Simultaneous and Sequential Contributions to Step-level Public Goods: One versus Two Provision Levels

Hans-Theo Normann¹ and Holger A. Rau²

Abstract
In a step-level public-good experiment, we investigate how the order of moves (simultaneous vs. sequential) and the number of step levels (one vs. two) 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 first movers who give less than half of the threshold contribution. The additional second step level—which is not feasible in standard Nash equilibrium—leads to higher contributions but does not improve public-good provision and lowers payoffs. We calibrate the parameters of Fehr and Schmidt’s model of inequality aversion to make quantitative predictions. We find that actual behavior fits remarkably well with several predictions in a quantitative sense.

Keywords
experimental economics, fund-raising, provision point public good, sequential play, threshold public good

¹Duesseldorf Institute for Competition Economics (DICE), Universitaet Duesseldorf, Duesseldorf, Germany
²University of Erlangen–Nuremberg, Nuremberg, Germany

Corresponding Author:
Hans-Theo Normann, Duesseldorf Institute for Competition Economics (DICE), Universitaet Duesseldorf, Duesseldorf, Germany.
Email: normann@dice.hhu.de
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. More generally, teamwork where the team has to meet a specific goal has step-level public-good character. Also, charities may have properties of step-level public goods if the underlying production of the public good is subject to nonconvexities (see Andreoni 1998).

Our article 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 standard Nash equilibrium and where the public good is provided at a higher level, improves public-good provision.

The issue of sequential versus simultaneous contributions is the subject of a growing literature. Following the theoretical works by Andreoni (1998, 2006) and Hermalin (1998), researchers have analyzed leading by example in experiments (Gächter et al., 2010b). 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 (Hermalin 1998) or about the quality of a charity Andreoni (2006).1

We study sequential versus simultaneous decisions in a step-level game with two players and with complete information. For such a setting, one would at first expect a sequential-move game to 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. In the unique subgame perfect equilibrium with selfish players, the first mover contributes such that a best responding follower merely breaks even by meeting the threshold with his or her contribution. In other words, the first mover actually gives a bad example by contributing less than the followers. In an experiment, this may reduce the 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.2 If such behavior occurs frequently, the higher efficacy of the sequential-move game will not materialize. Based on a calibrated model (see the later discussion), we hypothesize that the efficiency-enhancing effect dominates so that sequential moves improve public-good provision.

Now consider our second extension, the introduction of a second threshold. The general logic of multiple threshold public goods is that no return is obtained unless contributions meet the first level, and after this level, no additional return is earned until the second provision point is met. Multiple step levels have rarely been
analyzed before (Chewning, Coller, and Laury 2001; Hashim, Maximiano, and Kannan [2011], which we discuss later), but they seem realistic in many circumstances. For example, the successful development of a new product in teamwork will typically require a minimum level of efforts of the team members. Adding a further feature or quality level of the product may be subject to an effort threshold just as developing the main product and, accordingly, the management may set two threshold levels. Further examples include a public radio or TV station that may transmit more than one program, corresponding to multiple thresholds. Public bridges or highways may be built with one, two, or more lanes. Finally, any kind of public good may be provided at various quality levels and the production of these quality levels may be subject to nonconvexities, suggesting multiple thresholds.

The interaction of the two thresholds and the order of moves can be hypothesized as follows. In our experiments, first movers in the sequential-move game may aim for the second threshold since this yields higher payoffs—provided the threshold is met. Since such first-mover contributions must be higher than those required to meet the first threshold, second movers do not feel exploited and therefore do not punish first movers by making zero contributions. However, the second threshold is not a Nash equilibrium with selfish players. 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 is not an equilibrium (with either simultaneous or sequential moves). However, when players have Fehr and Schmidt (1999) preferences, the second threshold is a Nash equilibrium—and meeting the second threshold is of course efficient. In any event, even if some second movers exploit those first movers 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 be 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 second movers regularly punish first movers who contribute too little. This is in contrast to Gächter et al. (2010a), who find the opposite result, however, in an entirely different setting (see the later discussion). 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 or higher payoffs.

Our article also makes quantitative predictions for our experiment employing a fully calibrated Fehr and Schmidt (1999) model. Whereas Fehr and Schmidt’s model of inequality aversion has been used frequently in the literature, the predictions are almost always of a qualitative nature (“if players are sufficiently inequality averse, abc is an equilibrium”). We calibrate Fehr and Schmidt’s model on a (joint) distribution of the inequality parameters, and we make exact quantitative predictions (“w percentage of the first movers will contribute x,” or “given a
first-mover contribution of \( y \), the public good will be provided in \( z \) percentage of the cases \(^3\). We find that the calibrated Fehr and Schmidt (1999) model not only makes remarkably accurate quantitative predictions, but it also fails in two cases. The calibrated Fehr and Schmidt model predicts second-mover behavior (given first-mover behavior) in the sequential variant extremely well. Specifically, it accurately predicts 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, only slightly more than one-third of them do so. First movers behave “too greedily,” as has been observed in previous experiments (e.g., Huck, Muller, and Normann 2001). The calibrated Fehr and Schmidt (1999) model also predicts well in 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, as they exploit first movers significantly more frequently than predicted.

**Literature Review**

There are two major strands of the literature pertinent to our article. The first literature is about simultaneous versus sequential order of moves in public-good games. The second literature concerns public-good experiments with step-level character in general and, specifically, the small literature on experiments 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 article provides an explanation of how seed money, 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, a first mover may plausibly give an example to followers who rationally mimic the first mover’s behavior.

