

Does the Endowment of Contributors Make a Difference in Threshold Public Good Games?

Federica Alberti and Edward J. Cartwright

Keywords: Public Good, Threshold, Endowment

JEL Categories: C72, H41

25th August 2014

Abstract

We investigate experimentally whether the endowment of potential contributors changes outcomes in threshold public good games. We find that the variance in contributions is significantly reduced for low and high levels of endowment when compared to an intermediate level of endowment. This difference emerges with experience. Evidence on whether the level of endowment influences the success rate at providing public goods is more mixed. We provide, however, a measure of endowment size that correlates with success rate. Our interpretation suggests that people find it hardest to coordinate on the provision of threshold public goods when endowments are of intermediate size. By intermediate we mean that the endowment is small enough that people need to contribute relatively a lot to fund the good, but also large enough that no one person is critical in providing the good.

* Alberti: Max Planck Institute of Economics, Kahlaische Str. 10, 07745 Jena, Germany (alberti@econ.mpg.de);
Cartwright: School of Economics, University of Kent, Canterbury, Kent, CT2 7NP, United Kingdom (E.J.Cartwright@kent.ac.uk). This paper was supported by the Centre for Reasoning at the University of Kent under a project on 'Reasoning and focal points'. Financial support by the Max Planck Society is also acknowledged.

1. Introduction

A threshold public good is a good that is provided if and only if contributions reach some critical threshold. An archetypal example would be raising funds for, say, a new university music centre that costs €4 million. If, to a rough approximation, the €4 million is necessary and sufficient for the centre to be a success then the center is a threshold public good (Andreoni, 1998). A slightly more mundane example would be a group of house mates trying to purchase, between them, a new television that will cost €200. This time the television is the threshold public good. More generally, the fixed costs associated with running any charity or group activity typically require a minimum (and often quite large) amount be reached to make the activity viable (Bagnoli and McKee, 1991). This means threshold public goods are of wide practical interest.

A large experimental and theoretical literature has considered whether threshold public goods can be provided efficiently. The theory primarily suggests they should (e.g. Bagnoli and Lipman, 1989). In particular, it is a Nash equilibrium to provide a threshold public good, and so the trade-off between Pareto efficiency and individual rationality that we see for other types of public good does not exist. Experimental studies, however, provide a more depressing picture. The success rate at providing threshold public goods is far below 100 percent and more typically around 40 to 60 percent (Croson and Marks, 2000). This naturally raises the question of what factors are associated with increased efficiency. One strand of the literature addressing this question has looked at the consequences of institutions, such as refunds, rebates, and sequential contributions (e.g. Coats et al., 2009). A second strand has looked at the consequences of strategic parameters, such as the threshold level and returns to the public good. It is this second strand of literature that we take as our starting point.

The literature has highlighted two strategic measures that correlate with the success rate of providing the public good, the step return and net reward (Cadsby et al., 2008). Informally, the step return measures the return on the public good relative to its average per-capita cost, and the net reward measures the absolute difference between return and average per-capita cost. One thing notably absent from both these measures is the endowment of potential contributors. Surely, however, endowments may matter? For instance, it would seem more likely that a university can raise €4 million for the new music centre if there are some billionaire alumni to call upon. To the best of our knowledge the prior literature has little to say on this question, with one

exception. In a meta-analysis, Croson and Marks (2000) consider the endowment relative to threshold level as a potential explanatory variable of success and find a positive but statistically insignificant effect.¹ Thus, whether efficiency is related to the size of endowments remains an open and intriguing question.

Attention has been paid in the literature to the importance of the threshold level (Isaac et al., 1989; Rauchdobler et al., 2010).² Intuitively, an increase in endowment size may appear similar to a decrease in threshold and vice-versa. It is important, therefore, in motivating our work, and distinguishing it from the prior literature, to highlight why endowment size reflects something different to the threshold.³ Here is the crucial difference: endowment is solely a constraint on ability to provide the good while the threshold reflects on the cost or ‘price’ of the good. To clarify, suppose there is an increase in the threshold. Holding the return on the public good constant, this increases the opportunity cost of providing the public good; the ‘price’ of the good has increased. Suppose instead that endowment size decreases. This reduces ability to contribute towards the good but has no impact on opportunity cost; income is lower but the price of the good is the same. Clearly, opportunity cost can be expected to influence behavior and so we cannot isolate the effect of endowment size without focusing purely on endowment size. We shall return to this issue in section 4 when motivating our experimental design.

The objective of this paper is to ask whether the size of endowment influences success at providing threshold public goods. We address this question in two stages. First, we briefly review the literature, much of which appears after the meta-analysis by Croson and Marks (2000), and revisit the evidence for whether the endowment matters.⁴ Such an exercise is informative but ultimately limited by the fact that no study has explicitly focused on endowment size. Previous studies have instead focused on average per-capita cost and

¹ To clarify, Croson and Marks (2000) use threshold/endowment as the explanatory variable. We shall focus on endowment/threshold. Hence, they obtain a negative coefficient but we interpret this as suggestive of a positive effect.

² In the case of full refund, changes in the threshold appear to have little impact (Isaac et al., 1989). Rauchdobler et al. (2010) find a lower success rate for high levels of threshold. Their design, however, mixes changes in the value of the public good with changes in the threshold.

³ Although, we also note that, the existing literature has relatively little to say on the effect of changes in average per-capita cost (see table 1 for more details). In particular, the only study that allows comparison of threshold size is Cadsby and Maynes (1999), and this has very small sample sizes.

⁴ Croson and Marks (2000) analyzed the broad literature on threshold public goods while we shall focus on a specific sub-class, namely threshold public good games that are simultaneous and characterized by a refund and no rebate.

returns to the public good. Our second, and main, task is, therefore, to report on experiments that we performed in order to specifically focus on endowment size.

Our results point towards a strong effect of endowment size on the variance of contributions. For a relatively small or large level of endowment the variance of contributions decreases, after experience, to a low level. By contrast, with an intermediate level of endowment the variance remains high, despite experience. The evidence that endowment size influences the success rate of providing the public good is more mixed. We find that after experience the success rate for a small or large level of endowment is significantly higher than for an intermediate level of endowment. We also show that a measure of endowment size correlates with success rate and again points towards a lower success rate for an intermediate level of endowment. This is consistent with the idea that groups find it hardest to coordinate and provide the public good for an intermediate level of endowment. There is, however, no direct evidence of an effect on success rate when we compare across the subset of treatments that differ only in the level of endowment.

Before proceeding with the more detailed analysis we briefly want to emphasize that the relevance of the endowment is an important practical and policy question, as well as an intriguing theoretical one. It is socially efficient that threshold public goods be provided and so it is important to know what factors are likely to affect the success in providing such goods. This can inform a policy maker as to when intervention may be necessary and can inform the group as to when help is required in order to successfully provide the good.

We proceed as follows: in section 2 we define and explain a threshold public good, in section 3 we review the literature, in section 4 we explain our experimental design, in section 5 we provide our main results and in section 6 we conclude. Additional material is contained in an appendix.

