Experimental investigations of binary threshold public good games

Jörg Spiller

Friedel Bolle

European University Viadrina Frankfurt (Oder)

Department of Business Administration and Economics

Discussion Paper No. 393

January 2017

ISSN 1860 0921

Experimental investigations of binary threshold public good games

Jörg Spiller and Friedel Bolle

Europa-Universität Viadrina Frankfurt (Oder)

January 2017

Abstract

In Binary Threshold Public Good (BTPG) games n players contribute or not to the production of a public good which is produced if and only if there are at least k contributors. The BTPG games with the highest (k=n) and the lowest (k=1) threshold are the Stag Hunt game and the Volunteer's Dilemma. There is a plethora of equilibria in BTPG games. We experimentally investigate 16 different symmetric and asymmetric BTPG games with n=4 and k=1,2,3,4 and test general theoretical *attributes of equilibria* and whether equilibrium play can apply at all. As theory predicts, neither magnitude effects nor a negative instead of a positive frame affect behavior which is contrary to the bulk of the experimental literature. In the Stag Hunt game, which is often used to discriminate between payoff dominance and risk dominance, risk dominance as a description of behavior is clearly rejected and payoff dominance is moderately supported. We show that no theory with homogeneous players can describe behavior.

JEL codes: C72, D72, H41

Keywords: Binary threshold Public Goods, framing, equilibrium selection, payoff dominance, risk dominance, efficiency, experiment

Europa-Universität Viadrina Frankfurt (Oder)
D - 15207 Frankfurt (Oder), Germany

Corresponding author:
Friedel Bolle, bolle@europa-uni.de, Tel.: +49 173 9853143

Highlights

- A public good is produced if a sufficient number of players contribute to its production.
- Players with the same cost/benefit ratio contribute with the same frequency.
- Negative vs. positive costs and benefits do not cause a framing effect.
- In the Stag Hunt game, payoff dominance is a better predictor than risk dominance.
- No theory of behavior with identical subjects can apply.

1. Introduction

In order to substitute the empty toner cartridge of a publicly used printer only one volunteer is needed who is ready to bear the costs in terms of time lost and dirty hands. This is an example of the Volunteer's Dilemma, first analyzed by Diekmann (1985). More severe than this example are cases where victims of criminal violence or an accident need help, at least by someone calling the police[1]. Sometimes more than one volunteer is necessary for the production of a public good, for example when a large conference has to be organized or an office party has to be prepared or when a low income friend needs help because he moves to another apartment.

It requires all members of a cartel to keep their contract secret. With plausible assumptions about the profitability of the cartel and incentives for Whistle Blowers, this is an example of the Stag Hunt game, first described by Rousseau (1997 [1776]). A class of examples of these games are frontlines which have to be defended by individuals or units. These can be military frontlines or dykes or standards of behavior. In the case of dykes, in former times communities (villages) were responsible to keep their section of the dyke in order.

The Volunteer's Dilemma and the Stag Hunt game are extreme cases of Binary Threshold Public Good (BTPG) games where players have a dichotomous choice of either contributing to the production of a public good or not. The public god will be produced if and only if a certain threshold of contributions is reached or surpassed. In this paper the threshold is described as "at least k of n players must contribute". Player i bears costs $c_i > 0$ if he contributes and he enjoys benefits $G_i > c_i$ if the public good is produced. This structure is completely different from a linear public good game with binary contributions which has a unique equilibrium (no one contributes) while a BTPG game has a plethora of pure and mixed strategy equilibria. No player contributing is one of the equilibria if k>1, but for k ≤ n it is (strictly for k<n) Pareto-dominated by all other equilibria. If $G_i < c_i < 0$, then it is individually profitable to contribute but players provide a "public bad" when contributions surpass the threshold. An example is CO2 emissions if there is a threshold below which damages are bearable and beyond which catastrophe is

[1] An often cited example is **Kitty Genovese** who was stabbed to death in New York City, on March 13, 1964. According to the *The New York Times*, 37 or 38 witnesses saw or heard the attack and did not call the police. (https://en.wikipedia.org/wiki/Murder_of_Kitty_Genovese)

inevitable. Voting in parliaments and in committees as well as shareholder voting is an important example for BTPG games (see Bolle, 2015). In many other examples a *team with minimal requirements* concerning the number and perhaps also the complementary qualifications[2] of the members is necessary to launch a project or solve a problem for the best of their community.

Experiments. We report here about 16 BTPG experiments in four treatments. In the first four experiments (Treatment S+), players have different positive costs c_i and benefits G_i but their cost/benefit ratio is the same. In the next four experiments (Treatment S-), costs and benefits are negative but absolutely the same as in S+. In the other eight experiments, benefits are always the same but players have different costs. In four experiments (Treatment A), the differences in costs are small and completely mixed strategy equilibria exist, in four experiments (Treatment B), they are large and completely mixed strategy equilibria do not exist[3]. There are n=4 members of the experimental groups. In every treatment all possible thresholds k are investigated, i.e. k=1,2,3, and 4. Our goal is to describe behavior in these games and test game theoretic predictions and regularities found in other investigations.

Results. A surprising result is the lack of framing effects. For players with the *same cost/benefit ratio* we find:

(i) *Within a game, players with the same sign of costs contribute with the same probability.*

(ii) *Behavior in a game with a positive frame ($0 < c_i < G_i$) is, after the re-labeling of actions, the same as in a game with a negative frame ($G_i < c_i < 0$).*

Note that framing effects are regularly reported in economic experiments[4]. Two further fundamental questions concern the dependency of contribution probabilities on costs c_i and the threshold k. Intuitively we expect:

---

[2] "Complementary qualifications" require a generalization of the threshold definition. In Bolle (2015) the threshold is described by sufficient subsets of players as in cooperative games with binary characteristic functions (called simple cooperative games or voting games).

[3] Completely mixed strategy equilibria are plausible benchmarks because all other equilibria (except no player contributing for k>1 and all players contributing for k=n) require tacit agreement about some players playing mixed and others pure strategies.

[4] In linear Public Good experiments, it has always been found that the negative frame (linear static Common Pool experiments) evokes significantly less cooperation than the positive frame (Andreoni, 1995, Willinger and Ziegelmeyer, 1999; Park, 2000; Dufwenberg et al., 2011). This difference is confirmed in the only BTPG experiment with a positive and a negative frame (Sonnemans et al., 1998, see below).

(iii) *When benefits are the same then higher-costs players contribute with lower probability than lower-costs players.*

If we take the most efficient of the completely mixed strategy equilibria as a benchmark we should expect the contrary relationship; but (iii) is confirmed in experiments by Diekmann (1993), Franzen (1995), and Goeree (2005) for the Volunteer's Dilemma (k=1) and by this investigation for all thresholds k. Explanations can be efficiency concerns (Przepiorka and Diekmann, 2013, and Diekmann and Przepiorka, 2015, in a Volunteer's Dilemma experiment) or equilibria with role dependent social preferences as suggested by Bolle and Otto (2016) and Bolle (2016).

(iv) Higher thresholds go along with higher contribution probabilities. In the Stag Hunt game of the positive frame (k=n) contribution probabilities are close to 1, in the negative frame (k=1) they are close to 0.

This is an intuitive consequence of payoff dominance. If risk dominance applies then contribution probabilities should decrease with higher k. In the positive frame, a player must more and more rely on other players' contributions; for k=n contribution probabilities would be 0. In the negative frame, for decreasing k, a player must more and more rely on the non-contributions of others. Risk dominance predicts contribution probabilities of 1 for the case k=1.

In S+ and S- the Harsanyi and Selten (1988) equilibrium selection (called HS) can easily be applied. The selected equilibrium is characterized as the payoff dominant among the symmetric equilibria. Although some average contribution probabilities are close to the HS equilibrium the respective hypothesis will be rejected and, more than this:

(v) *Behavior cannot be described by identical players.*

Note that identical players do not mean identical play if cost/benefit ratios are different. Therefore any *quantitative* prediction by *unique* equilibria or non-equilibrium play of identical players is rejected. Nonetheless, behavior of a large fraction of subjects may be close to "plausible" equilibrium strategies which may then guide *qualitative* hypotheses. For treatments A and B it is difficult to derive the HS equilibrium. Nonetheless, completely mixed strategy equilibria have an intuitive appeal as benchmarks (Diekmann, 1985, 1993; see also footnote 2). In treatment A such equilibria exist, in Treatment B they do not exist. We might conclude that behavior in Treatment B is less stable than behavior in

A, but stability is difficult to measure without a reliable theory of behavior. If we take the fit of a regression equation for contributions as a measure of stability then behavior in B is as stable as that in A. An explanation may be social preferences under which completely mixed strategy equilibria exist also in B. However, we do not consider this a strong result.

