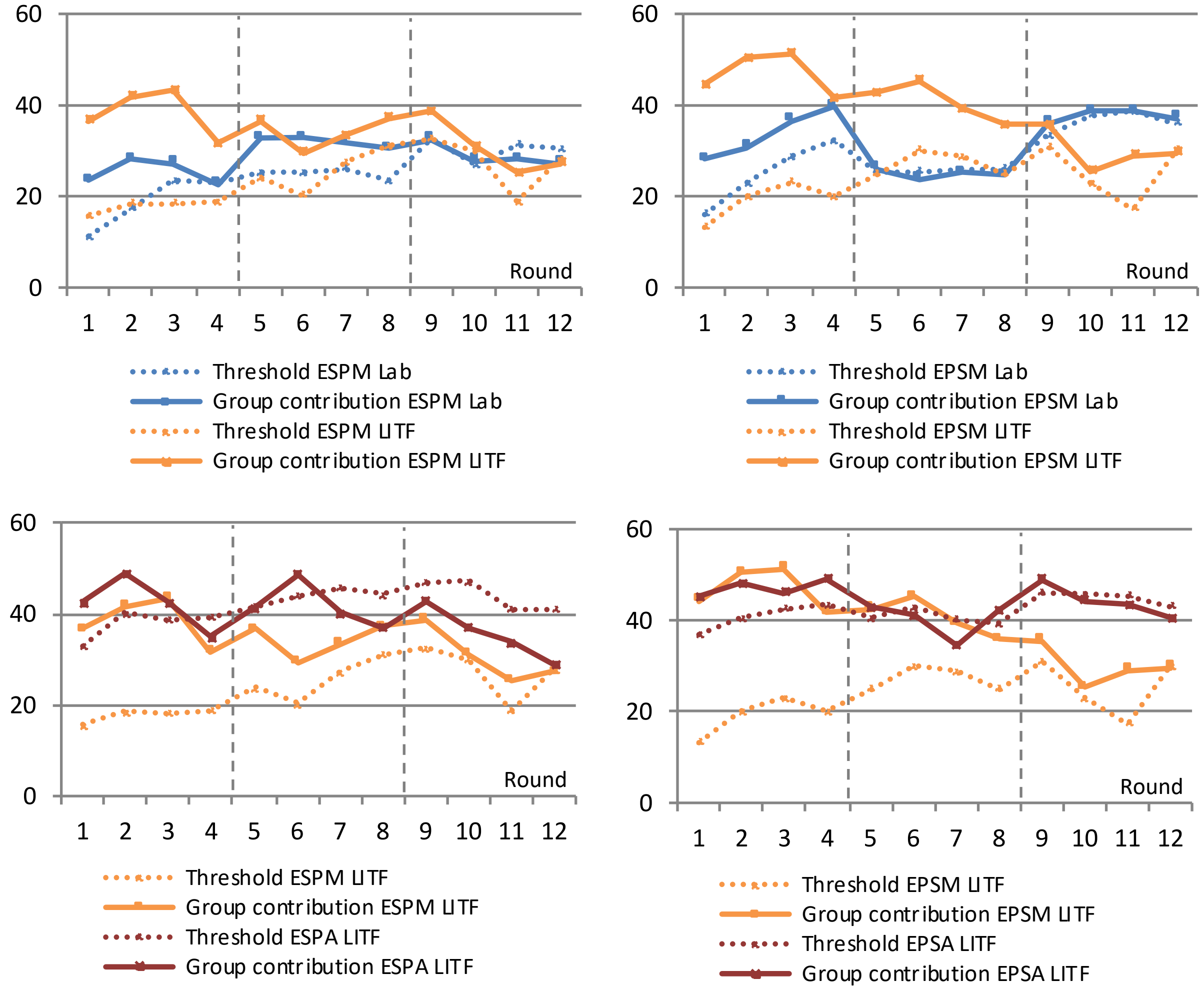


Figure 1 shows the average round by round development of thresholds and group contributions in all treatments in which the threshold was endogenously determined. The top panels show that average thresholds in the lab and in the lab-in-the-field (LITF) are rather close and seem to increase somewhat across rounds. Contributions are largely consistently above the thresholds and apart from the last phase of the top right panel, contributions are higher in the lab-in-the-field. Coordination seems to improve over time, as suggested by the decreasing gap between contributions and thresholds. The bottom panels of Figure 1 show that although the resulting thresholds in the average threshold treatments are higher than in the minimum threshold treatments (almost by definition), the difference between the contributions over time is small for these treatments, irrespective of the order of the distribution rule. This pattern is rather persistent and explains the high success rates and over-contributions rates in the two treatments with a minimum threshold. In the following, we perform pairwise comparisons of related treatments to test the effects of the variables of interest: distribution rule, threshold origin, endogenous choice of distribution rule, and subject characteristics.

4.2 Equal-payoff rule versus equal-sharing-of-the-bonus rule

We now compare group performance in phase 1 between (pairs of) treatments: the endogenous minimum threshold treatments ESPM versus EPSM (in both the lab and the lab-in-the-field), the exogenous threshold treatments XSPM versus XPSM in the lab, and the average threshold treatments ESPA versus EPSA lab-in-the field. Within each treatment, we carry out tests between phase 1 and phase 2 as observations differ only in distribution rule. Test results are presented in table 5, Annex B. We discuss the key observations here.

The equal-payoff rule seems to perform weakly better in between-subject comparisons. Between the two endogenous threshold treatments in phase 1 in the lab, groups in ESPM treatment (which played with the equal sharing of the bonus rule) make significantly lower group contributions ($p = 0.0636$, MW-U test), and obtain lower group payoffs ($p = 0.0145$, MW-U test) than groups in the EPSM treatment (which played with the equal-pay-off rule). There are no differences in the group threshold, the frequency of threshold met, or the over-contribution rate ($p \geq 0.1192$, MW-U tests). When comparing all treatments, there is no significant difference in any of the factors between treatments ($p \geq 0.1229$, MW-U tests), except for the threshold level between XSPM and XPSM treatments in the lab. However, these thresholds are selected exogenously based on the median threshold in comparable endogenous treatments and therefore they do not reflect subjects' behavior in the game.