2. Threshold Public Good Game

We shall consider a standard (symmetric, continuous) threshold public good game (e.g. Cadsby et al., 2008). In the game there are n players, each endowed with E units of a private good. Independently and simultaneously players must decide how much of their endowment to contribute towards a public good. If total contributions equal or exceed threshold T then each player receives an additional V units of the private good and we say that the group was successful in providing the public good. Amount V is interpreted as the individual value of the public good. If contributions are below the threshold each contribution is refunded and if contributions are

above the threshold no money is rebated. So, if we denote the contribution of player j by c_j for all j the payoff of player i is

$$u_i(c_1, \dots, c_n) = \begin{cases} E - c_i + V & \text{if } \sum_{j=1}^n c_j \geq T \\ E & \text{otherwise} \end{cases} \quad (1)$$

This game has a large set of Nash equilibria. Of primary importance are a set of equilibria where total contributions match the threshold. Specifically, vector of contributions (c_1, \dots, c_n) is a Nash equilibrium with public good provision if

$$\sum_{j=1}^n c_j = T \quad \text{and } c_i < V \text{ for all } i. \quad (2)$$

Note that player i would not gain from contributing less than c_i because the threshold would not be met. Similarly, he would not gain from contributing more than c_i because the extra contribution would be wasted. All of the Nash equilibria in this set are Pareto efficient and yield the same average payoff $E + V - T/n$.⁵ They differ in how the total payoff is distributed, with a player who contributes relatively less receiving a bigger share of the total payoff. This creates an interesting asymmetric coordination problem in which players want to collectively contribute enough to meet the threshold while individually contributing as little as possible. One potential solution is for players to split the payoff equally with each player contributing the average per-capita cost T/n . This is a Nash equilibrium that gives every player the same payoff. It is, however, only one of the many possible Nash equilibria with public good provision.⁶

3. Strategic Trade-offs and the Endowment

⁵ There also exists a set of equilibria where total contributions are less than the threshold. Specifically, vector of contributions (c_1, \dots, c_n) is a Nash equilibrium with no public good provision if $\sum_j c_j < T$ and $T - \sum_{j \neq i} c_j > \min\{E, V\}$ for all i . In this case every player gets payoff E .

⁶ The promise of a refund but no rebate creates a form of asymmetry whereby it cannot be a Nash equilibrium for total contributions to exceed the threshold but it can for them to be short of the threshold. The presence or not of refunds and rebates can make a difference (Marks and Croson, 1998; Spencer et al., 2009). There is always, however, the incentive for a player to contribute less in the hope that others will contribute more. The presence or not of a refund or rebate changes how risky this is to do, but not the incentive to do so. Prior results show a change in the variance of contributions when rebates and refunds are altered, but not much change in the success rate at providing the public good (e.g. Marks and Croson, 1998).

Numerous studies have shown that the success rate at providing threshold public goods is far below 100 percent and depends on the fundamental parameters of the game. Cadsby et al. (2008) discuss two summary measures of the strategic trade-off players face (see also Cadsby and Maynes, 1999; Croson and Marks, 2000).

The step return is the ratio between the individual value of the good and the average per-capita cost:

$$SR = \frac{V}{T/n}. \quad (3)$$

The net reward is the absolute difference between the individual value of the public good and the average per-capita cost:

$$NR = V - \frac{T}{n} \quad (4)$$

Both the step return and net reward appear to correlate with the success rate of providing the public good. More specifically, the higher the step return the higher are observed success rates, and the higher the net reward the higher are observed success rates after sufficient experience (Cadsby and Maynes, 1999; Croson and Marks, 2000; Cadsby et al., 2008).

One notable thing about both the step return and net reward is the absence of the endowment. Indeed, there are four fundamental parameters in a threshold public good game, E , T , V and n , and E is the one that is absent from both these measures. This is not to say, in any way, that the step and net reward are not good measures. It does, however, suggest that we might want to consider additional measures that do take into account the endowment. In doing this, we introduce three informal theoretical conjectures. Our first conjecture justifies a focus on T , V and n , at the expense of E .

(C1) The value of the public good V relative to average per-capita cost T/n captures the desire to provide the public good. The endowment, by contrast, is merely a constraint on contributions and, provided $E \geq T/n$, this constraint is non-binding. The size of endowment should, therefore, be irrelevant.

This line of reasoning suggests that the SR and NR are all we need. Given that the prior literature has directly or indirectly accepted conjecture (C1) we shall use it as the basis for our main hypothesis.

Hypothesis 1 (success): Ceteris paribus changes in the endowment do not significantly change the success rate of providing the public good.

Clearly, the current paper is motivated by a belief that, despite the arguments in (C1), endowment size may matter. We provide two reasons why it may do so.

(C2) The more a player is endowed with the more she may be willing to contribute towards the public good because of, say, diminishing marginal utility from the private good. Essentially, the higher the endowment the cheaper it is to contribute towards the public good.⁷

(C3) The less that each player is endowed with the more critical they individually are to the provision of the public good. The more critical a player the more they may be willing to contribute (e.g. Rapoport, 1987; Chen et al., 1996; De Cremer and van Dijk, 2002). Indeed, if the sum of endowments is just sufficient to fund the public good it is a weakly dominant strategy for a player to contribute her full endowment.⁸

Clearly, (C2) and (C3) work in opposite directions and so the relationship between the success rate and endowment may not be a simple one. We consider it, therefore, an interesting empirical issue whether the size of endowment does matter, and, if so, in what way.

The easiest way to see if the size of endowment matters is clearly to consider *ceteris paribus* changes in endowment and measure the change in success rate. This we did. Ultimately, however, any change in endowment needs to be anchored relative to the other structural parameters of the game in order to draw general inferences. We propose the following measure of endowment size, which we shall call the endowment remainder and is given by the endowment that will remain after contributing the average per-capita cost. Formally,

$$ER = E - \frac{T}{n}. \quad (5)$$

The parallel between the *ER* and *NR* is hopefully apparent.⁹ The *ER* captures in a stylized but simple way the relative size of the endowment. If the *ER* is large then contributing to the public good is relatively cheap,

⁷ In an experimental setting where the 'public good' is an extra V units of private good this argument needs some motivation. One could easily hypothesize, however, a 'psychological bonus' from knowing the group has succeeded in providing the public good, which is not measured in units of private money (Rapoport, 1987).

⁸ If she contributes her full endowment and someone else does not, she gets her contribution back and so has lost nothing. If she contributes her full endowment and everyone else does as well then she benefits from the public good and would not have done so if she had contributed any less. Note that the refund is crucial to making this argument.

⁹ Other possible measures can be considered (see Alberti and Cartwright, 2010). For simplicity, we focus exclusively here on the *ER* as it is our preferred measure (and also because our objective was not to distinguish between, what will inevitably, be highly correlated alternative measures of endowment size).

capturing possibility (C2). If the ER is small then each player is more critical, capturing possibility (C3), and if $ER = 0$ then it is a weakly dominant strategy for a player to contribute her entire endowment. In evaluating hypothesis 1 we shall also be interested in whether the ER can serve as an appropriate measure of endowment size.