Theoretical predictions are derived from completely informed and fully rational players who do not need to learn. Contrary to this hypothesis we find traces of dynamics in our games with eight repetitions in a stranger design.

(vi) *(a) There is a weak trend in S+ and S- towards more cooperation.* (b) *When a player was "pivotal" in the previous round then his contribution probability increases in games with a positive and decreases in games with a negative frame.*

Pivotality of player i means that k-1 of i's co-players contributed and therefore player i's decision was crucial for the outcome. Related findings in the literature (see below) stem from experiments with sequential contributions where pivotality means that, without a player's contribution, k cannot or might not be reached. The difference is that, with sequential contributions, players know when they are or could be pivotal while, with several rounds of the static game, players experience to be pivotal and seem to believe that, even in a stranger design, i.e. with new co-players, there is an increased probability of being pivotal also in the next round.

To the best of our knowledge, (i) and (iv) have never been investigated and (ii) only in one experiment (see below). (iii) confirms results from the literature and the results concerning the Stag Hunt game contribute to the literature with 4 *non*-2x2 games. (v) may be interpreted as HS and other equilibrium selection theories (applied with players with identical preferences) not really describing behavior. They may or may not provide us with correct qualitative predictions of behavior. Hansanyi's (1995) later favoring of risk dominance over payoff dominance as in HS leads to even worse predictions. The pivot player result (vi b) is new and surprising because it could not be expected in repetitions of a simultaneous moves game with a stranger design. At last, it should be emphasized that our results are derived from 16 games with different frames and a large variety of cost/benefit ratios. Therefore, we do not think that our results are driven by a special selection of experimental parameters. In Section 2 we formulate hypotheses concerning (i) to (v), sometimes with the contrary statement which is then rejected in Section 4.

Additional literature. Theoretical and experimental work concerning the Volunteer's Dilemma (k=1) has already been mentioned. Experimental studies of the Stag Hunt game (k=n) are mostly based on 2x2 games and are concentrated on the question which of the two pure strategy equilibria "no one contributes" (mostly selected by risk dominance) and "all contribute" (selected by payoff dominance) is played. Van Huyck et al. (1990) and Rydval und Ortmann (2005) find tendencies towards risk dominance; tendencies towards payoff dominance are found by Battalio et al. (2001), provided the "optimization premium" is high enough, and, in an experiment with chimpanzees, by Bullinger et al. (2011). Whiteman and Scholz (2010) find a positive influence of social capital. Al-Ubaydli et al. (2013) find that cognitive ability and risk aversion have no impact on successful coordination while patience does. Büyükboyacı (2014) shows that information about the risk attitude of others changes behavior which is, however, not affected by one's own risk attitude. Our experiments with four-player games are difficult to compare with all of these two-player games. Our results strongly support payoff dominance in the Stag Hunt game (72% - 97% average contributions in the four treatments). Feltovich and Grossman (2013) investigate the influence of group size (2 to 7 players) and communication on contributions. Without communication, contributions are independent of group size (between 37% and 45%).

Experiments with intermediate thresholds require contributions from two of three players up to six of ten. In all investigations only one or two different thresholds are considered. Goren et al. (2003) investigate a BTPG game with five players with different weights (5, 10, 15, 20, 25) and a threshold which requires the sum of weights to be at least 30. With the exception of Palfrey and Rosenthal (1991) all experiments are with complete information about monetary payoffs. Van de Kragt et al. (1983) and Palfrey and Rosenthal (1991) emphasize the importance of communication for successful coordination. Dawes et al. (1986) and Rose et al. (2002) investigate the (positive) influence of refunds of insufficient contributions and Dawes et al. (1986) also the punishment of successful free riding. Goren et al. (2003) find that the sequential-moves game leads to more efficient outcomes than the simultaneous-moves game and Erev and Rapoport (1990) show that, in addition, the information provided to the players in the sequential game matters. Erev and Rapoport (1990), Chen et al. (1996), and McEvoy (2010) find that, in sequential decisions, the pivotality (criticality) of players increases the contribution frequency. Bartling et al. (2015) find that pivotality increases responsibility *attribution*. Sonnemans et al. (1998) is, to the best of our knowledge, the only BTPG experiment where players contribute under a positive and under a negative frame. In the course of repetitions of games they find, in the negative frame, a trend towards less cooperation while we find a tendency towards HS, i.e., more cooperation. The main difference between the experiments is that Sonnemans et al. (1998) investigate 20 rounds with a partner design while we have eight rounds with a stranger design (i.e. a new random selection of groups in every round). In addition, their experiment has only one threshold, namely k=3 of n=5, while we investigate k=1,2,3,4 of n=4.

There are more experimental investigations of Threshold Public Good games with *non-binary* contributions and payoff functions with *two steps*. For an overview see Fischbacher et al. (2011) and Norman and Rau (2015).

Schedule. In the next section, the theory of BTPG games is presented (as far as necessary for the evaluation of the experimental results) and hypotheses are formulated. Section 3 describes the experiment and Section 4 provides results. Section 5 concludes.

2. Equilibria and equilibrium selection

The general theory of BTPG games is developed in Bolle (2015). Here we concentrate on players with equal importance for passing the threshold. In the positive frame, there is a set of players $N = \{1, \dots, n\}$ who can contribute (with costs $c_i > 0$) or not (without costs) to the production of a public good. If a certain threshold k of contributions is surpassed, the public good is produced (the project is launched) and the players earn $G_i > c_i$. There are $\binom{n}{k}$ pure strategy equilibria with the launch of the project where exactly k players contribute. For $k > 1$ there is one pure strategy equilibrium without the launch of the project where no one contributes. Only the latter equilibria and the “all contributing” equilibrium of the Stag Hunt game (k=n) are symmetric pure strategy equilibria. With equal (different) cost/benefit ratios mixed strategy equilibria are symmetric (asymmetric). The symmetry case is proved in Bolle (2015).

The case $G_i < c_i < 0$ is called the negative frame. It can be transformed into the positive frame.

Strategically neutral transformation: By renaming “contribution” as “non-contribution” (and vice versa), exchanging thresholds k and $n - k + 1$, and renormalizing utilities so that "non-contribution/non-launch" has a value of zero, the negative frame is transformed into the positive frame.

Let us assume that the players' contribution probabilities are $p = (p_i)_{i=1,\dots,n}$. $Q = Q(p)$ denotes the probability of success, i.e. that k or more players contribute to the production of the public good. Q_{-i} (Q_{+i}) denote the probability of success if i does not contribute (contributes). These probabilities dependent only on p_j , $j \neq i$. $q_i = Q_{+i} - Q_{-i}$ is the probability that i's contribution is crucial for the production of the public good. With these definitions player i's expected revenue is

(1) $$R_i(p) = G_i * Q(p) - p_i c_i$$
$$= G_i * Q_{-i} + p_i * [G_i * q_i - c_i] .$$

A mixed strategy equilibrium with $0 < p_i < 1$ requires that R_i is independent of p_i, i.e.

(2) $$\partial R_i / \partial p_i = G_i * q_i - c_i = 0.$$

This requirement has been derived verbally by Downs (1957, p. 244) for the binary decision of voting or not. If $G_i * q_i - c_i < (>) 0$ then player i contributes with $p_i = 0\ (1)$. Inserting q_i from (2) into (1) provides us with the equilibrium profit which i expects if she plays a mixed strategy.

(3) $$R_i = G_i * Q_{-i}$$
$$= G_i * Q_{+i} - c_i.$$

Proposition 1: The following statements apply in equilibrium:

(i) If i plays a strictly mixed strategy, then $q_i = r_i = c_i / G_i$.
(ii) $q_i > r_i$ implies $p_i = 1$ and $q_i < r_i$ implies $p_i = 0$.
(iii) In equilibrium, $R_i = G_i Q_{-i}$ applies for $p_i < 1$ and $R_i = G_i Q_{+i} - c_i$ for $p_i > 0$.

Proof: (1), (2) and (3).

