

# Evolution and the Social Dilemma\*

Susanne Lohmann<sup>†</sup>    Jörg Oechssler<sup>‡</sup>    Karl Wärneryd<sup>§</sup>

August 31, 2001

## Abstract

We study the evolution of preferences in games of voluntary contributions to public goods. We assume contributors sustain the same fitness cost but have different subjective costs of contributing. Given their preferences, individuals behave rationally. A contributor is always disadvantaged in fitness terms *vis-à-vis* noncontributors in any given interaction. Nevertheless, because a noncontributor runs the risk of ending up in a group in which the public good is not provided, altruists, whose subjective costs are lower than the fitness cost, may have above average fitness. This guarantees the survival of altruists in the long run.

---

\*We thank Mark Fey and audiences at the 2001 Public Choice Society Meeting in San Antonio, the University of Bonn, the University of Stockholm, and the Wissenschaftszentrum Berlin, for helpful comments on earlier versions. This is still a preliminary version.

<sup>†</sup>Department of Political Science, UCLA, Los Angeles, CA 90095-1472, USA.

<sup>‡</sup>Department of Economics, University of Bonn, Adenauerallee 24, D—53113 Bonn, Germany.

<sup>§</sup>Department of Economics, Stockholm School of Economics, Box 6501, S—113 83 Stockholm, Sweden. Email: [Karl.Warneryd@hhs.se](mailto:Karl.Warneryd@hhs.se).

The hero goes; and what the hero does, *all* feel that they ought to have done as well. The sophisms of the brain cannot resist the mutual-aid feeling, because this feeling has been nurtured by thousands of years of human social life and hundreds of thousands of years of pre-human life in societies.

Peter Kropotkin, *Mutual Aid: A Factor of Evolution*, 1914.

## 1 Introduction

We are not uncommonly in situations where the provision of a public good depends on the active, costly contributions of at least some of us. Sometimes provision fails spectacularly. In 1964, more than thirty New Yorkers watched without taking action as during the course of half an hour Kitty Genovese was sexually assaulted and stabbed to death in the street. Genovese's tortured death became a case study in urban apathy. It was widely seen as a symptom of the moral decay of New York City, the dehumanization and alienation produced by the urban environment, and the existential despair of the individual in the modern world.

In a laboratory experiment, Darley and Latané [4] studied the willingness of students to come to the aid of a person believed to be having an epileptic seizure. Two results stand out. One, that the speed at which a bystander helps or seeks help is strongly correlated with the number of bystanders—the fewer the number of bystanders, the more likely that the individual bystander will intervene, and if he intervenes, the faster he intervenes. That is, a free-rider problem seems to be operational. Second, the experimental data displays individual heterogeneity in response times that is unexplained by standard personal characteristics (for instance, women were not more likely to help than men). Different people responded differently even though they all faced the same situation, or the same incentives.

It is difficult not to conclude that people act differently in such settings because they have different preferences. Some individuals are altruists, per-

haps, and are more likely to take costly action to contribute to a public good. Yet economists and critics of rational choice theory alike typically frown on such explanations, since theories that allow deviations from narrow self-interest, whenever convenient, would make economics devoid of empirical content. But such criticisms do not apply if we also, as we shall do in this paper, ask where those preferences come from.

The Monday demonstrations in Leipzig, believed to have contributed to the downfall of the East German regime, shows individuals willing to risk arrest in order to bring about political change. Yet the same East Germans had remained passive for decades before exploding into action. Clearly, to explain why the same people contribute to public goods sometimes and not at other times, we also need to understand the dynamic interplay of preferences and incentives.

In this paper we adopt an evolutionary approach to such problems. We shall assume individuals are programmed, genetically or culturally, with subjective preferences. In particular, individuals have subjective costs of contributing to a public good. Given their preferences, they behave rationally. Our approach therefore differs from that of standard evolutionary game theory (see, e.g., Mailath [10] or Weibull [21] for surveys), in which individual agents are assumed to be programmed directly with behaviors.

