

Long-Run Selection and the Work Ethic*

Jens Josephson[†] Karl Wärneryd[‡]

September 21, 2004

*We thank audiences at Universitat Pompeu Fabra and at the Second World Congress of the Game Theory Society, Marseille 2004, for helpful comments. Josephson thanks the Jan Wallander and Tom Hedelius Foundation for financial support.

[†]Department of Economics and Business, Universitat Pompeu Fabra, Ramon Trias Fargas 25–27, 08005 Barcelona, Spain, and Centre de Referència d'Economia Analítica (CREA). Email: **Jens.Josephson@upf.edu**.

[‡]Department of Economics, Stockholm School of Economics, Box 6501, S–113 83 Stockholm, Sweden, and CESifo. Email: **Karl.Warneryd@hhs.se**.

Abstract

That individuals contribute in social dilemma interactions even when contributing is costly is a well-established observation in the experimental literature. Since a contributor is always strictly worse off than a non-contributor the question is raised if an intrinsic motivation to contribute can survive in an evolutionary setting. Using recent results on deterministic approximation of stochastic evolutionary dynamics we give conditions for equilibria with a positive number of contributors to be selected in the long run. *Journal of Economic Literature* Classification Numbers: C72, D23, H41, M14, Z13. Keywords: work ethic, evolution, group selection, public goods, stochastic dynamics.

Organizations endure [...] in proportion to the breadth of the morality by
which they are governed.

Chester I. Barnard, *The Functions of the Executive*, 1938.

1 Introduction

In a social dilemma the provision of a public good requires some costly effort from one or more individuals, but explicit contracting is not possible. Each participant in a dilemma therefore has an incentive to free-ride on the contributions of the others. How social institutions mitigate social dilemmas is the subject of a large literature both in economics and the social sciences more generally. In this paper, we instead study the evolutionary survival properties of an intrinsic motivation to contribute in dilemma situations.

Consider the case of cooperation within the firm. When individual contributions to output are not verifiably measurable, complete contracting within the firm is impossible, and a free-rider problem is present. This observation has been used as the basis for explanations of the organizational structure of the firm, such as the monitoring and budget-balance-breaking roles of an outside owner discussed by Alchian and Demsetz [1] and Holmström [7], respectively.

But free-rider problems in production may also be overcome if individual agents have an internalized work ethic. Experimental evidence of long standing suggests that

1. people *do* in fact contribute in dilemma situations even though contribution is not enforceable, and
2. the rate of contribution is declining in group size and the individual cost of contributing.

(See, e.g., the classic study of Latané, Williams, and Harkins [10]. For a survey specifically of the experiments on the underlying model of the present

paper, see Croson and Marks [6].) In a broad sense, laboratory behavior therefore more or less conforms to some of the possible predictions of the noncooperative public goods provision model introduced by Palfrey and Rosenthal [13, 14], which we shall take as our basic model in this paper.

A version of this model was originally proposed to explain phenomena such as positive turnout in elections. Given that other people vote, the probability of being pivotal is small. If voting is costly to the individual voter, how can it be rational to vote? The earlier political science literature puzzled over this. It was often suggested that observed voting behavior could only be explained by there being utility directly attached to the act of voting.

This notion is problematic for at least two reasons. First, it cannot be rational for everyone to abstain, either, since then the probability of being decisive is one for an individual who does decide to vote. Hence there must be some positive turnout in equilibrium. This is the point made by Palfrey and Rosenthal. But secondly, even if individuals are intrinsically motivated to vote, or, more generally, to contribute toward the provision of a public good, how could such a behavioral trait survive in the long run if it is costly to its carrier?

Since in any given dilemma interaction a contributor is always strictly worse off than a non-contributor, the intrinsic motivation to contribute—what we shall think of as a *work ethic*—could not evolve unless there were many interactions going on at the same time, and behavioral material from one interaction might end up in another later on. If the latter is the case, then the direct fitness disadvantage of contributors may be balanced by the indirect cost to noncontributors arising from the risk of being matched into an interaction where there are too few contributors and where the public good is therefore not provided. When this happens, we are dealing with group selection in the sense of, e.g., Sober and Wilson [19].

The general Palfrey-Rosenthal model typically has multiple equilibria, some of which involve positive contributions, others which do not. In this paper we embed the game in a dynamic evolutionary setting with random

mutations (along the lines of Young and Foster [23] and Kandori, Mailath, and Rob [8]), which allows us to predict the equilibria we would be likely to observe in the long run. By evolution we shall mean cultural evolution by means of imitation of successful behavior, but our model is also open to other interpretations. We show that for group sizes small enough, and a cost of contributing low enough, in the long run we are most likely to observe a positive amount of contributions even when also non-contribution is a stable state of the deterministic model.

The paper is organized as follows. We introduce the underlying model by means of a simple example, and note the relation to group selection theories, in Section 2. Section 3 studies the deterministic evolutionary dynamics of the more general model. We note that there are generally evolutionarily stable equilibria involving no contribution. Section 4 gives examples of all the possible dynamic behaviors of the deterministic model. In Section 5 we introduce mutations, and give conditions for equilibria with a positive measure of contributors to be selected in the long run. Section 6 concludes. All proofs of propositions are in the Appendix.

2 Group Selection and the Work Ethic

When contributors survive evolutionary selection in social dilemma interactions, as we shall show may happen, it is an example of *group selection* at work. Group selection is natural selection where what matters is the differential fitnesses of groups, not differences in fitness between individuals within the same group. Of special interest are situations in which a trait that is disadvantaged *vis-à-vis* other traits within a group can nevertheless grow in relative population representation over time.

Already Charles Darwin noted that the survival of traits seemingly disadvantageous to the individual organism might be explained by their benefits for the group. The central problem that group selection theories have faced

		Individual 2	
		Worker	Shirker
Individual 1	Worker	$1 - k, 1 - k$	$1 - k, 1$
	Shirker	$1, 1 - k$	$0, 0$

Table 1: An example group interaction.

since then is to define precisely what constitutes a group. Only fairly recently, thanks to the work of, e.g., Sober and Wilson [19], has some clarity been reached on this issue.

Although group selection effects have recently attracted some attention in economics (see, e.g., Sjöström and Weitzman [18], Bergstrom [5], Robson [15], and Kuzmics [9]), it is seldom pointed out that the potential for such effects is already built into the standard evolutionary game theory model.

Consider the following example, a special case of the more general model (due to Palfrey and Rosenthal [13]) that we shall study in this paper. A group consists of two individuals. An individual is programmed¹ to be either a *worker* or a *shirker*. If there is at least one worker in the group, the group is successful, and each individual gets a gross payoff, or *fitness*, of 1. Otherwise each individual gets a payoff of zero. A worker also always sustains the cost k , with $0 < k < 1$. That is, the payoff structure of the group interaction is as shown in Table 1.

