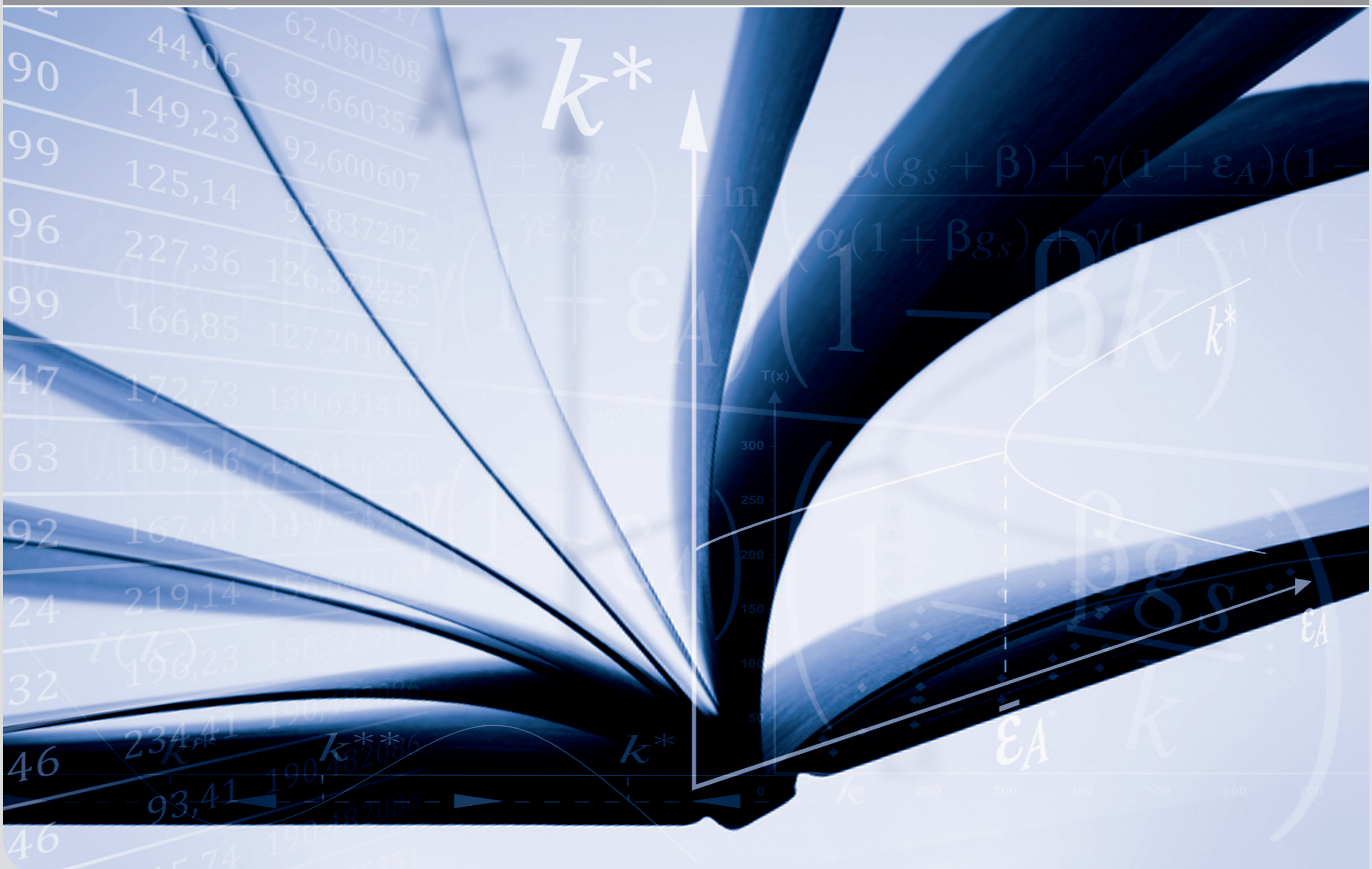
Karlsruher Institut für Technologie

Success rates in simplified threshold public goods games - a theoretical model

by Christian Feige

No. 70 | JUNE 2015

WORKING PAPER SERIES IN ECONOMICS



KIT – University of the State of Baden-Wuerttemberg and
National Laboratory of the Helmholtz Association

econpapers.wiwi.kit.edu

Impressum

Karlsruher Institut für Technologie (KIT)
Fakultät für Wirtschaftswissenschaften
Institut für Volkswirtschaftslehre (ECON)

Schlossbezirk 12
76131 Karlsruhe

KIT – Universität des Landes Baden-Württemberg und
nationales Forschungszentrum in der Helmholtz-Gemeinschaft

Working Paper Series in Economics
No. 70, June 2015

ISSN 2190-9806

econpapers.wiwi.kit.edu

Success rates in simplified threshold public goods games – a theoretical model

Christian Feige[a,*]

[a]*Karlsruhe Institute of Technology (KIT), Institute of Economics (ECON), Neuer Zirkel 3, 76131 Karlsruhe, Germany*

Abstract

This paper develops a theoretical model based on theories of equilibrium selection in order to predict success rates in threshold public goods games, i.e., the probability with which a group of players provides enough contribution in sum to exceed a predefined threshold value. For this purpose, a prototypical version of a threshold public goods game is simplified to a 2×2 normal-form game. The simplified game consists of only one focal pure strategy for positive contributions aiming at an efficient allocation of the threshold value. The game's second pure strategy, zero contributions, represents a safe choice for players who do not want to risk coordination failure. By calculating the stability sets of these two pure strategies, success rates can be put in explicit relation to the game parameters. It is also argued that this approach has similarities with determining the relative size of the strategies' basins of attraction in a stochastic dynamical system (cf. Kandori, Mailath, and Rob, 1993, Econometrica, Vol. 61, p. 29-56).

