Step-Level Public Goods: Experimental Evidence*

Hans-Theo Normann† and Holger A. Rau‡

Duesseldorf Institute for Competition Economics (DICE)

August 2011

Abstract

In a step-level public-good experiment, we investigate how the order of moves (simultaneous vs. sequential) and the introduction of a second step-level–which is not feasible in standard Nash equilibrium–affects public-good provision in a two-player game. We find that the sequential order of moves significantly improves public-good provision and payoffs, even though second movers often punish when first movers who give less than half of the threshold contribution. The additional step level leads to higher contributions but does not improve public-good provision and lowers payoffs. Based on an existing experimental data set, we calibrate Fehr and Schmidt's (1999) model of inequality aversion to make quantitative predictions. We find that actual behavior fits remarkably well with these predictions in a quantitative sense, but there are also two contradictions to the model's predictions.

JEL Classification: C92, D70, H41

Keywords: Experimental economics, sequential play, threshold public good, provision-point public good.

*We are grateful to Dirk Engelmann, Andreas Hildenbrand, Nikos Nikiforakis, and Theo Offerman for detailed comments on our paper. Volker Benndorf made very useful suggestions at an early stage of the project. We are also indebted to seminar audience at ESA Copenhagen, University of Frankfurt and Hohenheimer Oberseminar for helpful comments.

†Tel: +49 211 8115297, +49 211 8115499; email: *normann@dice.uni-duesseldorf.de*

‡Tel: +49 211 8110249, +49 211 8115499; email: *rau@dice.uni-duesseldorf.de*

1 Introduction

Public goods often have a step-level character, that is, the public good is provided only if some minimum threshold of contributions (or provision point) is met. Examples include the building of a bridge or a dike. Also charities often have properties of step-level public goods (see the examples in Andreoni, 1998).

Our paper makes two contributions to the literature on public goods with step levels. First, we analyze whether sequential contributions as opposed to simultaneous decisions improve public-good provision. Second, we analyze if an additional threshold which is not feasible in Nash equilibrium with standard preferences, where the public good is provided at a higher level, improves public-good provision.

The issue of sequential vs. simultaneous decisions is subject of a substantial and growing literature. Following the theory contributions by Andreoni (1998), Hermalin (1998) and Vesterlund (2003), researchers have analyzed "leading by example" in experiments. If a first-mover gives an example that is mimicked by the followers, sequential contributions to the public good may be superior to simultaneous decisions. This will particularly be the case when a first-mover is better informed about the return to contributions allocated to the common endeavor. This literature includes Erev and Rapoport (1990), Potters et al. (2005, 2007), Güth et al. (2007), Gächter et al. (2010a, 2010b), Figuières et al. (2010). (We review the literature below.)

We study sequential vs. simultaneous decisions in a step-level game with two players with complete imformation. For such a setting, one would at first expect a sequential-move game seem superior to a simultaneous-move setting. A threshold public-good game is foremost a coordination game. With simultaneous moves, there are multiple equilibria. Coordination failures may occur and, moreover, the public good is not provided in all equilibria. With sequential moves, there is a unique subgame perfect equilibrium in which the public good is provided. Hence, coordination and therefore public-good provision should be more frequent with sequential moves. There is, however, an aspect of sequential decision making that may reduce its alleged superiority. With symmetric players, the first mover is actually better off than the follower(s). In the unique subgame perfect equilibrium, the first mover contributes just enough such that a best responding follower just finds it (weakly) worthwhile to meet the threshold with her contribution. In other words, with selfish and rational players, the first mover actually gives a *bad* example by contributing *less* than the followers. In an experiment, this may reduce the alleged efficiency of the sequential-move setting. Players who try to exploit this first-mover advantage risk being "punished" by second-movers who do not best respond but contribute zero to the public good. If such behavior occurs frequently, the higher efficacy of the sequential-move game will not materialize. Based on a calibrated model (see below), we hypothesize that efficiency enhancing effect dominates so that sequential moves improve public-good provision.

Now consider our second extension, the introduction of a second thresholds. Multiple step-levels have been analyzed before (see Coller et al., 2001, Hashim et al., 2011). In our experiments, the second threshold is *not* a Nash equilibrium with selfish and rational players. Reaching the second step level is possible, but given one player aims at the second threshold by contributing a high amount, the best response of a second player is to contribute low such that the first level only is met. Thus, with standard preferences the second level does not make a difference. However, when players have Fehr and Schmidt (1999) preferences, the second threshold is a Nash equilibrium, and meeting the second threshold would of course be efficient. Even if some types exploit their opponents who aim for the second level, the public good is still provided at least at the first level and so no efficiency loss occurs. In other words, behaviorally, the existence of a second threshold might make it more likely that the *first* threshold will bet met. We thus hypothesize that the second step level improves public-good provision.

Our main findings regarding the two treatment variables are as follows. Sequential contribution decisions significantly improve public-good provision, even though first-movers frequently do contribute less than the followers and even though such behavior is regularly punished. Coordination rates and payoffs are higher whereas contributions are not higher with sequential moves. The existence of a second threshold causes significantly higher contributions but this does not result in higher public-good provision. To the contrary, payoffs are significantly lower when there are two step-levels.

Our paper also introduces a methodological innovation. We make *quantitative* predictions for our experiment based on a fully calibrated Fehr and Schmidt (1999) model. While Fehr and Schmidt's (1999) model of inequality aversion has frequently been used in the previous literature, the predictions are almost always of a qualitative nature ("if players are sufficiently inequality averse, *abc* is an equilibrium"). We will calibrate Fehr and Schmidt's (1999) model on an existing (joint) distribution of the inequality parameters, and we will make exact quantitative predictions ("v percent of the first-movers will contribute w"; or "given a first-mover contribution of x, the public good will be provided in y percent of the cases").[1]

We find that the calibrated Fehr and Schmidt (1999) model makes remarkably accurate quantitative predictions, but it also fails in two cases. The calibrated Fehr and Schmidt (1999) model predicts second-mover behavior (given first-mover behavior) in the sequential variant extremely

[1]Fehr, Kremhelmer and Schmidt (2008) also provide an analysis based on a calibration of Fehr and Schmidt's (1999) model. Their calibration is however based on a two-type categorization (40 percent fair players and 60 percent standard players).

well. Specifically, it predicts accurately the frequency of second-mover decisions (contribute such that the step level is met vs. punish first-movers by contributing zero). The prediction regarding the first-movers fails. First-movers should anticipate (or learn) that second-movers punish low contributions and thus always make the payoff-equalizing contribution. However, as has been observed in previous experiments (e.g., Huck et al., 2001), only slightly more than one third of them do so. The calibrated Fehr and Schmidt (1999) model also predicts well the case with simultaneous-move contributions where some players contribute whereas others do not. Finally, the model rather precisely predicts the share of first- movers who trust second-movers by making a high contribution in the sequential two-threshold case. Here, the prediction regarding the second-movers fails who exploit first-mover trust significantly more frequently than predicted.

2 Literature Review

There are two major strands of the literature pertinent to our paper. The first literature is about simultaneous vs. sequential order of moves in public-good games. The second literature concerns public-good experiments with step-level character in general and, specifically, those with more than one threshold.

As mentioned in the introduction, several researchers have analyzed "leading by example" theoretically. Andreoni (1998) examines the efficiency of leadership giving. The paper provides an explanation of how seed money, that may be raised from a group of "leadership givers," generates additional donations. In Hermalin (1998), a first-mover may be better informed about the return to contributions allocated to the common endeavor. Therefore, she may give plausibly an example to followers who rationally mimic the first mover's behavior. Vesterlund (2003) shows that in the presence of imperfect information on a charity's quality, an announcement strategy which leads to a sequential provision mechanism may be optimal.

An increasing experimental literature was triggered by these theory contributions. Following Hermalin (1998) and Vesterlund (2003), Potters et al. (2005) study an experimental voluntary contribution mechanism (VCM) where some donors do not know the true value of the good. The authors conclude that sequential moves result in higher contributions of the public good. They also have a treatment where the sequencing of choices emerges endogenously. Potters et al. (2007) report that the "leading by example" approach depends on whether there is incomplete information in the experiment. This explain why some experiments have not found sequential moves to be superior (Andreoni, Brown and Vesterlund, 2002) while Potters et al. (2005) did. Our experiments differ to those of Potters et al. (2005, 2007) in that we do not include information asymmetries, and we do not employ the VCM.[2, 3, 4]

Related to our study is also Gächter et al. (2010a). They experimentally study the effects of a simultaneous vs. a sequential choice mode in a test of Varian's (1994) VCM model. In contrast to Andreoni, Brown and Vesterlund (2002), the authors show that provision is generally higher in the simultaneous-move variant because second movers punish first movers who free ride in the sequential variant. One of the main difference to our approach is that the authors test the Varian (1994) model, whereas we study a step-level setup. Even though we observe similar punishing behavior, the sequential-move variant is more efficient in our data.