In the lab, within-subject comparisons show that the equal-payoff rule performs better than the equal-sharing-of-the-bonus rule. It does not matter whether the equal-payoff rule is played in phase 1 or in phase 2. In both phases, groups playing under the equal-payoff rule, in at least half of the treatments, generate significantly higher average group thresholds, make significantly higher group contributions, and receive significantly higher group payoffs ($p \leq 0.0937$, W S-R tests). In the rest of the treatments, there is no significant difference ($p \geq 0.2115$, W S-R tests). Success rates and over-contribution rates do not differ considerably between treatments in the lab ($p \geq 0.2277$, W S-R tests), except for the EPSM treatment in the lab. In this treatment, under equal-payoff rule (in phase 1), groups have significantly higher success rates ($p = 0.0038$, W S-R test), but also higher over-contribution rates ($p = 0.0023$, W S-R test). However, group payoffs are still higher under the equal-payoff rule. Therefore, playing under the equal-payoff rule is better or at least not worse than under the equal-sharing-of-the-bonus rule in all treatments in the lab.

In the lab-in-the-field, however, distribution rules do not have the same (systematic) impact on group performance. Comparing the impact of different distribution rules within treatments yields no difference in group contribution in any treatment ($p \geq 0.2249$, W S-R tests). Group threshold are either higher ($p \leq 0.0796$, W S-R tests) or about the same in phase 2, and over-contribution rates are always higher in phase 1 ($p \leq 0.0431$, W S-R tests). In fact, groups in the lab-in-the-field slightly increase their thresholds in phase 2 but keep group contributions at the same level, thus reducing over-contribution rates. This happens in all lab-in-the-field treatments no matter which rule is played first. Thus, it is hard to say which rule performs better. Success rates do not change significantly between the two phases, except for the endogenous average threshold EPSA treatment where the success rate decreases in phase 2 under equal-payoff rule ($p = 0.0755$, W S-R tests), with significantly lower group payoffs ($p = 0.0712$, W S-R tests). Another treatment where group payoffs are lower under equal-payoff rule is the endogenous minimum threshold EPSM treatment ($p = 0.0431$, W S-R tests): group payoffs in phase 1 suffer from higher over-contribution rates and low thresholds. Overall, if anything, the equal-sharing-of-the-bonus rule seems to result in better game performance in the lab-in-the-field.

4.3 Endogenous versus exogenous choices

In our experiment, endogenous choice has two dimensions: (i) subjects vote for a (minimum or average) threshold in the endogenous threshold treatments, and (ii) subjects vote for one of the two distribution rules before the start of phase 3. Here, we examine if either of these dimensions has an effect on subjects' behavior in the game.

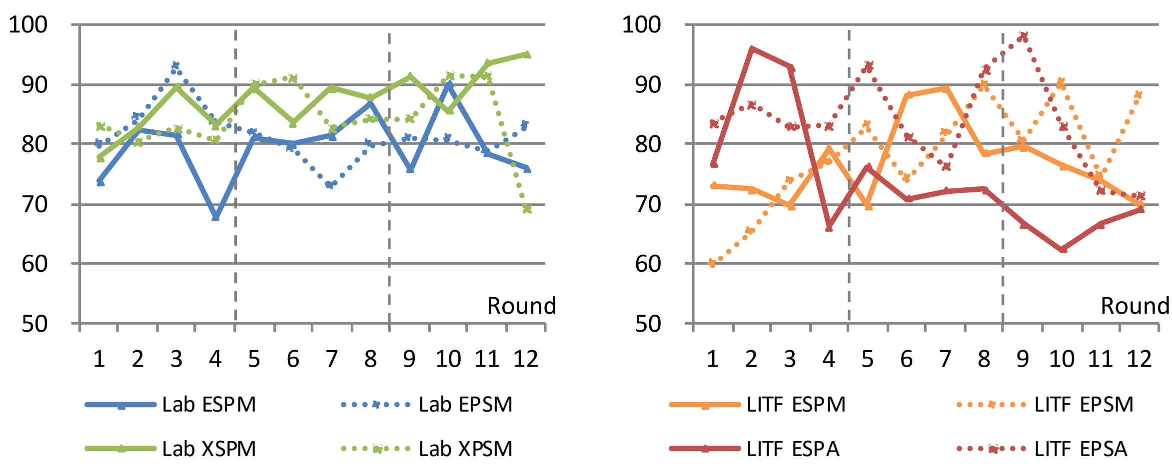
4.3.1 Endogenous choice of the threshold

Table 2 already showed that the success rates of the exogenous threshold treatments in the lab were higher than those of the corresponding endogenous threshold treatments. Also, in terms of group payoffs the endogenous threshold treatments do not perform better. Average group payoffs of the exogenous treatments XSPM and XPSM are 80.29 and 77.39 points, respectively, while average group payoffs of the endogenous treatments ESPM (72.47 points) and EPSM (74.97 points) are lower.

To test whether the endogenous threshold choice has a significant effect, we compare group contributions, success rates, over-contribution rates and group payoffs of the exogenous and endogenous treatments both over all rounds and in phase 1 of comparable treatments (i.e. ESPM versus XSPM, and EPSM versus XPSM). Note that in phase 1, the data are 'cleaner' because subjects' decisions have not been affected by previous experiences in the game. The test results show that the exogenous threshold treatments weakly dominate the endogenous ones: outcomes do not differ significantly ($p \geq 0.1370$, MW-U tests) except for group payoffs, which are significantly (but not substantially) higher for the XSPM exogenous threshold treatment ($p \leq 0.0454$, MW-U tests). Details are presented in table 6, Annex B. Thus, we find no evidence that endogenous choice of the threshold improves group performance.

4.3.2 Endogenous choice of the distribution rule

Figure 2: Average group payoffs, lab (left) and lab-in-the-field (LITF; right)



Subjects vote for their preferred distribution before the start of phase 3, the distribution rule being exogenously imposed in phases 1 and 2. Figure 2 illustrates that the trend in average group payoff is not significantly affected by the introduction of a new distribution rule, neither between phase 1 and phase 2, or in phase 3. It also shows that although group payoffs show no substantial differences between treatments, they are typically higher in treatments with exogenous thresholds as compared to those with endogenous thresholds.