Keywords: threshold public good, coordination games, learning,

equilibrium selection

JEL: C73, H41

*Phone: +49 721 608 43383, fax: +49 721 608 45471
Email address: christian.feige@kit.edu (Christian Feige)

Preprint submitted to Elsevier *June 26, 2015*

1. Introduction

The general idea of a threshold public goods game (ThrPG)[1] is that a group of people need to jointly provide a given amount of money to fund a project of "public interest", which can mean anything from stopping global warming or building a public library to developing a new gaming software that everybody in the group will enjoy. The problem is that the group members must not only come to an agreement (whether tacit or overt) on whether or not to provide this public good, but also on which player will provide which share of the total.

Roughly fifteen years ago, Croson and Marks (2000) published a meta-study on success rates in ThrPGs, i.e., whether or not the group's total contribution exceeds a predefined threshold value, postulating the "step return," which refers to the ratio of total valuation from reaching the threshold to the necessary threshold contribution, as one of the main explanatory variables in this game. Since then quite a number of additional experimental studies have been concerned with ThrPGs, but to my knowledge there is still no theoretical model which sufficiently explains the experimental data, i.e., why some subject groups reach the threshold consistently, while other groups appear to have no hopes of ever reaching this goal, leading them to converge on an outcome where nobody contributes anything.

In the present study, I will present just such a model, which I derive from previous work on equilibrium selection (Harsanyi and Selten, 1988; Harsanyi,

[1]This type of game is also discussed in the literature under the names "step-level public good" or "provision-point public good", the latter being more applicable to games with some kind of rebate of contributions in case of overcontribution.

1995). The idea is that the players in a ThrPG concentrate on the most focal (in the sense of Schelling, 1980) allocation of the threshold in their group and then decide whether or not this risky Nash equilibrium is to be preferred to the safe choice of contributing zero. I argue that the relative attractiveness of the most focal threshold allocation compared to zero contributions is the main determinant in a class of ThrPGs that give no or only a partial refund of contributions if the threshold is missed. It is only indirectly, via this relation, that the step return and other game parameters affect average success rates.

Admittedly, there have been a number of other attempts in the past to theoretically predict contribution behavior in ThrPGs, but they all have their limitations, if they make accurate predictions at all. A first attempt has been made by Palfrey and Rosenthal (1984), who calculate the equilibria for binary ThrPGs with and without a refund of contributions if the threshold is missed. In binary ThrPGs, each player has only two pure strategies – contribute or not contribute – which means that there is no symmetric pure-strategy equilibrium that exactly reaches the threshold (unless the threshold is equal to the total endowment of all players). Offerman et al. (1998) calculate the quantal response equilibria (McKelvey and Palfrey, 1995) for this type of ThrPG. Goeree and Holt (2005) use a similar approach and are even able to perform a comparative statics analysis for success rates dependent on the number of players and the step return. Yet despite the minimal strategy set, both models can provide only implicit characterizations of the success rate, which could be taken to mean that an explicit model for (binary) ThrPGs simply does not exist.

Recently, Alberti et al. (Unpublished) and Cartwright and Stepanova

(Unpublished) have applied impulse balance theory (see also Ockenfels and Selten, 2005; Selten and Chmura, 2008) to ThrPGs, theorizing that players learn from outcomes in previous rounds and experience a certain drive (impulse) to adapt their contributions afterwards. Just as the quantal response models, theirs ends up being only an implicit characterization of success rates, albeit with a more general applicability to larger individual strategy sets.

What all these models have in common, though, is that they ignore the possibility of equilibrium convergence, i.e., of the idea that the players learn to coordinate their behavior and then attain a stable outcome. Palfrey and Rosenthal (1984) at least mention that "the inefficient pure strategy equilibria of the [game with refund rule] are *weak*" (ibid., p. 180) and therefore inferior, but do not discuss the implications of this result in a repeated game.[2] With quantal response, a concept that *does* account for repeated interactions, the disregard of convergence arises from the assumption that the group composition changes after each round. Offerman et al. (1998) (see also Offerman et al., 1996, 2001) accordingly use a strangers procedure (i.e., randomly changing group compositions) in their accompanying experimental study, controlling for effects of learning through repeated interaction. Unfortunately, however, the large majority of experimental studies involving ThrPGs are repeated games with a *fixed* group composition (partners procedure), so that these models should not be able to predict more than the success rates in "initial" rounds of these experiments.

On the other hand, papers discussing convergence in ThrPGs (or its lack

[2]See also Bagnoli and Lipman (1989, 1992).

thereof) frequently point out that it may be hard to put this "initial" stage in terms of rounds. For instance, Cadsby and Maynes (1999) state that "14 periods did not appear to be sufficient in many cases for convergence to an equilibrium. [...] we increase the number of periods to 25." In contrast, other studies observe convergence to zero contributions after only seven rounds (Guillen et al., 2006) or ten rounds (Isaac et al., 1989; Feige et al., Unpublished). The convergence of total contributions to the threshold level is discussed in several studies by Croson and Marks (1998, 1999), as well as by Cadsby and Maynes (1999), but only on the basis of experimental data, not a theoretical model for why groups should coordinate on a particular threshold equilibrium (instead of, for example, zero contributions).

By developing such a convergence model, much can also be learned about success rates in ThrPGs, because the one is contingent on the other: In order to converge on zero contributions, a group must necessarily fail to reach the threshold. In contrast, the lower the probability of convergence to zero contributions, the higher the success rate. This is the general principle behind the model of a "simplified ThrPG" as it is presented here. The remainder of the paper is structured as follows. After a more detailed motivation of the theoretical approach in Section 2, the theoretical model is derived in Section 3. A comparative statics analysis based on this model is subsequently conducted in Section 4. Section 5 concludes with suggestions for future research.

2. Risk dominance and the probability of playing a particular equilibrium

The theoretical work on equilibrium selection, like Harsanyi and Selten (1988), rarely goes beyond discussing 2 × 2 normal-form games, clearly because a generalized analysis of more complicated games is, well, too complicated to be worthwhile. Having two players with two strategies each is sufficient to create the fundamental part of this problem. Assuming $x_i > y_i > z_i$ for each player $i = 1, 2$, the game shown in Figure 1 has two Nash equilibria in pure strategies: (X, X), which is payoff dominant because it yields the highest payoff x_i to each player i, and (Y, Y), which gives a "safe" payoff of y_i no matter what the other player does. This safe option becomes particularly attractive if $2y_i > x_i + z_i$ for all i, i.e., if (Y, Y) is risk dominant, which also means that there is a conflict between these two dominance criteria in this case. The game also has a unique mixed-strategy equilibrium, which figures prominently in the subsequent theoretical analysis.