We distinguish between the subjective preferences of agents, based upon which they act, and the objective fitness consequences of those acts. Equilibrium behavior at a point in time in turn determines the objective evolutionary fitnesses of the active preference types. Assuming all individuals who contribute to the public good sustain the same fitness cost, regardless of their subjective utility cost, it would seem as if individuals with subjective cost less than the fitness cost, whom we shall term *altruists*, would be doomed to extinction. Nevertheless, we show that if interactions are isolated and genetic or cultural material is dispersed over time, group selection effects imply the persistence of altruism in the long run.

Group selection arguments were long considered somewhat suspect. In

evolutionary biology, Wynne-Edwards [22] stirred controversy in the 1960s with examples that he argued showed animals acting for the benefit of their species or group rather than themselves as individuals. Quite apart from the fact that some of his examples were shown to be dubious, an intuitive argument against group selection is the observation that evolutionary selection necessarily operates at the level of the individual gene, not groups of organisms. In the social sciences, Friedrich Hayek argued that social institutions evolve by means of group selection (perhaps best expressed in Hayek [8]). Because of the notoriety of group selection arguments, he was not really taken seriously on this point even by his staunchest supporters.

Group selection has, however, recently been rehabilitated by, e.g., Sober and Wilson [18]. Sober and Wilson show that there is a perfectly legitimate and nonmysterious way in which group selection effects may operate. If individuals play games in groups that are isolated in interaction terms (in the sense that what happens in one group does not affect the outcome in another) and genetic material from one group is mixed into other groups in the future, then a trait that is relatively disadvantaged in every given group interaction may still proliferate in the population as a whole. Although such effects are perfectly possible within the standard evolutionary game theory framework (where groups are typically two-player matchings), it is seldom recognized since it does not happen in popular examples like the Prisoners' Dilemma.

In our setting, group selection happens because although a type who contributes in equilibrium is disadvantaged in fitness terms as compared to a noncontributor *in the same group*, a noncontributor runs the risk of being in a group where the public good is not provided. Hence from the perspective of the population as a whole there is an indirect fitness cost to having preferences that induce you not to contribute in equilibrium.

The paper is organized as follows. Section 2 introduces the basic public goods game. In Section 3, we study the evolutionary version. The relation of our results to the existing literature on preference evolution is discussed

in Section 4. Section 5 considers issues of interpretation and concludes.

## 2 Statics

A house is on fire. Unless at least one of the  $n \geq 2$  residents in the neighborhood calls the Fire Department, the fire will spread. If the Fire Department is alerted, everyone enjoys a utility of 1, and otherwise a utility of 0. An individual  $i$  who calls the Fire Department also sustains an individual, subjective utility cost of  $c \in [0, 1]$ . Thus calling the Fire Department is a public good, and each individual with  $c > 0$  would prefer that the Fire Department was in fact alerted—but by somebody else.

Public goods provision games of this type were introduced by Palfrey and Rosenthal [13] and further analyzed in, e.g., Palfrey and Rosenthal [14]. In the more general class of models  $\mu \geq 1$  contributors are necessary for provision of the public good to take place. For  $\mu > 1$ , such games typically have multiple symmetric equilibria. Since in our later evolutionary analysis we need to determine behavior at each moment in time, this gives rise to potentially controversial problems of equilibrium selection. To avoid this we focus on the case of  $\mu = 1$ .

Suppose individuals do not know the subjective costs of their neighbors, but *do* know that they are independently drawn from the commonly known, continuous distribution  $F$  on  $[0, 1]$ . We assume  $F$  has a density  $f$ , with  $f(c) > 0$  for all  $c$ .

This game has  $n$  asymmetric equilibria where one agent contributes and the others do not. We shall instead focus on symmetric equilibria, where all agents use the same strategy, which is a function of their cost type.