Erev and Rapoport (1990) were the first to study simultaneous vs. sequential moves in a step-level-public-good game with discrete choices. In their experiments, at least three of five players must contribute their endowments for the public good to be provided. Actions are minimal contribution sets, MCS, such that players either zero contribute or invest their whole endowment. They find a framing effect in that information about previous noncooperative choices is more effective in public-good provision than the sequential-move choices with only information about previous cooperative choices. The main differences to our experiment are the discrete action space and the number of players (two in our case). Discrete contributions in a step-level game have also been studied in Schram et al. (2008) where at least three of five or seven players need to contribute their full endowment to meet the threshold. Coats and Neilson (2005) and Coats et al. (2009) add refund policies to this setting. Both studies use groups of four players and analyze sequential moves, and Coats et al. (2009) furthermore analyzes a simultaneous-move order. The authors conclude that the sequential move order is superior to the simultaneous-move case. The refund mechanism stimulates efficiency. We do not offer a refund for excess contributions in our experiments.[5]

The literature on public-good games with multiple step-levels is much smaller. Coller et al.

---

[2] Serra-Garcia et al. (2010) extend the analysis of Potters et al. (2007) by comparing an the "action" treatment where a better informed first mover can give an example to a "word" treatment where the better informed player can verbally communicate the state of the world.

[3] Figuières et al. (2010) provide evidence that the "leadership effect" vanishes over time when subjects are randomly ordered in a sequence that differs from round to round. Gächter et al. (2010b) analyze the characteristics of effective leaders in a simple leader-follower public-good games. In a sequential VCM, they find that efficiency depends on the leaders' social preferences. Gächter and Renner (2005) demonstrate that there is a positive correlation between contributions of designated leaders (who act as first-movers) and second-movers' contributions. Güth et al. (2007) find evidence that leadership in the VCM setup is more efficient when the first mover has exclusion power.

[4] There also exist field experiments which demonstrate the efficiency of sequential designs (List and Lucking-Reily, 2002, Soetevent, 2005).

[5] Less relevant to our study is Cadsby and Maynes (1999b) which analyzes a simultaneous-move framework with a refund policy. The authors compare a continuous contribution mechanism with a MCS. Their results show that continuous contributions increase efficiency. The refund guarantee also encourages provision.

(2001) have a four-player experiment with one to five step levels. Compared to the baseline with one step level, treatments with multiple levels sometimes keep the social optimum constant and lower the Nash equilibrium contributions, sometimes they increase the group optima contributions but leave the Nash equilibria unchanged. In either case, contributions fall compared to the baseline, but complete free riding is prevented throughout. In our experiments, the second threshold implies a higher efficient outcome but is not tenable in a Nash equilibrium with selfish and rational players. We find that the second threshold increases contributions but lowers welfare.

Hashim et al. (2011) analyze a game with five levels and five players. Related to Croson and Marks (1998) study of the effect of information feedback with one step level, they vary information feedback about other members' contributions to a subsample of group members, specifically whom receives the information. Results show improvements in coordination when information targeting is used. Providing information randomly does not improve coordination and eventually degrades towards free-riding over time. The authors argue that a random information provision approximates strategies used in practice for educating consumers about digital piracy, information targeting may be useful.

3 Experimental Design and Procedures

In our experiments, there are two players, player 1 and player 2, who each have a money endowment $e = 10$. They can make a voluntary contribution, c_i, to the public good, where $0 \leq c_i \leq e$.

In half of our treatments, there is *one threshold* for the provision of the public good. If the sum of contributions is at least 12, this yields an additional payoff of 10 to both players. Any contributions between 1 and 11 and beyond 12 are wasted. More formally, if x_i denotes player i's monetary payoff, then

$$x_i = \begin{cases} e - c_i + 10 & \text{if } c_1 + c_2 \geq 12 \\ e - c_i & \text{if } c_1 + c_2 < 12 \end{cases}$$

The other treatments involve an additional *second threshold* of 18. If $c_1 + c_2 \geq 18$, both players receive an additional payment of 5. That is, in these treatments, we have

$$x_i = \begin{cases} e - c_i + 15 & \text{if} \quad 18 \leq \quad c_1 + c_2 \\ e - c_i + 10 & \text{iff} \quad 12 \leq \quad c_1 + c_2 \quad < 18 \\ e - c_i & \text{if} \quad\quad\quad\;\; c_1 + c_2 \quad < 12 \end{cases}$$

Since $2e > 18$, both thresholds of the public good are feasible, but, due to $e < 12$, no player can meet the threshold on her own. Further, because $2 \cdot 10 > 12$ and $2 \cdot 15 > 18$, the provision of the

		Order of moves	
		simultaneous	sequential
Step-levels	one	SIM_1	SEQ_1
	two	SIM_2	SEQ_2

Table 1: *Treatments*

public good at both provision points is collectively rational. Note that the return on contributing one euro at each of the two levels is the same, namely $12/20 = 0.6$ and $6/10 = 0.6$.

We have four treatments, labeled SIM_1, SIM_2, SEQ_1, and SEQ_2. The SIM labels refer to treatments where the two players make their decisions simultaneously whereas decisions are made sequentially (with player 1 moving first) in the SEQ treatments. The second treatment variable is the number of the thresholds (one or two). Table 1 summarizes our 2×2 treatments design.

Subjects play this game over 10 periods. The payoffs of the above game were denoted in euros in the experiments (so that the exchange rate was one to one). In each period, subjects were endowed with 10 euros but only one randomly chosen period was paid at the end of the experiment. Subjects were randomly matched in groups of two players. In the SEQ treatments, also the roles of first and second-movers were random.

We have three entirely independent matching groups per treatment. Each experimental session contained only one matching group. The size of the sessions or matching groups varied between 10 and 18 subjects. (We control for session size in our data analysis below). The subject pool consists of students from the University of Frankfurt from various fields. In total, we had 160 participants who earned on average 11.3 euros. The experiment was programmed in z-Tree (Fischbacher, 2007). Sessions last about ...

4 Predictions

Assumptions

We now derive the one-shot Nash equilibrium predictions for this public-good game. In addition to standard Nash predictions (selfish players who maximize their own monetary payoff), we will use Fehr and Schmidt's (1999) model, henceforth F&S. In their model, players are concerned not only about their own material payoff but also about the difference between their own payoff and other players' payoffs. Assumption 1 defines the two-player variant of their model.

Assumption 1. *Players' preferences can be represented by the utility function* $U_i(x_i, x_j) = x_i - \alpha_i \max[x_j - x_i, 0] - \beta_i \max[x_i - x_j, 0],\ \ x_i,\ x_j = 1, 2, i \neq j.$

Here, x_i and x_j denote the monetary payoffs to players i and j, and α_i and β_i denote i's aversion towards disadvantageous inequality (envy) and advantageous inequality (greed), respectively. Standard preferences occur for $\alpha = \beta = 0$. Following F&S, we assume $0 \leq \beta_i < 1$.

Using the specific functional forms of the step-level public good game for x_i above, we can write the F&S utilities as a function of contributions directly, so that we obtain $U_i(c_i, c_j)$. For the treatments with one step-level, we obtain

$$U_i(c_i, c_j) = 10 - c_i + 10\chi_1 - \alpha_i \max[c_i - c_j, 0] - \beta_i \max[c_j - c_i, 0] \tag{1}$$

whereas, for the two-step-levels treatments, we get

$$U_i(c_i, c_j) = 10 - c_i + 10\chi_1 + 5\chi_2 - \alpha_i \max[c_i - c_j, 0] - \beta_i \max[c_j - c_i, 0], \tag{2}$$

$c_i,\ c_j = 1, 2; i \neq j$. The χ_k are indicator functions indicating whether a step level has been reached: $\chi_1 = 1$ iff $c_1 + c_2 \geq 12$ and $\chi_2 = 1$ iff $c_1 + c_2 \geq 18$.