		Player 2	
		X	Y
Player 1	X	x_1, x_2	z_1, y_2
	Y	y_1, z_2	y_1, y_2

Figure 1: A 2 × 2 normal-form game with two pure strategy Nash equilibria ($x_i > y_i > z_i$ for each player $i = 1, 2$).

Harsanyi (1995), placing more importance on risk dominance here than in his and Selten's earlier equilibrium selection theory (Harsanyi and Selten, 1988), argues that the probability with which a particular pure strategy is chosen by player i in such a coordination game depends on this strategy's stability set, i.e., the set of mixed strategies of the *other* player against which this pure strategy is a best response for player i, which in turn is determined by the game's parameters x_i, y_i, and z_i. So we can calculate, for example, how increasing z_1 affects the probability that strategy X is chosen by player 1, and even derive the probability that (X, X) results as an outcome. If payoff dominance were the more important selection criterion, the relation between y_i and z_i should not matter at all, only that x_i is greater than both y_i and z_i for a given player i.

Following the reasoning of Harsanyi (1995), the probability that (X, X), i.e., the payoff-dominant equilibrium, results, depends on the relative distance between this equilibrium and the mixed-strategy equilibrium. Figure 2 illustrates this reasoning for the game described above. Subfigure a) shows the strategy space of this game, whereby **X** and **Y** refer to the two pure-strategy equilibria (X, X) and (Y, Y), whereas **M** denotes the mixed-strategy equilibrium. In symmetric games the mixed-strategy equilibrium **M** is located on the straight line from **X** to **Y**, but this need not be the case in a game with asymmetric payoffs. However, the simplex containing the stability sets of player i's pure strategies is indeed one-dimensional, as shown in Subfigure b). Here, the mixed-strategy equilibrium cleanly separates the stability set of strategy X from that of strategy Y. At any point closer to **X** on the simplex, player i will be better off switching to the pure strategy X. Similarly, at any point closer to **Y**, player i will prefer switching to strategy Y.

Let σ_i denote the probability with which player i plays X in the mixed-strategy equilibrium. The probability p that the associated pure-strategy equilibrium (X, X) is played, is then equal to the distance between (X, X)

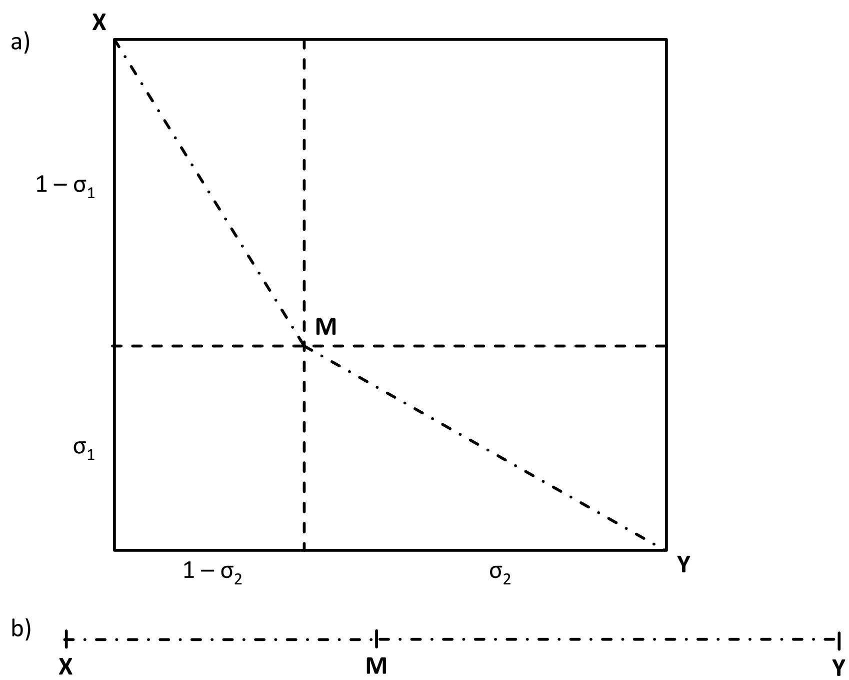


Figure 2: Strategy space (a) and simplex with stability sets (b) of a 2 × 2 normal-form game with two pure-strategy equilibria **X** and **Y** and a single mixed-strategy equilibrium **M**. The probability that equilibrium (X,X) (at point **X**) is played, is equal to the relative distance between **X** and **M**, i.e., $|\overline{XM}|/(|\overline{XM}|+|\overline{MY}|)$.

and **M** relative to the total distance between (X, X) and **M** as well as **M** and (Y, Y):

$$p = \frac{|\overline{XM}|}{|\overline{XM}| + |\overline{MY}|} = \frac{\sqrt{(1-\sigma_1)^2 + (1-\sigma_2)^2}}{\sqrt{(1-\sigma_1)^2 + (1-\sigma_2)^2} + \sqrt{\sigma_1^2 + \sigma_2^2}} \tag{1}$$

In a symmetric game with $x = x_1 = x_2$, $y = y_1 = y_2$, and $z = z_1 = z_2$, the mixed-strategy probabilities in equilibrium play are the same for both players. Accordingly, by letting $\sigma = \sigma_1 = \sigma_2$, (1) can be simplified to

$$p = 1 - \sigma = \frac{x - y}{x - z}. \tag{2}$$

If (X, X) is risk dominant in this symmetric game, we must have $2y < x + z$. (2) then implies that $p > 0.5$. Consequently, and as Harsanyi suggests as well (Harsanyi, 1995, p. 92), the equilibrium with the highest theoretical probability of being played is the risk-dominant outcome.