From the perspective of the single agent, let  $m$  denote the unknown number of *other* agents who contribute to the public good (i.e., call the Fire Department). An agent of type  $c$  then rationally contributes if we have that

$$1 - c \geq \text{Prob}(m = 0) \cdot 0 + \text{Prob}(m \geq 1) \cdot 1,$$

that is, if we have that

$$c \leq \text{Prob}(m = 0).$$

Let  $x$  be the equilibrium probability of an individual contributing, i.e., the probability of his having cost less than the threshold given by his probability of being decisive. Then his probability of being decisive is simply

$$\text{Prob}(m = 0) = (1 - x)^{n-1}.$$

In equilibrium, all individuals act rationally. Hence each individual's belief about the probability of others contributing must be consistent with the actual probabilities of contribution. The equilibrium probability  $x^*$  of a given individual contributing is therefore given by the fixpoint equation

$$x^* = F\left((1 - x^*)^{n-1}\right).$$

Equivalently, the symmetric equilibrium threshold cost  $c^*$ , which is such that all individuals with lower cost contribute, is given by

$$c^* = (1 - F(c^*))^{n-1}.$$

Define  $g(c) := (1 - F(c))^{n-1}$ . Since  $g$  is a continuous, decreasing function from the unit interval to itself, with the property that  $g(0) = 1$  and  $g(1) = 0$ , by the intermediate value theorem it has a unique interior fixpoint. That is, we have proved the following.

**Proposition 1** *There is a unique equilibrium threshold cost  $c^* \in (0, 1)$ .*

For purposes of the evolutionary analysis of the next section, we shall assume that each contributor also sustains an objective, material, or *fitness* cost  $k \in (0, 1)$ , identical for all individuals. Thus the fitness of a contributor is  $1 - k$ . The fitness of a noncontributor is 1 if the public good is provided and 0 otherwise.

It now seems natural to call individuals with  $c_i < k$  *altruists*, since they have a subjective bias to contribute even when it hurts them in material

terms. We note that it is not typically the case that only altruists contribute; if we have  $c^* > k$ , then some types with  $c > k$  contribute in equilibrium. Conversely, if we have  $c^* < k$ , not all altruists contribute in equilibrium.

The proportion of altruists in the population is  $F(k)$ . Note that we have  $F(k) = 1 - k^{1/(n-1)}$  if and only if we have  $g(k) = k$ . Suppose instead we have  $F(k) < 1 - k^{1/(n-1)}$ . Then  $g(k) = (1 - F(k))^{n-1} > k$ , which, since  $g$  is decreasing, implies  $c^* > k$ . Completing this line of reasoning for the remaining case yields the following Lemma, which will prove useful later.

**Lemma 1** *The proportion of altruists in the population is less than (equal to, greater than)  $1 - k^{1/(n-1)}$  if and only if  $c^*$  is greater than (equal to, less than)  $k$ .*

The quantity  $1 - k^{1/(n-1)}$  is each individual's equilibrium probability of contributing in the symmetric mixed strategy equilibrium of the degenerate game in which everyone is of type  $k$ . Naïve intuition perhaps suggests that in an evolutionary setting, this is the situation that would eventually emerge since, after all, only type  $k$  individuals actually maximize fitness. We shall see that although it is not true that, starting from an initial population distribution where types other than  $k$  are represented, only the  $k$ -types survive in the long run, the end result is very similar.

### 3 Selection Dynamics

Consider now what happens if the problem of the preceding section is played by consecutive generations of players, who inherit the preferences of their progenitors. Although Fire Departments are a fairly recent invention, situations similar to this game have been present throughout human social history. We are now interested in how the recurrent play of such games, and the evolutionary pressures thus generated, have shaped the current pool of preferences.

It is immediately obvious that in a society of  $n$  individuals playing the same game, a disposition that causes its bearer to contribute can never survive evolutionary selection, since the fitness of such an individual will always be less than that of noncontributors. Hence in the long run in such a setting, altruists must die out.