Note that if there is one worker and one shirker in the group, the fitness of the worker is strictly less than that of the shirker. Hence we may say that workers are strictly disadvantaged relative to shirkers in a group. (Note that this does not mean that being a worker is a strictly dominated strategy.) But if the group consists of two shirkers, they will both do less well than does a

¹A related paper, Lohmann, Oechssler, and Wärneryd [11], studies the survival properties of altruistic preferences in a social dilemma where individuals behave rationally given their preferences and their information about the interaction situation.

worker in a group with a shirker. A group with two shirkers is not successful. That is, although a worker always has lower fitness than a shirker *in the same group*, in a group with at least one worker everyone has greater fitness than individuals in a group with only shirkers. Hence workers increase group fitness at the expense of their own individual fitness.

Suppose now that there are many groups, composed randomly out of a large population where the overall share of workers is $x \in [0, 1]$. The average fitness of workers is then $1 - k$, and that of shirkers is $x \cdot 1 + (1 - x) \cdot 0 = x$. Hence if we have $x < 1 - k$, workers have greater average fitness than shirkers and would tend to grow in relative representation in the population. Conversely, if at any time we have $x > 1 - k$, the population proportion of shirkers would increase. Hence $x = 1 - k$ is the only stable proportion of workers. Although workers suffer the cost k , shirkers run the risk of ending up in a group that is not successful.

Sober and Wilson [19] identify two properties necessary for group selection of locally disadvantaged traits to occur—that group interactions are isolated from each other in payoff terms, i.e., that what happens in one group does not affect what happens in another, and that there is mixing of behavioral material over time, so that the offspring of players in one group may end up in another, differently composed, group. This also highlights the fact that evolutionary game theory in general (see, e.g., Mailath [12] for a survey) is, in essence, about group selection. The standard evolutionary game theory model concerns individual agents from a large population randomly matched and re-matched to play games (typically, two-player games). The match is, of course, a group.

As we have seen, the large-population version of our example game has a unique stable evolutionary equilibrium with a positive frequency of workers. In the following we shall show, however, that in a more general setting, with more than two players in a group and a contribution threshold greater than one, there is always a stable equilibrium with no workers as well. We shall eventually give conditions for an equilibrium with positive contribution

levels to be selected in the long run when the system is subjected to random mutations.

3 Deterministic Evolution

We now generalize the example from the preceding section. We assume there is a large number N of individuals, each of whom is either a *worker* or a *shirker*. The share of workers in the population is $x \in [0, 1]$.

Individuals are randomly matched² into *firms* that have $n \geq 2$ members each. Each firm's technology is such that in case there are at least m workers in the firm, the firm is successful and each firm member gets a gross payoff or fitness of 1. We assume $1 \leq m \leq n$. As before, workers also always sustain the fitness cost k , with $0 < k < 1$.

Letting μ be the random number of *other* workers in a matching, the expected fitness of a worker is therefore

$$u_w(x) := \text{Prob}(\mu \geq m - 1) \cdot 1 + \text{Prob}(\mu < m - 1) \cdot 0 - k,$$

and that of a shirker

$$u_s(x) := \text{Prob}(\mu \geq m) \cdot 1 + \text{Prob}(\mu < m) \cdot 0.$$

Hence the average expected fitness in the population is

$$\bar{u}(x) := xu_w + (1 - x)u_s = x(\text{Prob}(\mu \geq m - 1) - k) + (1 - x)\text{Prob}(\mu \geq m),$$

and the difference between the expected fitness of a worker and the population average is

$$u_w(x) - \bar{u}(x) = (1 - x)(\text{Prob}(\mu = m - 1) - k).$$

²In the following, we shall ignore complications arising from the fact that the population is actually finite, i.e., we shall assume the law of large numbers applies.

Since μ is binomially distributed we have that

$$\text{Prob}(\mu = m - 1) = g(x) := \binom{n-1}{m-1} x^{m-1} (1-x)^{n-m}.$$

Now assume actions are taken at discrete times $t \in \{1/N, 2/N, \dots\}$. In every period exactly one individual, drawn at random, has the opportunity to revise his strategy.

We shall assume the strategy updating proceeds by means of imitation. The individual who gets the opportunity to revise observes the action choice and current expected payoff of a randomly drawn other individual, and may choose to imitate the behavior of the observed individual.

It seems reasonable to assume that the probability of imitation is increasing in the current payoff advantage of the sampled action. In a non-strategic setting, Schlag [16, 17] shows that this is in fact the *optimal* scheme for imitation. Apestegua, Huck, and Oechssler [2] provide experimental results that suggest this model also agrees well with actual behavior.

We shall here assume the probability of imitation is in fact exactly equal to the approximate expected payoff difference between the strategy of the randomly drawn other player and that of the revising individual, provided this difference is positive. This behavior defines a Markov chain X^N on the space $\Delta^N X = \{0, 1/N, 2/N, \dots, 1\}$ with the transition probabilities

$$P(x, x + 1/N) = x(1-x) \max\{g(x) - k, 0\},$$

$$P(x, x - 1/N) = x(1-x) \max\{k - g(x), 0\},$$

and

$$P(x, y) = 0 \text{ for } |x - y| \geq \frac{2}{N}.$$

The difference between the first two transition probabilities gives the expected net increase in the population of workers from one transition time to the next, conditional upon the current state x , as

$$F_W(x) = P(x, x + 1/N) - P(x, x - 1/N) = x(1-x)(g(x) - k).$$

Note that this function is bounded and continuous in x . We are interested in the deterministic approximation of the above Markov chain when the population size is large and thus the time interval between successive transition times is short. The associated mean-field equation

$$\dot{x} = \varphi(x) := x(1-x)(g(x) - k)$$

gives the limiting deterministic dynamic as N tends to infinity. Benaïm and Weibull [4] prove that for large population size, the deterministic dynamic is a good approximation of the unperturbed Markov chain. In particular, by Proposition 4 of Benaïm and Weibull [4], as N tends to infinity, the asymptotic frequency distribution of X^N will almost surely put all probability mass on the set of stationary states of the deterministic dynamic.

The function φ is also the deterministic *replicator dynamics* of evolutionary game theory, since it says that the growth rate of workers is equal to the difference between their current average fitness and the population average. (See Taylor and Jonker [20].) The original purpose of the replicator dynamics was to serve as a model of asexual genetic reproduction. We have just seen that it may also be interpreted in terms of imitation behavior.

A *restpoint* of this dynamics is a point x such that $\varphi(x) = 0$. A restpoint x is *interior* if $0 < x < 1$.

Define

$$k_{\max} := \binom{n-1}{m-1} \left(\frac{m-1}{n-1}\right)^{m-1} \left(\frac{n-m}{n-1}\right)^{n-m}.$$

The following result about restpoints and all other results in the text are proved in the Appendix.