Note, from (2), that p is equal to the weight placed on strategy Y (i.e., the *other* strategy) in this game's unique mixed-strategy Nash equilibrium. While this may appear a bit confusing at a first glance, it is actually correct and consistent with Harsanyi's "proportionality requirement" for unanimity games[3] (cf. Harsanyi, 1995, p. 106f., Lemmas I and II): The more weight player i puts on strategy Y in the mixed-strategy equilibrium, the greater is this equilibrium's geometrical distance from (X, X), the larger is X's stability set, the higher is the probability that a player will choose X over Y.[4]

[3]For more on unanimity games see Harsanyi and Selten (1988), p. 213ff.

[4]Harsanyi (1995) shows (Lemma I) that using the size of the stability set directly as a proxy does not necessarily work if there are more than two available strategies, making this round-about approach necessary.

Extending the model to n players playing a "$2 \times \ldots \times 2$" game is straightforward. This game still has only two pure-strategy Nash equilibria, **X** and **Y**, in which either all players choose X or all choose Y, respectively, as well as a unique mixed equilibrium **M**.[5] Furthermore, the simplex of this game is once again one-dimensional, with **M** separating the stability sets of the two pure-strategy equilibria. Analogous to above, the theoretical probability p that all players choose X is therefore:

$$p = \frac{\sqrt{\sum_{i=1}^{n} (1-\sigma_i)^2}}{\sqrt{\sum_{i=1}^{n} (1-\sigma_i)^2} + \sqrt{\sum_{i=1}^{n} \sigma_i^2}} \tag{3}$$

At the mixed equilibrium **M**, player i is indifferent between the pure strategies X and Y, but if any other player were to change his own mixed strategy only slightly, either X or Y would immediately become a best response. Consequently, assuming that all other players choose their strategies independently, so that, e.g., player j plays X with probability σ_j, player i faces the following decision problem:

	$(X : \prod_{j \neq i} \sigma_j, Y : 1 - \prod_{j \neq i} \sigma_j)$
X	$(\prod_{j \neq i} \sigma_j) x_i + (1 - \prod_{j \neq i} \sigma_j) z_i$
Y	y_i

The break-even point, for which X and Y yield the same expected payoff to player i and which characterizes the mixed-strategy equilibrium, is given by the following set of equations:

[5]Cf. Kim (1996), Lemma 1, although technically the game described here does not belong to the set of games Π to which the lemma applies, since $\pi_k^H = \pi_{k-1}^H, \forall k < n$. Palfrey and Rosenthal (1984), show a similar result (Proposition 10) for a binary ThrPG.

$$\forall i = 1, \ldots, n : \prod_{j \neq i} \sigma_j = \frac{y_i - z_i}{x_i - z_i} \tag{4}$$

Solving this set of equations for an explicit expression for the mixed strategy σ_i yields the following result:

Lemma 1. *In the n-player two-strategy normal-form game defined above, the equilibrium mixed strategy for player i is given by:*

$$\sigma_i = \frac{x_i - z_i}{y_i - z_i} \sqrt[n-1]{\prod_{j=1}^{n} \frac{y_j - z_j}{x_j - z_j}} \tag{5}$$

Proof: For any two players i and j, divide the respective equations in (4) by each other to receive a new equation containing only σ_i and σ_j. Repeating the process for the same i, but in combination with other players, yields $n - 1$ such two-variable equations. Substituting these equations back into (4), namely into the equation generated from for player i's choice between X and Y, and solving for σ_i yields the above expression. □

For the homogeneous case with n players we can use (3) and (5) to derive a theoretical probability p that all players choose X of

$$p = 1 - \sigma = 1 - \sqrt[n-1]{\frac{y - z}{x - z}}. \tag{6}$$

Consequently, p decreases in larger groups, approaching zero if n approaches infinity, which conforms to the intuition that coordination should be more difficult with more players. In the geometric interpretation of the model, increasing the number of players moves **M** closer to **X**, implying that, for any particular player, Y's stability set becomes increasingly larger relative to X, so that this player has an increasingly lower probability of choosing X over Y.

This generalization to n players also yields a more general definition of risk dominance based on the size the pure strategies' stability sets:[6] Even for more than two players we can say that equilibrium **X** risk dominates equilibrium **Y** if $p > 0.5$, that is, if strategy X has the higher probability of being played. For the game discussed here, this is the case if for all players i, $x_i + (2^{n-1} - 1)z_i > 2^{n-1}y_i$.

3. The simplified ThrPG

While this approach seems to work well for games with only two pure strategies, this may still seem a long way away from ThrPGs with continuous contributions. However, I will argue that a model based on a 2×2 normal-form game is already rich enough to provide a basic understanding of what goes on in even a complicated game like a ThrPG.

3.1. Basic model

A ThrPG consists in a group of n players, each simultaneously choosing their individual contributions to a public account with a threshold T. Each player $i = 1, \ldots, n$ starts with an endowment $e_i > 0$ which can be used to pay for his contribution $q_i \in [0, \bar{q}_i]$ to the public good. The marginal costs of contribution, meaning the conversion rate from endowment to contribution,

[6] This corresponds to one of the characterizations of n-player risk dominance given by Kim (1996) which refers to the relative size of the pure-strategy equilibria's basins of attraction. See also Section 3.4.

is given by $c_i > 0$.[7] Usually, $c_i\bar{q}_i = e_i$ for all i, meaning that the players can spend their entire endowment on contributions, but not more than that.

If the total contribution $Q = \sum_{i=1}^{n} q_i$ is equal to or exceeds the threshold value $T > 0$, i.e., $Q \geqslant T$, each player i receives an individual benefit of $v_i > 0$. Otherwise, the contribution costs are returned to each contributing player at a refund rate of $0 \leqslant r \leqslant 1$. This means that, if $r = 1$, a full refund of contributions is granted, similar to a money-back guarantee. Let $\bar{q}_i < T$ for all players i, as well as $T \leqslant \sum_{j=1}^{n} \bar{q}_j$, so that one player alone cannot reach the threshold, but the entire group can. By assuming $c_i T < \sum_{j=1}^{n} v_j$ for all i, we ensure that reaching the threshold is collectively profitable for all feasible allocations of T among the players.

