Journal of Public Economics 70 (1998) 53–73

Social rewards, externalities and stable preferences

Chaim Fershtman[a,b,*], Yoram Weiss[b]

[a] *CentER, Tilburg University, Tilburg, The Netherlands*
[b] *The Eitan Berglas School of Economics, Tel Aviv University, Tel Aviv 69978, Israel*

Abstract

This paper examines the role of social rewards as a corrective mechanism for activities which generate externalities. The focus of this paper is on the circumstances under which social rewards provide effective and feasible incentive mechanism that may replace laws and regulations. In particular, social mechanism is effective only in a society in which individuals who care about their standing in the society can survive in the long run. We show that the nature of economic interaction between matched players influences whether the socially minded individuals survive in the long run and restricts the effectiveness of social rewards. However, circumstances exist where the socially minded survive, even though relative fitness is determined only by economic payoff.

Keywords: Social rewards; Incentive mechanism; Externalities; Stable preferences

1. Introduction

Generally speaking, there are three broad types of incentives that govern the behavior of individuals in society: (i) private rewards such as wages and profits, (ii) social rewards such as prestige and status, (iii) rules and laws that enforce certain types of behavior and penalize deviations. Casual observation indicates that societies differ in the mixture of incentives and rules they employ. Thus, in order to understand how societies function, one of the fundamental questions is why certain activities are subject to enforcement while others are governed by social rewards and conventions.

*Corresponding author. Fax: +972 3 6409908.

0047-2727/98/$ – see front matter
PII: S0047-2727(98)00060-7

As is well recognized in the economic literature, activities which affect other members of the society, but cannot be priced, are not efficiently regulated by private rewards. It was Arrow (1971) who first suggested the role of social norms as a mechanism designed to resolve the inefficiencies arising from externalities[1]. In this paper we consider a similar role for social rewards such as prestige and status. That is, an individual who chooses an action that has a positive externality is appreciated and esteemed by the other members of society, while an individual who causes a negative externality is treated with contempt. The use of such social mechanism is appealing as it implies that the problem of market inefficiency due to externalities may be resolved, even without legal rules. But social rewards are not always effective. The main concern of this paper is the limits of the effectiveness of social rewards as a corrective mechanism.

Social status is an effective mechanism only in a society in which individuals care about their standing in the society. Thus, as part of an analysis on the effectiveness of social mechanisms, one should address the question: ‘‘why should a selfish individual care about their social status or about what other people think about him?’’ To answer this question, we examine the circumstances under which evolution would lead to the survival of socially minded individuals, even though relative fitness is determined only by economic payoffs. In this respect, the approach of this paper is similar to Bester and Guth (1997) and Fershtman and Weiss (1997). In both these papers it is demonstrated that it is possible to have evolutionarily stable equilibrium in which individuals maximize an objective function that differs from the fitness criterion. What characterizes such situations is that when individuals are interacting strategically, the departure from individual maximization of fitness can induce favorable reaction by other players. This result has already been noted in the analysis of strategic delegation (e.g., Fershtman and Judd (1987)).

The approach taken in this paper is similar to the one taken in sociobiology in which preferences are determined endogenously, by evolutionary forces[2]. We consider a simple model in which individuals are randomly matched and are involved in a two-person interaction. Their actions generate externalities which influence all individuals in society. We assume that the social status of an individual is determined by his own action and the actions taken by the other members of society. Individuals, however, can differ in the importance which they assign to social rewards. Some may care about the opinion of others while others do not. We do not assume any initial profile of types but rather look for the one that emerges as an outcome of an evolutionary process.

[1]See Elster (1989) for a criticism on Arrow’s approach and Fershtman et al. (1996) for an analysis of the implications of social rewards for the allocation of talent in society.

[2]For evolutionary models that endogenize preferences, see Basu (1995); Bester and Guth (1997); Dekel and Scotchmer (1994); Fershtman and Weiss (1997); Hirshleifer (1980); Robson (1996); Rogers (1994).

Our analysis points out an interesting asymmetry. Depending on the properties of the payoff function, it is possible to use social rewards to either curtail or stimulate individual action, but not both. This result provides some insight on the complementary roles of social incentives and rules. For example, if the two players are engaged in a Cournot game, it would be possible to use social rewards to induce an increase of their output, but it is *not* possible to induce a reduction in output. The reason is that producers who care about social status and reduce their output will not survive in the long run. If it is desirable to reduce output, because the output causes pollution for instance, this must be done by other means such as legal enforcement. Social rewards will be effective only if it is desirable to increase output, because of positive externalities. Even then, social rewards may be insufficient to attain efficiency, and some enforcement may be required.