Suppose instead we consider a large (i.e., infinite) population randomly matched into subgroups of size  $n$  to play the public goods game. We assume each individual knows the actual population distribution  $F$ , which is therefore his prior when he makes his rational choice in the group game. This also, of course, implies that all individuals of the same type will behave the same way (unless they are indifferent between the actions). Depending on the composition of the group they end up in, individuals of the same type may still have different fitnesses. Since there are infinitely many group matchings, however, we shall assume the law of large numbers allows us to treat the average fitness of a preference type as equal to its expected fitness.<sup>1</sup>

The average fitness of type  $c$  individuals is therefore

$$E(c, F) := \begin{cases} 1 - k & \text{if } c \leq c^* \\ 1 - c^* & \text{otherwise,} \end{cases}$$

and average fitness in the population as a whole is

$$\bar{E}(F) := \int_0^1 E(c, F) dF = F(c^*)(1 - k) + (1 - F(c^*))(1 - c^*).$$

We shall now assume the population evolves according to the replicator dynamics (Taylor and Jonker [19]), the standard model of asexual genetic reproduction, adapted to handle a continuous type space. Let

$$\sigma(c, F) := E(c, F) - \bar{E}(F)$$

be the difference between the expected fitness of an individual of type  $c$  and the average population fitness. The idea of the replicator dynamics is that

---

<sup>1</sup>Although common, this assumption is not entirely innocuous. See, e.g., Judd [9] or Boylan [3].

the relative increment of the frequency of a set of types at a point in time is given exactly by the average differential fitness of types in that set. In the present setting, the replicator dynamics is then

$$\dot{F}\{A\} = \int_A \sigma(c, F) F\{dc\} \text{ for } A \in \mathcal{B}, \quad (1)$$

where  $\mathcal{B}$  is the Borel sigma-algebra on  $[0, 1]$ . Hence in particular we have that

$$\dot{F}(c) = \int_0^c \sigma(c, F) dF(c).$$

The most common version of the replicator dynamics is defined for finite trait spaces. Bomze [2] and Oechssler and Riedel [11] show how it may be generalized to continuous spaces. In the Appendix, we prove the following result, which guarantees the existence of a unique solution trajectory.

**Proposition 2** *The replicator dynamics is well defined and admits a unique solution.*

We are particularly interested in the rest points of this dynamics. A *stationary distribution* is a distribution  $F$  such that  $\dot{F}(c) = 0$  and hence  $\sigma(c, F) = 0$  for all  $c$ . We have that

$$\sigma(c, F) = E(c, F) - \bar{E}(F) = \begin{cases} (c^* - k)(1 - F(c^*)) & \text{if } c \leq c^* \\ (k - c^*)F(c^*) & \text{if } c > c^*. \end{cases} \quad (2)$$

Since we have  $0 < F(c^*) < 1$  by continuity of  $F$ , we have  $\dot{F}(c) = 0$  for all  $c$  if and only if  $c^* = k$ . By Lemma 1 this implies  $F(k) = 1 - k^{1/(n-1)}$ . We have therefore proved the following.

**Proposition 3** *A distribution  $F$  is stationary if and only if we have  $F(k) = 1 - k^{1/(n-1)}$  (and consequently  $c^* = k$ ).*

We next show that the set of stationary distributions is globally attracting.

**Proposition 4** *The replicator dynamics converges to a stationary distribution from every continuous starting distribution.*

**Proof.** Differentiating the fixpoint definition of the equilibrium threshold cost  $c^*$  with respect to time, we get

$$\dot{c}^* = -\frac{(n-1)\dot{F}(c^*)(1-F(c^*))^{n-2}}{1+(n-1)f(c^*)(1-F(c^*))^{n-2}}$$

Thus  $\dot{c}^*$  has the opposite sign to  $\dot{F}(c^*)$ . Suppose we have  $c^* > k$ . We then have  $E(c, F) = 1 - k > \bar{E}(F)$  for all  $c \leq c^*$ . Hence  $\sigma(c, F) > 0$  for all  $c \leq c^*$ ,  $\dot{F}(c^*) > 0$ , and  $\dot{c}^* < 0$ . The converse holds if we instead have  $c^* < k$ . Hence any time we have  $c^* \neq k$ ,  $c^*$  eventually converges to  $k$ , and  $F(c^*)$  converges to  $1 - k^{1/(n-1)}$ .  $\square$

To interpret this result, we note the following. Since we start with a continuous initial distributions where all types are represented, it cannot happen that any type disappears in finite time. Proposition 4 shows that altruists do not become extinct even as time goes to infinity. Specifically, the share of altruists in the population converges to  $F(k) = 1 - k^{1/(n-1)} > 0$ . Evolution allows for the coexistence of altruists and nonaltruists in the long run. In particular, it is not the case that only type  $k$  individuals, who maximize expected fitness, survive.