In the positive frame, the **case $k = n$** is the **Stag Hunt game,** first discussed by Rousseau (1997 [1762]). There are two symmetric pure strategy equilibria, namely $p = (0, \dots 0)$, $p = (1, \dots, 1)$ and, possibly, a completely mixed strategy equilibrium which is derived from (2) and $q_i = \prod_{j \neq i} p_j$. It follows $p_i = \left(\prod_j r_j\right)^{1/(n-1)} / r_i$. The condition of the existence of this equilibrium is $p_i < 1$ for all i. This condition is always fulfilled for n=2 or if all r_i are identical. Smaller r_i are connected with larger p_i. Because of (3) and $Q_{-i} = 0$ the mixed strategy equilibrium yields zero profits. There are possibly also pure/mixed strategy equilibria where some players contribute with probability 1 and the others play the mixed strategy equilibrium of a reduced Stag Hunt game. According to Proposition 1, those who contribute with probability 1 earn $R_i = G_i * Q_{+i} - c_i \geq 0$ (if $R_i < 0$, this is not an equilibrium) and the mixed strategy players earn zero. Because of Proposition 1 (iii), $p = (1, \dots, 1)$ is the payoff-dominant equilibrium.

Let us, for this case and certain parameters, determine the risk dominant equilibrium under the definition of Harsanyi and Selten (1988). In Bolle (2016) also the Global Games equilibrium selection (Carlsson and van Damme, 1993) is applied to the Stag Hunt game. For our experiments, both principles select $p = (0,0,0,0)$.

Proposition 2: In the case k=n, if $r_i > \prod_{j \neq i} r_j$ for all i then $(0, \dots, 0)$ risk dominates all other equilibria p.

Proof: Appendix A.

Corollary: In Treatments S+ and S- with identical cost/benefit ratios and in Treatments A and B with cost/benefit ratios (0.225, 0.25, 0.275, 0.3) and (0.1, 0.2, 0.3, 0.4), the risk dominant equilibrium in the game with k=4 is $p = (0, \dots, 0)$.

In the positive frame, the **case $\boldsymbol{k} = \mathbf{1}$** is the **Volunteer's Dilemma**, first investigated by Diekmann (1985, 1993). There are n pure strategy equilibria where exactly one player contributes. The only completely mixed strategy equilibrium is derived from (2) and $q_i = \prod_{j \neq i}(1 - p_j)$. It follows $p_i = 1 - \left(\prod_j r_j\right)^{1/(n-1)} / r_i$. Therefore this equilibrium exists under the same conditions as that of the Stag Hunt game. Smaller r_i are connected with smaller p_i (regarded as counter-intuitive by Diekmann 1993). Because of Proposition 1 and $Q_{+i} = 1$, in this equilibrium players earn $R_i = G_i - c_i$, i.e. as much as players who always contribute.

If 1<k<n, n>3, then completely mixed strategy equilibria, if they exist, can be determined only by numerical methods. For n=3, a quadratic equation has to be solved. If k of the n players are necessary for the production of the public good and if all $c_i / G_i = r_i = \rho$ are equal, then, in a completely mixed strategy equilibrium, all $p_i = \pi$ are equal (see Bolle, 2015) and π is derived from

(4) $\rho = q_i = \binom{n-1}{k-1} \pi^{k-1}(1-\pi)^{n-k}$.

For $1 < k < n$, the right hand side of (4) is a unimodal function of π with a maximum at $(k-1)/(n-1)$. Therefore (4) has either two solutions $\pi''(k) > \pi'$ (k) (for small enough ρ) or one solution (border case) or no solution; i.e., completely mixed strategy equilibria do not necessarily exist and, if they exist, generically there are two. In the positive frame, the equilibrium with π'' Pareto-dominates the one with π' and vice versa in the negative frame (Proposition 1 (iv)). In our experimental treatments S+ and S- two completely mixed strategy equilibria exist. All completely mixed or symmetric pure equilibria are reported in Table 2.

Threshold k	1	2	3	4
# pure str. equ.	4	7	5	2
# compl. mixed equ.	≤ 1	≤ 2	≤ 2	≤ 1
# pure/mixed equ.	≤10	≤ 24	≤ 24	≤ 10

Table 1: Number of equilibria if the threshold is "k contributions from 4 players".

Treat.	Type Equ.	k=1	k=2	k=3	k=4
S+	**Compl.Mix+**	**.26**	**0.46**	**.78**	**1**
	ComplMix -	-	.22	.54	.74
	0	-	0	0	0
S-	**Compl.Mix+**	**0**	**.22**	**.54**	**.74**
	ComplMix -	.26	.46	0.78	-
	1	1	1	1	-
A	Compl.Mix+	(.26,.33,.39,.44)	(.52,.59,.66,.74)	(.83,.87,.91,.96)	1
	ComplMix -	-	(.17,.13,.09,.04)	(.48,.41,.34,.26)	(.74,.67,.61,.56)
	0	-	0	0	0
B	Compl.Mix+	-	-	-	1
	ComplMix -	-	-	-	-
	0	-	0	0	0

Table 2: All Symmetric or completely mixed strategy equilibria.

Explanatory note: For symmetric equilibria only the identical contribution probability is reported, in asymmetric cases the vector of contribution probabilities for (player 1, player 2, player 3, player 4). Bold type means: selected by HS. For S+, A, and B ComplMix+ indicates, for k=1, the (if existent) unique completely mixed strategy equilibrium , for k=2 and 3 the Pareto-superior of the (if existent) two completely mixed strategy equilibria and, for k=4, always contributing. ComplMix- does not exist for k=1; for k=2 and 3, it indicates the Pareto-inferior of the (if existent) two completely mixed strategy equilibria and, for k=4, the (if existent) unique completely mixed strategy equilibrium. In the negative frame S- vice versa.

In symmetric games, Harsanyi and Selten (1988) restrict their selection to the set of symmetric equilibria. These can generically be ordered according to Pareto-dominance. Although S+ and S- have only "essentially symmetric" players (different c_i and G_i but identical c_i/G_i) we jnvestigate this selection criterion. Then, in S+, the unique completely mixed strategy of the Volunteer's dilemma is played for k=1, $\pi''(k)$ for k=2 and 3, and $p_i = 1$ for k=4. In S-, HS selects $p_i = 0$ for k=1, $\pi'(k) = 1 - \pi''(n - k + 1)$ for k=2 and 3 and the unique completely mixed strategy of the Stag Hunt game for k=4.

Also in our experimental treatments A with moderately different $c_i/G_i = r_i$ we find two completely mixed strategy equilibria. For the largely different r_i in Treatment B no completely mixed strategy equilibria exist. In all cases, there are many more pure/mixed strategy equilibria (see Table 1). In the case k=4, with parameters from treatments S+, S- or A, there are two pure strategy equilibria (all or no one contributes), there is one completely mixed strategy equilibrium, there are four equilibria where one player contributes with certainty and the others according to case k=n with n=3, and there are six equilibria where two players contribute with certainty and the other two according to case k=n with n=2. In Treatment B several of these equilibria do not exist.

Hypotheses. Every player plays games with eight repetitions in a stranger design and is characterized by his individual contribution frequency ICF, a number between 0 and 8. Aggregate behavior of player types in a game with threshold k can be characterized by frequency distributions $f_i(ICF, k)$ of player types i or, more aggregated, by the average contribution probabilities $ACP_i(k)$. The following hypotheses give (i) to (v) from the introduction an exact meaning. The hypothesis concerning (vi) is necessarily vague.

(H1) In Treatments S+ and S-, large and small players have the same $ACP_i(k)$ and $f_i(ICF, k)$.

(H2) If $ACP_i(k)$ and f_i (ICF) apply for S+ and $ACP_i'(k)$ and f_i' (ICF) for S- then (a) $ACP_i'(k) = 1 - ACP_i(n - k + 1)$ and (b) $f'_i(\mathrm{ICF}) = 8 - f_i(ICF, n - k + 1)$.

(H1) is motivated by Proposition 1 and (H2) by the Strategically Neutral Transformation (see above). Let us now ask whether at least properties of the HS equilibrium in S+ and S- and properties of the Pareto-superior of the two completely mixed strategy equilibria in Treatment A generally apply.

(H3) If, in Treatments A and B, $c_i < c_j$ then $ACP_i(k) < ACP_j(k)$.

(H4) (a) In all treatments, $ACP_i(k)$ increases with k. (b) In all treatments, in the Stag Hunt game (k=n) the payoff dominant equilibrium is selected.

H3 and H4 are inspired by the numerical result of Treatment A in Table 2. Diekmann (1993) has derived H3[5] for the Volunteer's dilemma, but found opposite relations in an experiment. The Pareto-inferior of the two completely mixed strategy equilibria show such an order of equilibrium probabilities (see Table 2) but they will turn out to be much lower than the empirical contribution probabilities.