One important source of evidence with which to evaluate hypothesis 1 is the prior literature. In table 1 we summarize the results of experiments from seven papers. The experiments are chosen to be directly comparable to the setting of interest here. In particular, there are refunds, no rebates, simultaneous contributions and repeated matching. The only exceptions are differences in the number of periods (25 or 14) and the second set of experiments by Croson and Marks (2001) where suggested contributions were made to players before they contributed.¹⁰ To the best of our knowledge these are the only directly comparable experiments reported in the literature. Comparing across experiments we see a lot of variation in the three measures of strategic trade-off. Note, however, that no study considers a *ceteris paribus* change in endowment or in ER .¹¹

[INSERT TABLE 1 AROUND HERE]

Table 2 details the results of weighted least square regressions in which we regress the success rate against combinations of SR and NR together with a quadratic formulation of ER , together with a dummy for whether n is five or ten. The quadratic formulation allows us to pick up any potential non-linearities that might occur given the competing reasons for why the endowment may matter. For now we ask you to focus solely on specifications (1) and (2). Consistent with previous studies, we observe a strong positive relationship between SR and success rate but no relationship between the NR and success rate (see Cadsby et al., 2008). More important for our purposes, there is no evidence that the endowment as measured by ER has a statistically significant effect. This supports hypothesis 1. It also fits with the results of Croson and Marks (2000) who, for a largely orthogonal set of data, find no evidence that endowment size has any effect. Less clear is whether this lack of evidence reflects the lack of an effect, or an inability to detect the effect because of insufficient statistical power. Given that no previous study has considered *ceteris paribus* changes in endowment size the

¹⁰ Given that the suggestions had little effect we include these experiments here.

¹¹ Changes in ER are always accompanied by changes in SR and/or NR . Cadsby et al. (2008) 'come closest' to considering a *ceteris paribus* change in ER , and we shall comment on their results in more detail below.

latter possibility is a definite concern. This motivates our experiments that more directly question whether endowment size matters.

[INSERT TABLE 2 AROUND HERE]

We finish this section by noting that while our primary interest is on the success rate at providing the public good other things are of interest. Two of these are the dynamics of contributions and the variance of contributions. Both of these have been observed to vary with measures of strategic trade-offs. For example, Cadsby et al. (2008) find that the success rate increases with experience the higher is the NR . Also, Croson and Marks (2000) observe that the variance of individual contributions increases with the SR .¹² Such effects are important to take account of in trying to understand better how and why the size of endowment may matter. This motivates two further hypotheses:

Hypothesis 2 (dynamics): Ceteris paribus changes in the endowment do not affect how the success rate changes with experience.

Hypothesis 3 (variation): Ceteris paribus changes in the endowment do not affect the variance in contributions.

4. Experimental Design

To evaluate hypotheses 1 to 3, we ran experiments with ceteris paribus changes in the endowment and ER . Given the linear dependence of SR , NR and ER it is difficult to vary one without varying the others. Indeed if we want to keep n constant and vary either T or V then we necessarily have to vary SR or NR . To keep things tractable we fixed the SR and considered five sets of parameters, where the NR did vary. Table 3 summarizes the sets of parameters considered. The baseline treatment is the same as the baseline treatment in Cadsby et al. (2008) and so provides a natural starting point. Both the NR and SR are held constant in the baseline, Low1 and High1 treatments giving a very simple ceteris paribus test of whether the endowment matters. To test the ER measure of endowment size we include treatments Low2 and High2 in which we change T and V while keeping the ER the same as in Low1 and High1, respectively. The change in T and V necessarily changes the NR but we shall try to control for this.

¹² On a related note, Marks and Croson (1998) and Croson and Marks (2001) find, respectively, that the rebate rule and suggestions/recommendations can have an effect on the variance of group contributions.

[INSERT TABLE 3 AROUND HERE]

In motivating and clarifying our experimental design it may be useful at this point to reaffirm why changes in endowment size are different to changes in the threshold or value of the public good. Compare, first, the Baseline parameters with the Low1 parameters. The, *ceteris paribus*, lower endowment makes each player's contribution more critical without, in any way, changing the rewards or incentive to provide the public good. This is formally captured by saying the *ER* changes while the *NR* and *SR* remain the same. Compare next the Baseline parameters with parameters $E = 55$, $V = 50$ and $T = 225$. The, *ceteris paribus*, higher threshold brings about a similar change in *ER* (as compared to the Low1 treatment), but only by drastically reducing the rewards or incentive to provide the public good. This would show up in changes to both the *SR* (from 2 to 1.11) and *NR* (from 25 to 5). Changes in endowment size cannot, therefore, be equated with changes in threshold size. To mitigate this effect, our Low2 treatment increases the value of the public good along with the threshold in order to keep the *SR* constant (which also reduces the change in *NR*).

The experiments were run at the University of Kent (United Kingdom) with subjects recruited from the general student population. We ran eight sessions in all. The 20 subjects in each session were assigned seats scattered in a large room, randomly assigned a group of 5, assigned a treatment, and then played the relevant threshold public good game for 25 periods. The groups remained the same throughout the 25 periods allowing us to look at dynamic group effects. In total there were 8 groups for the baseline treatment and 6 groups for each of the other treatments giving a total of 160 subjects.¹³ Subjects were not allowed to communicate with one another except through their decisions. Given the random assignment to groups interaction within each group was anonymous.

The experiments were computerized, using z-Tree (Fischbacher, 2007). The language used in the instructions and experiment was deliberately neutral, for example, asking subjects how many tokens they wanted to allocate to a group account. [The instructions are available in the appendix.] At the end of each period subjects were told the total contribution to the public good, whether or not the public good had been provided, and their payoff. At the end of the session, each subject was paid in cash the total amount earned in

¹³ Initially we ran 6 sessions giving 4 groups in the baseline treatment and 5 groups in the other four treatments. Following a suggestion of a referee we ran 2 additional sessions. These additional sessions were run under the exact same conditions, with new subjects. Given that no difference was observed between these sessions and the earlier ones, we shall report results based on the data for all 8 sessions.

the experiment. This was given by a £2 show-up fee plus 0.5 pence for every token accumulated over the 25 periods. The average payment was £10.70. Each experimental session lasted about 45 minutes.

5. Results

To give a first overall picture of the results, figure 1 plots the timeline of total contributions for each group, and table 4 summarizes the success rate in providing the public good by treatment. The fluctuations, clearly visible in figure 1, of total contributions around the threshold are what we have come to expect from the previous literature.¹⁴ More interesting, is the large difference in the variance of contributions across treatments. For example, this variance is noticeably larger in the baseline treatment than the Low1 or High1 treatments. We will look at this in more detail in section 5.2. Our focus before this will be on success rates.

[INSERT FIGURE 1 HERE AND TABLE 4 AROUND HERE]

5.1 Success Rates

The figures in table 4 suggest that there is no simple relationship between the success rate and endowment size in the first five to ten periods. In later periods, however, especially in the last five periods, a more simple relationship has seemingly emerged. Namely, the success rate in the Baseline treatment is relatively low when compared to the other four treatments. This difference is economically significant because a success rate of 52.5 percent is a big drop in efficiency compared to 76.7 percent. Note also how the success rate in the Baseline treatment is stable at a relatively low level, while in the Low1 and High1 treatments it is increasing, and in the Low2 and High2 treatments it is stable at a high level. This translates into overall success being highest in the Low2 and High2 treatments.