Player i's payoff $\pi_i(q_i)$ is given by:

$$\pi_i(q_i) = \begin{cases} e_i - c_i q_i + v_i & \text{if} \quad Q \geqslant T \\ e_i - (1-r)c_i q_i & \text{if} \quad Q < T \end{cases} \tag{7}$$

Any vector of individual contributions $\mathbf{q} = (q_1, \ldots, q_n)$, with $c_i q_i < v_i$ for all i, that exactly reaches a total contribution of $Q = T$ is a (strict) Nash equilibrium of this game. If any player decreases his own contribution below this amount, the threshold is missed and the player loses v_i, which is more than the contribution costs $c_i q_i$ that he could save in the process. And by increasing this contribution beyond q_i, the same player only manages to further reduce his endowment to no additional benefit.

[7] The idea of contribution costs, which may be different for different players, has been modeled by Palfrey and Rosenthal (1991) in the case of binary ThrPGs, yet their approach is more closely related to having heterogeneous endowments.

Another equilibrium is constituted by the zero-contribution vector $\mathbf{q^0} = (0, \ldots, 0)$, which arises from the assumption that no player alone can reach the threshold and accordingly should not contribute, if he believes to be the only contributor. Zero contributions is a strict equilibrium in the case of no or only partial refund if the threshold is missed ($r < 1$), but only a weak equilibrium if contributions are fully refunded ($r = 1$). In addition, a full refund establishes an entire set of "weak" Pareto inferior equilibria with a total contribution of $Q < T$. Because of the refund of contribution costs, a player is indifferent to changes of his individual contribution at any of these points, since whatever he contributes, the threshold will not be reached and his payoff will be the same.

3.2. Simplified ThrPG

Given that the general idea behind a ThrPG is very simple, it will be helpful to look separately at the two main components of this game:

1. Will the group reach the threshold or not?
2. Among the large number of possible threshold allocations, which one (if any) does the group choose?

You may notice that the first question is binary: the group will or will not reach the threshold. Assuming that "not reaching the threshold" is the same as an overall and individual contribution of zero (so as not to waste any contributions), this translates into the two strategies of Z (for "zero"), which is contribute zero, and Q_{α_i} (to indicate a positive total contribution "quantity"), which has a particular player i contribute his "fair" (or otherwise assigned) share of the threshold, denoted by α_i. If all players contribute their assigned share, the threshold value is reached exactly.

As long as all players agree on what a "fair share" is, we have already come a long way in simplifying this game. Based on the general model of a ThrPG presented above, we can then specify a simplified ThrPG[8] for a given threshold allocation $\alpha = (\alpha_1, \ldots, \alpha_i, \ldots, \alpha_n)$, where $0 < \alpha_i < 1$ is the relative share of the threshold value provided by player i, resulting in a contribution of $Q_{\alpha_i} = \alpha_i T$.[9] Obviously, $\sum_{i=1}^{n} \alpha_i = 1$. An example for a simplified ThrPG with only two players is given in Figure 3.

		Player 2: Q_{α_2}	Player 2: Z
Player 1	Q_{α_1}	$e_1 - c_1\alpha_1 T + v_1$, $e_2 - c_2\alpha_2 T + v_2$	$e_1 - (1-r)c_1\alpha_1 T$, e_2
	Z	e_1, $e_2 - (1-r)c_2\alpha_2 T$	e_1, e_2

Figure 3: A simplified two-player threshold public goods game with two pure-strategy Nash equilibria.

Note the similarity to the game shown in Figure 1, which effortlessly translates to the n-player case: If $c_i\alpha_i T < v_i$ for all i and $r < 1$, Z yields a guaranteed payoff of e_i to player i, whereas Q_{α_i} yields either strictly more than e_i if all players j choose the complement share Q_{α_j} or strictly less than

[8]This simplified game is not the same as the "reduced game" discussed by (Harsanyi and Selten, 1988) and (Harsanyi, 1995) which eliminates only clearly non-essential game components like dominated or duplicate strategies or players.

[9]Since no individual player can reach the threshold value on his own, $\alpha_i = 1$ is ruled out for any player i. Assigning to any particular player i a share of $\alpha_i = 0$ is tantamount to removing this player from the game and thus equivalent to reducing the number of players by one to only those with strictly positive shares.

e_i if even a single other player chooses Z.[10] The first qualifying assumption implies that the benefit from providing the public good, v_i, is strictly greater than the costs a player suffers from contributing his share, $c_i \alpha_i T$, which simply means that participation in this game must be individually rational. In addition, if $r = 1$, that is if a full refund is granted, any player i always earn at least e_i, even if he contributes and the others do not, making this a riskless choice between Q_{α_i} and Z and thus a case to which the model does not apply. For this reason, it is also assumed that $r < 1$.

Accordingly, the theoretical probability with which the threshold in a public goods game is reached, i.e., its predicted success rate, should likewise depend on the stability sets of two pure strategies. The predicted (or theoretical) success rate then corresponds to the probability with which the pure-strategy equilibrium associated with Q_{α_i} is played. For reasons of simplicity, I will call the equilibria associated with these two pure strategies **Z** and $\mathbf{Q}_\alpha$.

To be true, this approach assumes that the obstacle of selecting one of the large number of feasible threshold allocations, by itself an interesting problem, has already been overcome by the group, meaning that all players already know which one of the many threshold allocations is targeted and compared to **Z**. This is usually not the case in an experimental session where the group members have just come together for the first time for an unfamiliar task. However, theoretical concepts like focal points (Schelling, 1980) or

[10] Here it becomes most apparent what is behind this simplification process, and what is potentially lost in comparison to a more general analysis: In the original game, intermediate outcomes, with some players contributing and others not, may still entail a positive probability that the threshold is reached.

team reasoning (Sugden, 1995) as well as data from previous experimental studies can be used to single out the threshold allocation(s) that will be the most attractive to experimental subjects.[11] I simply assume that any feasible threshold allocation, i.e., any efficient pure-strategy equilibrium of the original ThrPG, is a possible candidate for a "fair" outcome.