An increasing experimental literature has been triggered by these theory contributions. Following Hermalin (1998), Potters, Sefton, and Vesterlund (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. Moxnes and Van der Heijden (2003) and Van der Heijden and Moxnes (2012) also highlight the importance of leaders in a public bad experiment and show that followers invest less in the public bad when leaders give a good example. Potters, Sefton, and Vesterlund (2007) report that the leading-by-example approach depends on whether there is incomplete...
information in the experiment. Levati, Sutter, and Van Der Heijden (2007) also find that incomplete information crowds out the effects of leadership. This explains why some experiments have not found sequential moves to be superior (Andreoni, Brown, and Vesterlund 2002), while Potters, Sefton, and Vesterlund (2005) did. Our experiments differ from those of Potters, Sefton, and Vesterlund (2005, 2007) in that we do not include information asymmetries, and we do not employ the standard VCM.4

Gächter et al. (2010a) is also related to our study. They experimentally study the effects of a simultaneous versus a sequential choice mode in a test of Varian’s (1994) VCM model. Varian models a public-good setting where sequential contributions are predicted to be lower than simultaneous contributions.5 In their experimental test of Varian’s model, Gächter et al. find that sequential contributions are indeed lower than simultaneous-move contributions, although the difference in aggregate contributions across the two move structures is not as large as predicted, in part because second movers punish first movers who free ride in the sequential variant. While this is in contrast to our results, note that one of the major differences in our approach is that Gächter et al. test the Varian 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 versus sequential moves in a step-level public-good game with discrete choices. In their experiments, at least three of the 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 that, with sequential-move choices, information about previous noncooperative choices only is more effective in public-good provision than information about previous cooperative choices. The main differences in our experiment are the discrete action space and the number of players (two in our case).

Schram, Offerman, and Sonnemans (2008) study a similar binary “all or nothing” setup and vary several relevant factors. They find that participants contribute significantly more when the value of the public good is increased. Increasing group size from five to seven players decreases the average contribution level, but the public good is provided more often when groups are large. Whether subjects play with random or fixed matching does not systematically affect behavior. Finally, subjects contribute significantly more in a “public-good” frame compared to a “public bad.” Coats, Gronberg, and Grosskopf (2009) analyze threshold public-good games with both simultaneous and sequential contributions and investigate how refund policies interact with the mechanism. Given either refund mechanism, efficiency is greater with sequential contributions. A full refund achieves higher efficiency but only for simultaneous contributions.

Cadsby and Maynes (1999) analyze a two by two factorial design with MCS (binary contributions) versus continuous contributions as one factor and refunds versus no refunds as the other. Continuous contributions turn out to significantly increase
contributions and public-good provision. A money-back guarantee further encourages provision. Cadsby and Maynes (1998b) also study the impact of binary versus continuous contributions but here the focus is on gender: female groups coordinate more closely on an equilibrium (whether it is a free riding or a threshold equilibrium) than male groups do. Cadsby and Maynes (1998a) find that, with continuous contributions, business and economics students make contributions converging to the free-riding equilibrium; by contrast, the contribution of nurses cycled around the efficient threshold equilibrium.

The literature on public-good games with multiple step levels is much less. Chewning, Coller, and Laury (2001) have a five-player experiment with one, two, three, or five step levels. Their experiment involves a simultaneous-move order. 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—as in our case—they increase the group optima contributions but leave the Nash equilibria unchanged. We discuss their design and results in detail subsequently. Recently, Hashim, Maximiano, and Kannan (2011) analyze a game with five levels and five players. The authors vary information feedback about other members’ contributions to a subsample of group members.

**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 two of our four 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

$$
x_i = \begin{cases} 
e - c_i + 10 & \text{if } c_1 + c_2 \geq 12 \\
e - c_i & \text{if } c_1 + c_2 < 12.
e - c_i + 10 & \text{if } c_1 + c_2 < 18
\end{cases}
$$

The other two treatments involve an additional second threshold of 18. If $c_1 + c_2 \geq 18$, both players receive 5 on top of the 10 euros they receive for hitting the first threshold. That is, in these treatments, we have

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

Since $2e > 18$, both thresholds of the public good are feasible, but, due to $e < 12$, no player can meet the threshold on his or her own. Further, because $2 \times 10 > 12$ and $2 \times 15 > 18$, the provision of the public good at both provision points maximizes
joint payoffs. Note that the return on contributing one euro at each of the two levels is the same.

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 in the SEQ treatments. First- and second-mover contributions or payoffs are indicated with subscripts \( F \) and \( S \), respectively, in the SEQ treatments. The second treatment variable is the number of the thresholds (one or two). Table 1 summarizes our \( 2 \times 2 \) treatments design.

Subjects play this game over ten periods. The payoffs of the previously mentioned game were denoted in euros in the experiments (so that the exchange rate was one to one). In each period, subjects were endowed with \( e = 10 \) euros. The final payoff at the end of the experiment was determined by the earnings of one randomly chosen period. (See also the instructions in the Online Appendix.)

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 ten and eighteen subjects. (We control for session size in our data analysis subsequently.) In each session and each period, subjects were randomly matched into groups of two players. In the SEQ treatments, the roles of first and second movers were also random.

The subject pool consists of students from various fields from the University of Frankfurt. In total, we had 191 participants. For the step-level public-good experiments, we had 160 participants who earned on average 11.3 euros. Further, we employed 31 subjects to replicate the results of a previous study we used for calibration (see the next section). They earned on average 12.3 euros. The experiment was programmed in z-Tree (Fischbacher 2007). Sessions lasted about sixty minutes.