In order to help interpret the relatively low success rate in the Baseline treatment, as compared to other treatments, we highlight that the 55 percent success rate we report is in line with the 44, 55, 58, 59 and 63 percent success rates previously reported in the literature (see table 1).¹⁵ That we obtain a weakly decreasing

¹⁴ This most likely reflects the fact that total contributions rarely equal T but more often exceed T . The proportion of times contributions equaled T were 0.12 in the Baseline, 0.05 in the Low1 treatment, 0.05 in the High1 treatment, 0.21 in the Low2 treatment, and 0.12 in the High2 treatment. In contrast, the proportion of times contributions exceeded T were 0.44 in the Baseline, 0.56 in the Low1 treatment, 0.6 in the High1 treatment, 0.52 in the Low2 treatment, and 0.61 in the High2 treatment.

¹⁵ These differences are significant ($p = 0.047$, Fisher's exact test, 2-tailed), but this is due to the 'outlier' success rates of 44% due to Cadsby et al. (2008) and 63% due to Marks and Croson (1998). Note, however, that the difference between our 55% and that of 63% is insignificant ($p = 0.206$), while the difference between our 55%

success rate over time contrasts with Cadsby et al. (2008), who found an increasing success rate over time, but is more similar to what is observed in Croson and Marks (2000, 2001). It does seem, therefore, that the low success rate we observe in the Baseline is no accident.

To more formally test the significance of endowment size on success rate we first report the results of random-effects logistic regressions with the probability of success as the dependent variable.¹⁶ Table 5 presents the results with treatment dummies as explanatory variables. The Baseline treatment is used as the comparator. Looking across all periods we see no significant difference between success rate in Low1 and High1 treatments relative to the Baseline. We do, though see a significantly higher success rate in the Low2 and High2 treatments compared to the Baseline. Given that both the Low2 and High2 treatments differ from the Baseline in more than endowment size these results are not evidence against hypothesis 1. Indeed, they point more towards the importance of the *NR* as the only strategic measure that systematically varies across the Baseline, Low2 and High2 treatments. Note, however, that this interpretation would conflict with the results of Cadsby et al. (2008) who found the *NR* had no effect on overall success rates. Also, note that the differences between the Low1 and Low2 treatments and that between the High1 and High2 treatments are insignificant.

If we look now at the results for the first five periods then the only significant difference we observe is the weakly higher success rate in High2 treatment. Noise is to be expected, however, in the first five periods as subjects learn about the game and other group members. More interesting is the final five periods. Here we observe a very similar effect across the High1 and High2 and Low1 and Low2 treatments. Treated separately, only the High1 and Low2 treatments show a weak effect. Collectively, however, the four treatments are significantly different from the baseline ($p = 0.04$). This is the strongest evidence so far that endowment size has a significant effect on success rates.

To complement the results in table 5 we report in table 6 regression results with *ER* and *ER*² as explanatory variables. Recall that the baseline, Low1, and High1 treatments provide the most direct test that the endowment matters and so we report results that use only these three treatments (1) as well as those that use all

and that of 44% is marginally significant ($p = 0.086$). Variations in success rate across studies are no surprise given different subject pools, instructions etc. Reassuringly, our estimate lies close to the average of previous studies.

¹⁶ The random-effects regression is to capture the fact that, since each group solves 25 choice problems, observations are not independent.

five treatments (2). The comparison of specifications (1) and (2), allow us to assess whether the *ER* is a reliable measure of endowment size.¹⁷ Looking over all periods we observe a significant effect of *ER* across all treatments. No significant effect is observed in specification (1) but the coefficients in specifications (1) and (2) are similar and so a lack of statistical power may be the explanation. In the final five periods we observe a marginally significant effect of *ER* in specification (1) and a more significant effect in specification (2).

To put the results reported in table 6 into context we can return to an analysis of the previous literature. Recall that the regression results summarized in columns (1) and (2) of table 2 did not provide evidence of a statistically significant correlation between success rates and the *ER*. In column (3) of table 2 we provide the results of a regression that includes the experiments reported in this paper, as well as those from the prior literature. While the coefficients on *SR* and *NR* remain unchanged, as we would hope, those on *ER* become significant.¹⁸ Moreover, the estimated coefficients are in line with those reported in table 6.¹⁹

Overall, the results summarized in tables 4, 5 and 6 are not inconsistent with hypothesis 1. In particular, we observe no strong effect across the Baseline, Low1 and High1 treatments. We do observe a strong effect across the Baseline, Low2 and High2 treatments but this may be due to factors other than endowment size, such as the *NR*. If we focus, however, on the last five periods then we find systematic differences across treatments that are suggestive endowment size matters. We also see that the *ER* is a predictor of success, even once we control for changes in the *NR*. To progress further we shall now widen the scope analysis to look at more than success rates.

[INSERT TABLES 5 AND 6 AROUND HERE]

5.2 Variance in Contributions

¹⁷ For completeness we report results for the first five periods even though the results in table 6 would suggest against using the *ER* in this case.

¹⁸ It is clear in table 1 that the success rate is lower with groups of size 10. In specifications (1) and (2) there is seemingly not enough statistical power to distinguish this effect and so the constant and $n = 10$ coefficient are insignificant. In specification (3) the effect is distinguished. This results in a larger constant and significantly negative $n = 10$ coefficient. The net effect is to increase the predicted success rate in groups of size five while leaving unchanged the predicted success rate in groups of size 10.

¹⁹ We can look briefly at the results of Cadsby et al. (2008), see table 1. These results are of particular note because they also allow comparison across a set of treatments where the *ER* ranges between 11 and 42 keeping the *SR* fixed to 2 (see table 1). They also find success is lower in the Baseline than treatments with a lower and higher *ER*, but these differences are statistically insignificant.

Figure 1 gives the impression that the variance in group contributions differed over the five treatments. Table 7 shows some descriptive statistics on this variance. We shall pay particular attention to the absolute deviation of group contribution from the threshold. It is noticeable that there is a decline in the absolute deviation in all treatments bar the baseline. The High1 treatment is particularly noteworthy in that we observe a relatively high absolute deviation in early periods but by the final periods this deviation is comparable to the other treatments, and much lower than in the baseline.

[INSERT TABLE 7 AROUND HERE]

Tables 8 and 9 report the results of OLS random-effects regressions with absolute deviation as the dependent variable. Table 8 presents results with treatment dummies as explanatory variables. Table 9 presents results with ER and ER^2 as explanatory variables. Across all periods we see a higher variance in the Baseline and High1 treatment than the other three treatments (see table 8). In the first five periods the variance in the High1 treatment is larger than that in the High2 treatment ($p = 0.01$). Note, however, that the period coefficient is strongly negative and the constant term relatively large. This is consistent with a learning effect that could explain why we observe no systematic relationship between the ER and success rate in the early periods (see table 9). In the final five periods we see a higher variance in the Baseline treatment in comparison with all other treatments (see table 8). This difference is highly significant, even if we restrict attention to the High1 and Low1 treatments.