In evolutionary equilibrium, since we have  $c^* = k$ , only altruists contribute to the public good. Although this hurts every altruist relative to nonaltruists in the same group, this fitness effect is balanced by nonaltruists running the risk of being in a group where the public good is not provided.

Equation (2) reveals another important feature of the evolutionary process. Type  $k$  players are always on the winners' side. In fact, depending on whether  $c^*(0)$  is larger or smaller than  $k$ , all types below or above  $k$ , respectively, have higher than average fitness and will increase in frequency for all  $t \geq 0$ . To be more precise, using Lemma 2 of Bomze [2] we know that the

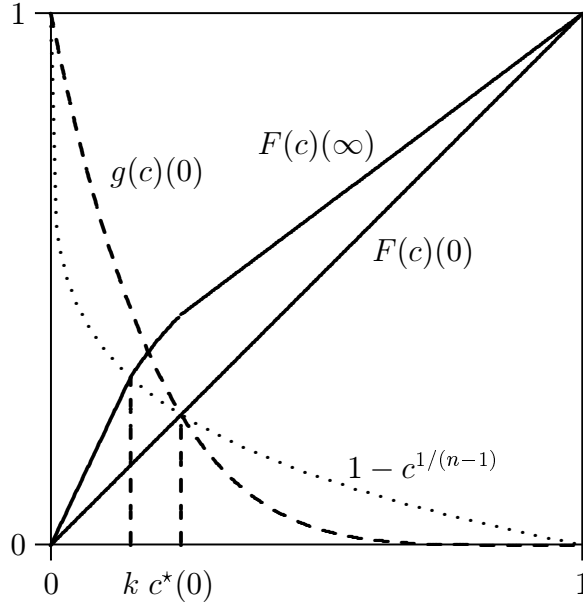


Figure 1: An example long-run distribution.

density evolves according to

$$f(c)(t) = \exp \left( \int_0^t \sigma(c, F(c)(s)) ds \right),$$

where  $\sigma(c, F(c)(s))$  is defined as in (2). In particular, if  $c^*(0)$  is larger than  $k$ , all types below  $k$  have the same growth rate at any moment of time, and similarly for the types above  $k$  when  $c^*(0)$  is smaller than  $k$ .

Figure 1 shows an example long-run population distribution. Here the initial distribution is uniform (and hence the cumulative distribution coincides with the 45° line) and we have  $k < c^*(0)$ . We can find  $c^*(0)$  at the intersection of  $g(c)(0)$  with the 45° line. Since all types below  $k$  grow with the same growth rate at all times, the lower part of the distribution function rotates around the origin until  $F(k) = 1 - k^{1/(n-1)}$ . Similarly, since all types above  $c^*(0)$  have the same negative growth rate the upper part of the distribution function rotates around  $(1, 1)$ .

The aggregate properties of a stationary state are effectively the same

as if everyone were of type  $k$ . The long-run population share  $F(k) = 1 - k^{1/(n-1)}$  corresponds to the equilibrium probability of contributing in an  $n$ -person game with only  $k$ -types. Equivalently,  $1 - k^{1/(n-1)}$  is the equilibrium proportion of contributors in the large-population matching game with only  $k$ -types. Unlike the standard approach to such asymmetric equilibria, which cannot suggest who should do what in equilibrium, our preference evolution approach prescribes the role allocation. Only altruists contribute.

The welfare properties of a stationary state are therefore also very different from what would be the case if everyone were of type  $k$ . Since only altruists, whose subjective costs of contributing are lower than the fitness cost, contribute, average subjective utility is greater than if everyone who contributed experienced a disutility of  $k$ .