### 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

\[
U_i(x_i, x_j) = x_i - \alpha_i \max[x_j - x_i, 0] - \beta_i \max[x_i - x_j, 0], \quad x_i, x_j = 1, 2, i \neq j.
\]

Here, \(x_i\) and \(x_j\) denote the monetary payoffs to players \(i\) and \(j\) and \(\alpha_i\) and \(\beta_i\) denote \(i\)’s aversion toward 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 \(x_i\) previously, 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

\[
U_i(c_i, c_j) = 10 - c_i + 10\chi_1 - \alpha_i \max[c_i - c_j, 0] - \beta_i \max[c_j - c_i, 0],
\]

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

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

\(c_i, c_j = 1, 2; i \neq j\). The \(\chi_k\) are indicator functions indicating whether a step level has been reached. We have \(\chi_1 = 1\) 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 \(\alpha\) and \(\beta\) parameters observed in Blanco, Engelmann, and Normann (2011). For each subject, they derive an \(\alpha_i\) from rejection behavior in the ultimatum game and a \(\beta_i\) from a modified dictator game. The distribution is reproduced in Table S1. On average, \(\alpha = 1.18\) and \(\beta = 0.47\).

There are several reasons to follow Blanco, Engelmann, and Normann (2011) here. First, while Fehr and Schmidt (1999) derive distributions for the \(\alpha\) and \(\beta\) 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 F&S model with the exception of Fehr, Kremhelmer, and Schmidt (2008), who assume that there are 60 percent players with \(\alpha = \beta = 0\) and 40 percent fair types with \(\alpha = 2\) and \(\beta = 0.6\), which seems too coarse for our purposes. Second, the joint \(\alpha - \beta\) distribution has been successfully replicated in, for instance, Dannenberg et al. (2007), Teyssier (2012), and Kölle, Sliwka, and Zhou (2011). Since our subject pool differs from the one used in Blanco, Engelmann, and Normann (2011) and the aforementioned experiments, we elicited a joint \(\alpha - \beta\) distribution using thirty-one participants from the current subject pool (not necessarily the same subjects). We find no significant differences between the two \(\alpha\) and \(\beta\) distributions according to Kolmogorov–Sminov tests (\(\alpha: D = 0.123, p = .872; \beta: D = 0.150, p = .663\)). So, we successfully replicate the data of Blanco, Engelmann, and Normann. Third, the use of this joint distribution (henceforth “\(\alpha - \beta\) data set”) seems promising, as it successfully predicts outcomes in several games (ultimatum game, sequential-move prisoner’s dilemma, public-good game) in
Assumption 2: Players’ inequality parameters are drawn from the joint $a/C_0$-$b/C_0$ distribution in Table S1. 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 the subgame perfect Nash equilibrium of 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 his or 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 zero, the second mover either obtains $U_S(12 - c_F, c_F) = 8 + c_F - \alpha_S(12 - 2c_F)$ or $U_S(0, c_F) = 10 - \beta_Sc_F$. We find that $U_S(12 - c_F, c_F) > U_S(0, c_F)$ iff

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

where we drop the $S$ subscripts of the inequality parameters for simplicity. The $\tilde{c}_F$ in equation (3) is the minimum acceptable first-mover contribution for a given set of individual inequality parameters. Any contribution at least as high as $\tilde{c}_F$ will be met

<table>
<thead>
<tr>
<th>Second-mover contribution</th>
<th>$c_F = 2$</th>
<th>$c_F = 3$</th>
<th>$c_F = 4$</th>
<th>$c_F = 5$</th>
<th>$c_F = 6$</th>
<th>$c_F = 8$</th>
<th>$c_F = 9$</th>
<th>$c_F = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_S = 12 - c_F$ (level 1 provided)</td>
<td>21.3%</td>
<td>37.7%</td>
<td>67.2%</td>
<td>83.6%</td>
<td>100%</td>
<td>60.7%</td>
<td>19.7%</td>
<td>19.7%</td>
</tr>
<tr>
<td>$c_S = 18 - c_F$ (level 2 provided)</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>39.3%</td>
<td>80.3%</td>
<td>80.3%</td>
</tr>
<tr>
<td>$c_S = 0$ (public good not provided)</td>
<td>78.7%</td>
<td>62.3%</td>
<td>32.8%</td>
<td>16.4%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td></td>
</tr>
<tr>
<td>Expected first-mover payoff</td>
<td>10.13</td>
<td>10.77</td>
<td>12.72</td>
<td>13.36</td>
<td>14.00</td>
<td>13.97</td>
<td>15.02</td>
<td>14.02</td>
</tr>
</tbody>
</table>
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 $\alpha$ and decreasing in $\beta$.

Based on our Assumptions 1 and 2, we now predict the frequencies of public-good provision as a function of $c_F$. For each player in that data set (see Table S1), we determine the $\tilde{c}_F$ as in equation (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 $\beta = .525$ 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 subjects in $\alpha - \beta$ data set allows us to predict how many players in our experiment will provide the public good as a function of $c_F$.

Table 2 shows the results of this calibration for the SEQ treatments in the left panel. 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 2 also reveals that the expected monetary utility of a risk-neutral first mover monotonically increases in $c_F$ and is maximized for $c_F = 6$ (the expected payoff from choosing $c_F = 0$ is 10). As $c_F < 6$ results in a lower likelihood of public-good provision, lower payoffs, and greater payoff inequality, both selfish and inequality-averse first movers will choose $c_F = 6$ in the perfect Bayesian equilibrium of this game. 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 $c_F = 2$ for the first movers. The calibrated F&S model predicts the frequencies of second-mover responses as in Table 2, and $c_F = 6$ 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 mentioned previously. But in the two-level game, the first mover may also choose his or 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 the subgame perfect 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 $7 + c_F$). 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 in the subgame perfect equilibrium with standard preferences.