The results in tables 8 and 9 are clearly inconsistent with hypothesis 3 in that we find significant differences in the variance of contributions across treatments. This difference emerges over time as the variance declines in all treatments except the Baseline. It seems reasonable to conjecture that the higher variance in contributions we observe in the Baseline is not unrelated to the relative low success rate that we also observe. If so, this would lend support to the argument that changes in endowment size do indeed influence success rates. One reason to make this connection is that a lower variance in contributions provides the opportunity for a subject's beliefs about the likely contributions of others to stabilize (Offerman et al., 2001). More accurately refined beliefs should improve success at coordinating. In particular, prior evidence suggests that most subjects contribute at or above the optimal level given their beliefs (Suleiman and Rapoport, 1992; Rauchdobler et al., 2010). If, therefore, subjects have accurate beliefs they should be successful in providing the public good.

Unfortunately, we did not elicit the beliefs of subjects in order to directly test this chain of logic. We can, however, explore the issue to some extent by looking at some data on individual contributions.

[INSERT TABLES 8 AND 9 AROUND HERE]

5.3 Individual Data

Table 10 summarizes the distribution of individual contributions relative to the average per-capita cost of providing the public good. In interpreting this data, let us start by looking at the Low1 and Low 2 treatments.²⁰

We see in table 10 that the distribution of choices was similar in both treatments and that the modal choice was to contribute the entire endowment.²¹ This is consistent with conjecture (C3) that increased criticality will increase success rate. Note that in order to obtain such criticality we require $E < V$ which means contributing the entire endowment can become a focal point.²² Note also, that the vast majority of contributions are within five of the average per-capita cost. This helps explain the low variance in contributions. It also should allow beliefs to become relatively accurate.

Consider now the Baseline, High1 and High2 treatments.²³ In these three treatments the modal choice was to contribute the average per-capita cost.²⁴ There are, however, notable differences between the treatments. In the High2 treatment most contributions are within five of the average per-capita cost. In the Baseline and High1 treatments contributions are noticeably more dispersed. In interpreting this difference it is crucial to recognize that dispersion in individual contributions need not translate into a high variance in group contributions. Indeed, we know that it does not in the High1 treatment. This is possible if groups ‘find’ a way to provide the public good where cost is not shared equally. In the High1 treatment groups were seemingly able to do this. In the Baseline treatment they were not. This is consistent with conjecture (C2) of increased success due to diminishing marginal utility of the private good.

²⁰ Recall that in the Low1 treatment subjects are endowed with 30 and need to contribute on average 25 while in the Low2 treatment they are endowed with 55 and need to contribute on average 50.

²¹ Consistently, the average individual contribution and the average individual proportion contributed, respectively, were 24.58 and 0.82 in the Low1 treatment and 49.75 and 0.90 in the Low2 treatment.

²² In order that a player’s contribution be critical we require nE to be close to T . So, in order for $V > T/n$, we essentially require $V > E$. In the Baseline, High1 and High2 treatments, where criticality is less, $E > V$. In these treatments only around four percent of individual contributions were above V .

²³ Recall that in the Baseline and High1 treatment subjects were endowed with 55 or 70 and needed to contribute on average 25. In the High2 treatment they were endowed with 55 and needed to contribute on average 10.

²⁴ The average individual contribution and the average individual proportion contributed were 25.06 and 0.46 in the Baseline, 25.58 and 0.37 in the High treatment and 10.63 and 0.19 in the High2 treatment.

[INSERT TABLE 10 AROUND HERE]

Given what we have seen it seems reasonable to suggest that the dispersion in contributions in the Baseline treatment is caused by: (i) the endowment being large enough that subjects can try to free-ride by contributing relatively little, which was not an option in the low treatments, and (ii) the endowment being small enough that the gain from free-riding is relatively large, which was not the case in the high treatments. This combination of a possibility to free ride and a significant gain from doing so may create too much conflict between group members for them to coordinate. Hence we see a large variance in contributions and little success in providing the public good, with no convergence towards a more efficient and stable outcome.

6. Conclusion

Many goods can be approximated as threshold public goods and so it is very important to know when such goods are likely to be produced at the efficient level, and when not. This question is made more pertinent by the consistent evidence that groups can be disappointingly bad at providing threshold public goods. The prior literature has highlighted strategic variables that seem to affect the success rate at providing such goods (e.g. Cadsby et al., 2008) but little attention has so far been given to the role of the endowment. The motivation behind this study was to question whether the endowment does matter.

We found evidence of a strong effect of endowment size on the variance of contributions. After experience, the variance is significantly lower with a small or large level of endowment when compared to an intermediate level. The evidence for an effect of endowment size on success rates is more mixed. But it is consistent with a relatively high success rate for a small or large level of endowment after experience. These relationships can be explained by strategic incentives: If the endowment is small then people have no alternative but to contribute if they want the public good provided. If the endowment is large then people gain little by trying to free ride. It is when the endowment is intermediate that a conflict of interest arises between group members – there is the incentive and possibility to free ride. Given this conflict it can prove hard for the group to coordinate as evidenced by the relatively large variance in contributions.

We feel that this is a robust story that shows the endowment matters. To put it in some perspective we can return to the music center example with which we began. If the University does have some billionaire

alumni to call upon then there will probably be little problem funding the center.²⁵ More interestingly, our results suggest that if it is known the music center can only be provided if everyone gives relatively a lot, then it will also be provided because potential contributors will feel critical. The difficult middle ground is where people are being asked to contribute relatively a lot but know that if they do not give someone else could fill that gap. Charities seem keenly aware of this by trying to give the impression that every donation is critical.

²⁵ This paper has not addressed the issue of endowment heterogeneity, which may be suggested by the example. We hope to consider this issue in future work.

Appendix

Here are the instructions for the baseline treatment. The instructions for all other treatments were identical except for the obvious change in parameter values.

In this experiment you will make decisions, and earn an amount of money that depends on what you and others choose. The money will be given to you at the end of the experiment in an envelope. Only you will know how much money you earned.

You have been organized into groups of 5. Each group will consist of the same 5 people for the duration of the session. The session will last for 25 periods. In each period you will be required to make a decision, and your total earnings will depend on your decisions in all periods.

At the beginning of every period you, and all other members of your group, will receive 55 tokens. Each of you must decide on your own how many of the 55 tokens to allocate to a group account.

If the total number of tokens allocated to the group account is 125 or more then you will each receive an additional 50 tokens.

If the total number of tokens allocated to the group account is less than 125 then you will receive no additional tokens but will get back any tokens you allocated to the group account.

So, at the end of the period:

If the total number of tokens allocated to the group account ≥ 125

your earnings = initial 55 tokens – tokens allocated to group account + additional 50 tokens

If the total number of tokens allocated to the group account < 125

your earnings = initial 55 tokens.