Using this model, we will make *quantitative* predictions. We fully calibrate the F&S model using the joint distribution of the α and β parameters observed in Blanco, Engelmann and Normann (2010). For each subject, they derive an α_i from rejection behavior in the ultimatum game and a β_i from a modified dictator game.[6] See Table 2. On average, $\alpha = 1.18$ and $\beta = 0.47$. There are no significant differences between the α distribution Blanco, Engelmann and Normann (2010) elicit and the one assumed in Fehr and Schmidt (1999). By contrast The β distribution differs significantly; however, one can argue the distributions still roughly compare and do not differ outlandishly.

Why rely on the specific α_i-β_i distribution of Blanco, Engelmann and Normann (2010)? While Fehr and Schmidt (1999) derive distributions for these parameters based on data from previous ultimatum-game experiments, here, we need the *joint* distribution of the parameters. We are not aware of any joint distribution of inequality-aversion parameters for the Fehr and Schmidt model—except, as mentioned above, Fehr, Kremhelmer and Schmidt (2008) make an assumption about

[6] In Blanco, Engelmann and Normann's (2010) modified dictator game, dictators chooses between 20-0 and equitable outcomes ranging from 0-0 to 20-20 (all denoted in £ (GBP)). A player i who is indifferent between payoff vectors (20, 0) and (x_i-x_i) has $\beta_i = 1 - x_i/20$.

Subject	α_i	β_i	Subject	α_i	β_i	Subject	α_i	β_i
1	0	0	22	0.409	0.175	42	0.929	0.8756
2	0	0.025	23	0.409	0.175	43	1.5	0.025
3	0	0.525	24	0.409	0.175	44	1.5	0.375
4	0	0.525	25	0.409	0.175	45	1.5	0.525
5	0	0.625	26	0.409	0.325	46	1.5	0.725
6	0	0.725	27	0.409	0.525	47	1.5	0.825
7	0	0.775	28	0.409	0.525	48	1.5	0.975
8	0	0.875	29	0.409	0.625	49	1.5	1
9	0	0.975	30	0.409	0.675	50	2.833	0.275
10	0.026	0	31	0.611	0.025	51	2.833	0.475
11	0.026	0	32	0.611	0.175	52	2.833	0.575
12	0.026	0.175	33	0.611	0.275	53	2.833	0.675
13	0.026	0.725	34	0.611	0.375	54	4.5	0
14	0.088	0.625	35	0.611	0.525	55	4.5	0
15	0.167	0.825	36	0.611	0.575	56	4.5	0.025
16	0.269	0.475	37	0.611	0.675	57	4.5	0.425
17	0.269	0.525	38	0.611	0.725	58	4.5	0.525
18	0.269	0.775	39	0.611	0.725	59	4.5	0.625
19	0.269	1	40	0.929	0.475	60	4.5	0.775
20	0.409	0	41	0.929	0.025	61	4.5	0.875
21	0.409	0.125						

Table 2: Blanco et al.'s (2010) joint α and β distribution

the joint distribution; namely that there are 60% players with $\alpha = \beta = 0$ and 40% fair types with $\alpha = 2$ and $\beta = 0.6$—which seems too coarse for our purposes. Following Blanco, Engelmann and Normann (2010), several papers have used the same (or very similar) games to elicit the Fehr and Schmidt parameters. See ... distribution has been successfully replicated in ...

Finally, the use of this joint distribution seems promising as it successfully predicts outcomes in several games (ultimatum game, sequential-move prisoner's dilemma, public-good game) which have a similar complexity as the present game (see Blanco, Engelmann and Normann 2010).[7] .

[7] Several papers () elicit inequality parameters following the method proposed in Blanco, Engelmann and Normann (2010) to make predictions at an individual level for decisions in other experimental decisions. Note, however, that the main conclusion of Blanco, Engelmann and Normann (2010) is that the distribution in Table 2 does not predict well across games at the individual level. Here, we want to propose using the existing distribution to make predictions for new games.

Assumption 2. *Players' inequality parameters are drawn from the joint α-β distribution in Table 2. This distribution is common knowledge. Players know their own type but not the type of the other player.*

Sequential moves, one threshold

We start with the sequential-move variant with one threshold (SEQ_1). In this treatment, a second-mover (S) with standard preferences will best respond to the first-mover's (F) contribution, c_F, by choosing zero if $c_F < 2$ and by contributing $12 - c_F$ if $c_F \geq 2$. Anticipating this, the first-mover will choose her payoff-maximizing contribution, which is $c_F = 2$.

Next consider players whose preferences and beliefs are consistent with Assumptions 1 and 2. Even if $c_F \geq 2$, second-movers with F&S preferences might choose $c_S = 0$ if the payoff inequality implied by c_F becomes too big. For $c_F \in [2, 6]$ and facing the decision between contributing $12 - c_F$ and $c_F = 0$, the second-mover either obtains $U_S(12 - c_F, c_F) = 8 + c_F - \alpha_i(12 - 2c_F)$ or $U_S(0, c_F) = 10 - \beta_i c_F$. We find that $U_S(12 - c_F, c_F) > U_S(0, c_F)$ iff

$$c_F \geq \frac{2(1 + 6\alpha)}{1 + 2\alpha + \beta} \equiv \tilde{c}_F. \tag{3}$$

The $\tilde{c}_F$ in (3) is the *minimum acceptable first-mover contribution* for a given set of individual inequality parameters. Any contribution as least as high as $\tilde{c}_F$ will be met by $c_S = 12 - c_F$ and will result in the public good being provided. Any contribution lower than this threshold will face $c_S = 0$ as the second mover's best reply. Intuitively, $\tilde{c}_F$ is increasing in α and decreasing in β.

Based on our Assumption 2 (the joint distribution of the α and β parameters observed in Blanco, Engelmann and Normann, 2010), we now predict the frequencies of public-good provision as a function of c_F. For each player in that data set (see Table 2), we determine the $\tilde{c}_F$ as in (3). For subject #1 with $\alpha = \beta = 0$, for example, we obtain $\tilde{c}_F = 2$ as the minimum acceptable first-mover contribution, whereas subject #58 with $\alpha = 4.5$ and $\beta = 0.525$ has $\tilde{c}_F = 5.32$ as the minimum acceptable first-mover contribution and will thus only accept $c_F = 6$. Doing this for all Blanco, Engelmann and Normann (2010) subjects allows us to predict how many players in *our* experiment will (not) provide the public as a function of c_F.

Table 3 shows the results of this calibration. In contrast to the game of players with standard preferences, the likelihood of public-good provision is strictly below 100 percent as long as $c_F < 6$. Table 3 also reveals that the expected monetary payoff of a risk-neutral first-mover monotonically increases in c_F and is maximized for $c_F = 6$.[8] First movers with a F&S utility function will a fortiori choose $c_F = 6$ as it minimizes the payoff inequality in addition.

[8]The expected payoff from choosing $c_F = 0$ is 10; $c_F > 6$ results in a lower likelihood of public-good provision, lower payoffs, and greater payoff inequality. Thus a selfish first-mover will choose $c_F = 6$.

Second-mover contribution	First-mover contribution $c_{FM} = 2$	$c_{FM} = 3$	$c_{FM} = 4$	$c_{FM} = 5$	$c_{FM} = 6$
$c_{SM} = 12 - c_{FM}$ (PG level 1 provided)	21.3%	37.7%	67.2%	83.6%	100%
$c_{SM} = 0$ (PG not provided)	78.7%	62.3%	32.8%	16.4%	0%
expected first-mover payoff	10.13	10.77	12.72	13.36	14.00

Table 3: Predicted second-mover responses conditional on first-mover choices and the resulting expected first-mover monetary payoff in SEQ_1 and SEQ_2

Thus we have

Proposition 1. *For treatment* SEQ_1, *the standard model predicts* $c_S = 0$ *if* $c_F < 2$, $c_S = 12 - c_F$ *if* $c_F \geq 2$, *and* $c_F = 2$. *The calibrated F&S model predicts the frequencies of second-mover responses as in Table 3, and* $c_{FM} = 6$ *for the first-movers.*

Sequential moves, two thresholds

Next, consider the *sequential-move variant with two thresholds* (SEQ_2). If the first-mover contributes $c_F \leq 6$, the analysis is as above. But in the two-level game, the first-mover may also choose her contribution in the range $c_F \in [8, 10]$ in order to make the second level feasible.