Now assume F&S players and begin with the second movers. With $c_F \in [8, 10]$, the second mover may choose $c_S = 18 - c_F$, $c_S = 12 - c_F$, or $c_S = 0$. 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 $c_S = 18 - c_F$ and $c_S = 12 - c_F$. First suppose $c_F = 8$. If the second mover chooses $c_S = 18 - c_F = 10$, we have $U_S(10, 8) = 15 - 2\alpha_S$. If the second mover chooses $c_S = 12 - c_F = 4$, we have $U_F(4, 8) = 16 - 4\beta_S$. We obtain
US(10, 8) < US(4, 8) iff 1 - 4\beta_S + 2\alpha_S > 0. This condition holds for 60.7 percent of the subjects in the \alpha - \beta data set. That is, if \( c_F = 8 \), the public good will be provided at level 1 with 60.7 percent probability and with 39.3 percent probability at level 2. 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 \( U_F(4, 8) = 17 - 6\beta_S \). We find that \( 16 < 17 - 6\beta_S \) iff \( 1 - 6\beta_S > 0 \). In the \alpha - \beta data set, 19.7 percent of the subjects meet this condition. That is, if \( c_F = 9 \), the public good will be provided at level 1 (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. The right panel of Table 2 summarizes the additional predictions in SEQ_2.

Consider next the first movers. \( c_F = 10 \) will never be chosen in a perfect Bayesian equilibrium 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 2) and accordingly an even lower F&S utility. Hence, a risk-neutral first mover will never choose \( c_F = 8 \) in a perfect Bayesian equilibrium. The remaining possibilities are that first movers will choose either \( c_F = 6 \) 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. Now 15.015 - 1.182\alpha > 14 iff \( \alpha < .859 \). This is predicted to hold for 36 percent of the subjects in the \alpha - \beta data set.

**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 Table 2, and that 64 percent of all first movers choose \( c_F = 6 \) and 36 percent choose \( c_F = 9 \).

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 \times 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
symmetric pure-strategy Nash 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 fifty-eight different types of players in the \( \alpha - \beta \) data set.\(^{10}\) We 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, 36 percent of the players contribute \( c_i = 0 \) whereas 64 percent choose \( c_i = 6 \). Hence, more types contributing \( c_i = 6 \) are required with F&S preferences to make players indifferent 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 less than 64 percent of players are expected to contribute \( c_i = 6 \), some players will still contribute. Learning is slower, and the shape of the best response correspondence differs from the standard case. We discuss this in detail subsequently.

**Proposition 3:** In treatment \( \text{SIM} \_1 \), 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 choose \( c_j = 6 \); 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* (\( \text{SIM} \_2 \)). As argued previously for \( \text{SEQ} \_2 \), meeting the second threshold is not a Nash equilibrium with standard preferences. As the equilibria derived previously for \( \text{SIM} \_1 \) are unaffected by the introduction of the second threshold; with standard preferences, \( \text{SIM} \_2 \) has the same Nash equilibria as \( \text{SIM} \_1 \).

We now look for a Bayesian Nash equilibrium of players with F&S utilities where the second level of the public good is provided. Suppose that some types choose \( c = 9 \). Previously, 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 \( c_j = 9 \). We 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 \( pU(9, 9) + (1 - p)U(9, 3) = 16p + (1 - p)(11 - 6x) \), 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 does contributing 3 iff

\[
p > \frac{6x - 4}{6x + 6\beta - 5}.
\]

For F&S players with \( x = \beta = 0 \), this condition is never met (as seen previously); 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 \( x - \beta \) data set, we find that for \( p = .72 \) exactly 72 percent of the players (forty-four players) have \( pU(9, 9) + (1 - p)U(9, 3) > pU(3, 9) + (1 - p)U(3, 3) \), whereas for 28 percent (seventeen players) 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 nine or three. The only possible deviation is to contribute \( c = 0 \) since any other contribution is dominated by either \( c = 0 \) or \( c = 3 \). Contributing \( c = 0 \) yields an expected F&S utility of \( 10 - 3\beta - 0.72 \times 6\beta \). But the equilibrium action \( c = 3 \) yields \( 0.72(17 - 6\beta) + (0.28)7 \) that is strictly larger for all \( \beta \in [0, 1] \). Thus, we 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 through 4, we now derive two hypotheses regarding the impact of our two treatment variables. Comparing the predicted public-good provision in SIM versus 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 1) in the unique equilibrium. This holds for both the one- and the two-threshold case. We maintain no hypothesis regarding contributions in the SIM versus SEQ treatments.

**Hypothesis 1:** The public good will be provided more frequently in the SEQ treatments compared to SIM.

Our second hypothesis, though, does depend on assuming F&S preferences. Propositions 1 through 4 show that public-good provision can be improved if there is the second threshold. There are multiple equilibria in the SIM treatments anyway, but there exists 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 but the other player exploits this, this does not harm total payoffs that much
as the first level of the public good is still provided. In both the simultaneous-move treatment and the sequential treatment with two levels, players may yield a higher payoff by achieving the second threshold level. Therefore, they have an incentive to make higher contributions, and public-good provision will be more likely in the presence of two thresholds. If first movers make higher contributions in SEQ_2, fewer punishments should occur and we should see more second movers who contribute. Both effects should cause increase in public-good provision at least at level 1.

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

### Overview of the Results

We present our results in three parts. In this section, we present a brief overview of the results. The next section presents tests of Hypotheses 1 and 2. The subsequent section presents a more detailed analysis of the predictive power of the calibrated F&S model.