As before, we only need to determine the mixed-strategy equilibrium, $\mathbf{M}_\alpha$, in order to calculate the theoretical probability p_α that the associated pure-strategy equilibrium $\mathbf{Q}_\alpha$ is played, which due to the impossibility of overcontribution in this simplified game also equals the theoretical success rate. Let $\sigma_i(\alpha)$ denote the probability with which player i plays Q_{α_i} in the mixed equilibrium $\mathbf{M}_\alpha$ given a particular allocation α. The theoretical success rate p_α is then equal to the distance between $\mathbf{Q}_\alpha$ and $\mathbf{M}_\alpha$ relative to the distance between $\mathbf{Q}_\alpha$ and $\mathbf{M}_\alpha$ as well as $\mathbf{M}_\alpha$ and $\mathbf{Z}$:

$$p_\alpha = \frac{\sqrt{\sum_{i=1}^{n} (1 - \sigma_i(\alpha))^2}}{\sqrt{\sum_{i=1}^{n} (1 - \sigma_i(\alpha))^2} + \sqrt{\sum_{i=1}^{n} \sigma_i^2(\alpha)}} \tag{8}$$

By letting $x_i = e_i - c_i\alpha_i T + v_i$, $y_i = e_i$, and $z_i = e_i - (1-r)c_i\alpha_i T$, we can use Lemma 1 to derive the following:

Corollary 1. *In a simplified ThrPG, the equilibrium mixed strategy for player i is given by:*

$$\sigma_i(\alpha) = \sqrt[n-1]{\frac{1-r}{\prod_{j=1}^{n} [\frac{v_j}{c_j\alpha_j T} - r]}} \left[\frac{v_i}{c_i\alpha_i T} - r\right] \tag{9}$$

[11] Conceivably, the method described by Harsanyi and Selten (1988) could be used just to single out a unique threshold allocation, but it is biased towards payoff-dominant equilibria (see ibid., Section 10.12) and would always predict that the threshold is reached.

Although the model is technically not defined for the case of a full refund, that is if $r = 1$, taking the limit of (9) for r approaching 1 gives $\sigma_i(\alpha) = 0$ and a success rate of $p_\alpha = 1$. This means that $\mathbf{Q}_\alpha$ is the only equilibrium predicted to occur by this theoretical approach in this special case, because $\mathbf{M}_\alpha$ moves closer and closer to $\mathbf{Z}$ as r approaches 1, being located at the same point as $\mathbf{Z}$ in the limit. This is consistent with the fact that $\mathbf{Z}$ is just a "weak" Nash equilibrium in this case and therefore presumably less attractive than the strict equilibrium $\mathbf{Q}_\alpha$, no matter how the game parameters are chosen. Obviously, this establishes a limitation of this model to the class of ThrPGs with no or only a partial refund.

3.3. Success rates for homogeneous games

If the players are homogeneous, so that $e = e_i = e_j, v = v_i = v_j$, and $c = c_i = c_j$ for all players i and j, it is once again possible to simplify (8) further in order to better identify the effects of the particular game elements on success rates. Symmetry can then be used as a justification to also assume $\alpha_i = \alpha_j = \frac{1}{n}$ as a (unique) focal allocation. Similar to the example in Section 2, the mixed-strategy probabilities will consequently be the same for all players as well, i.e., $\sigma = \sigma_i = \sigma_j$. Furthermore, the theoretical success rate is then given by

$$p_\alpha = 1 - \sigma = 1 - \sqrt[n-1]{\frac{1-r}{\frac{nv}{cT} - r}}. \tag{10}$$

Realizing that $\frac{nv}{cT}$ is just the step return SR (Croson and Marks, 2000), the theoretical success rate for a ThrPG with homogeneous players appears to depend only on the step return, the number of players, and the refund rate:

$$p_\alpha = 1 - \sqrt[n-1]{\frac{1-r}{SR-r}}. \tag{11}$$

Alternatively, we can define $\rho := \frac{T}{ne}$ as the proportion of total endowments required to provide the public good,[12] so that the success rate can also be stated as

$$p_\alpha = 1 - \sqrt[n-1]{\frac{1-r}{\frac{ne \cdot v}{T \cdot ec} - r}} = 1 - \sqrt[n-1]{\frac{1-r}{\frac{1}{\rho} \cdot \frac{v}{ce} - r}}. \tag{12}$$

The probability given in (11) also results by translating the "unanimity rule" variant in Palfrey and Rosenthal (1984) into the notation used in this paper.[13]

3.4. Equilibrium convergence

In the introduction to this chapter I have criticized previous theoretical approaches for ignoring the possibility of equilibrium convergence. At a first glance, the model presented above is similarly flawed, because Harsanyi

[12] Since $0 < T \leqslant ne$, we have $0 < \rho \leqslant 1$. By allowing $\rho = 0$ (or $T = 0$) it may also be possible to integrate linear public goods games into this model as an extreme case. However, for all i Q_{α_i} and Z are then indistinguishable, making the parameter v_i meaningless. Instead, linear public goods games grant a financial return on "overcontribution" (commonly referred to as a "rebate") to reward positive contributions. The fact that zero contributions is a dominant strategy in these games is nevertheless consistent with the idea that both "focal" pure strategies coincide in this case so that no coordination problem exists.

[13] Their model assumes that the valuation v is normalized to 1. By letting $q = \sigma$, $M = n$ and $c = 1/SR$, (11) follows from the equation $q = c^{1/(M-1)}$ in Proposition 10 (Palfrey and Rosenthal, 1984, p. 185) in the case of no refund ($r = 0$).

(1995) only discusses one-shot (normal-form) games and the mixed-strategy equilibrium is calculated under the assumption that the players act independently of each other. However, there has been extensive theoretical work in the literature on evolutionary game theory about the relation between risk dominance according to Harsanyi and Selten (1988) and stochastically stable strategies (e.g., Kandori et al., 1993; Kim, 1996; Samuelson, 1997).