We do not only want to test the hypothesis that behavior in S+ and S- can be described by the equilibrium selected by HS, but, more general, the hypothesis that behavior of all subjects of the same type (defined by cost/benefit ratio and sign of costs and benefits) is the same.

(H5) In each game the behavior of subjects with the same player type is the same.

Under this hypothesis we should find, for every player type, a binomial distribution of individual contribution frequencies. The basis for this requirement is the implicit assumption that also in finitely repeated games behavior is the same in every round. Therefore we introduce the following hypothesis with an admittedly vague part (b).

(H6) (a) There is no trend in the decisions. **(b)** There are no other dynamic effects.

3. Experiments

All our experimental games had n=4. If at least k players contributed, then every player received a benefit of G_i Lab Dollars. In Treatment 1 (positive frame), players 1 and 2 with $(c_i, G_i) = (4,10)$ are called small players; players 3 and 4 with $(c_i, G_i) = (8,20)$ are called

[5] Diekmann (1993) calls this property paradoxical. Bolle (2015) discusses the relations between costs and equilibrium probabilities in more detail and derives connections between orders of both and strategies being strategic complements and substitutes.

large players. All four players had the same $r_i = c_i/G_i = 0.4$. In Treatment 2 (negative frame), G_i and c_i have the same absolute values as in Treatment 1 but are both negative. In Treatments 3 and 4, every player received a benefit of $G_i = 20$ Lab Dollars. In Treatment 3, contribution costs (c_i) are (4.5, 5, 5.5, 6) Lab-Dollars and players thus have cost/benefit ratios (r_i)=(0.225, 0.25, 0.275, 0.3); in Treatment B costs are (2, 4, 6, 8) Lab-Dollars and cost/benefit ratios (0.1, 0.2, 0.3, 0.4). The (cost, benefit) combination of a player defined his *type*. A player kept his type during the whole experiment. Every subject participated in only one treatment but in four experiments with k=1,2,3,4.

We conducted a computerized laboratory experiment (implemented in a z-tree program design, Fischbacher, 2007) with 8 subjects per session who were randomly assigned to a player type. In Treatments 1 and 2, in every session there were four small and four large players. In Treatments 3 and 4, in every session there were two players of each type. Players participated in 32 games played in 32 periods. In every period, the eight subjects were randomly allocated to two groups with the restriction that the groups consisted of two small and two large players (Treatments 1 and 2) of exactly one player of each type (Treatments 3 and 4). Always after eight games the threshold k changed. During 32 periods all thresholds were adopted in different orders but with the restriction that, across the sessions, each *k* was played about the same times first, second, third, and fourth. Subjects were not informed about the order of the thresholds in the beginning, but only when the threshold was changed. As described above we used a *stranger design*, i.e. the composition of the groups was changed after each round and the co-players could not be identified. Subjects were informed about how many players contributed to the public good but not who contributed. Hence, players were unable to build a reputation. Most experiments took place in the laboratory of the university Viadrina in Frankffurt (Oder), but 18 of the 56 sessions were carried out in the laboratory of Technische Universität Berlin. (See Table 3.) We find a small subject pool effect.

Before subjects played the BTPG games, they were given printed instructions and had the possibility to ask questions. Instructions contained general information, the description of the threshold public good game and two example calculations (see Appendix C). Furthermore, they had to answer five on-screen comprehension questions to make sure that everyone understood the game. The experiment did not start before all subjects had answered the questions correctly. In cases of problems, personal advice was given. In every period, the subjects were reminded of the actual threshold and, every

8th period, the changing of the threshold was announced. In each period subjects were informed in the *decision screen* that the group composition had been changed and they were required to decide whether or not to contribute. In the *profit display screen* they were informed about the number of contributing players and whether the threshold was reached. They further received information about their payoff in the current period.

Treat-ment	Endow-ment	costs c_i	Benefits G_i	c_i/G_i	#sessions (at V, at TU)
S+	8	(4,4,8,8)	(10,10,20,20)	0.4	(10, -)
S-	20	(-4,-4,-8,-8)	(-10,-10,-20,-20)	0.4	(10, -)
A	8	(4.5, 5, 5.5, 6)	20	(0.225, 0.25, 0.275, 0.3)	(6, 12)
B	8	(2, 4, 6, 8)	20	(0.1, 0.2, 0.3, 0.4)	(10, 6)

Table 3: Game parameters (in lab dollars) in the four treatments for players i=1,2,3,4 and number of sessions with eight subjects either at TU (Technische Universität Berlin) or V (Europa-Universität Viadrina Frankfurt (Oder)).

Table 3 shows the parameters of all player types in all treatments. For each lab dollar earned, subjects were paid 4 Eurocents. After the experiment, subjects were presented three incentivized questions testing their understanding of probability calculus. For each correctly answered question (on average two), the subject was paid one additional Euro. Participants earned between 14 and 33 Euros with an average of 23.29 Euros. Sessions lasted roughly 45 minutes.

4. Results

4.1 Average contribution probabilities

In Table 3, 4, and 5, *average contribution frequencies ACPs* are reported. Tests are carried out with respect to our hypotheses.

Result 1 (small vs. large players): In Treatments S+ and S-, *small and large players show similar ACPs except for k=4.* (H1) is rejected for k=4 in both treatments; it is not rejected for other thresholds.

Result 2 (positive vs. negative frame): In Treatments S+ and S-, *ACPs in the positive and the negative frame are mirrored*, i.e. $ACP_i'(k) \ = 1 - ACP_i(n-k+1)$. (H2) (a) is not rejected in any of the eight comparisons.

k	S+			S-		
	SmPl	LaPl	HS	SmPl	LaPl	HS
1	0.35*	0.37*	0.26	0.30	0.26	0
2	0.49*	0.56	0.46	0.43*	0.39*	0.22
3	0.61*	0.63*	0.78	0.57*	0.49*	0.54
4	0.74§	0.81	1	0.75§	0.59	0.74

Table 3: Average contribution probabilities (ACPs) in Treatments S+ (positive frame) and S- (negative frame) and theoretical contribution probabilities according to Harsanyi and Selten (HS). Small player type SmPl with $(G_S,c_S)=(10,4)$ and large player type LaPl with $(G_L,c_L)=(20,8)$. k= threshold.

Explanatory note: § Significant (5% level) two-sided Wilcoxon matched pairs-tests for small vs. large players. * Significant (5% level) one-sided Wilcoxon matched-pairs test of non-increasing ACPs for k (position of *) vs. k+1. No significant results in two-sided Mann-Whitney tests between $ACP(k, c_i, G_i)$ and $1 - ACP(5 - k, -c_i, -G_i)$. All tests are based on averages in 10 sessions and $p<0.05$.

c_i/G_i k	0.225	0.25	0.275	0.3
1	0.389	**0.497***	0.333*	0.250*
2	0.622	0.625	**0.483***	0.483*
3	0.733	0.792	0.733*§	**0.559*§**
4	0.997	0.948	0.931	**0.944***

Table 4: Average contribution frequencies of the four player types in Exp A ($A_{TU} + A_V$).

Explanatory note: There are four significant differences (bold types) between V and TU subjects in two-sided Wilcoxon tests on the 5% level, in three cases higher probabilities in TU, in one case in V. All differences between threshold k and threshold k+1 are significant in two-sided Wilcoxon matched-pairs tests (except k=2, c_i/G_i =0.25 and k=4, c_i/G_i =0.25) on the 5% level. * (§) Significant differences between player types compared to type c_i/G_i =0.225 (0.25) in a two-sided Wilcoxon test on the 5% level.

c_i/G_i k	0.1	0.2	0.3	0.4
1	0.676	0.344*	0.227*	0.277*
2	0.781	0.613	0.398*	0.418*
3	0.930	0.840	0.688*§	0.637*§
4	0.984	0.945	0.918	0.883*

Table 5: Average contribution frequencies of the four player types in Exp B ($B_{TU} + B_V$)

Explanatory note: There are no significant differences between V and TU subjects in two-sided Wilcoxon tests on the 5% level. All differences between threshold k and threshold k+1 are significant in two-sided Wilcoxon matched-pairs tests (except k=1, c_i/G_i =0.1) on the 5% level. * (§) Significant differences between player types, compared to type c_i/G_i =0.1 (0.2) in a two-sided Wilcoxon test on the 5% level.

Results 1 and 2 are remarkable because they show that, in BTPG games contrary to linear public good games, framing does not play an important role. In particular Result 2 challenges the numerous examples where a negative frame reduces the cooperativeness of players.

Result 3 (Efficiency): In treatments A and B, players with lower costs contribute significantly (in 10 of 24 comparisons) more, i.e., the alternative to Hypothesis 3 is rejected.