Table 3 presents a summary statistics of the averages of our main variables of interest. (session-level data and variability measures can be found in Table S2). The second threshold level leads to a higher (sum of) contributions than the one-level variant both in the simultaneous and in the sequential treatments. The sequential-move order leads to a higher sum of payoffs compared to the simultaneous treatments. Public-good provision at the first level is most effective in the treatments with sequential moves. Public good level 1 is provided most frequently (85.56 percent) in SEQ_2. Only in 6 percent of the SIM_2 cases is the public good provided at the second threshold level. In SEQ_2, however, it does come out better (16.67 percent) in

**Table 3. Summary Statistics of Our Four Treatments.**

<table>
<thead>
<tr>
<th>Variable</th>
<th>SIM_1</th>
<th>SIM_2</th>
<th>SEQ_1</th>
<th>SEQ_2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum of contributions</td>
<td>10.44</td>
<td>11.99</td>
<td>9.92</td>
<td>12.14</td>
</tr>
<tr>
<td>First-mover contributions</td>
<td>—</td>
<td>—</td>
<td>4.76</td>
<td>6.41</td>
</tr>
<tr>
<td>Second-mover contributions</td>
<td>—</td>
<td>—</td>
<td>5.16</td>
<td>5.73</td>
</tr>
<tr>
<td>Second movers contributing $c_s = 0$, in %</td>
<td>—</td>
<td>—</td>
<td>18.57</td>
<td>10.00</td>
</tr>
<tr>
<td>Successful coordination, in %</td>
<td>49.05</td>
<td>17.00</td>
<td>77.62</td>
<td>81.11</td>
</tr>
<tr>
<td>Sum of payoffs</td>
<td>22.42</td>
<td>21.11</td>
<td>25.12</td>
<td>28.30</td>
</tr>
<tr>
<td>level 1 provided ($\chi_1 = 1$), in %</td>
<td>64.29</td>
<td>59.00</td>
<td>75.24</td>
<td>85.56</td>
</tr>
<tr>
<td>level 2 provided ($\chi_2 = 1$), in %</td>
<td>—</td>
<td>6.00</td>
<td>—</td>
<td>16.67</td>
</tr>
</tbody>
</table>

Note: Note that the public good is provided at level 2 ($\chi_2 = 1$) only if it is also provided at level 1 ($\chi_1 = 1$).
SEQ_2. Defining successful coordination as cases without wasteful contributions, we find that coordination is best in the environments with sequential moves.

Table 3 thus suggests that we do find tentative support for Hypothesis 1. Regarding Hypothesis 2, the second step level improves contributions; it also improves public-good provision (at level 1) in the SEQ treatments but not in the SIM settings.

**Main Treatment Effects**

We now report tests of Hypotheses 1 and 2. We mainly apply regression analysis where we take possible dependence of observations into account by clustering at the session level. We additionally report, in footnotes, nonparametric tests in which case we count each randomly matched session as one observation. In these cases, we report two-tailed \( p \) values.

As dependent variables we use **sequential** (a dummy which is equal to 1 if the move order is sequential), **twolevel** (a dummy which is equal to 1 if there are two levels), the interaction **sequential** × **twolevel**; 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. Regression 2 includes the interaction **sequential** × **twolevel**, and Regression 3 adds **period** and **sessionsize**. We ran further regressions where we add the interactions of **sequential**, **twolevel**, and **sequential** × **twolevel** with **period**. We briefly report whether these additional regressions (not fully reported here but are available upon request) lead to qualitatively different results in each of the following four sections.

**Sum of Contributions**

We first analyze the sum of contributions of the (randomly matched) two-player groups. The left panel of Table 4 reports a linear regression suggesting that the sum of contributions is not significantly influenced by the order of moves. Consistent with our Hypothesis 2, adding the second threshold leads to a significantly higher sum of contributions. The interaction of a sequential-move order and two levels does not lead to a further increased sum of contributions. **Period** is not significant, thus the sum of contributions is not affected by time dynamics. This still holds when we additionally employ the interactions of **sequential**, **twolevel**, and **sequential** × **twolevel** with **period**.

The variable **sessionsize** is significant; that is, in sessions with more participants, contributions are lower. While the coefficient is small, we note that this is consistent with findings in Botelho et al. (2009). In their article, repeated settings with “random strangers” and “perfect strangers” matching protocols are compared. The authors find that the assumption that subjects treat random stranger 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 again, has an impact on contributions.
Sum of Payoffs

In the right panel of Table 4, we report the results of a linear regression on the sum of payoffs of the two players. The table shows that the sequential contribution mechanism significantly improves payoffs. The second step level insignificantly reduces the payoffs. This can be explained by the fact that two thresholds increase contributions but, as we will see, the second level is rarely actually achieved. The size of the sessions is weakly significant, but, again, the coefficient is small.

When we add the interaction sequential × twolevel, we find that it significantly increases subjects’ payoff compared to the baseline SIM_1. The difference between SEQ_1 and SEQ_2 is, however, not significant as follows from a Wald’s test (p = .125). Further, when we add the interactions of sequential, twolevel, and sequential × twolevel with period to the regression, sequential × twolevel is insignificant. Instead, period × twolevel and period × sequential are negative and significant. This suggests an overall negative impact of the second threshold on payoffs.

Public-good Provision

Table 5 presents probit regressions of the frequency of public-good provision where the left panel is about public-good provision at level 1. The dependent variable equal to 1 if and only if the first threshold is met (i.e., if and only if $\chi_1 = 1$). The second
probit regression (right panel) has the dependent variable is equal to 1 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 show that sequential significantly improves the public-good provision at the first threshold. This supports Hypothesis 1. The implementation of a second threshold does not lead to a higher frequency of public-good provision. Interacting sequential with two thresholds suggests borderline significant support for an increased public-good provision which, however, disappears once we control for period and sessionsize. Overall, 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. We note, however, that the coefficient of sessionsize is small. Adding the interactions with period does not change the results qualitatively, although the treatment variable sequential turns out to be highly significant in this regression.