Players with standard preferences will not provide the public good at the second level in equilibrium. Given $c_F \in [8, 10]$, second-movers will respond with $c_S = 12 - c_F$ (yielding a monetary payoff of $8 + c_F$) but not with $c_S = 18 - c_F$ (which would yield only $7 + c_F$). By backward induction, first-movers will not choose $c_F \in [8, 10]$ but $c_F = 2$, as in the game with one step-level. The second threshold is irrelevant with standard preferences.

Now assume F&S players and begin with the second-movers. With $c_F \in [8, 10]$, the second-mover may chose $c_{SM} = 18 - c_{FM}$, $c_{SM} = 12 - c_{FM}$, or $c_{SM} = 0$. Since $U_S(12 - c_F, 0) > U_S(0, c_F)$ for $c_F \in [8, 10]$, we can restrict the second-mover choices to $c_{SM} = 18 - c_{FM}$ and $c_{SM} = 12 - c_{FM}$. First suppose $c_F = 8$. If the second-mover chooses $c_S = 18 - c_F$, we have $\chi_2 = 1$ and thus $U_F(10, 8) = 15 - 2\alpha_i$. If she chooses $c_S = 12 - c_F$, we have $\chi_1 = 1$ and $U_F(4, 8) = 16 - 4\beta_i$. We

Second-mover contribution	First-mover contribution $c_{FM} = 8$	$c_{FM} = 9$	$c_{FM} = 10$
$c_{SM} = 18 - c_{FM}$ (PG level 2 provided)	39.3%	80.3%	80.3%
$c_{SM} = 12 - c_{FM}$ (PG level 1 provided)	60.7%	19.7%	19.7%
$c_{SM} = 0$ (PG not provided)	0.0%	0.0%	0.0%
expected first-mover payoff	13.97	15.02	14.02

Table 4: Predicted second-mover responses conditional on first-mover choices between 8 and 10 and expected first-mover monetary payoff in Seq_2

obtain $U_S(10,8) < U_S(4,8)$ iff $1 - 4\beta + 2\alpha > 0$. This condition holds for for 60.7 percent of the subjects in Blanco, Engelmann and Normann (2010). That is, if $c_F = 8$, the public good will be provided at level one with 60.7 percent probability and with 39.3 percent probability at level two. Then consider $c_F = 9$. If $c_S = 18 - c_F$, we obtain $U_F(9,9) = 16$, whereas for $c_S = 12 - c_F$ we get $U_F(4,8) = 17 - 6\beta_i$. We find that $16 < 17 - 6\beta_i$ iff $1 - 6\beta > 0$. In the data of Blanco, Engelmann and Normann (2010), 19.7 percent of the subjects meet this condition. That is, if $c_F = 9$, the public good will be provided at level one (two) with 19.7 (80.3) percent probability. Finally, the case $c_F = 10$ turns out to be identical regarding the second-movers' incentives. That is, $c_F = 9$ and $c_F = 10$ are equally likely to be "exploited" by the second-mover, and the predicted frequencies of public good provision are hence the same. Table 4 summarizes the additional predictions in Seq_2.

Consider next the first-movers. $c_{FM} = 10$ will never be chosen by first-movers because $c_F = 9$ triggers the same second-mover response as $c_F = 10$ (in terms of public good provision) but $c_F = 9$ yields a higher expected payoff and higher F&S utility than $c_F = 10$. As for the choice between $c_F = 8$ or $c_F = 9$, we find that $c_F = 8$ yields a lower expected monetary payoff than $c_F = 6$ (see Table 4) and accordingly an even lower F&S utility. Hence, a risk neutral first-mover will never choose $c_F = 8$. The remaining possibilities are that first-movers will either choose $c_F = 6$ or $c_F = 9$. Contributing $c_F = 6$ yields an expected utility of 14 and $c_F = 9$ gives an expected utility of $15.015 - 1.182\alpha$. Now $15.015 - 1.182\alpha > 14$ iff $\alpha < 0.859$. This is predicted to hold for 36 percent of the Blanco, Engelmann and Normann (2010) subjects.

Proposition 2. *For treatment* SEQ_2, *the standard model makes the same predictions as for* SEQ_1. *The calibrated F&S model predicts the frequencies of second-mover responses as in Tables 1 and 2, and that 64% of all first-movers choose* $c_{FM} = 6$ *and 36% choose* $c_{FM} = 9$.

Taking second- and first-mover predictions together, we finally derive the prediction for the frequencies of public-good provision. We expect the public good to be provided at step-level 1 with a frequency of $0.64 + 0.36 \cdot 0.197 = 0.711$ and at step-level 2 in the rest of the cases.

Simultaneous moves, one threshold

With simultaneous moves, there are multiple equilibria both in the standard model and in the F&S model. With standard preferences, both players contributing nothing and all allocations where $c_1+c_2 = 12$ are the pure-strategy equilibria.[9] Perhaps somewhat surprisingly, all of these equilibria are also Nash equilibria with calibrated F&S preferences except for those where $(c_1 = 2, c_2 = 10)$ and $(c_1 = 10, c_2 = 2)$. (Proof available upon request.)

We believe that it is unlikely that entirely symmetric players will coordinate on asymmetric equilibria and we therefore focus on symmetric equilibria. The two pure-strategy equilibria are $c_i = c_j = 0$ and $c_i = c_j = 6$, and the symmetric mixed-strategy equilibrium has both players contribute $c_i = 0$ with 40 percent probability and $c_i = 6$ otherwise with standard preferences.

With the calibrated F&S model, the symmetric pure strategy (Bayesian-Nash) equilibria $c_i = c_j = 0$ and $c_i = c_j = 6$ are the same but the best response correspondence changes both quantitatively and qualitatively. First of all, note that we can "purify" the mixed-strategy equilibrium (Harsanyi, 1973) as we have a population of 58 different types of players in the Blanco, Engelmann and Normann (2010) data.[10] We will analyze the mixed equilibrium such that each of these players chooses a pure strategy. From Assumption 2, players know the distribution of types and thus they also know how many of the other players will play which strategy in equilibrium. In the (Bayesian-Nash) mixed-strategy equilibrium with calibrated F&S utilities, 64 percent of the players contribute $c_i = 6$ whereas 36 percent choose $c_i = 0$. Hence, more types contribute $c_i = 6$ with F&S preferences in the mixed-strategy equilibrium.

There is, however, also a qualitative difference to the standard case. With standard preferences, all players have the same best reply: if less than 60 percent of the players are expected to contribute, nobody will contribute (and vice versa if more than 60 percent contribute). With the calibrated F&S model, it is not the case that all players have the same best response. If there are less than 64 percent players expected to contribute $c_i = 6$, some players will still contribute. Learning will

[9] There are also numerous mixed-strategy equilibria.

[10] Among the 61 players reported in Table 2, three types occur twice so that there are 58 types in total.

be slower and the shape of the best response correspondence differs from the standard case. We discuss this in detail below.

Proposition 3. *In treatment* SIM_1, *the symmetric equilibria are* $c_i = c_j = 0$ *and* $c_i = c_j = 6$. *In the symmetric mixed-strategy equilibrium 60 percent of the players chooses* $c_j = 6$; *and 64 percent in the case of F&S preferences.*

Simultaneous moves, two thresholds

We turn to the variant with *simultaneous-move game with two thresholds* (SIM_2). As argued above for SEQ_2, meeting the second threshold is not a Nash equilibrium with standard preferences. As the equilibria derived above for SIM_1 are unaffected by the introduction of the second threshold; with standard preferences, SIM_2 has the same Nash equilibria as SIM_1.

We now look for a Bayesian Nash equilibrium of players with F&S utilities where the second level of the public good is provided. A Bayesian Nash equilibrium is a combination of type-dependent strategies such that every player maximizes her expected utility, given her F&S type and the strategies of all other players. Suppose that some types choose $c = 9$. Above, we have seen that, given $c_i = 9$, 80.3 percent of all types will reply with $c_j = 9$ whereas the rest plays $c_j = 3$. Hence, there cannot be a Bayesian Nash equilibrium where all types choose $c_i = 9$. We will therefore look for a Bayesian Nash equilibrium where p percent of all F&S types choose $c_i = 9$ whereas $1 - p$ choose $c_i = 3$.