It is remarkable that we find the same pattern of significant differences in A and B. The smaller k is, the higher is the significance level. The contribution probabilities of the least-costs player type are always significantly higher than those of the highest-costs player type.

Result 4 (Threshold): In all treatments *ACPs increase with the threshold k.* For the transition from k to k+1, the alternative to H4a, namely decreasing or constant ACPs, is rejected in 9 of 12 tests in *S*+ and S- and in all cases in treatments A and B. H4b is rejected but its prediction is far better than the prediction of risk dominance.

We find that ACPs increase from k to k+1, i.e., the alternative to (H5 a), non-increasing ACPs, is significantly rejected in 9 of 12 one-sided Wilcoxon matched pairs tests in S+ and S- and in all cases in A and B. In the positive frame, this means an attempt to meet the increasing threshold, which is apparently stronger than the fear to waste one's contribution if others do not cooperate. In S-, lower k make the players more reluctant to contribute and enjoy the negative costs. In the positive frame with k=1 a player can secure the benefit G_i by contributing; for larger k he has to rely on the cooperativeness when k increases. For k=4 (k=1 in S-), neither the predictions of payoff dominance (HS) nor those of risk dominance (RD) apply, but at least the empirical probabilities are closer to HS than to RD, i.e. to 1 (0 in S-) than to 0 (1 in S-).

Result S (Subject pool): In Treatment B, no significant differences between V and TU subjects are found when comparing the ACPs of a certain player type in a game with a certain threshold, i.e. in 16 two-sided Wilcoxon tests on the 5% level. In Exp A four significant differences are found.

4.2 The distribution of individual contribution frequencies

Every decision situation occurs 8 times so that every subject can contribute to the public good (for every threshold) between 0 and 8 times. We call this number the *individual contribution frequency (ICF)*. The distributions of these ICFs are provided in the appendix for every player type and, for S+ and S-, for the cases that a game had been one of the first two games (periods 1-16) or one of the last two games (periods 17-32). They are only used for second tests of (H1) and (H2) and for the question of whether theories with unique equilibrium selection can apply. In Figures 1 and 2 aggregations over player types and periods are presented. The question of whether there is a trend in the decisions is investigated in a regression analysis below. The individuals with 0 or 8 contributions can be assumed to play a pure strategy or to use mixture probabilities close to 0 or 1. According to this criterion, 40% of our subjects possibly play pure strategies in single games, 14.3% with zero contributions and 25.8% with full contributions.

Result[6] 1': *Players in the same frame with the same cost/benefit ratio show the same distribution of ICFs*. (H1) (b) is not rejected.

We test (H1) (b) by comparing the ICFs of small and large players (from all periods) in a chi-square test, i.e., we compare the frequencies of ICF=0,1,…,8 for k=1,2,3,4. In the positive frame we get $\chi^2 = 31.8$ (df=31[7], p=0.42), in the negative frame $\chi^2 = 34.3$ (df=32, p=0.36).

Result 2': *Behavior in the negative frame is equal to the "mirrored" behavior in the positive frame.* (H2) (b) is not rejected.

At first glance we notice that the frequency of ICF=8 in Treatment 1 with the threshold k=4 meets that of ICF=0 in Treatment 2 with the threshold k=1. If behaviors in Treatments 1 and 2 are completely mirrored, we should find these similarities also for other ICFs. A chi-square test ($\chi^2 = 34.4$, df=32, p=0.35) of this hypothesis, based on the frequencies in Figures 1 and 2, does not indicate significant differences.

[6] Results from chi-square tests based on individual contribution frequencies ICF are indicated with one stroke as 1'. Results from regression analyses in the next subsection are indicated with two strokes.

[7] There is one class (k=1, ICF=6) with zero contributions which is united with a neighboring class.

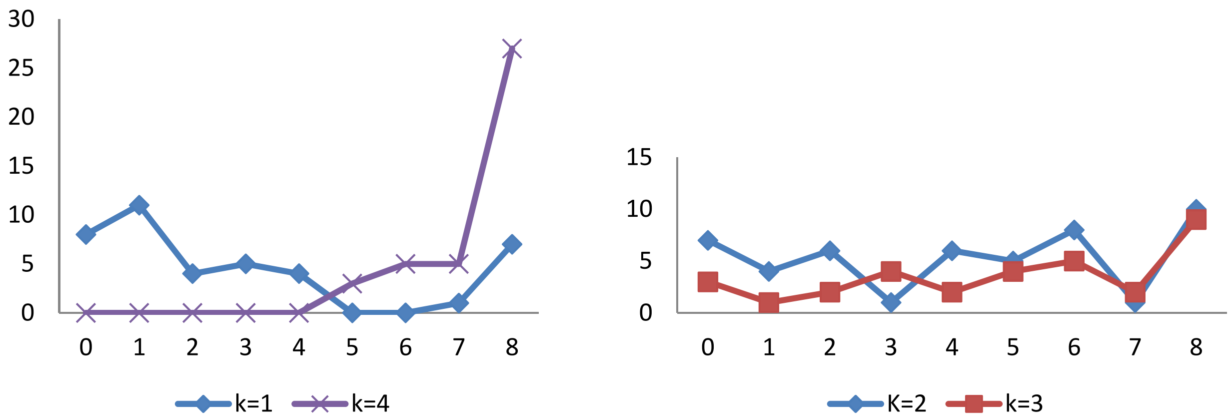

Figure 1: Frequency distribution of individual contribution frequencies (ICFs) of small and large players and for decisions in periods 17 – 32 in Treatment S+. k= threshold. For every k, 8 decisions by 40 individuals.

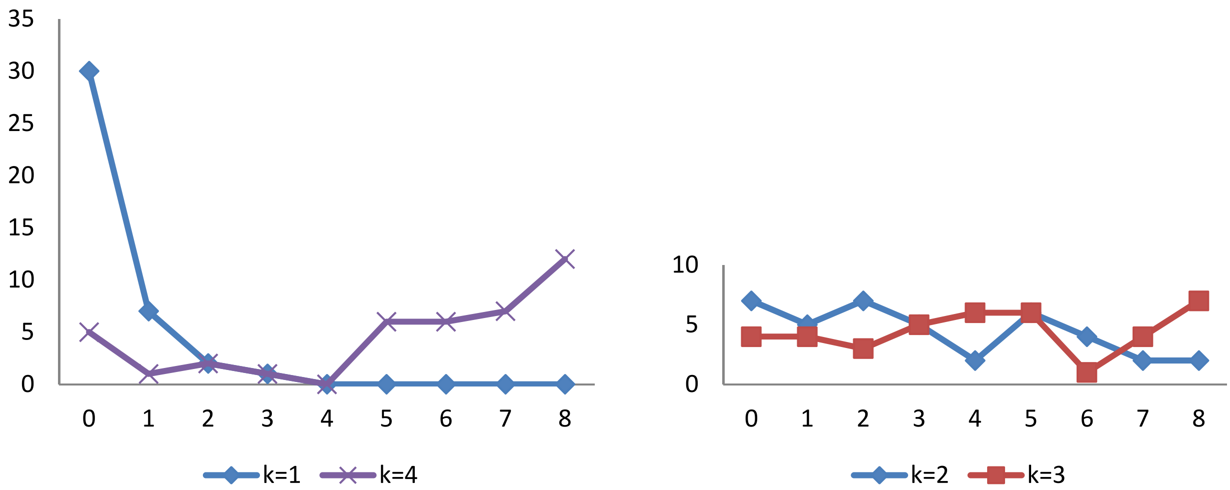

Figure 2: Frequency distribution of individual contribution frequencies (ICFs) of small and large players and for decisions in periods 17 – 32 in Treatment S-. k= threshold. For every k, 8 decisions by 40 individuals.

Result 5’: For every player type and every threshold (except k=4, in some cases) the hypothesis of a binomial distribution of ICFs is significantly rejected.

The distributions in Figures 1 and 2 are certainly not Binomial distributions except, perhaps, k=4 in Figure 1 and k=1 in Figure 2. The former is not significantly different from a binomial distribution (prob=0.044; $\chi^2 = 6.1$, df=7, p=0.53) but the latter is (prob=0.925; $\chi^2 = 16.2$, df=7, p=0.02). For all other distributions in Figures 1 and 2 the hypothesis of a binomial distribution is rejected with extremely small p-values. We get similar results for all the other distributions of ICFs recorded in the appendix. Any single equilibrium and every other hypothesis of identical behavior of the same types of players, however, predicts a binomial distribution of contributions if we assume identical social preferences and beliefs and, possibly, an additional constant error probability.