A social rewards mechanism requires that individuals' actions are observable to other members of society. It is also required that in each match the equilibrium outcome will depend on the true types of the two individuals. Otherwise, the socially minded individuals cannot gain fitness. Throughout our analysis, we assume that types are indeed observable to the partners of each match. This assumption can be justified by the fact that the equilibrium generated by perfect observability is identical to the steady state of a Cournot adjustment process, where each partner myopically determines his action as a best response to his rival's past action.

2. The model

Consider a society in which there is a large number of identical individuals. In each period, individuals are randomly matched into pairs and play the following symmetric game: Each player chooses an action $x_i \in R_+$. The monetary payoff of player i when matched with player j is given by

$$m_i = E(x^e)P(x_i,x_j), \tag{1}$$

where $P(x_i,x_j)$ is the direct payoff from the interaction of the two players and $E(x^e)$ is an externality term which depends on the average actions of *all* the players in society, x^e. The monetary payoff of player j is correspondingly given by $m_j = E(x^e)P(x_j,x_i)$.

In order to carry out our analysis, we restrict our attention to payoff functions that yield a unique Nash equilibrium. We assume that $P(x_i,x_j)$ is twice continuously differentiable, strictly concave in x_i and that $P_{11}(x_i,x_j)P_{11}(x_j,x_i) > P_{12}(x_i,x_j)P_{12}(x_j,x_i)$, where subscripts are used to denote partial derivatives.

There are two separate dimensions of the interaction between matched players. The *strategic* interaction, determined by the sign of $P_{12}(x_i,x_j)$, and the *payoff* interaction determined by the sign of $P_2(x_i,x_j)$. We assume that $P_{12}(x_i,x_j)$ and

$P_2(x_i,x_j)$ do not change sign. As we shall see, the analysis depends on whether or not the two terms have the *same* sign[3]. We shall discuss separately the cases in which $P_{12}(x_i,x_j)$ $P_2(x_i,x_j)$ is positive and negative. The first case corresponds, for instance, to Cournot quantity or price duopoly games. The second case applies, for instance, to tournaments.[4]

We define actions in such a way that the first partial derivative of $P(x_i,x_j)$ with respect to x_i is positive, when evaluated at (0,0). That is, an action of a player is defined to have a positive impact on his payoff, for a sufficiently low level of activity. This restriction allows us to define actions unambiguously. Thus, if x represents effort which increases one's payoff at a low level of activity, then refraining from effort, represented by $-x$, is *not* an action.

We assume that $E(x^e)$ is a differentiable function which is positive for all x^e and is either monotone increasing or monotone decreasing. When $E(x^e)$ is an increasing (decreasing) function, we say that there are positive (negative) externalities.

The presence of externalities, implies that even if the players in a given match could coordinate their activities, the equilibrium outcome may be socially inefficient. A possible mechanism to deal with this problem is to provide *social* rewards, in the form of social status to players whose actions conform to the social norm. In any given time, the social norm is represented by the *average* action, x^e. This norm is supported by a marginal social reward (punishment) parameter, σ, such that the social status of a person i who performs above (below) the norm is $\sigma(x_i - x^e)$. We thus write the social status of person i as

$$s_i = \sigma(x_i - x^e). \tag{2}$$

Each individual in society treats x^e and σ as given parameters and chooses his own level of activity, x_i. For society as a whole, only σ is a given parameter and x^e is determined endogenously by the aggregate behavior of all members of society[5]. Different values of the social reward parameter, σ, generate different aggregate behavior and thus change the social norm. In particular, a positive (negative) marginal rewards may encourage (discourage) activities with positive (negative) externalities.

We assume that individuals care about their social status, but allow for heterogeneity in the weight that individuals give to social rewards, relative to monetary considerations. The objective function of individual i is postulated to be of the following additive form:

[3]The assumption that $P_{12}(x_i,x_j)$ does not change sign is equivalent to the assumption that the reaction functions are either upward or downward sloping.

[4]In a tournament, when the winner takes all, the payoff is $P(x_1,x_2) = Pr\{x_1 + \epsilon_1 \geq x_2 + \epsilon_2\} - c(x_1)$ where ϵ_1 and ϵ_2 are i.i.d. and $c(x_1)$ is increasing and convex (see Lazear and Rosen, 1981). In this case an increase in x_2 raises the marginal payoff of x_1, but reduces the total payoff.

[5]Akerlof (1980) constructs a model in which the average action, in the steady state, is the social norm. In his model, deviations from this norm imply a loss of reputation or status.

$$U^{i}(x_i,x_j,x^{e}) \equiv m_i + \alpha s_i = E(x^{e})P(x_i,x_j) + \alpha\sigma(x_i - x^{e}), \tag{3}$$

where α, $\alpha \in \{0,1\}$, is a preference parameter such that $\alpha = 1$ identifies an individual who cares about what other individuals think about him, while an individual with $\alpha = 0$ does not care what others think about him. In Section 5, we modify this assumption and consider a finite number of types, α_1, α_2,...α_T where $\alpha_t \geq 0$, allowing more variation in the importance that different types assign to their social status.