Table 7 in Annex B presents the test results for within treatment tests of phase 1&2 averages (the first 8 rounds) versus the phase 3 averages (the last 4 rounds). We find that in phase 3: (i) group thresholds are not lower (they are significantly higher in three treatments in the lab, $p \leq 0.0257$, W S-R tests, but two of those are exogenous treatments where thresholds are not selected by subjects, and are thus not significantly different from the group thresholds in other treatments, $p \geq 0.1730$, W S-R tests); (ii) group contributions are only significantly higher in the lab XPSM treatment ($p = 0.0743$, W S-R test); (iii) both success rates and over-contribution rates are significantly lower ($p \leq 0.0469$, W S-R tests) or not different than in phase 1 and phase 2 ($p \geq 0.1599$, W S-R tests). The lower over-contribution rates can probably be attributed to learning. However, lower over-contribution rates only result in significantly higher group payoffs in the lab XSPM and the lab-in-the-field EPSM treatments.

Summarizing, there is no treatment where playing the game under the distribution rule that subjects voted for before the start of phase 3 significantly increases performance in term of group contribution, success rate, or group payoff. The results thus offer no support for the earlier prediction that endogenous choice would improve game outcomes.

Note that one explanation for not finding a strong positive effect of endogenous choice of the distribution rule could be that a substantial number of groups (about 25 to 50 percent in each treatment) ended up with a tied vote. In those cases, the computer randomly selected the distribution rule. Hence, it could be that subjects that got to play with a rule they did not vote for started contributing less, and subjects that got the rule they voted for contributing more, the two effects cancelling out. Hence, as a robustness check, we performed additional within-treatment tests of phase 1&2 versus phase 3 averages excluding those groups for which the votes were tied (see table 8, in Annex B). The results indicate that the earlier conclusion does not change: endogenous choice of the distribution rule has no significant impact on game outcomes. Group thresholds in phase 3 remain significantly higher in the exogenous treatments where thresholds are not selected by subjects ($p \leq 0.0192$, W S-R tests). The difference becomes insignificant for the lab EPSM ($p > 0.2328$, W S-R test) and makes the phase 3 threshold significantly higher for the lab-in-the-field EPSA treatment ($p < 0.0277$, W S-R test). However, there are no significant differences with respect to the group thresholds in most of the treatments ($p \geq 0.1420$, W S-R tests). Group contributions in phase 3 again are significantly higher only in the lab XPSM ($p < 0.0904$, W S-R test). Group payoffs and success and contribution rates also do not show major differences compared to the results in which groups with tied votes are not excluded from the data.

4.4 Influence of subject characteristics

Earlier, we observed different group outcomes between the lab and the lab-in-the-field sessions. Here, we use the post-experiment survey results to check whether subject characteristics and perceptions explain some of these differences.

The 204 subjects that participated in the lab sessions in Tilburg were BSc and MSc students at Tilburg University with different nationalities and academic backgrounds (economics, legal studies, management, social sciences). The 148 students that participated in the lab-in-the-field sessions in Leeuwarden were BSc students at the Van Hall Larenstein University of Applied Sciences, half of which were students dairy farm management, the other half being students of business administration. Of the Leeuwarden students, 72 percent indicated that their family was in the agricultural business, and another 15 percent indicated that their family worked in agriculture supply services. For the Tilburg students we have no information about family background.

Table 3 presents the survey findings. Note that the distribution of answers between the two group differs: subjects in the lab-in-the-field were less outspoken, with more neutral answers, possibly because they were less used to answering questions about their fairness considerations and motivations in playing the game.

Table 3. Summary statistics of subjects' game perceptions

	Lab results	Lab-in-the-field results
% wanted distribution of payoffs to be fair[i]	63	41
% wanted to maximize earnings[ii]	86	65
% found game explanation clear[iii]	94	86
% found the game complicated[iv]	12	38

[i] *Question whether respondent agrees with statement "I wanted the distribution of pay-offs (in the experiment) to be fair". We combined the answers of those that "Totally agree" and "Agree".*

[ii] *Question whether respondent agrees with statement "I wanted to maximize my earnings (in the experiment)". We combined the answers of those that "Totally agree" and "Agree".*

[iii] *Question whether the respondent believed the explanation of the experiment was clear. We combined the answers "Very clear" and "Clear".*

[iv] *Question whether the respondent found the game complicated. We combined the answers "Very complicated" and "Complicated".*

Test results show that all four game perception indicators in table 3 are significantly different across subject pools (p = 0.000, Fisher's exact tests). To determine the impact of subject's characteristics on individual contribution and individual payoff, we run regression models on individual contribution and individual payoffs in phase 1:

$$contribution_{it} = \alpha_0 + \alpha_1 threshold_{it} + \alpha_2 \min\_threshold_i + \alpha_3 exogenous_i + \alpha_4 equal\_bonus_{it} + \alpha_5 low_i + \alpha_6 X_i + \epsilon_{it} \quad (1)$$

$$payoff_{it} = \beta_0 + \beta_1 threshold_{it} + \beta_2 threshold\_met_{it} + \beta_3 \min\_threshold_i + \beta_4 exogenous_i + \beta_5 equal\_bonus_{it} + \beta_6 low_i + \beta_7 X_i + \epsilon_{it} \quad (2)$$

We run equations (1) and (2) for both subject pools. The lab experiment includes the *exogenous* threshold dummy variable and the lab-in-the-field experiment has the additional *min_threshold* dummy variable. Variables included in both regressions are: *threshold* (threshold level), *equal_bonus* (dummy variable for the distribution rule), *low* (dummy variable for player cost type), and *threshold_met* (dummy variable for achieving the threshold). X_i is the vector of personal characteristics of player *i* (see also Table 3), including: *earn_much* (subjects' wish to earn as much as possible), *fair_distribution* (subject's preference for fair payoff), *explanation_clear* (subjects' understanding about the instruction for the game), and *game_complexity* (subjects' perception about the game's complexity). The personal characteristics in X_i are dummy variables that are split based on the median value of each subject pool. Finally, to control for group correlation we cluster standard errors at the group level. The regression results are presented in Table 4.

We find that among the four game perception indicators, none has a significant effect on individual payoffs in phase 1. For individual contributions in phase 1, only *earn_much* has a significantly negative impact in the lab-in-the-field, and *game_complexity* has a negative impact in the lab. Cost type and threshold levels do have significant impacts on both individual contributions and individual payoffs in the experiment. Endogenous threshold treatments significantly increase contributions in both locations, but they significantly decrease individual payoffs because of high over-contribution rates.