Proposition 1 *Under the replicator dynamics, $x = 0$ and $x = 1$ are always restpoints. In addition,*

1. *if $m = 1$, then $x = 1 - k^{1/(n-1)}$ is a restpoint,*
2. *if $m = n$, then $k^{1/(n-1)}$ is a restpoint,*

3. if $1 < m < n$ and $k = k_{\max}$, then $x = (m - 1)/(n - 1)$ is a restpoint, and
4. if $1 < m < n$ and $k < k_{\max}$, then there are two interior restpoints, one strictly less than $(m - 1)/(n - 1)$, the other strictly greater than $(m - 1)/(n - 1)$.

There are no other restpoints.

The interior restpoints correspond to the symmetric mixed strategy equilibria identified by Palfrey and Rosenthal [13], who studied the static version of this model.

A restpoint is *asymptotically stable*—henceforth, *stable*—if, loosely speaking, a small perturbation causes the system eventually to return to the restpoint. In the present context a restpoint is stable if the derivative of $\varphi(x)$ with respect to x is negative at the restpoint. We can show the following about the stability of restpoints.

Proposition 2 *Under the replicator dynamics,*

1. if $m = 1$, then $x = 0$ and $x = 1$ are unstable restpoints, and $x = 1 - k^{1/(n-1)}$ is a stable restpoint,
2. if $m = n$, then $x = 0$ and $x = 1$ are stable restpoints, and $x = k^{1/(n-1)}$ is an unstable restpoint, and
3. if $1 < m < n$, then $x = 0$ is a stable restpoint, $x = 1$ is an unstable restpoint, and furthermore
 - (a) if $k = k_{\max}$, then $x = (m - 1)/(n - 1)$ is an unstable restpoint, and
 - (b) if $k < k_{\max}$, the lesser of the two interior restpoints is unstable and the greater is stable.

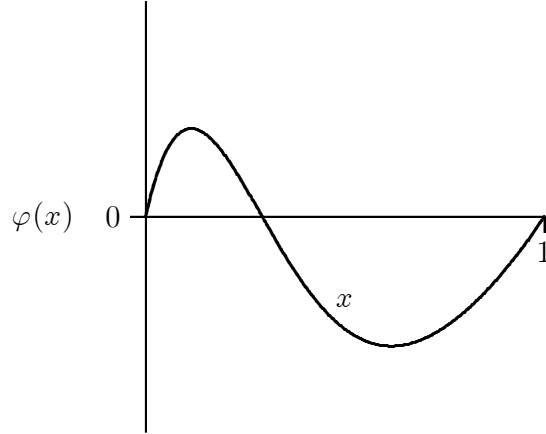


Figure 1: $n = 5, m = 1, k = .25$.

We note in particular that as long as we have $m > 1$, there is always a stable restpoint with no workers. When there is also a stable restpoint with a positive measure of workers, as is the case except when $m < n$ and $k \geq k_{\max}$, we need to consider perturbations of the system in order to make sharper predictions about the ultimate outcome.

4 Examples

In this section we discuss some examples that illustrate all the possible dynamic behaviors of the deterministic model.

Throughout, we shall set $n = 5$. In the phase diagram of Figure 1, we have $m = 1$ and $k = .25$. There is therefore a unique stable interior restpoint at $x \approx .29$.

Note that this restpoint is inefficient. In general, the expected number of workers in a firm when the population share is x is equal to nx . Hence in the present case it is approximately 1.45. It would be efficient to have exactly one participant working, but we have assumed that such a contract is not

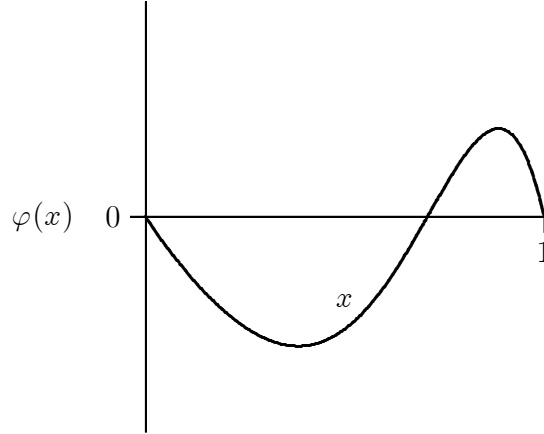


Figure 2: $n = 5, m = 5, k = .25$.

possible.

Figure 2 shows the situation when all participants in the firm have to be workers in order for the firm to be successful, still with $k = .25$. In this case, there is an unstable interior restpoint at $x \approx .71$. Starting points below $.71$ therefore ultimately converge to $x = 0$, starting points above to $x = 1$. The unstable interior restpoint measure of workers is thus the *critical mass* necessary for convergence to a population where everyone is a worker. We note that the latter situation is the unique efficient one.

The dynamical system in Figure 3 has an unstable restpoint at $x \approx .08$ and a stable one at $.5$. In this case, the lower interior restpoint functions as a critical mass for convergence to the higher one. For future reference, denote the lesser interior restpoint when there are two interior restpoints by x_L , and the greater interior restpoint by x_H . Again, the stable interior restpoint x_H is inefficient, since the expected number of workers in a firm is here $2.5 > 2$.

Figure 4 shows an example where we still have $m = 2$, but now we have $k = .5 > k_{\max} \approx .42$. Hence there are no interior restpoints, and the only stable state is $x = 0$.

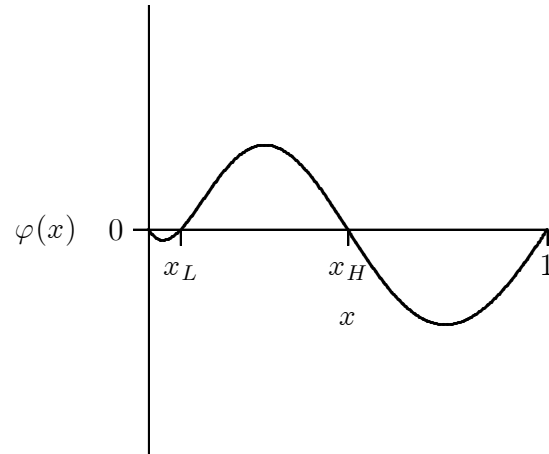


Figure 3: $n = 5, m = 2, k = .25$.

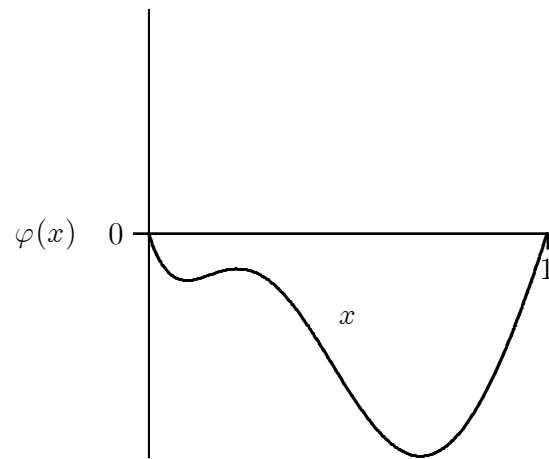


Figure 4: $n = 5, m = 2, k = .5$.