Result 6': In Treatments S+ and S-, there is a moderate tendency towards more cooperation from the first two to the second two experiments in a session. In chi-square tests for every threshold k, however, the distributions of ICFs are significantly different with $p<0.05$ only for k=4 in S+ ($p=0.0002$) and k=1 in S- ($p=10^{-7}$).

We compare, for treatments S+ and S- and for the aggregation of small and large player behavior, the ICFs in periods 1-16 and in periods 17-32 with a chi-square test, i.e., we ask whether a game with a threshold k is played differently when it was one of the first two games or whether it was one of the second two games. In Treatments A and B with their four *different* player types, there would have been only 10 players distributed to nine classes which would have made a chi-square test rather dubious. Also in the tests for Treatments S+ and S- some classes have zero entries in periods 1-16 as well as in periods 17-32. In such cases neighboring classes are aggregated.

So, learning takes place above all in the Stag Hunt game. The question of dynamic behavior will be posed anew in the next subsection.

4.3 Regression analysis

Table 6 shows the results of a regression analysis for treatments S+ and S- with clustered standard errors on the subject level. Table 7 shows the results for treatments A and B. We introduce the dummy variable PivotPlayer (=1 if k-1 of the other players contributed in the previous round) because Erev and Rapoport (1990), Chen et al. (1996), and McEvoy (2010) found that, in sequential decisions, the pivotality (criticality) of players increases the contribution frequency. Because of Result 6' we introduce variables Period x StH and PivotPl x StH, where StH (=StagHunt) is 1 if k=1 in S- and k=4 in the other treatments. Otherwise StH is 0. Stylized results of our regression analysis are, first, confirmations of previous results, namely:

Result 1'': The dummy for the large player is insignificant.

Result 2'': The "mirror image" character of Treatments 1 and 2 is supported because absolute coefficients of PivotPlayer, and Threshold k are quite similar.

Result 3”: In treatments A and B, the higher the costs the less frequent are contributions. In Treatment B with its larger spread of costs the significance levels are higher.

Result 4’’: The higher the threshold, the more frequent are contributions.

The subject pool effect which we found in 4.1 for Treatment B indicated no unidirectional differences and, thus, it is not confirmed as a significant dummy in the regression. The regression analysis, however, provides new insights concerning dynamic behavior. There is a significant trend only in treatments S- ; but *PivotPlayer* is significant in all treatments. The introduction of the interaction terms lead to lowest BIC values if only *PivotPl x StH was introduced. It showed that having been a Pivot Player in the previous period had a stronger effect in the Stag Hunt game than in the other games.*

Result 6’’: Cooperation increases in the course of a session only in Treatments S-. *Contributing is more (less in S-) probable if a player has been pivotal in the last period.This effect is particularly strong in the Stag Hunt game.*

Variable	S+	S-	S+	S-	S+	S-	S+	S-
Intercept	-1.72*** (0.28)	-0.45 (0.26)	-1.33*** (0.29)	-0.17 (0.27)	-1.42*** (0.29)	-0.05 (0.25)	-1.31*** (0.29)	-0.25 (0.27)
PivotPlayer	0.97*** (0.12)	-1.06*** (0.10)	0.93*** (0.12)	-1.01*** (0.12)	0.77*** (0.13)	-0.71*** (0.12)	0.81*** (0.13)	-0.67*** (0.12)
LargePlayer	0.23 (0.21)	-0.28 (0.20)	0.22 (0.22)	-0.28 (0.20)	0.23 (0.22)	-0.31 (0.21)	0.22 (0.22)	-0.31 (0.21)
Threshold k	0.51*** (0.07)	0.50*** (0.07)	0.36*** (0.08)	0.38*** (0.08)	0.41*** (0.07)	0.31*** (0.07)	0.36*** (0.08)	0.38*** (0.08)
Period	0.01 (0.01)	-0.02** (0.007)	0.01 (0.01)	-0.014* (0.007)	0.01 (0.01)	-0.014 (0.007)	0.01 (0.01)	-0.015* (0.007)
Period x StH	-	-	0.04*** (0.01)	-0.03* (0.01)	-	-	0.02 (0.01)	-0.03 (0.02)
PivotPl x StH	-	.	-	-	1.16*** (0.29)	-2.03*** (0.31)	0.85* (0.33)	-2.44*** (0.40)
Observations	2240	2240	2240	2240	2240	2240	2240	2240
-logL	1369.4	1370.1	1359.3	1364.9	1356.6	1333.9	1354.7	1331.5
BIC	2777.9	2778.5	2764.9	2776.1	**2759.5**	**2714.1**	2763.4	2717.0

Table 3: Logit-Regression of contribution decisions with standard errors (clustered with respect to subjects) in parentheses.

Explanatory notes: *(**,***) - significant at 5%(1%,0.1%)-level. In S+ (S-). StH is 1 for k=4 (k=1) and 0 otherwise. PivotPlayer is 1 if k-1 contributions by other players in previous period and 0 otherwise. Periods=2-8, 10-16, 18-24, 26-32.

	A_{TU}+A_V	B_{TU}+B_V	A_{TU}+A_V	B_{TU}+B_V	A_{TU}+A_V	B_{TU}+B_V	A_{TU}+A_V	B_{TU}+B_V
Intercept	-1.38***	-0.46	-1.09***	-0.31	-1.01***	-0.30	-1.03***	-0.29
	(0.30)	(0.34)	(0.28)	(0.32)	(0,29)	(0.32)	(0.28)	(0.32)
PivotPlayer	0.88***	0.88***	0.81***	0.85***	0.66***	0.77***	0.66***	0.77***
	(0.10)	(0.11)	(0.10)	(0.11)	(0.10)	(0.12)	(0.10)	(0.12)
Player2	0.17	-1.08***	0.17	-1.07***	0.16	-1.07***	0.16	-1.07***
	(0.30)	(0.30)	(0.29)	(0.30)	(0.29)	(0.29)	(0.29)	(0.29)
Player3	-0.38	-1.73***	-0.38	-1.72***	-0.38	-1.78***	-0.38	-1.73***
	(0.27)	(0.31)	(0.27)	(0.30)	(0.27)	(0.31)	(0.27)	(0.30)
Player4	-0.69*	-1.77***	-0.69*	-1.76***	-0.70**	-1.73***	-0.70**	-1.77***
	(0.27)	(0.31)	(0.27)	(0.31)	(0.27)	(0.31)	(0.27)	(0.31)
Threshold k	0.87***	0.92***	0.75***	0.85***	0.68***	0.84***	0.69***	0.84***
	(0.07)	(0.08)	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Period	-0.01	-0.00	-0.01	-0.00	-0.0	-0.00	-0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Period x StH	-	-	0.04*	0.02	-	-	-0.00	0.00
			(0.02)	(0.01)			(0.02)	(0.01)
PivotPl xStH	-	-			1.87***	0.73*	1.94***	0.67*
					(0.35)	(0.29)	(0.35)	(0.32)
Pool V	-0.28	-0.16	-0.31	-0.14	-0.26	-0.16	-0.26	-0.15
	(0.18)	(0.22)	(0.18)	(0.22)	(0.18)	(0.22)	(0.18)	(0.32)
Observations	4032	3584	4032	3584	4032	3584	4032	3584
-logL	2085.5	1731.5	2072.3	1728.2	2047.1	1725.5	2047.0	1725.5
-logL/observ	0.517	0.483	0.514	0.482	0.508	0.481	0.508	0.481
BIC	4237.4	3528.5	4219.3	3530.0	**4168.9**	**3524.6**	4177	3532.8

Table 7: Logit-Regression of contribution decisions with standard errors (clustered with respect to subjects) in parentheses.

Explanatory notes: *(**,***) - significant at 5%(1%,0.1%)-level. In S+ (S-). StH is 1 for k=4 (k=1) and 0 otherwise. PivotPlayer is 1 if k-1 contributions by other players in previous period and 0 otherwise. Periods=2-8, 10-16, 18-24, 26-32.

Moreover, we conducted several alternative regressions to control for various influences. We used sex and field of study and answers to incentivized questions about probability calculus as controls and we used the rounds (from 1 to 8) within a threshold level. We alternatively also conducted all regressions with an autoregressive term. None of these variations change the outcome of the analysis.

5. Conclusion

We have investigated BTPG games in four treatments with 16 different games. Our results cover framing effects and qualitative and quantitative equilibrium predictions.