Table 4. Individual contribution and individual payoff regressions (phase 1 only)

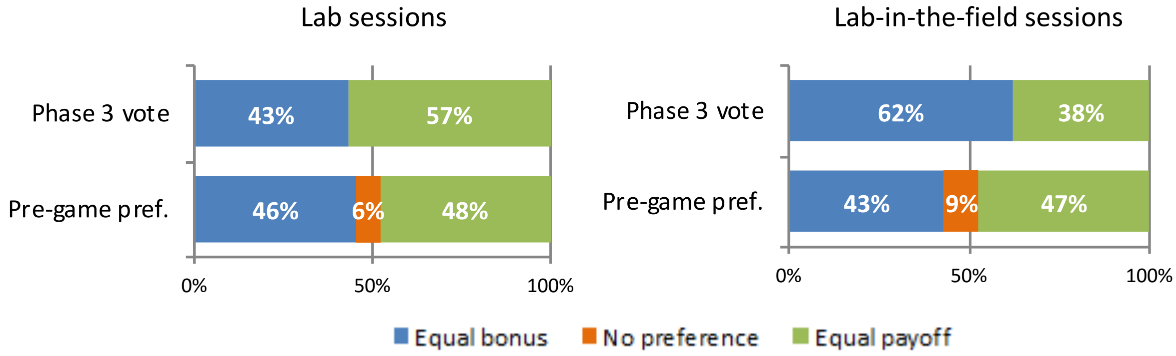
	Individual contribution			Individual payoff		
	Lab	LITF	Both locations	Lab	LITF	Both locations
Threshold level	0.180***	0.130***	0.158***	0.133***	0.102***	0.133***
	(0.014)	(0.027)	(0.014)	(0.021)	(0.039)	(0.021)
Threshold met or not				6.963***	12.170***	8.813***
				(0.884)	(1.245)	(0.829)
Minimum threshold treatment		1.939*	0.028		-3.976*	-0.075
		(1.028)	(0.571)		(1.109)	(0.562)
Endogenous threshold	0.433		1.448***	-0.487		-1.356**
	(0.487)		(0.559)	(0.493)		(0.562)
Equal-payoff rule	0.804	1.244**	0.910*	-0.118	-1.058	-0.407
	(0.489)	(0.667)	(0.465)	(0.534)	(0.724)	(0.527)
Low-cost type	2.192***	1.922***	1.972***	0.987***	3.342***	2.038***
	(0.406)	(0.614)	(0.351)	(0.323)	(0.549)	(0.324)
Fair-distribution	0.768	-0.015	0.386	-0.464	0.290	-0.102
	(0.522)	(0.610)	(0.417)	(0.385)	(0.624)	(0.404)
Earn-much	-0.112	-1.334**	-1.344***	0.196	0.642	1.176**
	(0.715)	(0.602)	(0.498)	(0.524)	(0.619)	(0.466)
Explanation-clear	-0.252	-0.856	-0.119	0.348	0.796	0.159
	(0.550)	(0.998)	(0.484)	(0.352)	(0.916)	(0.424)
Game-complexity	-0.939*	-0.221	-0.651*	0.244	0.556	0.407
	(0.488)	(0.653)	(0.388)	(0.384)	(0.600)	(0.329)
Constant	1.728	6.270***	3.257***	11.63***	5.398**	8.781***
	(1.086)	(1.538)	(1.090)	(1.076)	(2.187)	(1.341)
N	816	592	1408	816	592	1408
R-squared	0.319	0.164	0.296	0.383	0.533	0.389
Standard errors in parentheses	*p < 0.10, **p < 0.05, ***p < 0.01					

As expected, low cost subjects contribute about 2 points more and also earn more than high-cost subjects (payoffs are about 1 point higher in the lab and 3 points higher in the lab-in-the-field). Higher thresholds seem to lead to higher individual contributions and higher payoffs. Even though the impact is not very large, this result is in line with the broader literature (Croson and Marks 2000). Under the equal-payoff rule, subjects in the lab-in-the-field contribute significantly more compared to when they play under the equal-sharing-of-the-bonus rule. This confirms the earlier test results in section 4.2. However, the impact of the distribution rule on individual contributions and individual payoffs in the lab sessions is not as significant as the group contribution and group payoff test results shown before.

4.5 Distribution rule preference

Finally, we consider the individual preferences for one of the two distribution rules. Before the start of the game, we asked subjects which rule they preferred, and again when we asked them to vote before round 9. We find that subjects adapt their preferences on the basis of their game experiences. Figure 3 shows that across treatments there is no clear preference for either of the distribution rules at the start of the experiment, in both the lab and the lab-in-the-field ($p > 0.65$, binomial tests, using individual voting decisions as independent observations). At round 9, preferences have shifted: the equal-payoff rule is chosen more often in the lab, whereas in the lab-in-the-field this is the equal-sharing-of-the-bonus rule ($p < 0.10$, binomial tests, votes at group level as independent observations). This is in line with the finding that, in the lab, payoffs were higher under equal-payoff, whereas in the lab-in-the-field, payoffs were higher under the equal sharing-of-the-bonus rule.

Figure 3. Distribution rule preferences, pre-game vs. phase 3 vote



5. Conclusion

This paper assessed whether findings related to the sharing of benefits and endogenous choice from public good experiments carry over to a threshold public goods game experiment. The latter is more relevant from the perspective of environmental policymaking as payment schemes generally require individuals or groups to reach a certain threshold provision level before a payment is made.

The analysis considered three questions: i) does a differentiated bonus (that results in equal payoffs for heterogeneous subjects) lead to higher group payoffs and contribution levels than offering an undifferentiated, equal share of a group bonus; ii) does endogenous choice of thresholds and distribution rules enhance individual contributions and group payoffs and iii) do the results of the lab experiment carry over to the lab-in-the-field?

We find that the differentiated bonus improves game outcomes, but only in the lab experiment. In the lab-in-the-field sessions there are no significant differences between treatments, and we argue that one of the reasons why outcomes between the lab and the lab-in-the-field differ is that contributions are much less efficient in the lab-in the-field: over-contributions are significantly higher in the lab-in-the-field setting and although thresholds are met more often, payoffs are generally lower because of the failure to coordinate.