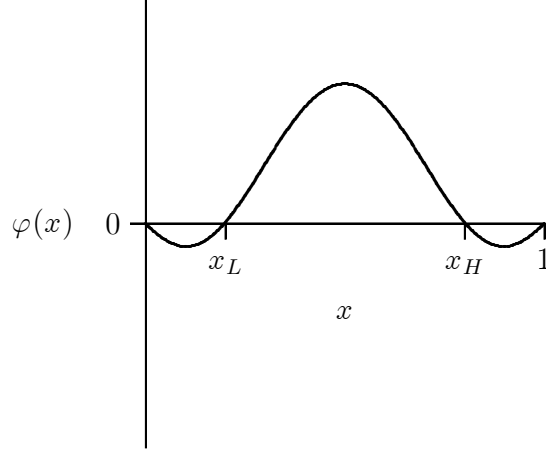


Figure 5: $n = 5, m = 3, k = .15$.

If we have n odd and $m = (n + 1)/2$, φ is symmetric around a maximum at $.5$. Since with this parameter relationship a firm is successful if more than half of its members are workers, we might call such technologies *simple-majority* technologies. Figure 5 shows an example with $n = 5$, $m = 3$, and $k = .15$.

When we have $k < k_{\max}$, simple-majority technologies allow for analytical solutions for the interior restpoints. They are

$$x_L = \frac{1}{2} - \frac{1}{2} \left(1 - 4 \left(k / \binom{n-1}{(n-1)/2} \right)^{2/(n-1)} \right)^{1/2}$$

and

$$x_H = \frac{1}{2} + \frac{1}{2} \left(1 - 4 \left(k / \binom{n-1}{(n-1)/2} \right)^{2/(n-1)} \right)^{1/2} .$$

In the example, we have $x_L \approx .20$ and $x_H \approx .80$.

5 Long-Run Dynamics

Biological evolution is subject to mutations, and in the context of cultural evolution by means of imitation of successful behavior it makes sense to assume that individuals may make mistakes in selecting their actions. Particularly when studying the behavior of an evolutionary system in the very long run it seems reasonable to take into account that there may be persistent random shocks. (See, e.g., Young and Foster [23], Kandori, Malaiti and Rob [8], Young [21], or Young [22].) This also turns out to allow us make more precise predictions about situations where the deterministic replicator dynamics has multiple stable restpoints.

Hence we now introduce perturbations to the Markov chain defined above. Specifically, we assume that with probability $\varepsilon \in (0, 1)$, the agent drawn to revise his strategy picks a strategy at random, irrespective of the payoff to the current strategy. This defines a *perturbed* Markov chain $X^{N,\varepsilon}$ with transition probabilities

$$P^\varepsilon(x, x + 1/N) = (1 - \varepsilon)x(1 - x) \max\{g(x) - k, 0\} + \frac{(1 - x)}{2}\varepsilon,$$

$$P^\varepsilon(x, x - 1/N) = (1 - \varepsilon)x(1 - x) \max\{k - g(x), 0\} + \frac{x}{2}\varepsilon,$$

and

$$P^\varepsilon(x, y) = 0 \text{ for } |x - y| \geq \frac{2}{N}.$$

The perturbations make the Markov chain both *irreducible* and *aperiodic*.³ Moreover, as ε tends to zero, its transition probabilities converge to those of the unperturbed Markov chain. For $X^{N,\varepsilon}$ the expected net increase in the

³A Markov chain is *irreducible* if there is a positive probability of moving from any state to any other state of the state space in a finite number of periods. Let N_z be the set of all integers $s \geq 1$ such that there is a positive probability of moving from the state z to z in exactly s periods. A Markov chain is *aperiodic* if, for every state z , the greatest common divisor of N_z is unity.

population of workers, from one transition time to the next, conditional upon the current state x becomes

$$F_W^\varepsilon(x) = x(1-x)(g(x) - k) + \left(\frac{1}{2} - x\right)\varepsilon.$$

The fact that $X^{N,\varepsilon}$ is irreducible and aperiodic implies the existence of a unique invariant probability measure $\nu^{N,\varepsilon}$ such that for any Borel set $B \subseteq [0, 1]$, $N \in \mathbb{N}$, and $\varepsilon \in (0, 1)$,

$$\nu^{N,\varepsilon}(B) = \lim_{T \rightarrow \infty} \Pr\left(X^{N,\varepsilon}(T) \in B\right) = \lim_{T \rightarrow \infty} V^{N,\varepsilon}(B, T),$$

where $V^{N,\varepsilon}(B, T)$ denotes the relative frequency with which B is visited by $X^{N,\varepsilon}$ during the first T periods. By computing the limit of $\nu^{N,\varepsilon}(B)$ as ε tends to zero and N tends to infinity, we are able to calculate the states that are most likely to be observed in the long run for large population size and small randomization probability. More precisely, we look for states with the following property.⁴

Definition 1 *We say that a state x is selected if, for any neighborhood U of x , we have that*

$$\lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow \infty} \nu^{N,\varepsilon}(U) = 1.$$

We can now show the following.

Proposition 3 *The following holds for the perturbed Markov Chain $X^{N,\varepsilon}$.*

1. *If $m = 1$, then $x = 1 - k^{1/(n-1)}$ is selected.*
2. *If $m = n$, then $x = 1$ is selected for $k < 1/2^{n-1}$, and $x = 0$ is selected for $k > 1/2^{n-1}$.*

⁴It can be shown that Proposition 3 below holds also if the order of limits is reversed in the definition of selection, i.e., if a state x is selected when, for any neighborhood U of x , $\lim_{N \rightarrow \infty} \lim_{\varepsilon \rightarrow 0} \nu^{N,\varepsilon}(U) = 1$.

3. (a) If $1 < n < m$ and $k \geq k_{\max}$, then $x = 0$ is selected.
- (b) If $1 < n < m$ and $k < k_{\max}$, then $x = x_H$ is selected if $x_H > 2x_L$, and $x = 0$ if $x_H < 2x_L$.

Consider, in particular, the case when there are two interior restpoints. The *basin of attraction* of the restpoint x_H under the deterministic replicator dynamics, i.e., the set of starting-points such that the trajectory eventually converges to x_H , is the interval $(x_L, 1)$. Let (x_L, x_H) and $(x_H, 1)$ be the *left* and *right sub-basins* of x_H , respectively.

We see that long-run selection of the high-effort equilibrium happens if the length of the left sub-basin of x_H is greater than the length of $(0, x_L)$, the basin of attraction of x_L . Hence it could happen that x_H has a larger basin of attraction than x_L but is nevertheless not selected in the long run.