Note the different roles of the parameters α and σ. The parameter σ describes what other members of society think about an individual, while the parameter α describes whether an individual cares about what other individuals think about him. The distribution of α in the population determines the effectiveness of the social reward, σ, in manipulating aggregate behavior.

A key assumption in our model is that both the actions taken by individuals and their types are fully observable. Clearly, if actions are unobservable then status considerations cannot influence the individuals' choice of action and social rewards are ineffective. We also assume that when two players are matched, each player recognizes the type of the player he is matched with. One possible interpretation of our observability assumption is that each pair of individuals play the one shot game many (but finite) rounds. In each round, both players myopically react to the action chosen by their rival at the previous round. Given our assumptions on the payoff function, the players' strategies converge to the Nash equilibrium strategies of the game with observed types (see Fudenberg and Tirole, 1992, pp. 23–29)). Assuming that this convergence is fast relative to the number of rounds that each pair plays, we can use the Nash equilibrium payoffs to approximate the average payoff of each individual during the period he is matched with a certain type of opponent[6].

Consider a society with a given social rewards parameter, σ. Let q be the proportion of individuals in the society who care about social status, with $\alpha = 1$. We restrict our attention to symmetric equilibria such that all agents of a given type choose the same strategy. We denote by $x(\alpha,\alpha',x^{e})$ the strategy of a player of type α when matched with a player of type α', when the average action in the population is x^{e}.

Given $q \in [0,1]$, we define equilibrium as a triplet consisting of x^{e}, and strategies for players of type 1 and 0, $x^*(1,\alpha,x^{e})$ and $x^*(0,\alpha,x^{e})$, respectively, such that:

(i) The pair of strategies $(x^*(\alpha,\alpha',x^{e}),x^*(\alpha',\alpha,x^{e}))$ is a Nash equilibrium in the game between players of types α and α', $\alpha,\alpha' \in \{0,1\}$, when the average action is x^{e}.

[6] An apparent drawback of this adjustment process is that players react myopically without considering the impact of their choice of action on the behavior of their opponents. However, in the context of evolutionary models, it is customary to endow agents with only limited foresight (or rationality).

(ii) The average action x^e is consistent with the equilibrium strategies and the distribution of types in the population. Specifically,

$$x^e = \frac{1}{2}[q^2 2x^*(1,1,x^e) + 2q(1-q)(x^*(0,1,x^e) + x^*(1,0,x^e)) + (1-q)^2 2x^*(0,0,x^e)]. \tag{4}$$

Our assumptions on $P(x_i,x_j)$ guarantee the uniqueness of the Nash equilibrium for a given x^e. We shall also assume that the impact of x^e on the aggregate output of each pair is less than two, i.e., $\partial x^*(\alpha,\alpha',x^e)/\partial x^e + \partial x^*(\alpha,\alpha',x^e)/\partial x^e < 2$ for all $\alpha,\alpha' \in \{0,1\}$ whenever x^e satisfies Eq. (4). This condition is sufficient to guarantee a unique solution for x^e in Eq. (4).

For a given q, we denote by $x(\alpha,\alpha',q)$ the equilibrium action of type α who is matched with type α', i.e., $x(\alpha,\alpha',q) = x^*(\alpha,\alpha',x^e)$, where x^e satisfies Eq. (4). The equilibrium monetary payoff of type α when matched with type α' is denoted by $M(\alpha,\alpha',q)$.

3. The determination of preferences

Why would anyone care about the opinion of others? To answer this question, we consider the evolutionary formation of preferences. We assume that the proportion of individuals of a given type in the population increases if their expected monetary payoff exceeds the average payoff in the population[7]. We thus define the fitness of a particular type in terms of his *monetary* payoffs, rather than his utility which takes into account also social rewards. The underlying assumption is that even when people care about social rewards, their fitness is determined only by their economic success. It is possible, of course, that status contributes directly to fitness, as it is sometimes assumed in the biological literature. Our assumption that fitness is derived only from monetary payoffs is somewhat unrealistic but it helps us to sharpen the discussion of the circumstances under which it is possible for a society with *socially* minded individuals to be evolutionarily stable.

In considering the evolutionary process of preferences, imitation cannot be the main engine of transmission, as in the discussion of the evolution of strategies. Instead, we consider the transmission of preferences across generations. A possible mechanism is one in which parents spend resources to shape the preferences of their children. Wealthy parents can spend more and are, therefore, more successful in reproducing their own preferences (see Becker (1992) and Becker and Mulligan

[7]See Maynard Smith (1982) for the biological foundation and the surveys of economic applications by Hammerstein and Selten (1994); Weibull (1995).