Endogenous choice has two dimensions in our experiment: (i) subjects vote for a (minimum or average) threshold in the endogenous threshold treatments, and (ii) subjects vote for one of the two distribution rules. For either of the two dimensions, we find no significant impact on game outcomes, contrary to the results of Gallier et al. (2017), Dal Bó et al. (2010) and Haigner and Wakolbinger (2010). Subjects in the endogenous threshold treatments did not perform better than in the exogenous benchmarks, most possibly because the exogenous threshold offers a clear focal point. This interpretation seems to be backed by the significantly negative treatment effect of the endogenous threshold on individual payoffs.

With regard to the transferability of findings, we find that game coordination failures were especially significant in the lab-in-the-field. Although subjects in the lab-in-the-field more often considered the experiment to be complicated, this did not explain game behaviour significantly.

Still, although students in the lab and the-lab-in-the-field were generally quite satisfied about the clarity of game explanations, in the lab-in-the-field almost 40 percent of students perceived the game as complicated, as compared to little over 10 percent in the lab. This may be related to the fact that most students in the lab had previous experience with lab experiments, while for the participants in the lab-in-the-field sessions it was their first time.

Interestingly, most participants of the lab-in-the-field sessions that perceived the game as complicated were students of the dairy farming and management track. In fact, these students indicated that they had found the game abstract and difficult to follow, whereas the students of the business administration track believed the game to be a good representation of agri-environmental management schemes. This could be an indication that exposure alone is not the key difference, but rather experience with abstract reasoning. Whereas the students of farm management were generally looking for cues how to behave (trying to look at each other's screen, asking for suggestions and advice) the students of business administration did not ask for advice and played the game individually.[5]

Finally, and in line with the broader literature, our results show that, on average, higher thresholds lead to higher contributions and higher payoffs. As environmental payment schemes are typically conditional on reaching a certain threshold of group effort, this suggests that setting higher exogenous thresholds may improve outcomes overall. This is an important finding from the perspective of agricultural policy making, but further testing is required to assess whether these results hold in the field. For this, it may be necessary to introduce more context and examples in game explanation, although this comes at the cost of increased noise and heterogeneity (Schram 2005). Still, from a policy perspective, it is important to be able to use lab experiments to test the design of the policy mechanisms (Ludwig et al. 2011), for subjects that lack experience with abstract decision-making problems in a game setting, the learning curve may otherwise simply be too steep.

References

Abeler, J., Falk, A., Gotte, L. & Huffman, D. (2009). Reference Points and Effort Provision. *CESifo Working Paper,* No.2585, Category 4: Labour Markets.

Alberti, F. & Cartwright, E.J. (2016). Full agreement and the provision of threshold public goods. *Public Choice,* 166, 205-233.

Anderson, L.R., Mellor, J.M. & Milyo, J. (2008). Inequality and public good provision: An experimental analysis. *The Journal of Socio-Economics*, 37(3), 1010-1028.

Ansink, E. & Bouma, J. (2013). Effective support for community resource management. *Forest Policy and Economics*, 37, 94-103.

Balafoutas, L., Kocher, M., Putterman, L. and Sutter, M. (2013). Equality, Equity and Incentives: An experiment. *European Economic Review*, 60, 32-51.

Bauman, Y. & Rose, E. (2011). Selection or indoctrination: Why do economics students donate less than the rest? *Journal of Economic Behavior & Organization*, 79(3), 318-327.

Brick, K. & Visser, M. (2015). What is fair? An experimental guide to climate negotiations. *European Economic Review*, 74, 79-95.

[5] Clearly, no advice was given and we made sure group members could not look at each other's screen.

Cadsby, C. B. & Maynes, E. (1999). Voluntary provision of threshold public goods with continuous contributions: experimental evidence. *Journal of Public Economics*, 71(1), 53-73.

Cadsby, C. B. & Maynes, E. (1998a). Choosing between a socially efficient and free-riding equilibrium: Nurses versus economics and business students. *Journal of Economic Behavior & Organization*, 37(2), 183-192.

Cadsby, C. B. & Maynes, E. (1998b). Gender and free riding in a threshold public goods game: experimental evidence. *Journal of Economic Behavior & Organization*, 34(4), 603-620.

Carlsson, F. Daruvala, D., & Johansson-Stenman, O. (2005). Are People Inequality-Averse, or Just Risk-Averse? *Economica*, 72(287), 375-396.

Cartwright, E. & Stepanova, A. (2017). Efficiency in a forced contribution threshold public good game. *International Journal of Game Theory*, 46(4), 1163-1191.

Chan, K.S., Mestelman, S. & Moir, R. (1999). Heterogeneity and the Voluntary Provision of Public Goods. *Experimental Economics,* 2(1), 5-30.

Chen, D.L., Schonger, M. & Wickens, C. (2016). oTree - An open-source platform for laboratory, online, and field experiments, *Journal of Behavioral and Experimental Finance*, 9, 88-97.

Cherry, T.L., Kroll, S. & Shogren, J.F. (2005). The impact of endowment heterogeneity and origin on public good contributions: evidence from the lab. *Journal of Economic Behavior & Organization*, 57, 357-365.

Cipriani, G. P., Lubian, D. & Zago, A. (2009). Natural born economists? *Journal of Economic Psychology*, 30(3), 455-468.

Cooper, D. J., Kagel, J. H., Lo, W. & Gu, Q. L. (1999). Gaming against managers in incentive systems: Experimental results with Chinese students and Chinese managers. *American Economic Review*, 89(4), 781-804.

Cooper, D. J. (2006). Are experienced managers experts at overcoming coordination failure? *Advances in Economic Analysis & Policy*, 5(2).

Corazzini, L., Cotton, C. & Valbonesi, P. (2015). Donor coordination in project funding: Evidence from a threshold public goods experiment. *Journal of Public Economics*, 128, 16-29.

Croson, R.T.A. & Marks, M.B. (2000). Step Returns in Threshold Public Goods : A Meta- and Experimental Analysis. *Experimental Economics*, 2, 239-259.

Dal Bó, P. D., Foster, A. & Putterman, L. (2010). Institutions and behavior: Experimental evidence on the effect of democracy. *The American Economic Review*, 100(5), 2205-2229.

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

Davis, D. & Holt, C. (1993). Experimental economics: Methods, problems and promise. *Estudios Económicos*, 8(2), 179-212.

Fiegenbaum, A., Hart, S. & Schendel, D. (1996). Strategic Reference Point Theory. *Strategic Management Journal*, 17(3), 219-235.