In the special case of the simple-majority technologies defined earlier, we can solve explicitly for the conditions for selection of x_H . We need that

$$k < \binom{n-1}{(n-1)/2} \left(\frac{2}{9}\right)^{(n-1)/2}.$$

More generally, we can prove some limited comparative statics results relating the cost k to long-run equilibrium selection.

Proposition 4 *If $1 < m < n$, x_H/x_L is strictly decreasing in k for $k \in (0, k_{\max})$. Moreover, there exists a $k^* \in (0, k_{\max})$ such that for $k \in (0, k^*)$, $x_H/x_L > 2$, for $k = k^*$, $x_H/x_L = 2$, and for $k \in (k^*, k_{\max})$, $x_H/x_L < 2$.*

Proposition 5 *Let \hat{m} , \hat{n} , and s be positive integers. If $k \in (0, k_{\max})$, $m-1 = (\hat{m}-1)s$, $n-1 = (\hat{n}-1)s$, and $1 < \hat{m} < \hat{n}$, then there exists an \hat{s} such that for $1 \leq s \leq \hat{s}$, x_H/x_L is decreasing in s , and for $s > \hat{s}$, there are no interior stationary states.*

Consider the effects of scaling the game up or down by multiplying the firm's size and the threshold number of workers by some constant. Propositions 3, 4, and 5 together imply that if we have $1 < m < n$, the interior

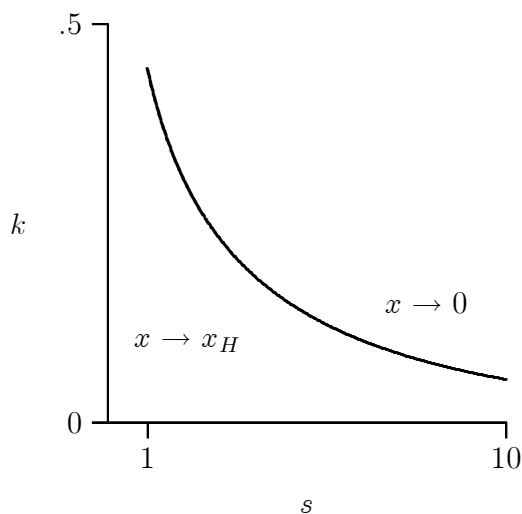


Figure 6: Example selection boundary in cost-scale space.

equilibrium x_H is more likely for low cost and low scale, and the equilibrium with no workers for high cost and high scale.

Figure 6 shows an example with $m - 1 = s$, $n - 1 = 2s$, $x_L = 1/3$, and $x_H = 2/3$. Each point on the curve corresponds to a combination of k and s such that $x_H/x_L = 2$. The equilibrium x_H is selected for combinations to the southwest of the graph, and $x = 0$ for combinations to the northeast of the graph.

To sum up, we have shown that for a cost of contributing low enough, and a firm size small enough, an equilibrium with a substantial frequency of contributors is the situation most likely to be observed in the long run. In the case of simple-majority technologies, the condition can be made precise.

6 Concluding Remarks

We know from experimental work that individuals contribute in dilemma games even when contributing is not enforceable. We have shown in this paper how such behavior may be understood as the long-run stable equilibrium

of an evolutionary system subject to mutations.

We have thought of the underlying public goods game as one in which individuals expend irretrievable effort. The model can easily be extended to a setting where contributions are instead monetary and fully refunded if the public good is not provided, a case that Palfrey and Rosenthal [13] also study from a static point of view. In the evolutionary version, one can then show that there is a unique interior restpoint with a larger share of contributors than in the interior restpoints without refund. Moreover, this interior restpoint is the unique stable state of the replicator dynamic and thus also the unique long-run prediction of the stochastic model.

A different extension of the model would be to allow for worker and firm heterogeneity. Workers could differ in terms of their fitness cost of contributing and firms in terms of their size and contribution threshold levels. It is also possible to make the size and contribution thresholds of firms determined endogenously through evolution. We leave these questions for future research.

Appendix

Proof of Proposition 1. That $x = 0$ and $x = 1$ are restpoints is immediate. At an interior restpoint we must have that

$$\binom{n-1}{m-1} x^{m-1} (1-x)^{n-m} = k. \quad (1)$$

For the cases where $m = 1$ and $m = n$, (1) may be solved explicitly for the interior restpoints. When we have $1 < m < n$, the left hand side of (1) is a single-peaked function of x with a maximum at $x = (m-1)/(n-1)$. Hence if we have that

$$k \leq \binom{n-1}{m-1} \left(\frac{m-1}{n-1}\right)^{m-1} \left(\frac{n-m}{n-1}\right)^{n-m} = k_{\max},$$

the equation (1) has at least one solution. In case the relation holds with equality, the unique solution is $x = (m - 1)/(n - 1)$. Otherwise (1) has exactly two roots, where one is necessarily strictly less than $(m - 1)/(n - 1)$, the other strictly greater. \square

Proof of Proposition 2. We have that

$$\varphi'(x) = k(2x - 1) - \binom{n-1}{m-1} x^{m-1} (1-x)^{n-m} ((n+1)x - m).$$

Suppose we have $m = 1$. Then we have $\varphi'(0) = 1 - k > 0$, $\varphi'(1) = k > 0$, and $\varphi'(1 - k^{1/(n-1)}) = (n - 1)(k^{n/(n-1)} - k) < 0$.

Suppose instead we have $m = n$. Then we have $\varphi'(0) = -k < 0$, $\varphi'(1) = k - 1 < 0$, and $\varphi'(k^{1/(n-1)}) = (n - 1)(k - k^{n/(n-1)}) > 0$.

Finally, suppose we have $1 < m < n$. Then we have $\varphi'(0) = -k < 0$ and $\varphi'(1) = k > 0$. Let x_* , with $0 < x_* < 1$, be such that

$$\binom{n-1}{m-1} x_*^{m-1} (1 - x_*)^{n-m} = k.$$

Substituting, we find that

$$\varphi'(x_*) = k(m - 1 - (n - 1)x_*) \stackrel{\leq}{\geq} 0 \text{ as } x_* \stackrel{\geq}{\leq} \frac{m-1}{n-1}.$$

Note that if we have $x_* = (m - 1)/(n - 1)$, then x_* is a saddle point that attracts in one direction and repels in the other, and is therefore categorized as unstable. \square

The idea behind the proof of Proposition 3 is to calculate the limit of the unique invariant probability measure $\nu^{N,\varepsilon}$ as N tends to infinity and ε tends to 0. When the deterministic dynamic has a *unique* stationary state, we rely upon the following lemma, which holds for any $\varepsilon \in [0, 1)$, to determine this limit.