(1997)[8]. Specifically, let $W^1(q)$ and $\underline{W}^0(q)$ be the expected equilibrium payoffs of types 1 and 0, respectively, and let $W(q)$ be the average payoff in the population. The difference $W^\alpha(q) - W(q)$ is a measure of the (relative) fitness of type α. Following Maynard Smith (1982), p.14), a type α_t is evolutionary stable if, when almost all members of the population are of this type, the fitness of these typical members is greater than that of any possible type. Thus, the type $\alpha = 1$ is evolutionarily stable if $W^1(q) > W^0(q)$ for every q close to 1. Similarly, the type $\alpha = 0$ is evolutionarily stable if $W^1(q) < W^0(q)$ for every q close to 0.

Necessary and sufficient conditions for the evolutionary stability of type α are that for every $\alpha' \neq \alpha$:[9]

(i) $M(\alpha, \alpha, q_\alpha) \geq M(\alpha', \alpha, q_\alpha)$,

(ii) $M(\alpha, \alpha', q_\alpha) > M(\alpha', \alpha', q_\alpha)$, whenever $M(\alpha, \alpha, q_\alpha) = M(\alpha', \alpha, q_\alpha)$,

where q_α is close to 1 if $\alpha = 1$ and q_α is close to 0 if $\alpha = 0$[10].

While we allow the profile of individual preferences to vary over time, we hold the social status function constant. That is, a society is characterized by its social reward parameter, σ, and *all* individuals within the society, irrespective of their α, evaluate the actions of their colleagues according to this parameter.

4. The effectiveness of social rewards

Social status can be an effective reward mechanism only in a society in which some individuals care about social status. Such individuals will be present, in the long run, only if social preferences (with $\alpha = 1$) are evolutionarily stable. We shall refer to a social reward, σ, as effective if a population, consisting of individuals who care about social status, is evolutionarily stable.

To determine the stable preference profiles, we need to consider the equilibrium fitness of different types of individuals.

Lemma 1: Consider an action which yields positive (negative) social rewards, then, a socially minded individual with $\alpha = 1$ chooses, in equilibrium, a higher

[8]An alternative, but probably less realistic, hypothesis is that wealthy individuals have higher reproduction rate and that preferences are transmitted within families through a process of imitation (see Basu (1995)).

[9]The first condition requires that α is a best reply against itself. The second condition requires that if α' is doing as well as α against α, then α is doing better against α' than α' itself.

[10]This formulation differs from the standard formulation in that the payoffs in a particular match depend on q. This reflects the presence of externalities. In the standard formulation, conditions (i) and (ii) are independent of the distribution of types in the population. One implication of this difference is that it is possible to have interior stationary points with $W^1(q) = W^0(q)$. We shall assume throughout our analysis that such interior points are not evolutionarily stable, because $W^1(q)$ is likely to be more responsive than $W^0(q)$ to an increase in q. We have not established sufficient conditions for this result, in terms of restrictions on $E(.)$ And $P(.,.)$.

(lower) level of such action than an asocial individual with $\alpha=0$, irrespective of the type of the matched partner. That is, for $\sigma>0$, $x(1,0,q)>x(0,0)$ and $x(1,1,q)>x(0,1,q)$, while for $\sigma<0$, $x(1,0,q)<x(0,0)$ and $x(1,1,q)<x(0,1,q)$.

Proof: The result follows directly from the first order conditions characterizing the Nash equilibrium and our assumptions regarding the payoff function $P(x_i,x_j)$. ■

Intuitively, when $\sigma>0$, a socially minded individual has the incentive to raise his action x which implies an outward shift of the individual's reaction function. When the reaction functions are either both downward sloping or both upward sloping such a shift implies an *increase* of the equilibrium level of action.

The evolutionary stability of a society with socially minded individuals depends on the equilibrium relationships between preferences and fitness, which in turns depends on the nature of the interaction between the matched players. For a given σ and x^e, let $c=\sigma/E(x^e)$. Then the reaction curve of a type α player is $R_\alpha(x)\equiv \text{ArgMax}_y\ (P(y,x)+\alpha cy)$. Let $\Psi_\alpha(x_i)=P(x_i,R_\alpha(x_i))$ denote the fitness of player i (who may be of any type) when he chooses action x_i against a player of type α who reacts optimally. We assume that for any constant, c, $\Psi_\alpha(x_i)$ is single peaked[11].