At the end of the session, you will be asked to fill in a short questionnaire. You will be paid in cash the total amount that you earned for all periods in the session plus £2. Each token will be worth 0.5p[ence].

References

- Alberti, F., and Cartwright, E. J. (2010), Does the Endowment of Contributors Make a Difference in Threshold Public Good Games?, University of Kent School of Economics Discussion Paper No. 1009.
- Andreoni, J. (1998), Toward a Theory of Charitable Fund Raising, *Journal of Political Economy* 106, 1186-1213.
- Bagnoli, M., and Lipman, B. L. (1989), Provision of Public Goods: Fully Implementing the Core Through Private Contributions, *Review of Economic Studies* 56, 583-601.
- Bagnoli, M., and McKee, M. (1991), Voluntary Contribution Games: Efficient Private Provision of Public Goods, *Economic Inquiry* 29, 351-366.
- Cadsby, C. B., Croson, R., Marks, M., and Maynes, E. (2008), Step Return Versus Net Reward in the Voluntary Provision of a Threshold Public Good: An Adversarial Collaboration, *Public Choice* 135, 277-289.
- Cadsby, C. B., and Maynes, E. (1999), Voluntary Provision of Threshold Public Goods with Continuous Contributions: Experimental Evidence, *Journal of Public Economics* 71, 53-73.
- Chen, X. P., Au, W. T., and Komorita, S. S. (1996), Sequential Choice in a Step-level Public Goods Dilemma: The Effects of Criticality and Uncertainty, *Organizational Behavior and Human Decision Processes* 65, 37-47.
- Coats, J., Gronberg, T., and Grosskopf, B. (2009), Simultaneous Versus Sequential Public Good Provision and the Role of Refunds - An Experimental Study, *Journal of Public Economics* 93, 326-335.
- Croson, R., and Marks, M. (1998), Identifiability of Individual Contributions in a Threshold Public Goods Experiment, *Journal of Mathematical Psychology* 42, 167-190.
- Croson, R., and Marks, M. (2000), Step Returns in Threshold Public Goods: A Meta- and Experimental Analysis, *Experimental Economics* 2, 239-259.
- Croson, R., and Marks, M. (2001), The Effect of Recommended Contributions in the Voluntary Provision of Public Goods, *Economic Inquiry* 39, 238-249.
- De Cremer, D., and van Dijk, E. (2002), Perceived Criticality and Contributions in Public Good Dilemmas: A Matter of Feeling Responsible to All?, *Group Processes and Intergroup Relations* 5, 319-332.
- Fischbacher, U. (2007), z-Tree: Zurich Toolbox for Ready-Made Economic Experiments, *Experimental Economics* 10, 171-178.

Isaac, R. M., Schmidt, D., and Walker, J. M. (1989), The Assurance Problem in a Laboratory Market, *Public Choice* 62, 217-236.

Marks, M., and Croson, R. (1998), Alternative Rebate Rules in the Provision of a Threshold Public Good: An Experimental Investigation, *Journal of Public Economics* 67, 195-220.

Offerman, T., Sonnemans, J., and Schram, A. (2001), Expectation Formation in Step-level Public Good Games, *Economic Inquiry* 39, 250-269.

Rapoport, A. (1987), Research Paradigms and Expected Utility Models for the Provision of Step-level Public Goods, *Psychological Review* 94, 74-83.

Rauchdobler, J., Sausgruber, R., and Tyran, J. R. (2010), Voting on Thresholds for Public Goods: Experimental Evidence, *FinanzArchiv / Public Finance Analysis* 66, 34-64.

Spencer, M., Swallow, S., Shogren, J., and List, J. (2009), Rebate Rules in Threshold Public Good Provision, *Journal of Public Economics* 93, 798-806.

Suleiman, R., and Rapoport, A. (1992), Provision of Step-level Public Goods with Continuous Contribution, *Journal of Behavioral Decision Making* 5, 133-153.

Table 1*Prior Results in the Literature*

Paper	n	E	T	V	Periods	SR	NR	ER	Success Rate %	Groups
Cadsby et al. (2008)	5	55	65	26	25	2	13	42	54	4
	5	55	125	50	25	2	25	30	44	4
	5	55	65	38	25	2.92	25	42	79	4
	5	55	220	69	25	1.56	25	11	30	4
	5	55	220	88	25	2	44	11	54	4
Croson and Marks (2000)	5	55	125	30	25	1.2	5	30	33	4
	5	55	125	50	25	2	25	30	59	4
	5	55	125	75	25	3	50	30	63	4
Cadsby and Maynes (1999)	10	10	25	20	25	8	17.5	7.5	48	1
	10	10	25	20	14	8	17.5	7.5	71	1
	10	10	25	5	25	2	2.5	7.5	16	1
	10	10	50	7.5	25	1.5	2.5	5	28	1
	10	10	50	11	25	2.2	6	5	20	1
	10	10	50	20	25	4	15	5	44	1
	10	10	50	30	25	6	25	5	60	1
	10	10	75	10	25	1.33	2.5	2.5	10	2
	10	10	75	22.5	25	3	15	2.5	40	1
	10	10	75	32.5	25	4.33	25	2.5	36	1
Marks and Croson (1998)	5	55	125	50	25	2	25	30	63	4
Bagnoli and McKee (1991)	5	11	12.5	5	14	2	2.5	8.5	86	1
	10	11	25	5	14	2	2.5	8.5	64	1
Croson and Marks (2001)	5	55	125	50	25	2	25	30	55	5
	5	55	125	50	25	2	25	30	58	5
Croson and Marks (1998)	5	55	125	50	25	2	25	30	55	5

Table 2*The Results of Weighted Least Square Regressions with Success Rate as the Dependent Variable*

	(1)	(2)	(3)
Constant	21.41* (12.26)	28.45 (17.10)	46.624*** (10.327)
<i>SR</i>	6.71** (2.69)	6.71** (2.74)	7.005** (2.501)
<i>NR</i>	0.19 (0.28)	0.21 (0.29)	0.235 (0.224)
<i>ER</i>	0.51 (0.33)	-0.22 (1.25)	-1.514** (0.643)
<i>ER</i> ²	-	0.015 (0.025)	0.036** (0.013)
<i>n</i> = 10	-13.33 (14.11)	-17.30 (15.79)	-30.891** (11.383)
<i>R</i> -squared	0.555	0.564	0.607
Adj. <i>R</i> -squared	0.461	0.443	0.521
Observations	24	24	29

Note: Regressions (1) and (2) use exclusively the prior literature summarized in table 1. Regression (3) includes the data from the experiments we report in this paper. Standard errors in brackets; * indicates significant at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 3*The Experimental Parameters*

Treatment	<i>n</i>	<i>E</i>	<i>T</i>	<i>V</i>	<i>NR</i>	<i>SR</i>	<i>ER</i>	Groups
Baseline	5	55	125	50	25	2	30	8
Low1	5	30	125	50	25	2	5	6
High1	5	70	125	50	25	2	45	6
Low2	5	55	250	100	50	2	5	6
High2	5	55	50	20	10	2	45	6

Note: *n* gives the number of players and ‘groups’ gives the number of experimental groups for each treatment.