Finus, M. (2008). Game theoretic research on the design of international environmental agreements: insights, critical remarks, and future challenges. *International Review of Environmental and Resource Economics* 2(1), 29-67.

Fischbacher, U. (2007). z-Tree: Zurich Toolbox for Ready-made Economic Experiments, *Experimental Economics* 10(2), 171-178.

Gallier, C., M. Kesternich & B. Sturm (2017). Voting for burden sharing rules in public goods games. *Environmental and Resource Economics* 67(3), 1-23.

Goossens, B.H.J. (2017), Development of a Threshold Public Goods Game in oTree. Bachelor thesis Tilburg University, Tilburg School of Economics and Management.
Haigner, S.D. & Wakolbinger, F. (2010). To lead or not to lead: Endogenous sequencing in public goods games. *Economics Letters,* 108(1), 93-95.
Harrison, G. W. & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, *42*(4), 1009-1055.
Henrich, J., Heine, S. J. & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, 466(7302), 29.
Isaac, R. M., Walker, J. M. & Thomas, S.H. (1984). Divergent evidence on free riding: An experimental examination of possible explanations. *Public choice*, 43(2), 113-149.
Kallbekken, S., Kroll, S. & Cherry, T.L. (2011). Do you not like Pigou, or do you not understand him? Tax aversion and revenue recycling in the lab. *Journal of Environmental Economics and Management* 62(1): 53-64.
Kesternich, M., Lange A. & Sturm B. (2014). The impact of burden sharing rules on the voluntary provision of public goods. *Journal of Economic Behavior and Organization* 105, 107-123.
Kesternich, M., Lange, A. & Sturm, B. (2018). On the performance of rule-based contribution schemes under endowment heterogeneity. *Experimental Economics*, 21(1), 180-204.
Kocher, M., Martinsson, P., Persson, E. & Wang, X. (2016). Is there a hidden cost of imposing a minimum contribution level for public good contributions? *Working Paper 654*, Department of Economics, University of Gothenburg.
Koszegi, B. & Rabin, M. (2006). A model of reference-dependent preferences. *Quarterly Journal of Economics,* 121(4), 1133-1165.
Ledyard, J. (1995). Public goods: A survey of experimental research. In J. Kagel & A. Roth, eds. *Handbook of experimental economics*. Princeton: Princeton University Press., 111-194.
Levati, M.V, Sutter, M. & van der Heijden, E. (2007). Leading by example in a public goods experiment with heterogeneity and incomplete information. *Journal of Conflict Resolution*, 51(5), 793-818.
Levitt, S. D. & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *The Journal of Economic Perspectives*, 21(2), 153-174.
Levitt, S. D. & List, J. A. (2009). Field experiments in economics: The past, the present, and the future. *European Economic Review*, 53(1), 1-18.
List, J. A. (2003). Does market experience eliminate market anomalies? *The Quarterly Journal of Economics*, 118(1), 41-71.
List, J.A. & Price, M.K. (2016). The Use of Field Experiments in Environmental and Resource Economics. *Review Environmental Economic Policy,* 10(2), 206-225
Ludwig, J., Kling, J.R. & Mullainathan, S., 2011. Mechanism experiments and policy evaluations. *The Journal of Economic Perspectives*, 25(3), 17-38.
Martinsson, P. & Persson, E. (2016). Public Goods and Minimum Provision Levels: Does the institutional formation affect cooperation? *Working Papers in Economics,* No. 655*,* University of Gothenburg, Department of Economics.
Potters, J. & Van Winden, F. (2000). Professionals and students in a lobbying experiment: Professional rules of conduct and subject surrogacy. *Journal of Economic Behavior & Organization*, 43(4), 499-522.
Putterman, L., Tyran, J-R. & Kamei, K. (2011). Public goods and voting on formal sanction schemes. *Journal of Public Economics,* 95(9), 1213-1222.
Rapoport, A. & Suleiman, R. (1993). Incremental contribution in step-level public goods games with asymmetric players*. Organizational Behavior & Human Decision Processes,* 55(2), 171-194.

Rouaix, A., Figuières, C. & Willinger, M. (2015). The trade-off between welfare and equality in a public good experiment. *Social Choice and Welfare*, 45(3), 601-623.

Schram, A. (2005). Artificiality: The tension between internal and external validity in economic experiments. *Journal of Economic Methodology*, 12(2), 225-237.

Sutter, M. & Weck-Hannemann, H.(2003). On the effects of asymmetric and endogenous taxation in experimental public goods games. *Economics Letters,* 79(1), 59-67.

Sutter, M. & Weck-Hannemann, H. (2004). An experimental test of the public goods crowding out hypothesis when taxation is endogenous. *FinanzArchiv: Public Finance Analysis* 60(1), 94-110.

Tavoni, A., Dannenberg, A., Kallis, G. & Löschel, A., (2011). Inequality, communication and the avoidance of disastrous climate change in a public goods game. *Proceedings of the National Academy of Sciences* 108(29), 11825-11829.

Walker, J., Gardner, R., Herr, A. & Ostrom, E. (2000). Collective Choice in the Commons: Experimental Results on Proposed Allocation Rules and Votes. *The Economic Journal*, 110, 212-234.

Westerink, J., D.C.P. Melman & R.A.M. Schrijver (2015). Scale and self-governance in agri-environmental schemes: experiences with two alternative approaches in the Netherlands. *Journal of Environmental Planning and Management,* 58(8), 1490-1508.

Zelmer, J. (2003). Linear Public Goods Experiments: A Meta-Analysis. *Experimental Economics,* 6(3), 299-310.

Annex A: Game protocol (with framing in bold) Endogenous minimum threshold treatment (ESPM)

Translated from Dutch to English

WELCOME

We are about to start with a simulation game, or experiment about collective decision-making. We have developed the experiment to resemble collective agri-environmental management: from the 1st of January 2016 onwards, farmers can only receive a subsidy for agri-environmental management when they are part of a collective. Collectives of farmers jointly prepare a plan for agri-environmental management, including a certain ambition level, for which they can receive a subsidy. In the experiment we will talk in more abstract terms about the key elements of the collective decision-making process.