Lemma 1 *For any open set $U \subset [0, 1]$ containing all the stationary states of the deterministic dynamic, we have that*

$$\lim_{N \rightarrow \infty} \left(\lim_{T \rightarrow \infty} V^{N, \varepsilon}(U, T) \right) = 1 \text{ almost surely.}$$

Proof. This follows directly from Proposition 4 of Benaïm and Weibull [4] by observing that the deterministic dynamic has no periodic orbits. \square

When there is more than one stationary state, we proceed by estimating and comparing the minimum probability of reaching and the maximum probability of leaving the basins of attraction of every asymptotically stable state. Formally, let x_0 be any asymptotically stable state of the deterministic flow and let (x_0^-, x_0^+) denote its basin of attraction. By Proposition 8 in Benaïm and Weibull [3], for any $x \in [x_0^-, x_0^+]$, the *cost* of the transition from $\{x_0\}$ to $\{x\}$ is equal to

$$c(\{x_0\}, \{x\}) = \int_x^{x_0} (\log P^\varepsilon(y, y + 1/N) - \log P^\varepsilon(y, y - 1/N)) dy.$$

Using this concept we can define the *radius* of x_0 , $R(\{x_0\})$, as the lowest cost of going from $\{x_0\}$ to any state outside its basin of attraction, i.e., as

$$R(\{x_0\}) := \min \left\{ c(\{x_0\}, \{x_0^-\}), c(\{x_0\}, \{x_0^+\}) \right\}.$$

Likewise, the *co-radius* of x_0 can be defined as the highest cost to go to the basin of attraction of x_0 from anywhere outside its basin of attraction, i.e., as

$$CR(\{x_0\}) := \max_{y \notin (x_0^-, x_0^+)} \min \left\{ c(\{y\}, \{x_0^-\}), c(\{y\}, \{x_0^+\}) \right\}.$$

Lemma 2 *Let x be an asymptotically stable state of the deterministic dynamic and suppose $U \subset [1, 0]$ is an open neighborhood of x . If $R(\{x\}) > CR(\{x\})$, then*

$$\lim_{N \rightarrow \infty} \lim_{T \rightarrow \infty} V^{N, \varepsilon}(U, T) = 1 \text{ almost surely}$$

and

$$\lim_{N \rightarrow \infty} \lim_{T \rightarrow \infty} \Pr \left(X^{N, \varepsilon}(T) \in U \right) = 1.$$

Proof. This follows immediately from Proposition 6 in Benaïm and Weibull [4].

□

The following lemmata are useful for the calculation of the radius and co-radius.

Lemma 3 *Let x_0^ε and x_P^ε be two arbitrary states such that $x_0^\varepsilon \rightarrow x_0$ and $x_P^\varepsilon \rightarrow x_P$ as ε tends to 0. Then*

$$\lim_{\varepsilon \rightarrow 0} \frac{c(\{x_0^\varepsilon\}, \{x_P^\varepsilon\})}{-\log \varepsilon} = \int_{x_0}^x \lim_{\varepsilon \rightarrow 0} \frac{\log P^\varepsilon(y, y + 1/N) - \log P^\varepsilon(y, y - 1/N)}{\log \varepsilon} dy.$$

Proof. Assume without loss of generality that $x_0^\varepsilon \leq x_P^\varepsilon$. Let $\varepsilon \in (0, e^{-1/2})$ and define the indicator function $I_{x_0^\varepsilon, x_P^\varepsilon}$ that takes on the value 1 for $x \in [x_0^\varepsilon, x_P^\varepsilon]$ and 0 for all other x . Then

$$\begin{aligned} 0 &\leq \frac{1}{\log \varepsilon} \int_{x_0^\varepsilon}^{x_P^\varepsilon} \log P^\varepsilon(x, x + 1/N) dy \\ &= \int I_{x_0^\varepsilon, x_P^\varepsilon} \frac{1}{\log \varepsilon} \log((1 - \varepsilon)y(1 - y) \max\{g(y) - k, 0\} + (1 - y)\varepsilon/2) dy \\ &\leq \int I_{x_0^\varepsilon, x_P^\varepsilon} \frac{1}{\log \varepsilon} \log((1 - y)\varepsilon/2) dy < \int I_{0,1} (1 - \log(1 - y)) dy = 2. \end{aligned}$$

Hence, by Lebesgue's dominated convergence theorem we have that

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \int_{x_0^\varepsilon}^{x_P^\varepsilon} \frac{\log P^\varepsilon(y, y + 1/N)}{\log \varepsilon} dy &= \int \lim_{\varepsilon \rightarrow 0} I_{x_0^\varepsilon, x_P^\varepsilon} \frac{\log P^\varepsilon(y, y + 1/N)}{\log \varepsilon} dy \\ &= \int_{x_0}^{x_P} \lim_{\varepsilon \rightarrow 0} \frac{\log P^\varepsilon(y, y + 1/N)}{\log \varepsilon} dy \end{aligned}$$

By the same argument follows that

$$\lim_{\varepsilon \rightarrow 0} \int_{x_0^\varepsilon}^{x_P^\varepsilon} \frac{\log P^\varepsilon(y, y - 1/N)}{\log \varepsilon} dy = \int_{x_0}^{x_P} \lim_{\varepsilon \rightarrow 0} \frac{\log P^\varepsilon(y, y - 1/N)}{\log \varepsilon} dy,$$

which proves the statement. □

Lemma 4 *Let (x_0^ε) be a sequence of asymptotically stable states such that $\lim_{\varepsilon \rightarrow 0} x_0^\varepsilon = x_0$, and let (x_P^ε) be a sequence of states such that x_P^ε is in the*

closure of the basin of attraction of x_0^ε for every $\varepsilon \in (0, 1)$ and $\lim_{\varepsilon \rightarrow 0} x_P^\varepsilon = x_P$. Then

$$\lim_{\varepsilon \rightarrow 0} \frac{c(\{x_0^\varepsilon\}, \{x_P^\varepsilon\})}{-\log \varepsilon} = x_P - x_0.$$

Proof. Suppose, without loss of generality, that $x_P - x_0 > 0$. Then

$$\begin{aligned} & \lim_{\varepsilon \rightarrow 0} \frac{1}{\log \varepsilon} \int_{x_0^\varepsilon}^{x_P^\varepsilon} \log P^\varepsilon(y, y + 1/N) dy \\ &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\log \varepsilon} \int_{x_0^\varepsilon}^{x_P^\varepsilon} \log((1 - \varepsilon)y(1 - y) \max\{g(y) - k, 0\} + (1 - y)\varepsilon/2) dy. \end{aligned}$$

Define the indicator function $I_{P_+=0}$ that takes on the value 1 for x such that $P(x, x + 1/N) = 0$ and 0 for all other x . Using this function and Lemma 3, we can write the above expression as

$$\begin{aligned} & \int_{x_0}^{x_P} \lim_{\varepsilon \rightarrow 0} I_{P_+=0} \frac{\log((1 - y)\varepsilon/2)}{\log \varepsilon} dy \\ &+ \int_{x_0}^{x_P} \lim_{\varepsilon \rightarrow 0} (1 - I_{P_+=0}) \frac{\log((1 - \varepsilon)y(1 - y)[g(y) - k] + (1 - y)\varepsilon/2)}{\log \varepsilon} dy \\ &= \int_{x_0}^{x_P} I_{P_+=0} dy + 0 = x_P - x_0. \end{aligned}$$