Table 5 also presents a probit regression of the frequency of public-good 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. The dummy sessionsize is not significant. Regression 2 reveals that public-good provision at level 2 moderately decreases over time. Adding the period interactions in

<table>
<thead>
<tr>
<th>Table 5. Probit Regressions of Public-good Provision.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
</tr>
<tr>
<td>Sequential</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Twolevel</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Sequential × twolevel</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Period</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Sessionsize</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Constant</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
</tr>
<tr>
<td>$R^2$</td>
</tr>
</tbody>
</table>

Note: Robust standard errors in parentheses. *p < .1, **p < .05, ***p < .01.
an additional regression (not reported here) show that the negative time trend is driven by *sequential*. In this regression, *period* is insignificant but *period × sequential* is.

**Coordination Rates**

We define $C = c_1 + c_2$ and cases of successful coordination as those where $C \in \{0, 12\}$ and $C \in \{0, 12, 18\}$ in the one- and two-step treatments, respectively. To economize on space, we report descriptive statistics and simple nonparametric tests here only. A regression analysis of successful coordination is qualitatively very similar to the one on payoffs reported previously.

Figure S1 compares coordination in the simultaneous and the sequential treatment with one threshold. In SEQ_1, nonwasteful coordination on $C = 12$ is the most frequent outcome (74 percent). By contrast, coordination on $C = 12$ occurs only in 47 percent of the observations in SIM_1 ($p = .05$, rank-sum test). The difficulty of coordinating in SIM_1 is also documented by the higher number of coordination failures 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 SIM_1, 51 percent of the subjects do not manage to coordinate. The remaining cases are those where $C = 0$ where coordination is successful in that no contributions are wasted but no public good is being provided. In SEQ_1, there are only 22 percent cases with coordination failure. Mainly, these involve second movers punishing low first-mover contributions.

Figure S2 compares coordination rates in the treatments with two thresholds. This plot again documents the superiority of the sequential- over the simultaneous-move variant. In SEQ_2, about 80 percent of the contributions subjects manage to coordinate on the first ($C = 12$) or the second threshold ($C = 18$) without generating wasteful excess contributions. This is in contrast to the coordination rates in SIM_2, where a significantly smaller fraction of the contribution sums (16 percent) is efficient ($p = .05$, rank-sum test). In SIM_2, subjects seem to face a great deal of difficulty in terms of coordination. This leads to a high amount of wasteful contribution sums (the sum of all cases where $0 < C < 12$, $12 < C < 18$, or $C > 18$) of 83 percent. Figure S2 therefore serves as an explanation of the fact that the second level leads to smaller payoffs. Especially in SIM_2, the second level leads to costly miscoordination of the players.

However, two levels are efficient in the environment with sequential moves that explains the significance of our interaction term *sequential × twolevel*. The result is driven by first movers contributing higher amounts in SEQ_2 compared to first movers in the one-level treatment. This is shown in Table 3 where the average first-mover contributions of SEQ_1 and SEQ_2 are presented. It shows that first movers on average make higher contributions in the sequential treatment with two thresholds. In SEQ_2, first movers contribute on average to more than half of the first threshold (6.41). Thus, second movers are not “exploited” that frequently, and they
only punish first-mover behavior in 10 percent of all cases. This is in contrast to the one-level treatment where first movers make average contributions below six (4.97) and second movers punish in 19 percent of all cases.

**What Improves Payoffs in the Sequential Case?**

We saw that both public-good provision and coordination are significantly better in the sequential treatments. We also saw that higher payoffs occur in SEQ_1 (25.12) compared to SIM_1 (22.42). Can we say more about the sources of higher payoffs in the sequential case?

In order to investigate this question, we analyze the sum of payoffs conditional on the public-good provision. First, conditional on $w_1 = 1$, we find that the payoff sums are nearly the same in SEQ_1 (27.99, 316 observations) and SIM_1 (27.53, 270 observations). They do not differ much from the conditional mean of $27.76|\chi_1=1$. The explanation is that coordination is rather good in either case, provided $w_1 = 1$. In SEQ_1, players coordinate successfully in 312 of the 316 cases (99 percent), so there is almost no waste. In SIM_1, successful coordination occurs only in 196 (73 percent) of the 270 cases. However, excess contributions conditional on $w_1 = 1$ are small: in the seventy-four cases of unsuccessful coordination, only 1.7 cents are wasted on average.

Now we condition on $w_1 = 0$; that is, we analyze the data where the public good was not provided. We find that subjects earn substantially more in SEQ_1 (16.42, 104 observations) compared to SIM_1 (13.21, 150 observations) and also compared to the conditional mean of $14.52|\chi_1=0$. Successful coordination is rare here in both treatments: 14 (13.5 percent) of the 104 in SEQ_1 and 10 (6.7 percent) of the 150 cases in SIM_1, respectively. However, provided coordination is not successful, a lot more money is wasted with simultaneous moves: conditional on $\chi_1 = 0$ and unsuccessful coordination, 7.27 cents are wasted in SIM_1 as opposed to 4.13 cents waste in SEQ_1.

To sum up, if the public good is provided, payoffs and coordination are similar in SEQ_1 and SIM_1 but the public good is provided more often in SEQ_1. The main source of the payoff difference of 2.7 therefore appears to be the wasteful contribution in SIM_1 when the public good is not provided.

**The Predictive Power of the Calibrated F&S Model**

We now discuss the quantitative predictions of the F&S model in more detail. We begin with Proposition 1. Figure 1 contrasts the predictions made in Table 2 to the observations of the frequency of second movers who contribute $c_s = 12 - c_F$ in reply to first-mover contributions. The data underlying Figure 1 pool the $c_F$ in both sequential treatments SEQ_1 and SEQ_2. Using one-sample chi-square tests, we
cannot reject that predicted and observed frequencies are the same (all $\chi^2_{(1)} < 2.38$ and $p > .123$). The F&S model predicts the second-mover responses remarkably well.