The stability sets of a pure strategy are closely relate to the respective equilibrium's basin of attraction, meaning that a measure of their size will also be a predictor of equilibrium convergence. In other words, if there is a high probability that equilibrium $\mathbf{Q}_\alpha$ is played in initial rounds of the experiment, game-play will likely also converge to $\mathbf{Q}_\alpha$ in the long run. Convergence to a threshold equilibrium may or may not increase success rates, though, depending on the remaining volatility of total contributions. Even groups that are very efficient in terms of total contributions may have only low success rates, because the total contribution can just as likely be marginally above or below the threshold. Similarly, some groups may converge more quickly than others and therefore make fewer coordination errors which also affects empirical success rates.

The model presented here cannot capture this kind of convergence behavior (and the associated effect on success rates), because this behavior appears to be concerned with the coordination process for how exactly the threshold should be allocated among the group members. Convergence to zero contributions, on the other hand, is a clear indicator for a collective unwillingness to take the risk involved in providing the public good, and it will obviously lead to significantly lower success rates than if game-play converges to the threshold.

With respect to the original ThrPG, it should be pointed out that the basins of attraction of the threshold equilibria are actually infinitesimally small if contributions are continuous – even though these equilibria are strict, which is usually taken to imply asymptotic stability under a replicator dynamic (cf. Samuelson, 1997, Proposition 2.11 on p.75) – because the set of threshold allocations is convex and the next closest equilibrium is reached in just two infinitesimally small steps.[14] On the other hand, $\mathbf{Z}$'s basin of attraction has a measurable extension (unless $r = 1$), which again varies with the location of the mixed-strategy equilibrium, suggesting this basin's relative size as a suitable measure for the predicted success rate, that is, the smaller the basin, the higher the success rate.

4. Comparative statics

Is the theoretical success rate consistent with the results reported in the experimental literature? A first benchmark in this regard is the meta-study by Croson and Marks (2000). Table 1 translates their main empirical findings[15] into the notation used in this model, whereby + and − denote, respectively, a positive or negative effect on success rates and * denotes statistical significance ($p < 0.05$).

Except for SR, n, and ρ, the independent variables in this meta-analysis are dummies, indicating whether or not a particular treatment has this prop-

[14] The first subtracts a negligible amount $\epsilon > 0$ from the contributions of player i, the second adds the same amount ϵ to the contributions of any other player j.

[15] See Croson and Marks (2000), Table 2.

Table 1: Empirical findings by Croson and Marks (2000).

	SR	n	ρ	Binary
Success rate	$+^*$	$-^*$	$-$	$-^*$

	Refund (r)	Rebate	Homogeneous	Communication
Success rate	$+^*$	$+$	$+$	$+^*$

* denotes statistical significance ($p < 0.05$)

erty. "Binary" refers to treatments that allow only binary contributions. "Refund" applies only to treatments with a full refund ($r = 1$). "Rebate" refers to a return on contributions beyond the required threshold value. Obviously, "Homogeneous" indicates groups with homogeneous players. In the Croson and Marks (2000) meta-study, "Communication" applies to any treatment in which the groups have a face-to-face discussion about the individual contributions, which at that point in time had only been done in two treatments from the study by van de Kragt et al. (1983), however.

As most of the literature is concerned with homogeneous groups (and marginal costs of $c = 1$), I shall restrict the comparative statics analysis to this special case. In (11), the step return SR is in the denominator of a negative term and therefore positively correlated with the success rate. Letting the refund rate r approach 1 makes the fraction it is contained in converge to 0, so that the success rate converges to 1.

The effect of the number of players n is more difficult to determine, because it is also a component of the step return. SR increases in n, because more players receive the same valuation v at the same cost T. However, this upward impulse on success rates is more than compensated in larger groups by the increasing risk of coordination failure. Mathematically, the increasing power of the root term means that the success rate ceteris paribus decreases in larger groups, approaching 0 as n approaches infinity.

In order to determine the effect of the proportion of the threshold value to total endowments ρ, we need to refer to (12) for p_α. Realizing that ρ is inversely proportional to the step return, but otherwise placed in a similar position as SR is in (11), we should expect lower success rates if this proportion is increased.

The results are accordingly quite consistent with the meta-study by Croson and Marks (2000), as shown in Table 1. Although they do not find a statistically significant effect of ρ on the success rate, this does not necessarily mean that (12) is wrong. For one thing, Croson and Marks (2000) include a large number of treatments in their sample that grant a full refund ($r = 1$) to the groups that do not reach the threshold. As, strictly speaking, the model presented in this study does not apply to these treatments, it is still possible that a significant effect of ρ on success rates appears in only the sub-sample of treatments with partial or no refund.

The other dummy variables included in the Croson and Marks (2000) analysis are not fully accounted for in the simplified ThrPG. Although it seems plausible that binary contributions reduce success rates, because of the lack of focal threshold equilibria in pure strategies, this result cannot be derived from the model. Simplifying the ThrPG also abstracts from any effects of a rebate rule on success rates if the Nash equilibrium is unaffected, simply because overcontribution cannot occur.

Homogeneous groups may have an advantage over heterogeneous groups playing a simplified ThrPG as well, given that the only focal allocation in a symmetric game is equal contributions, which is likely to have a comparatively high success rate due to its central location.[16] Interestingly, though, the most frequent type of heterogeneity investigated in the literature, heterogeneous endowments, leads to the same theoretical success rate in the simplified ThrPG as the homogeneous case, provided that the players also coordinate on equal contributions. Since experimental subjects usually do not coordinate on this "risk-minimizing" allocation, the frequently observed lower success rates in heterogeneous groups are in part compatible with the theoretical model.

5. Conclusion

In summary, the main finding of this paper is that success rates in ThrPGs appear to be determined by three different major components:

1. the relative size of the basin of attraction of the zero contribution equilibrium, or respectively this strategy's stability sets
2. the selection process of a unique (focal) equilibrium from the set of threshold equilibria
3. the convergence process (speed and volatility) towards coordination on a specific equilibrium

[16] Put briefly, assigning one player a larger contribution share makes it more risky for him to contribute, so that he is less likely to do so. Even though another player becomes more likely to contribute at the same time as the result of a reduced contribution share, the trade-off of the individual probabilities of contribution is not one-to-one (cf. (4)). A mathematical analysis in fact shows that equalizing the individual probabilities of contribution maximizes the theoretical success rate.