A surprising result is the lack of framing effects concerning (i) magnitude effects and (ii) positive vs. negative frames. Results[8] 1, 1', and 1'' show that players with the same sign of costs and benefits and the same cost/benefit ratio contribute with the same probability. After substituting thresholds k by n-k+1 and contribution probabilities q by 1-q in the negative frame (mirroring), players contribute with the same probability as players in the positive frame with the same cost/benefit ratio. Note that this is a bit surprising because framing effects are not exceptions but the rule in economic experiments.

Contribution probabilities (iii) decrease with costs and (iv) increase with the threshold k (Results 3, 3'', 4, 4''). If we take the Pareto-superior of the completely mixed strategy equilibria as a benchmark then (iii) contradicts theory and (iv) follows theory. For $1<k<n$ there is often more than one completely mixed strategy equilibrium (or no one) and in every case there are other equilibria so that the benchmark heavily depends on equilibrium selection. The Pareto-superior of the completely mixed strategy equilibria is the HS selection for treatments S+ and S-. Neither in treatments S+ and S- nor in treatments A and B is it a quantitative predictor of behavior. Even more general, (v) no theory with homogeneous players can describe behavior. (Result 5').

The question of whether in the Stag Hunt game the payoff dominant or the risk dominant equilibrium applies is investigated in number of experiments, usually with 2x2 games. In our investigation with 4x2 games, payoff dominance performs considerably better than risk dominance, but 77%, 72%, 97%, and 93% contributions in treatments S+, S-, A, and B are less than the predicted 100%.

Deviations from (vi) static behavior are small in our finitely repeated games experiment with a stranger design; but in addition to a significant trend towards more cooperation in treatment S- an unexpected effect is observed in all treatments. When a player was "pivotal" in the previous period then his contribution probability increases in the positive and decreases in the negative frame. *Pivotality* of player i means that k-1 of i's co-players had contributed.

BTPG is an important class of games with many applications, such as forming teams for producing a public good or preventing a public bad. The most severe obstacle for the application of theory is the plethora of equilibria and the question of their relative and

[8] Rmember that results without stroke are from non-parametric tests comparing the average contribution frequencies ACF in different sessions. Results with one stroke are from chi-square tests based on individual contribution frequencies ICF. Results with two strokes are from regression analyses.

absolute importance. *Concerning equilibrium play* our most important message is (v) that no single equilibrium (nor any other constant behavior) with subjects having identical preferences and beliefs can describe behavior. The conclusion is that, necessarily, we must introduce individual differences. In a game theoretic approach we can assume a distribution of preferences or beliefs or a combination of both. With the same data set as this paper, Bolle (2016) investigates a model with identical preferences for each player type (role dependent preferences, Bolle and Otto, 2015) but different beliefs about the appropriate equilibrium.

All our results mean progress for our understanding of how BTPG games are played and provide also lessons beyond BTPG games. To the best of our knowledge, (i) and (iv) have never been investigated and (ii) only in one paper. (v) is fundamentally important. (vi) is new and surprising because it could not be expected in repetitions of a simultaneous moves game with a stranger design.

Our results are derived from 16 games with different frames and a large variety of cost/benefit ratios. Therefore we do not think that they are driven by a special selection of experimental parameters. The largest challenge for BTPG games is a general theory which meets the reported qualitative and quantitative results.

Acknowledgement: We are grateful for the funding by the German Ministry of Education and Research (FKZ 01LA1139A) and the Deutsche Forschungsgemeinschaft (project BO 747/14-1)

References

Al-Ubaydli, O., Jones, G., Weel, J. (2013) "Patience, cognitive skill, and coordination in the repeated stag hunt" *Journal of Neuroscience, Psychology, and Economics*, *6*(2), 71.

Andreoni, J. (1989) "Giving with impure altruism: applications to charity and Ricardian equivalence" *The Journal of Political Economy*, 1447-1458.

Andreoni, J. (1990) "Impure altruism and donations to public goods: a theory of warm-glow giving" *The Economic Journal*, 464-477.

Andreoni, J. (1995). Warm-glow versus cold-prickle: the effects of positive and negative framing on cooperation in experiments. *The Quarterly Journal of Economics*, 1-21.

Bartling, B., Fischbacher, U., & Schudy, S. (2015). Pivotality and responsibility attribution in sequential voting. *Journal of Public Economics*, *128*, 133-139.

Battalio, R., Samuelson, L., Van Huyck, J. (2001) "Optimization incentives and coordination failure in laboratory stag hunt games" *Econometrica*, 69(3), 749-764.

Bilodeau, M. and Slivinsky, A. (1996) "Toilet Cleaning and Department Chairing: Volunteering a Public Service" *Journal of Public Economics* 59, 299-308.

Bolle, F. (2015) "Costly Voting - A General Theory of Binary Threshold Public Goods" *Discussion Paper, Frankfurt (Oder)*.

Bolle , F. (2016) "A Behavioral Theory of Equilibrium Selection", Manuscript, Frankfurt (Oder),

Bolle, F., Otto, P.E. (2016). "Role-dependent social preferences" Economica, DOI: 10.1111/ecca.12180.

Bullinger, A. F., Wyman, E., Melis, A. P., Tomasello, M. (2011) "Coordination of chimpanzees (Pan troglodytes) in a stag hunt game" *International Journal of Primatology*, *32*(6), 1296-1310.

Büyükboyacı, M. (2014) "Risk attitudes and the stag-hunt game" *Economics Letters*, *124*(3), 323-325.

Chen, X. P., Au, W. T., Komorita, S. S. (1996). Sequential choice in a step-level public goods dilemma: The effects of criticality and uncertainty. *Organizational Behavior and Human Decision Processes*, *65*(1), 37-47.

Dawes, R., Orbell, J., Simmons, R., van de Kragt, A. (1986) "Organizing Groups for Collective Action", *American Political Science Review* 80, 1171-1185.

Diekmann, A. (1985) "Volunteer's Dilemma" *Journal of Conflict Resolution*, Vol. 29/4, 605-610.

Diekmann, A. (1993) "Cooperation in an Asymmetric Volunteer's Dilemma Game – Theory and Experimental Evidence" *International Journal of Game Theory* 22, 75-85

Diekmann, A., Przepiorka, W. (2015). “Punitive preferences, monetary incentives and tacit coordination in the punishment of defectors promote cooperation in humans” *Scientific reports*, 5.

Downs, A. (1957) “An Economic Theory of Democracy” *New York: Harper*.

Dufwenberg, M., Gächter, S., Hennig-Schmidt, H. (2011). The framing of games and the psychology of play. *Games and Economic Behavior*, *73*(2), 459-478.

Erev, I., Rapoport, A. (1990) “Provision of Step-Level Public Goods: The Sequential Contribution Mechanism” *Journal of Conflict Resolution*, Vol. 34/3, 401-425.

Feltovich, N., & Grossman, P. J. (2013) “The effect of group size and cheap talk in the multi-player stag hunt: experimental evidence” *(No. 52-13). Monash University, Department of Economics*.

Fischbacher, U. (2007) “Z-tree: Zurich toolbox for ready-made economic experiments” *Experimental Economics* 10, 10(2):171 – 178.

Fischbacher, U., Güth, W., Levati, M. V. (2011) “Crossing the point of no return: A public goods experiment *(No. 2011, 059). Jena economic research papers*.

Franzen, A. (1995) “Group size and one-shot collective action” *Rationality and Society*, 7, 183-200.

Goeree, J.K., Holt, C.,Moore, A. (2005) “An Experimental Examination of the Volunteer’s Dilemma”, *www.people.virginia.edu/~akm9a/Volunteers Dilemma/vg_paperAug1.pdf.*

Goren, H., Kurzban, R., Rapoport, A. (2003) “Social loafing vs. social enhancement: Public goods provisioning in real-time with irrevocable commitments” *Organizational Behavior and Human Decision Processes*, *90*(2), 277-290.

Harsanyi, J. C., Selten, R. (1988) A general theory of equilibrium selection in games. *MIT Press Books*.

Harsanyi, J.C. (1995) "A New Theory of Equilibrium Selection for Games with Complete Information", *Games and Econonmic Behavior* 8, 91–122.

McEvoy, D. (2010) “Not it: opting out of voluntary coalitions that provide a public good” *Public Choice, Springer*, vol. 142(1), 9-23.

Normann, H.T., Holger A. Rau, H.A.. ”Simultaneous and Sequential Contributions to Step-Level Public Goods: One vs. Two Provision Levels”. *Journal of Conflict Resolution*, 59 (2015), 1273-1300.