Now consider a Nash equilibrium and suppose that player i increases x_i slightly, because his preferences for status *change*. The impact on his monetary payoff, $\Psi_\alpha(x_i)$, is given by $\Psi'_\alpha(x_i)=P_1(x_i,R_\alpha(x_i))+P_2(x_i,R_\alpha(x_i))R'_\alpha(x_i)$. There are two effects on i's monetary payoff: The direct effect, $P_1(x_i,R_\alpha(x_i))$, resulting from increased action and the indirect effect, $P_2(x_i,R_\alpha(x_i))R'_\alpha(x_i)$, resulting from the reaction of its rival. The sign of the indirect effect is fully determined by the sign of the product $P_2(x_i,x_j)\ P_{12}(x_i,x_j)$, because $R'_\alpha(x_i)$ and $P_{12}(x_i,x_j)$ are of the same sign. If the payoff and the strategic interactions are of the same sign then the indirect effect is positive, irrespective of whether the actions are strategic complements or substitutes[12]. If the strategic and the payoff interactions are of opposite sign then the indirect effect is negative. The direct effect of a small increase in x depends on σ. If $\sigma=0$ then, in equilibrium, the marginal fitness is zero and the direct effect is negligible. If $\sigma<0$ then, in equilibrium, the marginal fitness is positive, and the direct effect is positive. If $\sigma>0$, then, in equilibrium, the marginal fitness is negative and the direct effect is to reduce i's fitness, but for a small σ, the indirect effect dominates. Based on these considerations, we can prove the following.

[11]This condition is sufficient to guarantee a unique Stackelberg solution and it is satisfied, for instance, in the standard Cournot game (see Appendix B).

[12]If the actions are strategic complements, and the payoff interaction effect is positive, then when player i raises his action, the rival increases his action and player i gains fitness. Similarly, if the actions are strategic substitutes, and the payoff interaction term is negative, then when player i raises his action, the rival reduces his action and player i gains fitness.

Proposition 1: When $P(x_i,x_j)$ is such that the effects of rival's action on the total and the marginal payoffs are both of the same sign, i.e., $P_2(x_i,x_j)$ and $P_{12}(x_i,x_j)$ are either both positive or both negative, then:

(i) When $\sigma<0$, a society in which all individuals are asocial (i.e., $\alpha=0$) is evolutionarily stable, while a society in which all individuals are socially minded (i.e., $\alpha=1$) is not evolutionarily stable.

(ii) When $\sigma>0$, there exists a positive σ_0, such that, for all $0<\sigma\leq\sigma_0$, a society in which all individuals are socially minded (i.e., $\alpha=1$) is evolutionarily stable, while a society in which all individuals are asocial (i.e., $\alpha=0$), is not evolutionarily stable.

Proof: See Appendix A.

Proposition 2: When $P(x_i,x_j)$ is such that the effects of rival's action on the total and the marginal payoffs are of the opposite sign, i.e., $P_2(x_i,x_j)P_{12}(x_i,x_j)<0$, then:

(i) When $\sigma>0$, a society in which all individuals are asocial (i.e., $\alpha=0$) is evolutionarily stable, while a society in which all individuals are socially minded (i.e., $\alpha=1$) is not evolutionarily stable.

(ii) When $\sigma<0$, there exists a negative σ_0, such that, for all $\sigma_0\leq\sigma<0$, a society in which all individuals are socially minded (i.e., $\alpha=1$) is evolutionarily stable, while a society in which all individuals are asocial (i.e., $\alpha=0$), is not evolutionarily stable.

Proof: The proof is analogous to the proof of Proposition 1.

Propositions 1 and 2 are based on the observation that individuals who do *not* maximize fitness, but some other objective, may end up, in equilibrium, with *higher* fitness than those who maximize fitness. This occurs because the departure from individual maximization of fitness can induce favorable reactions by matched partners.

We can use Fig. 1 and Fig. 2 to illustrate Proposition 1, where by assumption, the product $P_2(x_i,R_\alpha(x_i))R'_\alpha(x_i)$ is always positive. Fig. 1 is used to illustrate part (i) of Proposition 1. Point a^- represents the equilibrium action and payoffs of a type 0 who is matched with another individual of type 0. Since in such equilibrium $P_1(x(0,0),R_0(x(0,0))=0$, the slope $\Psi'_0(x(0,0))$ is positive which implies that this point is to the left of the peak. From Lemma 1, $\sigma<0$ implies that $x(1,0,q)<x(0,0)$. Thus, point b^-, which represents the equilibrium action and payoff of a type 1 who is matched with an individual of type 0, must be on the left of the point a^-. As seen in Fig. 1, $P(x(1,0,q),\ x(0,1,q))<P(x(0,0),x(0,0))$.