The expected utility from playing $c = 9$ is $pU(9,9) + (1-p)U(9,3) = 16p + (1-p)(11-6\alpha)$, and the expected utility from playing $c = 3$ is $pU(3,9) + (1-p)U(3,3) = p(17-6\beta) + (1-p)7$. Contributing 9 yields a higher expected F&S utility than contributing 3 iff

$$p > \frac{6\alpha - 4}{6\alpha + 6\beta - 5}.$$

For F&S players with $\alpha = \beta = 0$, this condition is never met (as seen above); that is, selfish own utility maximizers will always choose $c = 3$. If p is sufficiently large, however, inequality averse players prefer $c = 9$. In the Blanco, Engelmann and Normann (2010) data, we find that for $p = 0.72$ exactly 72 percent of the players have $pU(9,9) + (1-p)U(9,3) > pU(3,9) + (1-p)U(3,3)$ whereas for 28 percent the inequality is reversed. Thus these strategies constitute a Bayesian Nash equilibrium. It remains to check, though, whether it pays to deviate to any contribution other than 9 or 3. The only possible deviation is to contribute $c = 0$ since any other contribution dominated either by $c = 0$ or $c = 3$. Contributing $c = 0$ yields an expected F&S utility of $10 - 3\beta - 0.72 \cdot 6\beta$.

But the equilibrium action $c = 3$ yields $0.72(17 - 6\beta) + (0.28)7$ which is strictly larger for all $\beta \in [0, 1]$. Thus we have established

Proposition 4. *The Bayesian Nash equilibria of* Sim_1 *are also equilibria in treatment* Sim_2. *With standard preferences, there are no additional equilibria. With the calibrated F&S preferences, 72 percent of the F&S types choosing* $c = 9$ *and the rest* $c = 3$ *is a Bayesian Nash equilibrium.*

Hypotheses

Based on Propositions 1 to 4, we will now derive two hypotheses regarding the impact of our two treatment variables. Will will return to the propositions and the performance of the F&S model below.

Comparing the predicted public-good provision in Sim vs. Seq, we note that there are multiple equilibria in the Sim treatments and that the public good is not provided in all equilibria. By contrast, in the Seq treatments, the equilibrium is unique and the public good is provided (at least at level one) in the unique equilibrium. This holds for both the one and the two-threshold case.

Hypothesis 1. *The public good will be provided more frequently in the* Seq *treatments compared to* Sim.

Note that this hypothesis does not depend on assuming that players have F&S preferences. The F&S model makes the point that the $c_F = 2$ predicting with standard preferences in the Seq treatments will be punished regularly by the second-movers, but it also predicts that this case will not arise because first-movers anticipate this. Our second hypothesis does depend on assuming F&S preferences.

Propositions 1 to 4 show that public-good provision can be improved if there is the second threshold. In the Seq treatments, the public good will always be provided at level one but in 29% of the cases also at step-level two. The case for improved public good provision in the Sim treatments is as follows. There are multiple equilibria in the Sim treatments anyway but there exist an equilibrium in which the second level is met with positive probability. For both Seq_2 and Sim_2, we note that even if one player attempts to reach the second level by choosing e.g. $c_i = 9$ and the other player exploits this with $c_j = 3$, this does not harm payoffs as the first level of the public good is still provided. Thus the two-level treatments should be weakly superior.

Hypothesis 2. *The public good will be provided more frequently in the treatments with two threshold compared to one-threshold treatments.*

5 Main Treatment Effects

We present our results in two parts. Section 5 presents tests of Hypotheses 1 and 2. In addition to public good provision, we will also analyze contributions and payoffs (or efficiency). The next section presents a more detailed analysis of the predictive power of the calibrated F&S model.

When we apply regressions analysis, we use Generalized Linear Latent and Mixed Models (gllamm; see Rabe-Hesketh and Skrondal, 2005) regressions, taking possible dependence of observations at the level of a (randomly matched) group and at the individual level into account. As dependent variables we use *sequential* (a dummy which is equal to one if the move order is sequential), *twolevel* (a dummy which is equal to one if there are two levels), *seq2* (an interaction term for the sequential treatment with two levels), furthermore we control for *period* and the *sessionsize*.

We typically report three regressions. Regression (1) reports the impact of the treatment variables *sequential* and *twolevel* only. Regressions (2) includes the interaction *seq2*, and (3) adds *period* and *sessionsize*.

Overview

We start with a summary statistics of our four treatments in Table 5. It shows public good provision contributions, frequency of coordination, and the resulting payoffs. Note in our treatments with two threshold levels we also count the cases where the second level has been achieved as successful provision of *PG level 1*.

As can be seen, public good provision at the first level is most effective in the treatments with sequential-move order. *PG level 1* is provided most frequently (85.56%) with the sequential-move order and two thresholds and thus *PG level 1* provision is also more effective in SEQ_2 compared to SEQ_1 where only 75.24% subjects manage to provide the public good. Only in 6% of SIM_2'S cases is the public good provided at the second threshold level. However, the second threshold level does come out better with sequential-move order (16.67% of PG level 2 in SEQ_2). The second threshold level leads to higher contributions in the simultaneous as well as in the sequential treatment. We define successful coordination as cases without wasteful contributions (that is cases where $c_1 + c_2 \in \{0, 12\}$, or $c_1 + c_2 \in \{0, 12, 18\}$ in the two step-level cases). Coordination is best in the environment of sequential moves. Furthermore the sequential-move order also leads to higher payoffs compared to the simultaneous treatments.

A first look at the data in Table 5 thus suggest that we do find tentative support for Hypothesis 1. Regarding Hypothesis 2, the effect is ambiguous as ambiguous since the second level improves public-good provision (at level one) in the SEQ treatments but not in the SIM settings.

	Treatment			
Variable	SIM_1	SIM_2	SEQ_1	SEQ_2
PG level 1 provided (χ_1=1) in %	64.29	59.00	75.24	85.56
PG level 2 provided (χ_2=1) in %	-	6.00	-	16.67
PG not provided (in %)	35.71	41.00	24.76	14.44
Contributions	5.22	5.99	4.96	6.07
	(2.23)	(2.88)	(2.36)	(2.57)
Successful coordination (in %)	49.05	17.00	77.62	81.11
Payoff	11.21	10.30	12.56	13.32
	(3.86)	(4.27)	(2.92)	(3.18)

Table 5: Summary statistics of our four treatments. Note that the public good is provided at level 2 ($\chi_2 = 1$) only if it is also provided at level 1 ($\chi_1 = 1$).

Public-good provision

Table 6 presents gllamm probit regressions of the frequency of PG provision. The first probit regression (left panel) is about public-good provision at level 1. The dependent variable is equal to one if and only if the first threshold is met (that is, if and only if $\chi_1 = 1$). The second probit regression (right panel) has the dependent variable is equal to one if and only if the second level ($\chi_2 = 1$) is met. Note that the public good is provided at level 2 ($\chi_2 = 1$) only if it is also provided at level 1 ($\chi_1 = 1$).

The regressions in the left panel shows that *sequential* is significant. That is, the sequential-move order improves the PG provision at the first threshold. This is support for Hypothesis 1. The implementation of a second threshold does not lead to a higher frequency of public good provision. The same is true for the sequential treatment with two thresholds. That is, we do not find support for Hypothesis 2 which predicts that the second threshold leads to more public good provision.

In regression (3), we find that the coefficient of *sessionsize* is negative and weakly significant. That is, sessions with a higher numbers of subjects exhibit lower public-good provision. This is consistent with findings in Botelho et al. (2009). Botelho et al. (2009) compare repeated settings with “random strangers” and “perfect strangers” matching protocols and find that the assumption that subjects treat Random Strangers designs as if they were one-shot experiments is false. Our results indicate that the session size and hence the likelihood of meeting a random stranger once more has an impact on cooperation. We note, however, that the coefficient of *sessionsize* is small.