Table 4*A Summary of the Results*

Treatment	<i>NR</i>	<i>ER</i>	Success Rate for Provision % (Periods)					ALL
			1-5	6-10	11-15	16-20	21-25	
Baseline	25	30	62.5	57.5	52.5	50	52.5	55
Low1	25	5	43.3	50	76.7	63.3	73.3	61.3
High1	25	45	53.3	56.7	66.7	70	76.7	64.7
Low2	50	5	63.3	70	83.3	70	76.7	72.7
High2	10	45	83.3	66.7	70	73	70	72.7

Table 5*The Results of Random Effects Logistic Regressions with Success as the Dependent Variable*

	Periods		
	ALL	1-5	21-25
Constant	-0.068 (0.277)	-0.390 (0.505)	-2.249 (3.054)
Low1	0.285 (0.368)	-0.833 (0.541)	1.032 (0.707)
High1	0.443 (0.370)	-0.403 (0.537)	1.213* (0.717)
Low2	0.814** (0.375)	0.038 (0.545)	1.213* (0.717)
High2	0.821** (0.375)	1.161* (0.633)	0.846 (0.696)
Period	0.022** (0.011)	0.312** (0.127)	0.103 (0.131)
Log-likelihood	-500.115	-97.721	-94.089
Observations	800	160	160
Groups	32	32	32

Note: Standard errors in brackets; * indicates significant at the 10% level, and ** at the 5% level.

Table 6*The Results of Random Effects Logistic Regressions with Success as the Dependent Variable*

	Periods					
	ALL		1-5		21-25	
	(1)	(2)	(1)	(2)	(1)	(2)
Constant	0.335 (0.492)	0.471 (0.664)	-1.465* (0.782)	-0.940 (1.122)	-2.597 (3.938)	-0.950 (3.242)
<i>ER</i>	-0.049 (0.047)	-0.073* (0.039)	0.086 (0.063)	0.014 (0.065)	-0.154* (0.087)	-0.138* (0.071)
<i>ER</i> ²	0.001 (0.001)	0.002** (0.001)	-0.002 (0.001)	0.000 (0.001)	0.003* (0.002)	0.003** (0.001)
<i>NR</i>	-	0.009 (0.014)	-	0.002 (0.023)	-	0.012 (0.027)
Period	0.030** (0.013)	0.022** (0.011)	0.262* (0.152)	0.313** (0.127)	0.195 (0.169)	0.103 (0.131)
Log-likelihood	-322.898	-501.174	-66.086	-101.568	-59.110	-94.130
Observations	500	800	100	160	100	160
Groups	20	32	20	32	20	32

Note: Specification (1) includes data from the Baseline, High1 and Low1 treatments. Specification (2) includes data from all five treatments. Standard errors in brackets; * indicates significant at the 10% level, and ** at the 5% level.

Table 7

The Mean Total Contribution per Group and its Standard Deviation, and the Variances in Contributions across Periods as Measured by Absolute Deviation of Total Contribution from the Threshold

Treatment	Mean	Std. Dev.	Absolute Deviation from the Threshold (Periods)					
			1-5	6-10	11-15	16-20	21-25	ALL
Baseline	125.3	17.4	16.4	9.4	13.3	9.5	11.1	11.9
Low1	122.9	11.2	15.3	6.4	4.5	5.4	4.2	7.2
High1	127.9	19.5	24.7	15.3	11.6	8.8	4.2	12.9
Low2	248.8	12.6	10.8	8.6	3.4	6.5	3.3	6.5
High2	53.1	7.52	10.3	5.5	3.8	3.5	3.2	5.3

Table 8*The Results of Random Effects Ordinary Least Square Regressions of Absolute Deviation from the Threshold*

	Periods		
	ALL	1-5	21-25
Constant	18.066*** (1.931)	29.078*** (3.854)	14.788* (8.333)
Low1	-4.738* (2.795)	-1.075 (5.080)	-6.85** (3.175)
High1	1.022 (2.795)	8.358 (5.080)	-6.883** (3.175)
Low2	-5.378* (2.795)	-5.575 (5.080)	-7.717** (3.175)
High2	-6.645** (2.795)	-6.108 (5.080)	-7.816** (3.175)
Period	-0.474*** (0.048)	-4.234*** (0.649)	-0.163 (0.351)
<i>R</i> -squared Within	0.115	0.251	0.002
<i>R</i> -squared Between	0.294	0.254	0.259
<i>R</i> -squared Overall	0.157	0.252	0.145
Observations	800	160	160
Groups	32	32	32

Note: Standard errors in brackets; * indicates significant at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 9*The Results of Random Effects Ordinary Least Square Regressions of Absolute Deviation from the Threshold*

	Periods					
	ALL		1-5		21-25	
	(1)	(2)	(1)	(2)	(1)	(2)
Constant	12.814*** (3.733)	5.876 (5.103)	28.968*** (7.314)	19.953** (9.687)	15.837 (12.707)	3.478 (9.639)
<i>ER</i>	0.296 (0.370)	0.602* (0.308)	-0.407 (0.693)	0.254 (0.578)	0.915** (0.437)	0.970*** (0.321)
<i>ER</i> ²	-0.003 (0.007)	-0.010 (0.006)	0.013 (0.014)	-0.001 (0.011)	-0.018** (0.009)	-0.020*** (0.006)
<i>NR</i>	-	0.116 (0.110)	-	0.123 (0.206)	-	-0.009 (0.115)
Period	-0.542*** (0.065)	-0.474*** (0.048)	-3.985*** (0.897)	-4.243*** (0.649)	-0.685 (0.520)	-0.163 (0.351)
<i>R</i> -squared Within	0.126	0.115	0.200	0.251	0.023	0.002
<i>R</i> -squared Between	0.157	0.154	0.131	0.051	0.206	0.255
<i>R</i> -squared Overall	0.133	0.124	0.170	0.169	0.124	0.143
Observations	500	800	100	160	100	160
Groups	20	32	20	32	20	32