Fig. 2 is used to illustrate part (ii) of Proposition 1. Point a^+ represents the equilibrium action and the payoffs of a type 1 individual who is matched with type 1. It can be shown (see Appendix A) that for a small positive σ, the slope $\Psi'_1(x(1,1,q))$ is positive, which implies that a^+ is to the left of the peak. From Lemma 1, $\sigma>0$ implies that $x(1,1,q)>x(0,1,q)$. Thus point b^+, which represents

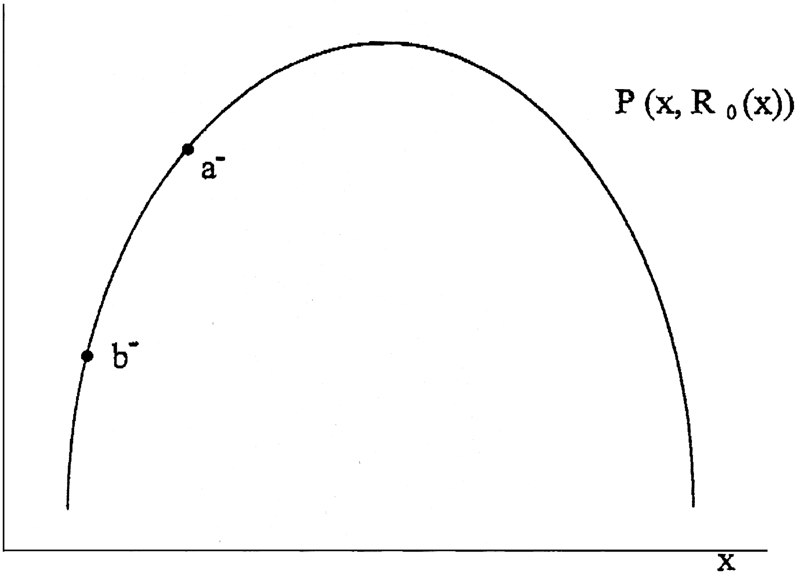

Fig. 1.

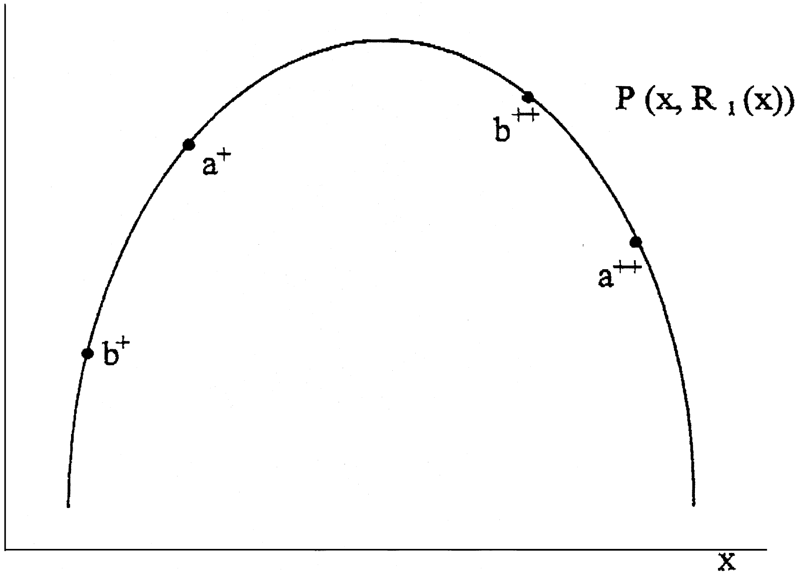

Fig. 2.

the equilibrium action and payoff of a type 0 who is matched with an individual of type 1, must be to the left of point a^{+}. As seen in Fig. 2, $P(x(0,1,q), x(1,0,q)) < P(x(1,1,q),x(1,1,q))$.

Note that Proposition 1 does not imply that every increasing social status function leads to an evolutionarily stable society in which everyone is socially minded. If the marginal social reward, σ, is positive but too high, the socially minded individual may select an action which reduces his fitness and, although his rival is induced to act in a favorable way, the net impact on fitness can be negative. This situation is illustrated by points a^{++} and b^{++} in Fig. 2. Point a^{++} is the equilibrium action and payoffs of type 1 individual who is matched with another individual of type 1. A large social reward σ induces the players to choose an action beyond the peak of $P(x,R_1(x))$. Since Lemma 1 implies that for $\sigma > 0$, $x(0,1,q) < x(1,1,q)$, point b^{++}, which represents the equilibrium action and payoff of a type 0 who is matched with type 1, is on the left of a^{++} and may yield a higher payoff.