The instructions of the experiment will be read out aloud now, and you are invited to read along. You are about to participate in an experiment. Before we start, we would like to ask that you do not communicate with other people during this session. Please also turn off your mobile phone. You receive a show-up fee of €5 for being here today. You can earn additional money depending on your decisions and the decisions of other participants. The additional payoff in the experiment will be calculated in points, where the exchange rate is 1 point = €0.10.

INSTRUCTIONS

In this experiment, you decide how you want to spend an endowment that will be given to you. In each round, the endowment you and all other participants receive is 20 points, the value of which corresponds to 2 euro. All the decisions that you make during the experiment will remain anonymous.

In the experiment, you will form a group with three other participants. The group composition will remain the same in the entire experiment, so you will be in a group with the same participants. The choice you and your group members will make is whether you want to contribute (part of) your endowment to a group project. If sufficient contributions are made to the group project, that is, if total contributions in the group are equal to or higher than a group threshold contribution level, the group receives a bonus that is shared between the group members. However, if the threshold contribution level is not met (that is, total group contributions are below the group threshold level), the group project will not be executed and your contribution to the group project will not be transferred back to you.

You may consider the group project as collective agri-environmental management, the threshold as the group ambition level for agri-environmental management and the group bonus as the subsidy paid when the collective reaches its ambition level, e.g the agreed targets for agri-environmental management.

In total, the experiment has 3 phases, with each 4 rounds, hence a total of 12 rounds. In each round you have to make two decisions: a voting decision and a contribution decision.

THRESHOLD VOTING

First, you and your group members will be asked to vote for a threshold contribution level for your group project. You can vote for a level of 0, 1, 2, 3, … up to 80 points (which is the sum of the endowments). The lowest threshold proposed by yourself and your group members will be the group threshold level. The group bonus is equal to 1.6 times the group threshold.

For example, if you vote for a threshold of 40, two group members vote for a threshold of 20 and one group member votes for 10, the group threshold contribution level will be 10 points. If the threshold is met, the group bonus will be 1.6 × 10 = 16 points, and this group bonus will be shared among the group members.

CONTRIBUTION CHOICE

Once all votes for the threshold contribution level have been collected, you and your group members will be informed about the result: the threshold that will represent the group's ambition level. Next, you will be asked to decide how much of your endowment you want to contribute to the group project. Remember, each player has an initial endowment of 20 points. 1 point in the experiment is worth €0.10. You can contribute between 0 and 20 points to your group project.

If the group threshold is met, you receive your share of the group bonus. Recall that the group bonus is equal to 1.6 times the group threshold. If the group threshold is not met, you lose your contribution to the group project, and no bonus is paid. The same rules apply to your group members.

PAYOFFS

For every player, payoffs are calculated as follows.

If the group threshold is not met, then your

Payoff = (20 – "cost parameter" × your contribution)

If the group threshold is met, then your

Payoff = (20 – "cost parameter" × your contribution) + your share of group bonus.

COST PARAMETER

In each group, two players face a relatively high costs of contributing to the group project and two players face a relatively low cost of contributing to the group project. The "cost parameter" represents this difference: for the two high-cost players the cost parameter = 1.2 and for the two low-cost players the cost parameter = 0.6. The computer will randomly assign two low-cost and two high-cost players in each group. You and your group members will be informed about this at the beginning of the experiment.

Throughout the experiment, all players remain the same player type (high-cost or low-cost).

DISTRIBUTION RULE

For the sharing of the group bonus there are two different distribution rules. You will play the first phase of 4 rounds with one distribution rule, and the next phase of 4 rounds with the other distribution rule. In the last phase of the experiment you can vote for the distribution rule you prefer. The two distribution rules are:

- *Equal-sharing-of-the-bonus rule (phase 1, rounds 1 - 4)*: There is no differentiation in group bonus share. If the group threshold level is met, each group member receives an equal share of the group bonus;

- *Equal-payoff rule (phase 2, rounds 5 - 8)*: There is differentiation in group bonus share. Group members receive different shares of the group bonus in such a way that every group member receives in principle the same payoff.

EXAMPLES
Suppose the group threshold is 50 and total group contributions are equal to or higher than 50 such that the threshold is met. The group bonus is equal to 80 points (= 1.6 × 50).

Example 1: Equal-sharing-of-the-bonus rule.
Under this rule each group member receives the same share of the group bonus, 20 (=80/4).

- If a *low-cost player type* has contributed 5 points, then her payoff will be:
  Payoff = (20 – 0.6 × 5) + 20 = 37 points.
- If a *low-cost player type* has contributed 20 points, then her payoff will be:
  Payoff = (20 – 0.6 × 20) + 20 = 28 points.
- If a *high-cost player* has contributed 9 points, then his payoff will be:
  Payoff = (20 – 1.2 × 9) + 20 = 29.2 points.
- If a *high-cost player* has contributed 17 points, then his payoff will be:
  Payoff = (20 – 1.2 × 17) + 20 = 19.6 points.

Example 2: Equal-payoff rule.
Under this rule, the group bonus of 80 points is distributed in such a way that each group member receives in principle the same payoff.

Suppose a low-cost player type has contributed 20 points, the other low-cost player contributed 10 points, one high-cost player contributed 14 points and the other 8 points. The group bonus of 80 points will then be divided as follows: 20.9, 14.9, 25.7, and 18.5 points, such that payoffs will be equal for all group members.
The resulting payoffs will be:

Payoff low-cost player 1 = (20 – 0.6 × 20) + 20.9 = 28.9 points.
Payoff low-cost player 2 = (20 – 0.6 × 10) + 14.9 = 28.9 points.
Payoff high-cost player 1= (20 – 1.2 × 14) + 25.7 = 28.9 points
Payoff high-cost player 2= (20 – 1.2 × 8) + 18.5 = 28.9 points

Note: Under the equal-payoff rule it may happen (but chances are very slim) that the group threshold level is met, but it is not possible to distribute the group bonus such that all group members receive exactly the same payoff. If this happens the group bonus will be shared equally among the group members. This implies that you always receive at least 20 – cost parameter × your contribution.

DISTRIBUTION RULE VOTING

In the first two phases of the experiment the distribution rule is pre-determined: in phase 1 (rounds 1-4) the distribution rule is Equal-sharing-of-the-bonus rule, and in phase 2 (round 5-8), the distribution rule is Equal-payoff rule.