Similarly, by defining the indicator function $I_{P_-=0}$ that takes on the value 1 for x such that $P(x, x - 1/N) = 0$, and 0 for all other x , we can write

$$\begin{aligned} & \lim_{\varepsilon \rightarrow 0} \frac{1}{\log \varepsilon} \int_{x_0^\varepsilon}^{x_P^\varepsilon} \log P^\varepsilon(x, x - 1/N) dy \\ &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\log \varepsilon} \int_{x_0^\varepsilon}^{x_P^\varepsilon} \log((1 - \varepsilon)y(1 - y) \max\{k - g(y), 0\} + y\varepsilon/2) dy \\ &= \int_{x_0}^{x_P} \lim_{\varepsilon \rightarrow 0} I_{P_-=0} \frac{\log(\varepsilon y/2)}{\log \varepsilon} dy \\ &+ \int_{x_0}^{x_P} \lim_{\varepsilon \rightarrow 0} (1 - I_{P_-=0}) \frac{\log((1 - \varepsilon)y(1 - y)(k - g(y)) + y\varepsilon/2)}{\log \varepsilon} dy \\ &= 0. \end{aligned}$$

□

Proof of Proposition 3. Case 1: $m = 1$. For positive ε , the function F_W^ε has a unique root on the unit interval; x_M^ε . As ε tends to zero, x_M^ε tends to $x_M = 1 - k^{1/(n-1)}$ and hence, the statement follows directly by Lemma 1.

Case 2: $m = n$. For sufficiently small ε , the function F_W^ε has three roots on the unit interval; x_0^ε , x_P^ε , and x_1^ε , but only x_0^ε and x_1^ε are asymptotically stable states. As ε tends to zero, the three points converge to 0, $x_P = k^{1/(n-1)}$, and 1 respectively. It is clear that the radius of x_0^ε is equal to $c(\{x_0^\varepsilon\}, \{x_P^\varepsilon\})$ and the co-radius $c(\{x_1^\varepsilon\}, \{x_P^\varepsilon\})$. Similarly, the radius of x_1^ε is equal to $c(\{x_1^\varepsilon\}, \{x_P^\varepsilon\})$ and the co-radius $c(\{x_0^\varepsilon\}, \{x_P^\varepsilon\})$. By Lemma 2, in the limit, as ε tends to zero, the ratio of costs becomes

$$\lim_{\varepsilon \rightarrow 0} \frac{c(\{x_1^\varepsilon\}, \{x_P^\varepsilon\})}{c(\{x_0^\varepsilon\}, \{x_P^\varepsilon\})} = \frac{x_1 - x_P}{x_P - x_0} = k^{1/(1-n)} - 1.$$

This implies that if $k < 2^{1-n}$, there exists an $\hat{\varepsilon}$, such that for all $\varepsilon \in (0, \hat{\varepsilon})$, $\text{CR}(\{1\}) < \text{R}(\{1\})$ and if $k > 2^{1-n}$, there exists an $\check{\varepsilon}$, such that for all $\varepsilon \in (0, \check{\varepsilon})$, $\text{CR}(\{0\}) < \text{R}(\{0\})$.

Case 3: $1 < n < m$ **and** $k < k_{\max}$. For sufficiently small ε , the function F_W^ε has three roots on the unit interval; x_0^ε , x_L^ε , and x_H^ε . The first and the last of these points are asymptotically stable. As ε tends to zero, the three points converge to 0, x_L , and x_H respectively. It is clear that the radius of x_0^ε is equal to $c(\{x_0^\varepsilon\}, \{x_L^\varepsilon\})$ and the co-radius $c(\{x_H^\varepsilon\}, \{x_L^\varepsilon\})$. Similarly, the radius of x_H^ε is equal to $c(\{x_H^\varepsilon\}, \{x_L^\varepsilon\})$ and the co-radius $c(\{x_0^\varepsilon\}, \{x_L^\varepsilon\})$. By Lemma 2, in the limit, as ε tends to zero, the ratio of costs becomes

$$\lim_{\varepsilon \rightarrow 0} \frac{c(\{x_H^\varepsilon\}, \{x_L^\varepsilon\})}{c(\{x_0^\varepsilon\}, \{x_L^\varepsilon\})} = \frac{x_H}{x_L} - 1.$$

This implies that if $x_H/x_L > 2$, there exists an $\hat{\varepsilon}$, such that for all $\varepsilon \in (0, \hat{\varepsilon})$, $\text{CR}(\{x_H\}) < \text{R}(\{x_H\})$, and if $x_H/x_L < 2$, there exists an $\check{\varepsilon}$, such that for all $\varepsilon \in (0, \check{\varepsilon})$, $\text{CR}(\{0\}) < \text{R}(\{0\})$.

Case 4: $1 < m < n$ **and** $k \geq k_{\max}$. If $x_\star = (m-1)/(n-1) < 1/2$ and $k = k_{\max}$, for sufficiently small ε , the function F_W^ε has three roots on the unit interval; x_0^ε , x_L^ε , and x_H^ε (if $(m-1)/(n-1) = 1/2$, the two latter coincide

with x_*). Of these, x_0^ε and x_H^ε are asymptotically stable states. As ε tends to zero, x_0^ε converges to 0, and x_L^ε and x_H^ε both converge to x_* . This case is identical to the above where $1 < m < n$ and $k \in (0, k_{\max})$, and since

$$\lim_{\varepsilon \rightarrow 0} \frac{c(\{x_H^\varepsilon\}, \{x_L^\varepsilon\})}{c(\{x_0^\varepsilon\}, \{x_L^\varepsilon\})} = \frac{x_*}{x_*} - 1 = 0,$$

$\text{CR}(\{0\}) < \text{R}(\{0\})$ for sufficiently small ε .

If $x_* = (m-1)/(n-1) > 1/2$ and $k = k_{\max}$, or if $k > k_{\max}$, for sufficiently small ε , the function F_W^ε has one root on the unit interval; x_0^ε . Since this point converges to 0 as ε tends to zero, the statement follows directly by Lemma 1. \square

Proof of Proposition 4. First, we shall prove that $\partial(x_H/x_L)/\partial k < 0$ for all $k \in (0, k_{\max})$. For $x \in \{x_L, x_H\}$, $g(x) - k = 0$ and

$$\frac{\partial(g(x) - k)}{\partial x} = k \left(\frac{m-1}{x} - \frac{n-m}{1-x} \right) \neq 0.$$

Hence, the implicit function theorem applies and we have that

$$\begin{aligned} \frac{\partial x}{\partial k} &= \frac{-\partial(g(x) - k)/\partial k}{\partial(g(x) - k)/\partial x} \\ &= k^{-1} \left(\frac{m-1}{x} - \frac{n-m}{1-x} \right)^{-1}. \end{aligned}$$

The last expression is negative for $x > (m-1)(n-1)$ and positive for $x < (m-1)(n-1)$. Since $x_L < (m-1)(n-1) < x_H$, it follows that $\partial x_L/\partial k > 0$ and $\partial x_H/\partial k < 0$, and thus,

$$\frac{\partial(x_H/x_L)}{\partial k} = \frac{x_L \partial x_H/\partial k - x_H \partial x_L/\partial k}{x_L^2} < 0.$$

This implies that x_H/x_L is a continuous and strictly decreasing function of k on $(0, k_{\max})$.