5. Extended preferences sets

So far, we restricted our attention to a given set of possible preferences, allowing the weight given to social status, α, to be either 0 or 1. We now wish to extend our discussion and let α take any non-negative value, allowing players to vary in the importance they assign to social status. Specifically, let $A = \{\alpha_1, \alpha_2, \ldots \alpha_T\}$ be a finite set of non-negative real numbers. Each element α_t represents a possible weight given to social status in individual preferences. Let q_t be the proportion in the population of individuals with preference parameter α_t. Note that, by assuming that $\alpha \geq 0$, we exclude the possibility that social status is viewed as a ‘‘bad’’. Maintaining the assumption that relative fitness determine the growth rate of each type, we may inquire again which type survives in the long run.

For given σ and x^e, let $x^*_\sigma(\alpha,\alpha',x^e)$ be the equilibrium action of a player of type α when he plays with a player of type α'. Let $R_{\alpha,\sigma}(x,x^e)$ be the reaction function of a player with preference parameter α, given that the average action in the population is x^e, and given the social reward parameter, σ.[13] Now consider the pair (α^*, x^*) which satisfies the following two conditions:

$$E(x^*)P(x^*,x^*) > E(x^*)P(x,x^*) + \alpha^*\sigma(x - x^*) \forall x \neq x^*. \tag{5}$$

$$\forall \alpha \neq \alpha^*, P(x^*, R_{\alpha^*,\sigma}(x^*,x^*)) > P(x, R_{\alpha^*,\sigma}(x,x^*)) \textit{ where } x = x^*_\sigma(\alpha,\alpha^*,x^*). \tag{6}$$

Eq. (5) states that x^* is a strict Nash outcome within pairs in a society populated exclusively by players of type α^*. Eq. (6) states that x^* maximizes the fitness of

[13]We continue to assume that $\Psi_\alpha(x_i)$ is singled peaked for every α.

any player who engages in a game with a player of type α^*. Thus, if a person who enters a society populated by players of type α^* could choose his own preferences, he would maximize his fitness by having the preference parameter α^*. In this sense, α^* is the best response against α^* in terms of fitness. Although the existence of a solution (α^*, x^*) which satisfies Eqs. (5) and (6) is not guaranteed by our assumptions on the payoff function and the externality term, we provide an example of such a solution in Appendix B.

Proposition 4:

(i) If $\alpha^* \in A$, then a population consisting exclusively of this type is evolutionarily stable, *irrespective* of the values of the other members of A.

(ii) For every $\alpha \neq \alpha^*$, there is α' such that if both α and α' are members of A then α is not evolutionarily stable.

Proof: (i) Consider a population of individuals who are all of type α^*. By Eq. (5), playing x^* against a player of type α^* yields the highest economic payoffs. If another player, with $\alpha \neq \alpha^*$, is introduced into a society consisting mainly of type α^* players, his chosen action will differ from x^* and will provide him lower economic payoffs.

(ii) Consider a population of individuals who are all of type α, where $\alpha \neq \alpha^*$. Because $P(x,R_\alpha(x,x^e))$ is single peaked, it is possible to find α' which would induce players of this type to choose an action which yields them higher fitness than the fitness of players type α. ■

We may also consider some simultaneous changes in the social reward parameter, σ and the set of possible preferences, A. Recall that σ represents what others think about what one does, while α represents the importance that a person gives to the opinions of others. Despite the different interpretation of the two parameters, it is evident from equation Eq. (3) that individual choice depends only on the product $\alpha\sigma$.[14] That is, the same incentive is provided if the social reward is high but one cares little about it as if the reward is low but one care much about social status. We therefore obtain the following *invariance* result:

Proposition 5: Compare two societies, denoted by a and b, with the same monetary payoff function, $E(x^e)P(x_i,x_j)$, but different social reward parameters, σ_a and σ_b, respectively, such that $\sigma_a > \sigma_b > 0$. If (α_a^*, x^*) satisfies Eqs. (5) and (6) for society a, so that α_a^* is evolutionarily stable under σ_a, then (α_b^*, x^*) satisfies Eqs. (5) and (6) for society b, so that α_b^* is evolutionarily stable under σ_b, where $\alpha^*_b = \alpha^*_a \ \sigma_a/\sigma_b$.

[14] This property holds in our model because of the linear specifications of the utility function. Under a more general formulation, it is still correct that there are combinations of α and σ which provide the same incentives (locally), and maintain the same equilibrium action. These combinations will not necessarily satisfy $\alpha\sigma$=constant, but the negative relation between α and σ will be maintained.

Proposition 5 suggests that if we allow sufficient variation in the set of preferences, the effectiveness of social rewards is extremely limited. Any variation in the social reward will be eventually offset by a corresponding change in preferences so as to maintain the *same* action.