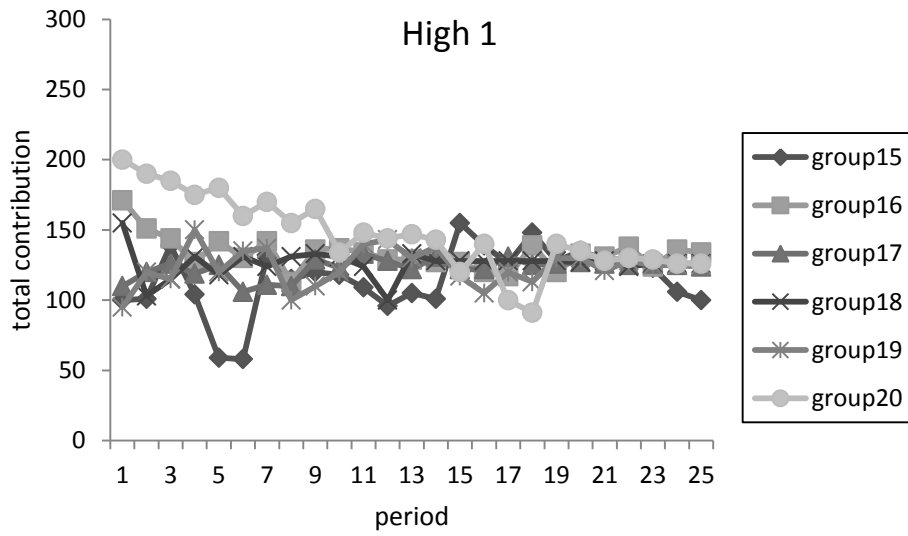
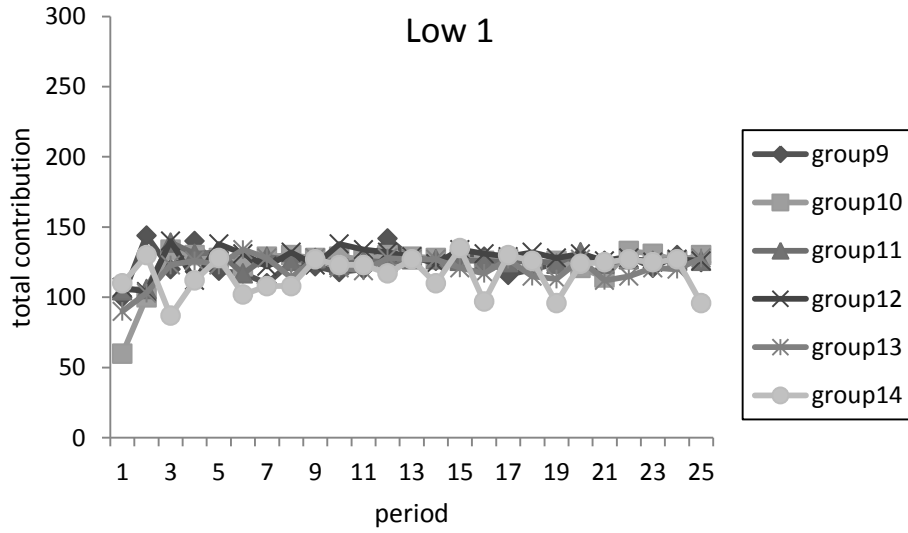
Note: Specification (1) includes data from the Baseline, High1 and Low1 treatments. Specification (2) includes data from all five treatments. Standard errors in brackets; * indicates significant at the 10% level, ** at the 5% level, and *** at the 1% level.

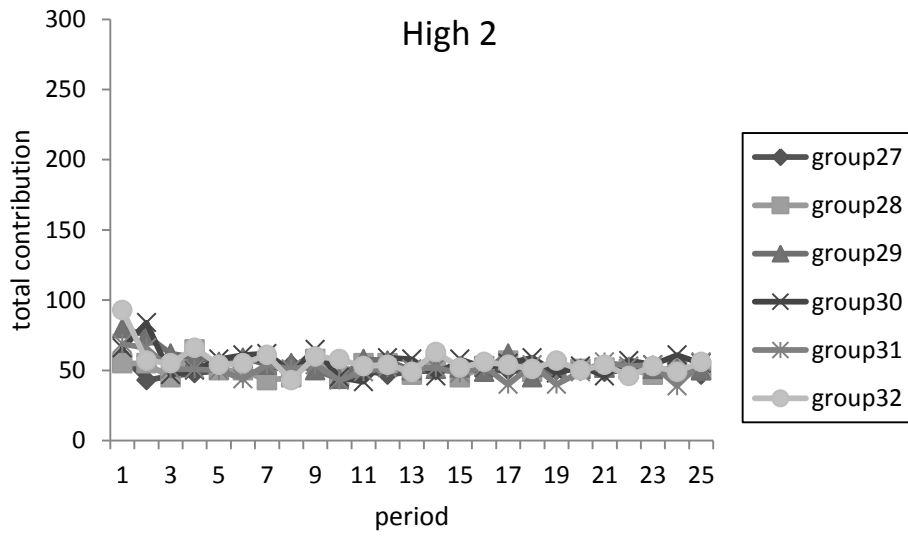
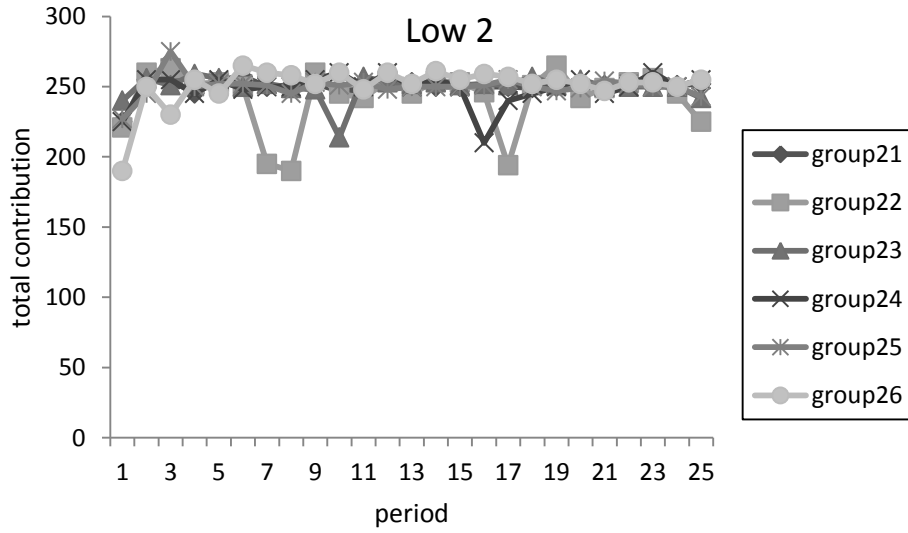
Table 10

Proportion of Individual Contributions Relative to the Average (Avg.) Per-capita Cost of Providing the Public Good

Contribution Relative to Avg. Per-capita Cost	Periods									
	ALL					21-25				
	Baseline	Low1	High1	Low2	High2	Baseline	Low1	High1	Low2	High2
Less 6 or more	21.5	11.7	24.8	10.5	4.3	25.5	12.0	23.3	14.7	5.3
Less 5	5.4	8.1	10.1	6.8	8.5	3.5	3.3	12.0	2.0	5.3
Less 1 to 4	3.8	17.2	4.8	10.5	22.1	4.0	23.3	5.3	11.3	29.3
Per-capita cost	28.7	22.1	25.2	28.5	29.1	27.0	18.7	31.3	22.0	25.3
Plus 1 to 4	10.2	10.4	5.2	10.8	17.1	6.5	10.7	2.0	14.0	18.0
Plus 5	10.5	30.4	13.1	32.8	10.0	9.0	32.0	12.7	36.0	9.3
Plus 6 or more	19.9	-	16.8	-	8.9	24.5	-	13.3	-	7.3

Note: For example, in the Baseline treatment, over all 25 periods, 21.5 percent of contributions were 6 or more below the average per-capita cost of 25 (meaning between 0 and 19).





Note: Each line represents one group. There are 32 groups in total.