Secondly, we shall prove that there exists a \hat{k} such that $x_H > 2x_L$. Let $\alpha \in (1, (n-1)/(m-1))$ and define

$$\hat{k} := \binom{n-1}{m-1} \left(\frac{\alpha m-1}{2n-1} \right)^{m-1} \left(1 - \alpha \frac{m-1}{n-1} \right)^{n-m}.$$

Clearly, $\hat{k} \in (0, k_{\max})$ and for all $x \in [(\alpha/2)(m-1)/(n-1), \alpha(m-1)/(n-1)]$, we have that

$$g(x) > \hat{k}.$$

It follows that

$$\frac{x_H}{x_L} > \left(\alpha \frac{m-1}{n-1} \right) \left(\frac{\alpha}{2} \frac{m-1}{n-1} \right)^{-1} = 2.$$

Since $\lim_{k \rightarrow k_{\max}} x_H/x_L = 1 < 2$, the continuity and strict monotonicity of x_H/x_L on $(0, k_{\max})$ imply that there exists a k^* such that for $k \in (0, k^*)$, $x_H/x_L > 2$, for $k = k^*$, $x_H/x_L = 2$, and for $k \in (k^*, k_{\max})$, $x_H/x_L < 2$. \square

Proof of Proposition 5. Denote the probability function of the binomial distribution by

$$g(x, a, b) := \binom{b}{a} x^a (1-x)^{b-a}.$$

Suppose that, for $x \in \{x_L, x_H\}$ and $s \geq 1$,

$$g(x, s(m-1), s(n-1)) = k_s.$$

Then there exists a $k_{s+1} > 0$ such that, for $x \in \{x_L, x_H\}$,

$$g(x, (s+1)(m-1), (s+1)(n-1)) = k_{s+1}.$$

This k_{s+1} is strictly smaller than k_s since

$$\begin{aligned} \frac{k_{s+1}}{k_s} &= \frac{g(x, (s+1)(m-1), (s+1)(n-1))}{g(x, s(m-1), s(n-1))} \\ &\leq \frac{g((m-1)/(n-1), (s+1)(m-1), (s+1)(n-1))}{g((m-1)/(n-1), s(m-1), s(n-1))} < 1, \end{aligned}$$

where the last inequality follows from a standard property of the binomial distribution. Since, by Proposition 4, $\partial(x_H/x_L)/\partial k < 0$, this implies that x_H/x_L is strictly decreasing in s for s such that

$$g((m-1)/(n-1), (s+1)(m-1), (s+1)(n-1)) \geq k.$$

Finally, since

$$\lim_{s \rightarrow \infty} g(x, s(m-1), s(n-1)) = 0$$

for any $x \in [0, 1]$, it follows that there exists an $\hat{s} > 1$ such that for $s > \hat{s}$,

$$g((m-1)/(n-1), s(m-1), s(n-1)) < k.$$

Hence, for such s there are no interior stationary states. \square

References

- [1] Armen A Alchian and Harold Demsetz. Production, information costs, and economic organization. *American Economic Review*, 62:777–795, 1972.
- [2] Jose Apesteguia, Steffen Huck, and Jörg Oechssler. Imitation: Theory and experimental evidence. Working paper, 2003.
- [3] Michel Benaïm and Jörgen W Weibull. Deterministic approximation of stochastic evolution in games. Working paper, Stockholm School of Economics, 2001.
- [4] Michel Benaïm and Jörgen W Weibull. Deterministic approximation of stochastic evolution in games. *Econometrica*, 71:873–903, 2003.
- [5] Theodore C Bergstrom. Evolution of social behavior: Individual and group selection. *Journal of Economic Perspectives*, 16:67–88, 2002.
- [6] Rachel T A Croson and Melanie Beth Marks. Step returns in threshold public goods: A meta- and experimental analysis. *Experimental Economics*, 2:239–59, 2000.
- [7] Bengt Holmström. Moral hazard in teams. *Bell Journal of Economics*, 13:324–340, 1982.

- [8] Michihiro Kandori, George J Mailath, and Rafael Rob. Learning, mutation, and long run equilibria in games. *Econometrica*, 61:29–56, 1993.
- [9] Christoph Kuzmics. Individual and group selection in symmetric 2-player games. Working paper, Northwestern University, 2003.
- [10] Bibb Latané, Kipling Williams, and Stephen Harkins. Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality & Social Psychology*, 37:822–832, 1979.
- [11] Susanne Lohmann, Jörg Oechssler, and Karl Wärneryd. Evolution and the social dilemma. Working paper, Stockholm School of Economics, 2001.
- [12] George J Mailath. Do people play Nash equilibrium? Lessons from evolutionary game theory. *Journal of Economic Literature*, 36:1347–1374, 1998.
- [13] Thomas R Palfrey and Howard Rosenthal. Participation and the provision of discrete public goods: A strategic analysis. *Journal of Public Economics*, 24:171–193, 1984.
- [14] Thomas R Palfrey and Howard Rosenthal. Private incentives in social dilemmas: The effects of incomplete information and altruism. *Journal of Public Economics*, 35:309–332, 1988.
- [15] Arthur J Robson. Evolution and human nature. *Journal of Economic Perspectives*, 16:89–106, 2002.
- [16] Karl H Schlag. Why imitate, and if so, how? A boundedly rational approach to multi-armed bandits. *Journal of Economic Theory*, 78:130–56, 1998.
- [17] Karl H Schlag. Which one should I imitate? *Journal of Mathematical Economics*, 31:493–522, 1999.

- [18] Tomas Sjöström and Martin L Weitzman. Competition and the evolution of efficiency. *Journal of Economic Behavior and Organization*, 30:25–43, 1996.
- [19] Elliot Sober and David Sloan Wilson. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Harvard University Press, Cambridge, 1998.
- [20] Peter D Taylor and Leo B Jonker. Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences*, 40:145–156, 1978.
- [21] H Peyton Young. The evolution of conventions. *Econometrica*, 61:57–84, 1993.
- [22] H Peyton Young. *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton University Press, Princeton, 1998.
- [23] H Peyton Young and Dean Foster. Cooperation in the short and in the long run. *Games and Economic Behavior*, 3:145–56, 1991.