In SEQ_1, all 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. In our SEQ_1 data, it turns out that $c_F = 5$ is the (ex post) payoff maximizing strategy (yielding an expected payoff of 14.26, as opposed to 13.76 with $c_F = 6$) and that it is chosen in 25.7 percent of the cases. While this rejects the F&S prediction, we note that similar observations where first movers behave suboptimally, given second-mover behavior has been made before (see the following discussion).

Figure 2 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 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.

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 the first mover chooses $c_F = 9$, he or 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 percent are. Predicted and observed share differ significantly (binomial
test, \( p < .05 \). The failure of the theory seems surprising since the costs 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. (See the discussion at the end of the section).

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 \( c = 0 \). 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 respond?

Figure S3 (see the online supplement) 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 players, the best reply correspondence has a “bang-bang” property. If the belief is that player \( j \) chooses \( c_j = 6 \) 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 percent 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 \( c_j = 6 \), where 25 percent still choose \( c_j = 0 \).

As mentioned in Proposition 3, the share of players choosing \( c_F = 6 \) required such that \( c_F = 6 \) is a best reply that is slightly larger with F&S players. Inequality aversion has an effect similar, in fact a stronger effect, than does risk aversion (on
average, players in Holt and Laury [2002] 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.

What can we conclude from the analysis of the calibrated F&S model? First, we find remarkable confirmations of the predictions of the model. One may argue that, regarding SEQ_1, these are not so surprising because of the partial similarity of SEQ_1 to the ultimatum game (from which the αs were elicited). However, the SEQ_1 prediction also depends on the joint distribution and not on the α only. Moreover, we also found confirmation of the calibrated F&S model for SEQ_2 and SIM_1. Hence, we conclude that the model is particularly powerful in our setup.

How about the two contradictions to the calibrated F&S model then? First, we found that first movers behaved too greedily to be consistent with Assumptions 1 and 2, providing \( c_F < 6 \) too often. This finding is not new. 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 proposers offer less than the equal split.\(^{19}\) Huck, Muller, and Normann (2001) show that, in quantity-setting duopoly, Stackelberg followers are inequality averse, but the Stackelberg leaders still choose too high an output. The payoff maximizing (and inequality minimizing) output in that data set was the symmetric Cournot-Nash solution. In ultimatum games, the Stackelberg game, and this study, risk-loving behavior can explain the first-mover behavior. However, it could also be that first movers feel entitled to more than 50 percent of the pie, whereas second movers regard the equal split as fair. Social norms may be perceived differently by first and second movers.

We secondly saw that second movers exploit first-mover trust (that is, \( c_F = 9 \)) too often in SEQ_2. We consider the following explanation plausible. In SEQ_2, first movers frequently choose \( c_F < 6 \) and, just as in SEQ_1, the second movers are in the weaker position. Whenever \( c_F = 9 \), second movers are suddenly in the stronger position. They can now ensure themselves the higher payoff, and they often do so. It could be the low degree of trustworthiness is second movers scoring off greedy first movers, with a “now it is my turn” attitude (recall the game is repeated ten times). In contrast to costly punishments of \( c_F < 6 \), responding with \( c_S = 3 \) to any \( c_F = 9 \) is free, in fact yields an even higher payoff. If so, second movers do not reflect that first movers contributing \( c_F = 9 \) are unlikely to be the same first movers who offered \( c_F < 6 \) in a previous round.

**Conclusion**

How should, say, two academics organize their joint work when the goal is that a certain threshold in terms of quality of the journal has to be met? Our experiments indicate two answers to this question. First, we find that an additional second step level (say, targeting an A- rather than a B-level journal) does lead to significantly
higher contributions (efforts in the case of teamwork), although the impact on payoffs is ambiguous and insignificant. The effect on public-good provision is positive, especially in the sequential case (where the effect is significant). The logic is as follows: first movers often contribute high\textsuperscript{20} such that the second step level can be met. Second movers may exploit this by contributing less, however, they still contribute enough so that the probability of meeting the first step-level increases. For academic teamwork, say, a strategy like “we need to invest this much effort for a B journal, but with that much more work we could go for an A-level journal” might work out. With simultaneous moves, however, coordination failure becomes more frequent.

Second, we find that the sequential-move variant yields more frequent provision of the public good and higher payoffs. This confirms the literature on leading by example where, in our setup, it is mainly the better coordination that renders the sequential mechanism superior in the threshold public-good game. Even though some low-contributing first movers (who actually give a bad example) are punished by second movers, higher provision rates and payoffs emerge. The finding is in contrast to Gächter et al.’s (2010a) test of Varian’s (1994) model. They find that sequential contributions are lower with sequential moves, but the difference is not as big as predicted. One reason for this is that, as in our setting, second movers sometimes punish first movers. In our setting, the sequential-move variant is more efficient.

As mentioned in the literature survey, Chewning, Coller, and Laury (2001) run five-player step-level public-good games with simultaneous moves and different numbers of provision points (ranging from 1 to 5). Their treatments with two and three thresholds can be compared to our experiments with simultaneous moves because in these treatments, the Nash equilibria are the same (0 and 7.5 in their case), whereas the Pareto optimum is higher with three than with two thresholds. This is the same in our experiment.\textsuperscript{21} Comparing two and three thresholds, Chewning, Coller, and Laury find that contributions increase in the first five periods with three thresholds. However, in periods eleven to fifteen, they are below the two thresholds case. This is consistent with the findings in our treatment with two levels where two levels initially lead to higher contributions. We also find that contributions decrease over time. They are, however, higher than that in the one-level treatment throughout. Chewning, Coller, and Laury conclude that the decrease in the treatment with three thresholds is due to coordination problems. Thus, the high degree of coordination failure in our SIM\_2 treatment is in line with these findings.