The first of these components is analyzed in more detail, resulting in a model that sets the most prominent parameters of the game in explicit relation to the success rate. As a secondary finding it follows that, in theory, granting a full refund of contributions, if the threshold is not reached, removes the possibility of convergence to zero contributions, suggesting that games with this parameter setting should be treated as an altogether different type of game and be investigated separately.

It should be pointed out that this mathematical model can be nothing more than an approximation, a factor that correlates with observed success rates, but does not provide a formula to directly calculate these rates (like a physical model), let alone explain why some groups are successful, while others are not. What it *can* do is give support for more general behavioral theories which might predict that coordination is more difficult in larger groups or that larger incentives increase the willingness to contribute, but do not exactly state how these two factors will interact. As a consequence, this model may give rise to additional experimental work, in particular examining the effect of the step return in larger groups of, say, thirty or even forty players, or at least methodically varying the number of players in smaller groups.

Future work should also extend the model to cover the other two components as well as other design variations. This extension is likely to create additional "novel" predictions (cf. Lakatos, 1970) to be tested experimentally in order to corroborate (or refute) the model. As this model assumes that a focal threshold allocation can be reached by playing pure strategies, it complements the studies by Palfrey and Rosenthal (1984), Offerman et al. (1998), and Goeree and Holt (2005) who (at least implicitly) assume that the only (focal) symmetric equilibria are in mixed strategies, a fact which makes it difficult for groups to coordinate their behavior and reach the threshold.

References

Alberti, F., Cartwright, E.J., Stepanova, A., Unpublished. Explaining success rates at providing threshold public goods: An approach based on impulse balance theory. Available at SSRN: http://ssrn.com/abstract=2309361.

Bagnoli, M., Lipman, B.L., 1989. Provision of public goods: Fully implementing the core through private contributions. The Review of Economic Studies 56, 583–601.

Bagnoli, M., Lipman, B.L., 1992. Private provision of public goods can be efficient. Public Choice 74, 59–78.

Cadsby, C.B., Maynes, E., 1999. Voluntary provision of threshold public goods with continuous contributions: experimental evidence. Journal of Public Economics 71, 53–73.

Cartwright, E.J., Stepanova, A., Unpublished. Efficiency in a forced contribution threshold public good game. University of Kent, School of Economics Discussion Paper No. 1506.

Croson, R.T., Marks, M.B., 1998. Identifiability of individual contributions in a threshold public goods experiment. Journal of Mathematical Psychology 42, 167–190.

Croson, R.T., Marks, M.B., 1999. The effect of heterogeneous valuations for threshold public goods: an experimental study. Risk Decision and Policy 4, 99–115.

Croson, R.T., Marks, M.B., 2000. Step returns in threshold public goods: A meta- and experimental analysis. Experimental Economics 2, 239–259.

Feige, C., Ehrhart, K.M., Krämer, J., Unpublished. Voting on contributions to a threshold public good - an experimental investigation. KIT Working Paper Series in Economics, Paper 60.

Goeree, J.K., Holt, C.A., 2005. An explanation of anomalous behavior in models of political participation. American Political Science Review 99, 201–213.

Guillen, P., Schwieren, C., Staffiero, G., 2006. Why feed the Leviathan? Public Choice 130, 115–128.

Harsanyi, J.C., 1995. A new theory of equilibrium selection for games with complete information. Games and Economic Behavior 8, 91–122.

Harsanyi, J.C., Selten, R., 1988. A General Theory of Equilibrium Selection in Games. MIT Press.

Isaac, R.M., Schmidtz, D., Walker, J.M., 1989. The assurance problem in a laboratory market. Public Choice 62, 217–236.

Kandori, M., Mailath, G.J., Rob, R., 1993. Learning, mutation, and long run equilibria in games. Econometrica 61, 29–56.

Kim, Y., 1996. Equilibrium selection in n-person coordination games. Games and Economic Behavior 15, 203–227.

van de Kragt, A.J.C., Orbell, J.M., Dawes, R.M., 1983. The minimal contributing set as a solution to public goods problems. The American Political Science Review 77, 112–122.

Lakatos, I., 1970. Falsification and the methodology of scientific research programmes, in: Lakatos, I., Musgrave, A. (Eds.), Criticism and the Growth of Knowledge. Cambridge University Press, Cambridge, and others, pp. 91–196.

McKelvey, R.D., Palfrey, T.R., 1995. Quantal response equilibria for normal form games. Games and Economic Behavior 10, 6–38.

Ockenfels, A., Selten, R., 2005. Impulse balance equilibrium and feedback in first price auctions. Games and Economic Behavior 51, 155–170.

Offerman, T., Schram, A., Sonnemans, J., 1996. Value orientations, expectations and voluntary contributions in public goods. The Economic Journal 106, 817–845.

Offerman, T., Schram, A., Sonnemans, J., 1998. Quantal response models in step-level public good games. European Journal of Political Economy 14, 89–100.

Offerman, T., Schram, A., Sonnemans, J., 2001. Expectation formation in step-level public good games. Economic Inquiry 39, 250–269.

Palfrey, T.R., Rosenthal, H., 1984. Participation and the provision of discrete public goods: A strategic analysis. Journal of Public Economics 24, 171–193.

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, 183–220.

Samuelson, L., 1997. Evolutionary Games and Equilibrium Selection. MIT Press, Cambridge, Mass., London.

Schelling, T.C., 1980. The Strategy of Conflict. Harvard University Press.

Selten, R., Chmura, T., 2008. Stationary concepts for experiental 2×2-games. The American Economic Review 98, 938–966.

Sugden, R., 1995. A theory of focal points. The Economic Journal 105, 533–550.

Working Paper Series in Economics

recent issues