Finally, we note that our result tells us that if a stationary distribution is perturbed, over time the system will return to a stationary state. The latter may not be the *same* distribution as the original one, but its aggregate properties (i.e., the proportion of contributors in the game equilibrium associated with the distribution) will be identical.

## 4 Related Literature

This paper is not the first to assume that players behave rationally given their preferences but that their preferences change through an evolutionary process. Earlier contributions to the theory of preference evolution include, e.g., Güth and Yaari [7], Sandholm [17], and the papers surveyed in Robson [15] and Samuelson [16]. Bester and Güth [1] study the evolution of altruistic preferences that explicitly take into account the payoff of an opponent. For an economics example of group selection at work, see Vega-Redondo [20].

Our paper innovates in several respects relative to the received literature: We study  $n$ -person games of nontrivial incomplete information and consider stationary distributions with more than one type of preferences. We also

show that preferences may survive evolutionary selection even though—like those of the altruists in our model—they do not equate utility with fitness, and even under incomplete information.

Call preferences *distorted* if they lead to choices different from those a fitness maximizing individual would take. With few exceptions all models in which distorted preferences prosper are based on individuals having complete information about the preferences of their opponents. The intuitive reason for this is that only if your opponents know your type can it affect their behavior. Fairly general results by Ely and Yilankaya [6] show that in finite, symmetric games, a stable outcome of evolution must yield behavior that corresponds to a Nash equilibrium of the fitness game (i.e., the game where everyone is of type  $k$ ) if individuals have only incomplete information about preferences. Furthermore, Ok and Vega-Redondo [12] show that under incomplete information and some additional assumptions, distorted preferences must necessarily vanish.

Nevertheless, altruists may in fact grow in numbers in our model. Our results do not qualitatively depart from those of Ely and Yilankaya, since every stationary distribution in our model corresponds to the mixed equilibrium of the fitness game. Nor do they contradict the more general intuition that distorted preferences cannot do better than undistorted ones; in our model  $k$ -types always have maximal fitness. Our model does not, however, satisfy the more demanding assumptions of Ok and Vega-Redondo, which entail that distorted and undistorted types never optimally choose the same action. Since there are only two possible actions in our framework, all types who choose the same action as the  $k$ -types will also have maximal fitness.

## 5 Concluding Remarks

If contributing to a public good is costly in fitness terms, it might appear that evolutionary selection would weed out any preference types that have

a bias in favor of contributing. As we have seen, this argument depends on there being only one group playing the game. If instead we consider societies where such games are played in local subgroups, where fitnesses are not affected by what happens in other group interactions, and the genetic or cultural offspring from one group is randomly matched into new groups in the future, then a form of group selection ensures that altruists survive. Furthermore, we have shown that altruists and nonaltruists will coexist in the long run.

Although for technical reasons we consider a model where only one contributor is necessary for the public good to be provided, we might want to speculate about the implications of our results for more involved settings. For instance, at least since Downs [5] students of democracy have puzzled over why people vote in elections. Even a small cost of voting would seem to outweigh the expected gain, given that an individual's probability of being pivotal is typically very small. Does our model say something about the celebrated "paradox of voting"?

If we think of the relevant selection process as being biological evolution, certain restrictions on the possible significance of our results are imposed by the historical environments in which the human mind was formed. A standard presumption is that very complex developments in the brain cannot happen in a few generations, or even 5000 years, the accepted estimate of the length of time during which human beings have lived in large societies. Such developments may take tens or even hundreds of thousands of years to evolve. It is probably safe to say that human emotions (such as those that control our contributions to public goods) were mainly shaped in a hunter-gatherer environment where societies consisted of about 150 individuals.

Hence our model provides a good framework for reasoning about situations such as the Kitty Genovese case discussed in the Introduction, where, according to newspaper articles of the time,  $n$  was equal to 38. But in our ancestral environment, there were no social situations even remotely similar to mass elections involving millions of voters. For this reason, it is not clear

that people would conceptualize the “mass” situations in the same way as they might conceptualize the Kitty Genovese situation, i.e., as a free-rider problem. Although the explanation for large turnouts in elections is likely to ultimately be based on an evolutionary foundation, we are perhaps not quite there yet.