Otsubo, H., Rapoport, A. (2008) “Dynamic Volunteer’s Dilemmas over a Finite Horizon – An Experimental Study” *Journal of Conflict Resolution* 52/6, 961-984.

Palfrey, T. R., Rosenthal, H. (1991) “Testing for effects of cheap talk in a public goods game with private information” *Games and economic behavior*, 3(2), 183-220.

Park, E. S. (2000). Warm-glow versus cold-prickle: a further experimental study of framing effects on free-riding. *Journal of Economic Behavior & Organization*, *43*(4), 405-421.

Przepiorka, W., Diekmann, A. (2013) “Individual heterogeneity and costly punishment: a volunteer's dilemma” *Proceedings of the Royal Society of London B: Biological Sciences*, 280(1759), 20130247.

Rapoport, A., Eshed-Levy, D. (1989). Provision of step-level public goods: Effects of greed and fear of being gypped. *Organizational Behavior and Human Decision Processes*, *44*(3), 325-344.

Rose, S. K., Clark, J., Poe, G. L., Rondeau, D., Schulze, W. D. (2002) “The private provision of public goods: tests of a provision point mechanism for funding green power programs” *Resource and energy economics*, 24(1), 131-155.

Rousseau, J.J. (1997 [1776]) “The Social Contract' and Other Later Political Writings, trans. Victor Gourevitch” *Cambridge University Press*.

Rydval, O., Ortmann, A. (2005) “Loss avoidance as selection principle: evidence from simple stag-hunt games” *Economics Letters*, 88(1), 101-107.

Sonnemans, J., Schram, A., Offerman, T. (1998) “Public good provision and public bad prevention: The effect of framing” *Journal of Economic Behavior & Organization*, 34(1), 143-161.

Van de Kragt, A. J., Orbell, J. M., Dawes, R. M. (1983). The minimal contributing set as a solution to public goods problems. *American Political Science Review*, *77*(01), 112-122.

Van Huyck, J. B., Battalio, R. C., Beil, R. O. (1990) “Tacit coordination games, strategic uncertainty, and coordination failure” *American Economic Review*, 80(1), 234-248.

Whiteman, M. A., Scholz, J. T. (2010) “Social Capital in Coordination Experiments: Risk, Trust and Position”. Working Paper, Southern Illinois University Carbondale.

Willinger, M., Ziegelmeyer, A. (1999). Framing and cooperation in public good games: an experiment with an interior solution. *Economics letters*, *65*(3), 323-328.

Appendix

A. Proof of Proposition 4

Risk domination according to HS: For the question whether a mixed or pure strategy equilibrium p risk dominates another equilibrium p' first the *bicentric prior* of p and p' is derived. For BTPG games we have to determine, for every $0 \leq t \leq 1$, whether $a_i = 1$ or $a_i = 0$ is a best response of player i to the other players contributing with probabilities $t * p_{-i} + (1-t) * p'_{-i}$. The shares of t with $a_i = 1$ constitute a vector x of prior probabilities. With these priors the *tracing procedure* is carried out where for every $0 \leq t' \leq 1$ equilibria are determined in a game where player i assumes that, with probability t', the BTPG game is played and with 1-t' the other players decide according to the prior probability. If there is a continuous path of equilibria from t'=0 to t'=1 then the corresponding equilibrium for t'= 1 is selected.

Proposition 4: In the case k=n, if $r_i > \prod_{j \neq i} r_j$ for all i then $(0, \ldots, 0)$ risk dominates all other equilibria p.

Proof: The bicentric priors of the equilibria $(1, \ldots, 1)$ and $(0, \ldots, 0)$ are $(x_i^*) = (r_i)$ and are at least as large as the bicentric priors of any strategy profile p and $(0, \ldots, 0)$. Because of $q_i = \prod_{j \neq i} r_j$ the best response to these priors is $p_i = 0$ (Proposition 1 (ii)). Then there is a constant path of equilibria $(0, \ldots, 0)$ for all t which constitutes the generically unique risk dominant equilibrium. (Lemma 4.17.7 in Harsanyi and Selten, 1988).

B. Data

		Periods 1- 16									Periods 17 - 32								
k	ICF	0	1	2	3	4	5	6	7	8	0	1	2	3	4	5	6	7	8
1		5	4	2	3	1	0	2	1	2	6	5	0	2	3	0	0	0	4
2		2	1	3	1	2	2	1	1	3	3	2	4	1	2	3	4	1	4
3		1	1	1	3	5	4	3	1	5	3	1	1	4	1	1	0	0	5
4		2	1	3	0	4	0	3	4	3	0	0	0	0	0	2	2	3	13

Table A1: Frequency distribution of ICFs (individual contribution frequencies) in Treatment S+, small player. k= threshold. For every k, 8 decisions by 40 individuals.

		Periods 1 - 16									Periods 17 - 32								
k	ICF	0	1	2	3	4	5	6	7	8	0	1	2	3	4	5	6	7	8
1		5	2	2	3	4	0	1	2	1	2	6	4	3	1	0	0	1	3
2		0	2	3	2	2	2	2	0	3	4	2	2	0	4	2	4	0	6
3		1	0	0	6	3	5	3	1	5	0	0	1	0	1	3	5	2	4
4		0	0	1	4	0	6	2	4	3	0	0	0	0	0	1	3	2	14

Table A2: Frequency distribution of ICFs (individual contribution frequencies) in Treatment 1, large player. k= threshold. For every k, 8 decisions by 40 individuals.

		Periods 1 - 16									Periods 17 - 32								
k	ICF	0	1	2	3	4	5	6	7	8	0	1	2	3	4	5	6	7	8
1		3	1	0	3	3	4	1	4	1	16	3	1	0	0	0	0	0	0
2		2	4	2	2	3	4	1	1	1	3	2	4	3	1	4	1	1	1
3		1	1	4	2	5	1	1	3	2	0	2	2	3	5	3	0	1	4
4		0	0	0	0	5	4	7	4	0	2	1	0	0	0	4	5	2	6

Table A3: Frequency distribution of ICFs (individual contribution frequencies) in Treatment 2, small player. k= threshold. For every k, 8 decisions by 40 individuals.

		Periods 1 – 16									Periods 17 - 32								
k	ICF	0	1	2	3	4	5	6	7	8	0	1	2	3	4	5	6	7	8
1		3	1	1	3	5	1	1	2	3	14	4	1	1	0	0	0	0	0
2		4	2	3	2	1	4	1	1	2	4	3	3	2	1	2	3	1	1
3		3	0	3	2	1	2	5	2	2	4	2	1	2	1	3	1	3	3
4		2	1	1	1	3	2	4	1	5	3	0	2	1	0	2	1	5	6

Table A4: Frequency distribution of ICFs (individual contribution frequencies) in Treatment 2, large player. k= threshold. For every k, 8 decisions by 40 individuals.

C. Instructions

Welcome

You are participating in an economic experiment. You will receive your payoff personally and directly after the experiment. The payoff depends on your own decisions and the decisions of your co-players.

Please, turn off your cellphone and similar devices. The entire experiment is conducted on the computer. During the course of the experiment, please do not speak and do not communicate with other participants in any other way.

Below you will find an explanation of the experiment. Please read it carefully. If you have questions notify the experimenter. The experimenter will then answer them. After reading these instructions you will answer several test questions. If you have problems answering these questions, please also notify the experimenter.

Instructions for Treatment 1

- In this experiment you have to make decisions in several periods.
- In each period **groups of 4 players** are built. **You are always player 1** in your group. [In other instructions: Player 2, 3, or 4]
- Each period **each player is endowed with 8 points**.
- Each player can either choose **A** or **B**.
- For now **choosing B** has **no impact** on your points.
- **Choosing A costs**
    - **you and player 2** **4 points each**
    - **player 3 and 4** **8 points each**
- If a **threshold of players choosing A** is reached then
    - **you and player 2** **get 10 points each**
    - **player 3 and 4** **get 20 points each**
- The **level** of this **threshold** is changed **every 8th round**. It is displayed on the screen.
- Each 25 points pays you 1 Euro.

Example

At the beginning of the period you get 8 points. The threshold is 1. Your 3 co-players choose B.

In case you choose A:

	you	player 2	player 3	player 4
points at the beginning of the period	8	8	8	8
costs for choosing A	-4	0	0	0
profit for reaching the threshold	+10	+10	+20	+20
period payoff	14	18	28	28

In case you choose B:

	you	player 2	player 3	player 4
points at the beginning of the period	8	8	8	8
costs for choosing A	0	0	0	0
profit for reaching the threshold	0	0	0	0
period payoff	8	8	8	8