Public Good Provision 1 Level				Public Good Provision 2 Levels	
	(1)	(2)	(3)	(1)	(2)
Sequential	0.657***	0.390*	0.395**	0.638**	0.550**
	(0.184)	(0.229)	(0.200)	(0.300)	(0.277)
Twolevel	0.123	-0.137	-0.151		
	(0.183)	(0.227)	(0.199)		
Seq2		0.541	0.429		
		(0.330)	(0.297)		
Period			-0.00633		-0.0466*
			(0.0124)		(0.0241)
Sessionsize			-0.0643*		-0.0669
			(0.0346)		(0.0507)
Constant	0.281*	0.408**	1.342***	-1.773***	-0.630
	0.154	(0.160)	(0.509)	(0.230)	(0.716)
Obs.	1,600	1,600	1,600	760	760

*** p<0.01, ** p<0.05, * p<0.1

Table 6: Gllamm probit regressions of Public Good provision 1 and 2 Level(s)

Table 6 also presents a gllamm probit regression of the frequency of PG provision of level 2. (Here, *twolevel* cannot be part of the regression analysis, of course.) *sequential* is again significant, that is, sequential-move contributions also stimulate the provision of the second level which is additional support for Hypothesis 1. Regression (2) reveals that PG provision at level two moderately decreases over time. The dummy *sessionsize* is not significant here.

Contributions

We now analyze subjects contribution levels. The left panel of Table 7 reports a gllam linear regression of the players' contribution. Contributions are *not* significantly influenced by the order of moves. Interestingly, adding the second threshold leads to significantly higher contributions. The interaction of a sequential move order and two levels does not lead to further increased contributions. Over time, contributions get weakly smaller. *sessionsize* is significant, that is, in sessions with more participants contributions are slightly lower.

Contribution				Payoff		
	(1)	(2)	(3)	(1)	(2)	
Sequential	-0.00397	-0.126	-0.0959	2.140***	1.390***	1.389***
	(0.254)	(0.304)	(0.315)	(0.379)	(0.446)	(0.386)
Twolevel	1.202***	0.904*	0.684***	-0.0103	-0.867**	-0.898**
	(0.257)	(0.424)	(0.244)	(0.393)	(0.439)	(0.392)
Seq2		0.446	0.390		1.595***	1.404**
		(0.528)	(0.436)		0.618	(0.565)
Period			-0.0330*			-0.00250
			(0.0175)			(0.0310)
Sessionsize			-0.135***			-0.115*
			(0.0434)			(0.0671)
Constant	5.039***	5.076***	7.151***	10.84***	11.21***	12.83***
	(0.169)	(0.174)	(0.639)	(0.254)	(0.323)	(0.99)
Obs.	1,600	1,600	1,600	1,600	1,600	

*** p<0.01, ** p<0.05, * p<0.1

Table 7: Gllamm linear regressions of Contribution and Payoff

Payoffs

In Table 7 (right panel), we report the results of a linear regression on subjects' payoffs. First, Table 7 shows that the sequential contribution mechanism significantly improves subjects' payoffs. This is due to the fact that public-good provision is improved by the sequential-move order. The second step-level significantly reduces the payoffs. This can be explained by the fact that, on the one hand, two thresholds increase contributions but, on the other hand, the second level is rarely actually achieved. When we add the interaction *seq2*, we find that it significantly boosts subjects' payoff by 1.4 compared to the baseline SIM_1. The difference between SEQ_1 and SEQ_2 is, however, not significant as follows from a Wald test ($p = 0.22$). This emphasizes the overall negative impact of the second threshold on payoffs. Indeed, payoffs are worst in SIM_2. Furthermore the size of the sessions is weakly significant, but again the coefficient is small. The time trend is insignificant here.

The payoff variable is the variable a social planner would ultimately be interested in. Payoffs reflect the combined effect of contributions, public-good provision and avoiding excess contributions (coordination). The above regression confirms that the payoff differences reported in our summary statistics are significant. Specifically, it follows that SEQ_2 has the highest payoffs, followed by SEQ_1 and SIM_1, and SIM_2 has the lowest payoffs.

Coordination Rates

Figure 1 compares coordination in the simultaneous and the sequential treatment with one threshold. As we are interested in the sum of contributions, we define $C = c_1 + c_2$. In SEQ1, $C = 12$ is the most frequent outcome (contribution sum). That is, subjects coordinate on $C = 12$ in 74.29% of all cases. By contrast, the simultaneous contribution mechanism only guarantees efficient contributions in 46.67% of the time. The difficulty of coordinating in SIM1 is also documented by the aggregation of the cases where the contribution sum is either too low ($0 < C < 12$) or too high ($C > 12$). As for the sum of these inefficient cases, we find that in SIM1, 50% of the subjects do not manage to contribute efficiently. The remaining cases are those where $C = 0$, which is efficient in that there are no contributions wasted. In SEQ1, there are only 22.34% inefficient cases. Mainly, these involve second movers punishing low first-mover contributions. Figure 1 documents that the increased payoff in SEQ1 is mainly due to the fact that subjects' less contributions are wasted in the sequential environment.

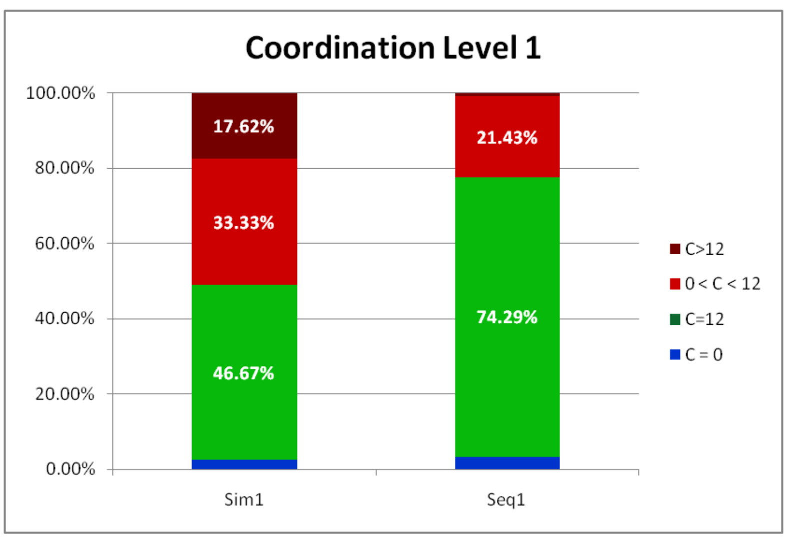

Figure 1: Frequency of the contribution sums (C) of both players in the simultaneous and sequential step-level public good game with one threshold.

Figure 2 compares coordination in the simultaneous and sequential treatment with one threshold. This plot again documents the superiority of the sequential- over the simultaneous-move variant. In Seq2, about 80% of all contribution sums are efficient. That is, subjects manage to exactly coordinate on the first threshold ($C = 12$) or on the second threshold ($C = 18$) without generating wasteful excess contributions.

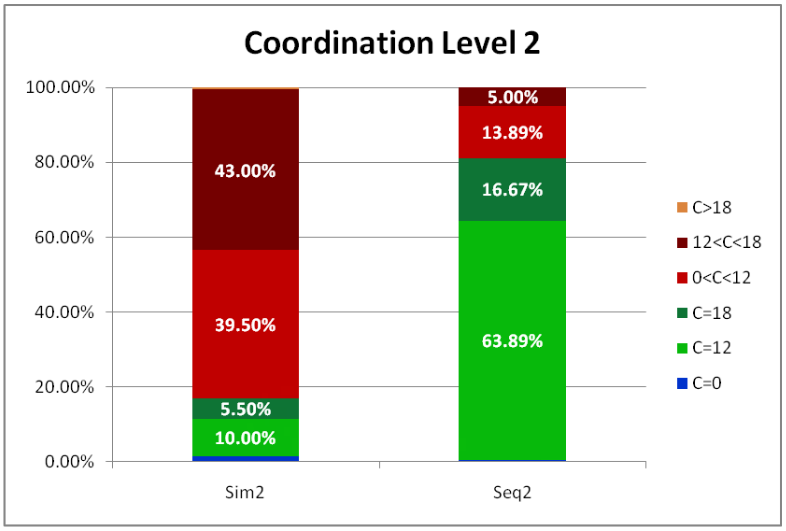

Figure 2: Frequency of the contribution sums (C) of both players in the simultaneous and sequential step-level public good game with two threshold.

This stands in strong contrast to the efficient cases in Sim2 where only 15.5% of the contribution sums are efficient. In Sim2, subjects seem to face big difficulties in terms of coordination. This leads to a high amount of wasteful contribution sums (82.50% miscoordination). Figure 2 therefore serves as an explanation of the fact that the second level leads to smaller payoffs. Especially in Sim2 the second level leads to costly miscoordination of the players. However, two levels are efficient in the environment with sequential moves which expalins the significance of our interaction term *Seq2*.

6 The predictive power of the calibrated F&S model

We now discuss the quantitative predictions of the F&S model in more detail. While our Hypothesis 2 was based on the F&S model and largely failed, we will see that some of the model's predictions materialize rather well.