At the beginning of phase 3, in round 9, players are allowed to vote on the distribution rule they prefer in phase 3 (rounds 9-12). In each group, the distribution rule outcome will be based on majority rule voting. So if the majority in your group votes for Equal-sharing-of-the-bonus rule, the Equal-sharing-of-the-bonus rule will apply in rounds 9-12, if the majority in your group votes for Equal-payoff rule, the Equal-payoff rule will apply in rounds 9-12. In the case that two group members vote for the "Equal-sharing-of-the-bonus rule" and two group members vote for the "Equal-payoff rule", the vote is tied and the computer will randomly determine the distribution rule for your group. You will be informed about the voting outcome in your group and about the distribution rule that will be applied in phase 3.

EXPERIMENTAL EARNINGS

If the experiment is complete, the computer will randomly select three rounds to be paid, one in each of the three phases. You will receive the payoff of these three rounds in Euros. This payoff will be added to the show-up fee of €5.

Are there any questions at this point? Please raise your hand, and we will come to you to answer your questions in private. If there are no further questions, we now continue with a short test. Please answer the questions on the following page. When you have problems answering them, please raise your hand and we will come by to check your answers.

TEST QUESTIONS

1) What will be the threshold contribution level if you vote for a threshold of 50, one group member for a threshold of 35, another group member votes for a threshold of 20, and the last group member votes for a threshold of 80 points?

   …… points

2) If the threshold contribution level is not reached, will your contribution to the group project be transferred back to you? YES / NO

3) If you don't contribute, but the threshold contribution level is reached, do you receive a share of the group bonus? YES / NO

4) Suppose you are a low-cost type. What will be your payoff under the equal sharing of the group bonus rule if the threshold contribution level of 50 points is reached and you contributed 10 points?

   …… points

5) Suppose you are a high-cost type. What will be your payoff under the equal payoff rule if the threshold contribution level of 50 points is not reached and you contributed 10 points?

   …… points

Annex B: Overview of non-parametric test results

Table 5: Non-parametric test results, distribution rule comparison

p-value	Group threshold	Group contribution	Success rates	Over-contribution rates	Group payoff
Phase 1 Lab ESPM vs Lab EPSM	0.1494	0.0636	0.1192	0.6312	0.0145
Phase 1 Lab XSPM vs Lab XPSM	0.0134	0.1469	0.7699	0.8951	0.7920
Phase 1 LITF ESPM - LITF EPSM	0.4143	0.1229	0.3980	0.9353	0.8075
Phase 1 LITF ESPG - LITF EPSG	0.9566	0.2647	0.1518	0.2534	0.1917
Phase 1 vs Phase 2 Lab ESPM	0.0937	0.2115	0.2277	0.9547	0.0884
Phase 1 vs Phase 2 Lab EPSM	0.8753	0.0354	0.0038	0.0023	0.0843
Phase 1 vs Phase 2 Lab XSPM	0.0455	0.4098	0.7098	0.6098	0.0712
Phase 1 vs Phase 2 Lab XPSM	0.0039	0.0248	0.9121	0.2845	0.3329
Phase 1 vs Phase 2 LITF ESPM	0.2367	0.5534	0.1573	0.0180	0.4990
Phase 1 vs Phase 2 LITF EPSM	0.0796	0.2249	0.3173	0.0431	0.0431
Phase 1 vs Phase 2 LITF ESPA	0.0060	0.7536	0.0755	0.0150	0.0712
Phase 1 vs Phase 2 LITF EPSA	0.7007	0.2484	0.1978	0.0231	0.9721

Table 6: Non-parametric test results, endogenous-exogenous threshold comparison

p-value	Group threshold	Group contribution	Success rates	Over-contribution rates	Group payoffs
Phase 1 Lab ESPM vs Lab XSPM	0.6205	0.3798	0.3017	0.9223	0.0454
Phase 1 Lab EPSM vs Lab XPSM	0.9059	0.4641	0.9473	0.7697	0.6605
Three phases Lab ESPM vs Lab XSPM	0.8063	0.5582	0.4013	0.8836	0.0318
Three phases Lab EPSM vs Lab XPSM	0.5181	0.4823	0.1370	0.7253	0.2695

Table 7: Non-parametric test results, endogenous-exogenous distribution rule comparison

p-value	Group threshold	Group contribution	Success rates	Over-contribution rates	Group payoffs
Phase 1,2 vs Phase 3 Lab ESPM	0.3201	0.9547	0.0204	0.0007	0.3942
Phase 1,2 vs Phase 3 Lab EPSM	0.0257	0.1240	0.7764	0.7299	0.8753
Phase 1,2 vs Phase 3 Lab XSPM	0.0087	0.3078	0.1743	0.0229	0.0843
Phase 1,2 vs Phase 3 Lab XPSM	0.0046	0.0743	0.2192	0.0469	0.8785
Phase 1,2 vs Phase 3 LITF ESPM	0.1763	0.1282	0.0299	0.0425	0.7353
Phase 1,2 vs Phase 3 LITF EPSM	0.6858	0.0431	0.1599	0.0431	0.0796
Phase 1,2 vs Phase 3 LITF ESPA	0.1823	0.0994	0.0073	0.0096	0.0414
Phase 1,2 vs Phase 3 LITF EPSA	0.1730	0.7797	0.3057	0.3109	0.5525

Table 8: Non-parametric test results, endogenous-exogenous distribution rule comparison (excluding groups with tied votes)

p-value	Group threshold	Group contribution	Success rates	Over-contribution rates	Group payoffs
Phase 1,2 vs Phase 3 Lab ESPM	0.1420	0.5337	0.0084	0.0033	0.9292
Phase 1,2 vs Phase 3 Lab EPSM	0.2328	0.5754	0.3598	0.4838	0.2626
Phase 1,2 vs Phase 3 Lab XSPM	0.0192	0.3139	0.1325	0.0077	0.1097
Phase 1,2 vs Phase 3 Lab XPSM	0.0173	0.0904	0.6089	0.3105	1.0000
Phase 1,2 vs Phase 3 LITF ESPM	0.5930	0.2850	0.1573	0.2850	1.0000
Phase 1,2 vs Phase 3 LITF EPSM	0.2850	0.1088	0.3173	0.1088	0.2850
Phase 1,2 vs Phase 3 LITF ESPA	0.4838	0.0116	0.0138	0.0117	0.0173
Phase 1,2 vs Phase 3 LITF EPSA	0.0277	0.4631	0.3441	0.1730	0.3454