Based on a fully calibrated Fehr and Schmidt’s (1999) model, we make accurate 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 model also predicts behavior well in the sequential treatment with two step levels, and in the simultaneous-move case with one level.
Acknowledgments
We are grateful to Volker Benndorf, Dirk Engelmann, Nikos Nikiforakis, Theo Offerman, Martin Sefton, and Marta Serra Garcia for detailed comments on our article. We are also indebted to audiences at ESA Innsbruck, University of Frankfurt, University of Heidelberg, Centre for European Economic Research Mannheim, Hohenheimer Oberseminar, the Annual Congress of the European Economic Association in Oslo, and the Annual Meeting of the Verein für Socialpolitik in Frankfurt for helpful comments.

Declaration of Conflicting Interests
The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The authors received no financial support for the research, authorship, and/or publication of this article.

Supplementary Material
The online supplementary tables and figures are available at http://jcr.sagepub.com/supplemental.

Notes
1. The experimental literature on these issues (which we review in detail subsequently) includes Erev and Rapoport (1990); Potters, Sefton, and Vesterlund (2005); Gächter et al. (2010a, 2010b); and Figuieres, Masclet, and Willinger (2011).
3. Fehr, Kremhelmer, and Schmidt (2008) also provide a calibration of Fehr and Schmidt’s (F&S; 1999) model. Their calibration is based on a two-type categorization (40 percent fair players and 60 percent standard players). See the following discussion.
4. See also Figuieres, Masclet, and Willinger (2011); Gächter et al. (2010b); and Teyssier (2012) for related experiments. For related field experiments, see Glazer and Konrad (1996), List and Lucking-Reiley (2002), Soetevent (2005), or Huck, Muller, and Normann (2001).
5. This prediction also holds in Andreoni, Brown, and Vesterlund (2002), who study the same model as Gächter et al. (2010a), although with different parameters.
6. Rauchdobler, Sausgruber, and Tyran (2010) study how different thresholds affect contributions in a voluntary contribution mechanism (VCM) variant. Thresholds differ between \( T = 0 \) and \( T = 57 \); however, there is always only one threshold at a time. Moreover, higher thresholds do not improve efficiency per se here but merely serve as a minimum target for players which may be imposed exogenously or endogenously.
7. In Blanco, Engelmann, and Normann’s (2011) modified dictator game, dictators choose between 20-0 and equitable outcomes ranging from 0-0 to 20-20 (all denoted in £ [Pounds
A player \( i \) who is indifferent between payoff vectors \((20, 0)\) and \((x_i - x_i)\) has 
\[
\beta_i = 1 - x_i/20.
\]
8. There are no significant differences between the \( z \) Blanco, Engelmann, and Normann
(2011) elicit and the one assumed in F&S (1999). The \( \beta \) distribution differs significantly; however, one can argue that distributions still roughly compare and do not differ
outlandishly.
9. There are also numerous mixed-strategy equilibria.
10. Among the sixty-one players reported in Table S1, three types occur twice so that there
are fifty-eight types in total.
11. Note that, in our treatments with two threshold levels, when the second level is met, this
also counts as successful provision of public good level 1 by definition.
12. Since we have three sessions per treatment, these tests are either significant at the 5 per-
cent level or insignificant: when comparing two treatments with three observations per
treatment, there are \( \binom{6}{3} = 20 \) different possibilities to rank the observations. Thus, the
probability to observe the outcome where the lowest observation in one treatment is still
higher than the highest observation in another treatment is \( p = 1/20 = .05 \). All other out-
comes are not significant.
13. Hypothesis 2 is also supported in that rank-sum tests suggest that both SIM_2 and SEQ_2
have significantly higher contributions than SEQ_1 \( (p = .05) \); however, the comparisons
to SIM_1 are not significant.
14. The according rank-sum tests indicate that both SEQ_1 and SEQ_2 have significantly
higher payoffs than SIM_1 and that SEQ_2 has higher payoffs than SIM_2 \( (p = .05) \). Other
comparisons are not significant.
15. Rank-sum tests consistent with this finding are that both SEQ_1 and SEQ_2 have signifi-
cantly higher public-good provision than does SIM_2 and that SEQ_2 has higher payoffs
than does SIM_1 \( (p = .05) \). Other comparisons are not significant.
16. This is warranted because, first, the F&S model does not predict any differences and sec-
ond, we do not observe differences—with the exception of \( c_F = 5 \), where contribution
rates differ significantly (two-sample chi-square test, \( \chi^2(1) = 8.579 \), \( p < .01 \)). Important-
antly, the minor differences we observe are not systematic. Contributions of \( c_S = 12 - c_F \) are more frequent for \( c_F = \{3, 4\} \) in SEQ_1 than in SEQ_2 but the other way
round for \( c_F = \{5, 6\} \). 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 public good was provided).
17. Here, we cannot include SEQ_2 data in the figure because behavior is predicted (and does)
differ whenever \( c_F \geq 8 \).
18. These percentages are based on data from periods six to ten where we observe less heter-
egeneity in the data. Some subjects indeed choose \( c_i > 6 \) but—for our data—
\( 0 < c_j < 6 \) is never a best reply with standard or F&S preferences. Thus, we focus
on \( c \geq 6 \) and \( c = 0 \).
19. In Blanco, Engelmann, and Normann (2011), 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.
20. A similar pattern is observed by Laury, Walker, and Williams (1999), who find in a public-good setting with diminishing returns that players contribute more than the Nash prediction.

21. Their treatments with one and two thresholds, by contrast, do not compare easily to our setup because introducing a second threshold changes Nash equilibria (0 and 7.5 rather than 0 and 12.5) but maintains the Pareto optimum (at 12.5). Their treatment with five thresholds can neither be compared to our two-threshold treatment. In this treatment again the Nash equilibrium is 7.5. Moreover, two additional thresholds exist which both guarantee a higher Pareto optimum. By contrast, in our two-threshold setup this is only the case for one additional threshold.

References