We begin with Proposition 1. Figure **??** contrasts the predictions made in Table 3 to the observations of the frequency of second-movers who contribute $c_{SM} = 12 - c_{FM}$ in reply to first-mover contributions. The data underlying Figure **??** pools the c_{FM} in both sequential treatments Seq_1 and Seq_2.[11] Using one-sample chi-square tests, we cannot reject that predicted and

[11] This is warranted because, firstly, the F&S model does not predict any differences and, secondly, we do not

observed frequencies are the the same (all $\chi_(1) < 2.38$, $p > 0.123$). The F&S model predicts the second-mover responses amazingly well.

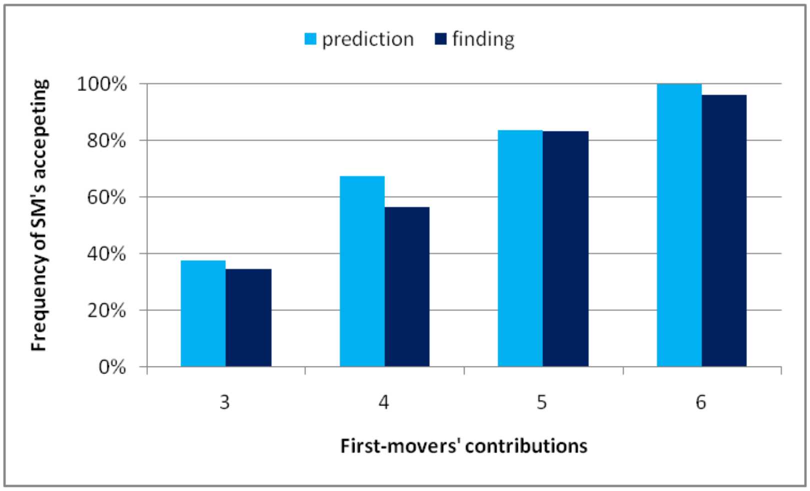

Figure 3: Frequency of the contribution sum ($c_{FM} + c_{SM} = 12$) of both players in the sequential step-level public good game in Seq_1 and Seq_2.

Proposition 1 also states that first-movers should choose $c_F = 6$ in order to maximize payoffs and F&S utilities. This is not the case as $c_F = 6$ is chosen only in 37.1 percent of the cases. While this rejects the F&S prediction, we note that similar observations have been made before. For standard ultimatum-game experiments, it can be argued that offering the equal split may be payoff maximizing (assuming risk neutrality), but about half the the proposers offer less than the equal split.[12] Huck, Müller and Normann (2001) show that, in quantity-setting duopoly, Stackelberg followers are inequality averse but the Stackelberg leaders still choose too high an output (to be precise, the Stackelberg leader output is below the standard prediction but above what would be payoff maximizing). Risk-loving behavior can explain the first-mover behavior.

Figure 4 is a bubble plot of first- and second-movers in Seq_1. The modal outcome is (6, 6) as predicted, and many observations are on the Pareto frontier where $c_F + c_S = 12$. One can identify the "punishing" second-movers on the vertical axis where $c_S = 0$. For the first-movers in Seq_2, Proposition 2 predicts that 36 percent contribute $c_F = 9$ and 64 percent should choose $c_F = 6$. In our data, 36.7 percent of the first-movers choose 9—which seems a remarkable confirmation of the

observe differences—with the exception of $c_F = 5$ where contribution rates () differ significantly (two-sample chi-square test, $\chi_(1) = 8.579$, $p < 0.01$). Importantly, the differences we observe are not systematic. Contributions of $c_{SM} = 12 - c_{FM}$ are more frequent for $c_{FM} = 3, 4$ in Seq_1 than in Seq_2 but the other way round for $c_{FM} = 5, 6$. Note that, for $c_F = 6$, we cannot apply a test because the predicted frequency is 100 percent. Regarding $c_F = 2$, we only have two observations so we cannot test either (in one case the PG was provided).

[12]In Blanco, Engelmann and Normann (2010), offering the equal split is actually (expected) payoff maximizing, but their ultimatum game was done with the strategy method which typically induces higher rejection rates.

prediction. The remaining 63.3 percent choose $c_F \in [2, 6]$. While we do not find that 64 percent choose $c_F = 6$, this only restates the previous finding that first-movers do not always choose the risk-neutral payoff maximizing action.

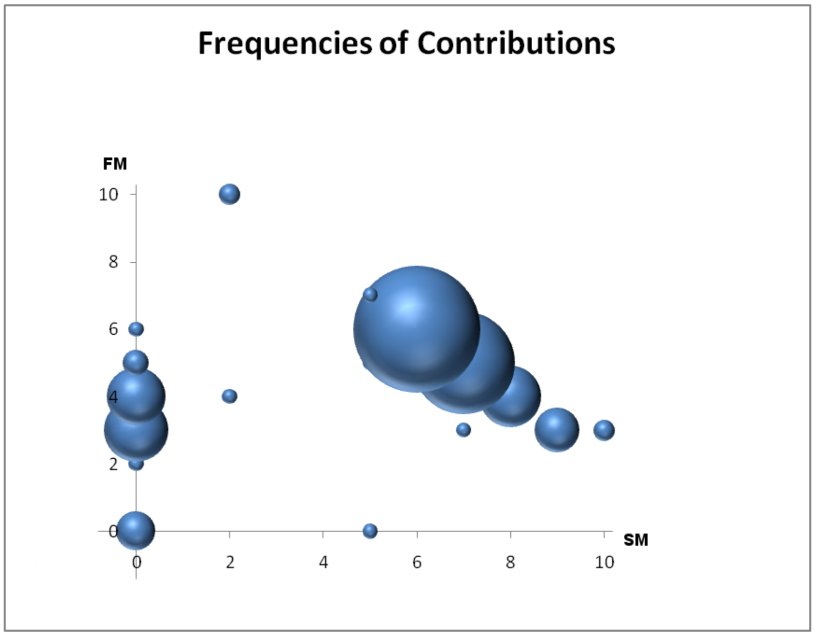

Figure 4: Predicted frequencies (based on the calibrated F&S model) and observed frequencies of second-movers contributing such that the PG at level 1 is provided.

Intriguingly, the second-mover prediction of Proposition 2 fails (whereas it was the first-mover prediction of Proposition 1 that failed). The first-mover in the two-level case is in a trust-game like situation. If she chooses $c_F = 9$, she can be exploited by second-movers. While the calibrated F&S model predicts that more than 80.3 percent of the second-movers will be trustworthy, it turns out only 50.9 are. Predicted and observed share differ significantly (binomial test, $p < 0.05$). The failure of the theory seems surprising since the cost of being trustworthy are low here: second-movers gain only one additional euro by exploiting the first-mover, but this costs the first-mover five euros.

We finally turn to Proposition 3, the SIM_1 case. In SIM_1, we observe that in 81.4 percent of the cases subjects choose $c \geq 6$ and in 13.8 of the cases they choose $c = 0$.[13] Hence, both the standard model and the calibrated F&S model would predict that play converges to the pure-strategy equilibrium where both players choose $c = 6$. This is, however, not the case. There is no positive time trend, and some players persistently choose $c = 0$. Why do subjects not best

[13] These percentages are are based on data from periods 6 to 10 where we observe less heterogeneity in the data. Some subjects indeed choose $c_i > 6$, but $0 < c_j < 6$ is never a best reply with standard or F&S preferences. Thus we focus on $c \geq 6$ and $c = 0$

respond?

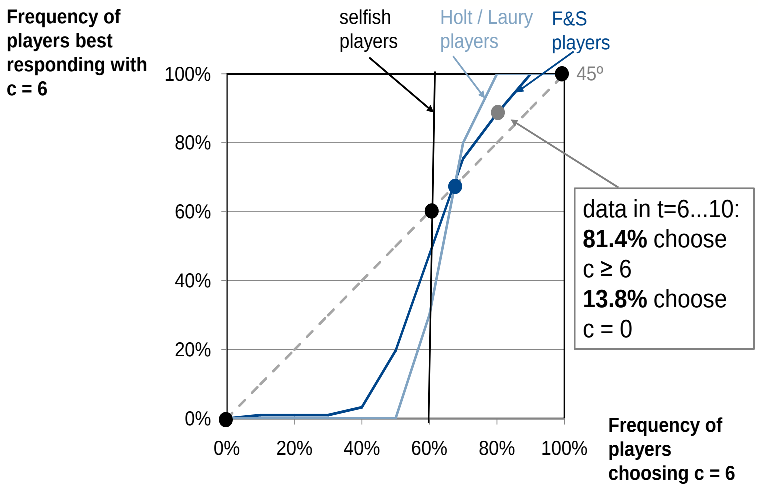

Figure 5: Best-reply correspondences for standard players, F&S players and Holt-Laury players in Sim_1.