## Appendix

**Proof of Proposition 2.** For the purpose of this proof let  $Q$  denote the probability measure corresponding to distribution function  $F$ , i.e.  $Q([0, c]) := F(c)$ . Similarly, let  $R$  be the measure corresponding to some other continuous distribution  $G$  on  $[0, 1]$ . The proposition follows from Theorem 2 and Lemma 3 of Oechssler and Riedel [11] once we show that the Lipschitz condition

$$\|Q\|, \|R\| \leq 2 \Rightarrow \sup_c |\sigma(c, F) - \sigma(c, G)| \leq L \|Q - R\| \quad (3)$$

and the boundedness condition

$$\sup_{Q: \|Q\| \leq 2} |\sigma(x, F)| \leq \sigma_\infty, \quad (4)$$

hold for  $\sigma$ , where  $L$  and  $\sigma_\infty$  are some constants with  $L, \sigma_\infty < \infty$ , and  $\|\cdot\|$  denotes the variational norm. When the variational norm is used, the distance between two probability measures  $P$  and  $Q$  is given by  $\|P - Q\| = 2 \sup_{A \in \mathcal{B}} |P(A) - Q(A)|$ .

Let  $c_H^*$  be the equilibrium threshold cost associated with the population distribution  $H = F, G$ . To confirm (4), note that

$$\begin{aligned} |\sigma(x, F)| &= |E(c, F) - \bar{E}(F)| \leq \max\{1 - k, 1 - c_F^*\} + \int_0^1 E(x, F) dF(c) \\ &\leq \max\{1 - k, 1 - c_F^*\} (1 + \|Q\|) \leq 3 \end{aligned}$$

for  $\|Q\| \leq 2$ . Hence (4) is satisfied.

To establish (3) we need to utilize that

$$|F(c_F^*) - G(c_G^*)| \leq |F(c_F^*) - G(c_F^*)| \leq \sup_{A \in \mathcal{B}} |Q(A) - R(A)| = \frac{1}{2} \|Q - R\|, \quad (5)$$

where the first inequality follows from the fixed-point construction of  $c_F^*$  and  $c_G^*$ .

The left hand side of (3) can be decomposed into

$$\begin{aligned} |\sigma(c, F) - \sigma(c, G)| &= |E(c, F) - E(c, G) + \bar{E}(F) - \bar{E}(G)| \\ &\leq |c_F^* - c_G^*| + |\bar{E}(F) - \bar{E}(G)|. \end{aligned} \quad (6)$$

We shall treat the two terms of (6) separately. The first yields

$$|c_F^* - c_G^*| = \left| [1 - F(c_F^*)]^{n-1} - [1 - G(c_G^*)]^{n-1} \right|.$$

We shall make use of the following fact. If  $0 \leq a, b \leq 1$ ,  $n = 1, 2, \dots$ , then  $|a^n - b^n| \leq 2|a - b|$ . Without loss of generality let  $a \geq b$ . Then  $|a^n - b^n| = a^n - b^n \leq a^n + ab^{n-1} - ba^{n-1} - b^n = (a - b)(a^n - b^n) \leq 2(a - b) = 2|a - b|$ .

Thus

$$|c_F^* - c_G^*| \leq |G(c_G^*) - F(c_F^*)| \leq \frac{1}{2} \|Q - R\| \quad (7)$$

by (5).