Figure **??** illustrates what might be going on. It shows the best-reply correspondences for standard selfish players, for F&S players and also for players with standard preferences but with a degree of risk aversion according to the findings in Holt and Laury (2002). With selfish and rational player, the best reply correspondence has a "bang-bang" property. If the belief is that player j chooses $c_i = 6$ less than 60 percent, *all* players will best respond with $c_i = 0$, and vice versa for a belief of more than 60 percent. With the calibrated F&S model, this is not the case. For beliefs between (roughly) 40 and 80 percent, the best replies of the various F&S types differ. For example, given a belief that 70 percent of all players choose $c_i = 6$, only 75 percent of the players will best respond with $c_i = 6$ where 25 percent still choose $c_i = 0$.

As mentioned in Proposition 3, the share of players choosing $c_F = 6$ required such that $c_F = 6$ is a best reply is slightly larger with F&S players. Inequality aversion has an effect similar, in fact a stronger effect, than risk aversion (on average, players in Holt and Laury are slightly risk averse). We also see that the best replies differ from the case with standard preferences. Around the mixed-strategy equilibrium, the best replies are not vertical but somewhat "flat", implying that not all players will best reply once the fixed point of the mixed strategy is exceeded. This is what we see in the data.

7 Conclusion

We analyze the provision of a step-level public good in an experiment. Specifically, we investigate how the order of moves (simultaneous vs. sequential) and the introduction of a second step-level (which is not feasible in standard Nash equilibrium) affects public-good provision. We find that the sequential-move game yields more frequent provision of the public good and higher payoffs. An additional step-level does lead to higher contributions but the effect on public-good provision is ambiguous and it lowers payoffs.

Based on the existing experimental data of Blanco, Engelmann and Normann (2010), we fully calibrate Fehr and Schmidt's (1999) model of inequality aversion to make ex ante predictions. We find that actual behavior fits quantitatively well with these predictions. Specifically, the F&S model predicts the second-mover responses amazingly well. While the predictive power on first-mover behavior is less impressive, similar findings have been observed before in other sequential games. The calibrated Fehr and Schmidt (1999) model also predicts behavior well in the sequential treatment with two step-levels, and in the simultaneous-move case with one level.

References

[1] Andreoni, J. (1998) "Toward a Theory of Charitable Fund-Raising", The Journal of Political Economy 106, 1186-1213.

[2] Andreoni, J., Brown, P., Vesterlund, L. (2002) "What makes an allocation fair? Some experimental evidence." Games and Economic Behavior 40, 1-24.

[3] Berg, J., Dickhaut, J. and McCabe, K. (1995). "Trust, Reciprocity and Social History", Games and Economic Behavior 10, 122-142.

[4] Blanco, M., Engelmann, D., and Normann, H., T. (2010). "A within-subject analysis of other-regarding preferences", Games and Economic Behavior forthcoming.

[5] Botelho A., Harrison, G., W., Costa Pinto, L., M., and Rutström, E., E. (2009). "Testing static game theory with dynamic experiments: A case study of public goods", Games and Economic Behavior 67, 253-265

[6] Cadsby, C., B., Maynes, E. (1998a) "Gender and free riding in a threshold public goods game: Experimental evidence", Journal of Economic Behavior and Organization 34, 603-620.

[7] 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 and Organization 37, 183-192.

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

[9] Coats, J., C., Neilson, W., S. (2005) "Beliefs About Other-Regarding Preferences in a Sequential Public Goods Game", Economic Inquiry.

[10] Coats, J., C., Gronberg, T., J., Grosskopf, B. (2009) "Simultaneous versus sequential public good provision and the role of refunds Ů An experimental study", Journal of Public Economics 93, 326-335.

[11] Coller M., Chewning E., and Laury, S. (2001). "Voluntary Contributions to a Multiple Threshold Public Good", Research in Experimental Economics 8, 47-83.

[12] Croson, R. and Marks, M. (1998). "Identifiability of Individual Contributions in a Threshold Public Goods Experiment", *Journal of Mathematical Psychology* 42, 167-190.

[13] Erev, I., Rapoport, A. (1990) "Provision of Step-Level Public Goods : The Sequential Contribution Mechanism", The Journal of Conflict Resolution 34, 401-425.

[14] Fehr, E., Gächter, S. "Cooperation and Punishment in Public Goods Experiments". The American Economic Review 90, 980-994.

[15] Fehr, E., Kremhelmer, S. and Schmidt, K. (2008). "Fairness and the Optimal Allocation of Ownership Rights". Economic Journal 118, 1262-84.

[16] Fehr, E., Schmidt, K. (1999). "A Theory of Fairness, Competition and Co-operation", Quarterly Journal of Economics 114, 817-868.

[17] Figuières, C., Masclet, D., and Willinger, M. (2010), "Vanishing leadership and declining reciprocity in a sequential contribution experiment", Economic Inquiry forthcoming.

[18] Fischbacher, U. (1999). "z-Tree: Zurich Toolbox for Readymade Economic Experiments - Experimenter's Manual", mimeo, Institute for Empirical Research in Economics, University of Zurich.

[19] Gächter, S., Renner, E. (2005), "Leading by Example in the Presence of Free-Riders", mimeo, Nottingham University.

[20] Gächter, S., Nosenzo, D., Renner, E., and Sefton, M. (2010a), "Sequential vs. simultaneous contributions to public goods: Experimental evidence", Journal of Public Economics 94, 515-522.

[21] Gächter, S., Nosenzo, D., Renner, E., and Sefton, M. (2010b), "Who makes a Good Leader? Cooperativeness, Optimism, and Leading-by-Example", Economic Inquiry.

[22] Güth, W., Levati, M., V., Sutter, M., and van der Heijden, E. (2007) "Leading by example with and without exclusion power in voluntary contribution experiments", Journal of Public Economics 91, 1023-1042.

[23] Harsanyi, J.C. (1973). "Games with randomly disturbed payoffs: a new rationale for mixed-strategy equilibrium points", International Journal of Game Theory 2, 1-23.

[24] Hashim, M.J., Maximiano, S., and Kannan, K.N. (2011). "Information Targeting and Coordination: An Experimental Study", mimeo, Purdue University.

[25] Hermalin, B., E. (1998) "Toward an Economic Theory of Leadership: Leading by Example", The American Economic Review 88, 1188-1206.

[26] Holt, C., Laury, S., (2002) "Risk Aversion and Incentive Effects", The American Economic Review 92, 1644-1655.

[27] Huck, S., Müller, and Normann, H., T. (2001), "Stackelberg Beats Cournot: On Collusion and Efficiency in Experimental Markets", The Economic Journal 111, 749-765.

[28] Isaac, R., M., Walker, J. (1988) "Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism", The Quarterly Journal of Economics 103, 179-199.

[29] Potters, J., Sefton, M., and Vesterlund, L. (2005) "After You - Endogenous Sequencing in Voluntary Contribution Games", Journal of Public Economics 89, 1399-1419.

[30] Potters, J., Sefton, M., and Vesterlund, L. (2007) "Leading-by-example and signaling in voluntary contribution games: an experimental study", Economic Theory 33, 169-182.

[31] Rabe-Hesketh, Sophia, and Anders Skrondal (2005) "Multilevel and longitudinal modeling using Stata". College Station, TX: Stata Press.

[32] Schram, A., Offerman, T. and Sonnemans, J. (2008) "Explaining the comparative statistics in step-level public good games", in: C.R. Plott and V.L. Smith (eds.), The Handbook of Experimental Economics Results, Volume 1, Amsterdam: North-Holland

[33] Serra-Garcia, M., van Damme, E., Potters, J. (2010) "Which Words Bond? An Experiment on Signaling in a Public Good Game", CentER Working Paper 2010 33, Tilburg University.

[34] Versterlund, L. (2003) "The informational value of sequential fundraising ". Journal of Public Economics 87, 627-657.