The second term of (6) yields

$$\begin{aligned} |\bar{E}(F) - \bar{E}(G)| &= \left| \int_0^1 E(c, F) dF(c) - \int_0^1 E(c, G) dG(c) \right| \\ &= |(1 - k)F(c_F^*) + (1 - c_F^*)(1 - F(c_F^*)) \\ &\quad - (1 - k)G(c_G^*) - (1 - c_G^*)(1 - G(c_G^*))| \\ &\leq (1 - k) |F(c_F^*) - G(c_G^*)| \\ &\quad + |(1 - c_F^*)(1 - F(c_F^*)) - (1 - c_G^*)(1 - G(c_G^*))| \\ &\leq \|Q - R\| + |c_G^* - c_F^*| + |G(c_G^*) - F(c_F^*)| + |c_F^*F(c_F^*) - c_G^*G(c_G^*)| \\ &\leq 2\|Q - R\| + |c_F^*F(c_F^*) - c_G^*G(c_G^*)|. \end{aligned} \quad (8)$$

Concerning the last term, note that

$$\begin{aligned}
|c_F^* F(c_F^*) - c_G^* G(c_G^*)| &\leq \max\{c_F^*, c_G^*\} |F(c_F^*) - G(c_G^*)| \\
&\quad + \max\{F(c_F^*), G(c_G^*)\} |c_F^* - c_G^*| \\
&\leq \frac{3}{2} \|Q - R\|. \tag{9}
\end{aligned}$$

Using (7)–(9) we see that  $|\sigma(c, F) - \sigma(c, G)| \leq 4 \|Q - R\|$ , which is the desired result.  $\square$

## References

- [1] Helmut Bester and Werner Güth. Is altruism evolutionarily stable? *Journal of Economic Behavior and Organization*, 34:193–209, 1998.
- [2] Immanuel Bomze. Cross entropy minimization in uninventable states of complex populations. *Journal of Mathematical Biology*, 30:73–87, 1991.
- [3] Richard T Boylan. Laws of large numbers for dynamical systems with randomly matched individuals. *Journal of Economic Theory*, 57:473–504, 1992.
- [4] John M Darley and Bibb Latané. Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, 8:377–383, 1968.
- [5] Anthony Downs. *An Economic Theory of Democracy*. Harper and Row, New York, 1957.
- [6] Jeffrey C Ely and Okan Yilankaya. Nash equilibrium and the evolution of preferences. *Journal of Economic Theory*, 97:255–272, 2001.
- [7] Werner Güth and Menachem Yaari. An evolutionary approach to explain reciprocal behavior in a simple strategic game. In Ulrich Witt, editor, *Explaining Process and Change: Approaches to Evolutionary Economics*, pages 23–34. University of Michigan Press, Ann Arbor, 1992.

- [8] Friedrich A Hayek. *Law, Legislation and Liberty*. The University of Chicago Press, Chicago, 1973–1979. 3 volumes.
- [9] Kenneth L Judd. The law of large numbers with a continuum of IID random variables. *Journal of Economic Theory*, 35:19–25, 1985.
- [10] George J Mailath. Do people play Nash equilibrium? Lessons from evolutionary game theory. *Journal of Economic Literature*, 36:1347–1374, 1998.
- [11] Jörg Oechssler and Frank Riedel. Evolutionary dynamics on infinite strategy spaces. *Economic Theory*, 17:141–162, 2001.
- [12] Efe A Ok and Fernando Vega-Redondo. On the evolution of individualistic preferences: An incomplete information scenario. *Journal of Economic Theory*, 97:255–272, 2001.
- [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. The biological basis of economic behavior. *Journal of Economic Literature*, XXXIX:11–33, 2001.
- [16] Larry Samuelson. Introduction to the evolution of preferences. *Journal of Economic Theory*, 97:225–230, 2001.
- [17] William H Sandholm. Preference evolution, two-speed dynamics, and rapid social change. Working paper, University of Wisconsin-Madison, 2000.

- [18] Elliot Sober and David Sloan Wilson. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Harvard University Press, Cambridge, 1998.
- [19] Peter D Taylor and Leo B Jonker. Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences*, 40:145–156, 1978.
- [20] Fernando Vega-Redondo. Competition and culture in an evolutionary process of equilibrium selection: A simple example. *Games and Economic Behavior*, 5:618–631, 1993.
- [21] Jörgen W Weibull. *Evolutionary Game Theory*. MIT Press, Cambridge, MA, 1995.
- [22] V C Wynne-Edwards. *Animal Dispersion in Relation to Social Behavior*. Oliver and Boyd, Edinburgh, 1962.