6. Status and externalities

So far, the externality term $E(x^e)$ did not enter into the analysis. Because the evolution of preferences depends on the *relative* fitness of types, the multiplicative externality effect, $E(x^e)$, has no effect on which type survives. However, the presence of externalities influences the long term welfare of the surviving type. It is this consideration which justifies our interest in social rewards as a corrective mechanism. It seems obvious that with a *given* profile of preferences, social rewards can induce individuals to internalize their effects on others and thereby eliminate the inefficiency caused by externalities. The question is whether social rewards can be effective when preferences are determined by an evolutionary process.

In our model, there are two external effects. The first externality represents coordination failure between the partners of each match. The second externality arises because the partners in each match do not consider the impact of their own actions on other members of society with whom they do not interact directly. These externalities can be clearly seen if we start from a situation with no social rewards, $\sigma=0$. Since, in this case, all individuals are identical, and there is only one action for each player, we may define an efficient action as a maximizer of the common economic payoff, i.e., $x^m = \text{ArgMax } E(x)P(x,x)$. Lack of coordination within each pair implies that the equilibrium outcome need not maximize $P(x,x)$. The neglect of external effects, means that players ignore their impact on $E(x)$. These two externalities may work in opposite directions, but, in general, the equilibrium action differs from x^m. We refer to the situation in which the equilibrium action is below (above) x^m as under (over) provision[15,16].

To raise efficiency, a society can use positive (negative) social rewards to increase (decrease) the equilibrium outcome. The question is whether individuals who respond to social rewards can survive. We consider this question, for

[15] It is also possible that without social status there is under provision but introducing social status will lead to over provision.

[16] A single efficient action arises only if all individuals are of the same type. If the population is not homogenous, there will be a multiplicity of (Pareto) undominated actions for the two types. Moreover, in the context of evolution, the composition of the population changes and tastes vary. However, x^* remains efficient as long as all individuals are identical, irrespective of whether they care about status. This property is a consequence of our assumption that status is measured in terms of deviations from the mean.

simplicity, under the assumption that there are only two types $\alpha \in \{0,1\}$. Based on the analysis of Section 4, we conclude:

Proposition 6: (i) When $P(x_i,x_j)$ is such that the effects of rival's action on the total and the marginal payoffs are of the same sign, i.e., $P_2(x_i,x_j)\ P_{12}(x_i,x_j)>0$, then under provision of x, can be regulated by social rewards, while over provision cannot be regulated by social rewards and requires an enforcement mechanism.

(ii) When $P(x_i,x_j)$ is such that the effects of rival's action on the total and the marginal payoffs are of the opposite sign, i.e., $P_2(x_i,x_j)\ P_{12}(x_i,x_j)<0$, then over provision of x can be regulated by social rewards, while under provision cannot be regulated by social rewards and requires an enforcement mechanism.

Proof: These results follow directly from Propositions 1 and 2. ■

Proposition 6 points out an interesting asymmetry in dealing with over and under provision of x. Depending on the characteristics of the payoff function, it is possible to use social rewards to deal either with over provision or with under provision of x but not with both. This result suggests that whether society relies on enforcement or social rewards depends on the nature of the economic interaction between matched players. It common in economics to represent interactions as the product of the activity levels of the two players (e.g., a Cournot game with linear demands). This specification, which we analyze in detail in Appendix B, has the feature that the effect of a player's action on the total and the marginal payoffs of his rival are of the same sign. In this case, Proposition 6 implies that social rewards will be effective in dealing with under provision, but not with over provision, in which case legal enforcement will be required.

It is important to note that Proposition 6, does *not* imply that by using social rewards one can attain the efficient level, x^m. We only claim that *some* improvement can be made by the use of positive social rewards. This qualification arises because it is possible that the σ required to support x^m is too large and, therefore, socially minded individuals will not survive in the long run. Thus, some legal intervention may be required even if social rewards are effective. Under the conditions discussed in Section 5, where we extend the range of values that the preference parameter α may take, the evolutionarily stable action, x^* is *independent* of σ (as long as σ does not change sign) and, therefore, any relation between x^* and x^m is possible. That is, the evolutionarily stable equilibrium is generally inefficient and, depending on the parameters, there may be either under or over provision (see the example in Appendix B).

7. Concluding remarks

This paper examined the role of social status as a corrective mechanism for externalities. To raise efficiency, it is possible to use legal rules, subsidies or taxes.

However, each of these means is costly either because of direct loss of resources (e.g., jailing) or because of negative effects on incentives (dead weight loss). We have shown that, under some conditions, society can use social rewards, such as status, which are relatively cheap, to regulate externalities. The main requirement is that socially minded individuals, who care about social status, will not be driven away by asocial individuals who selfishly maximize their fitness. We show that, depending on the nature of economic interactions, a player who cares about status may end up with higher fitness, in equilibrium. In our model, the only way a person who cares about status can gain fitness is by inducing his rivals in economic interactions to select actions which are beneficial to his fitness. This idea is quite distinct from a possible effect of status on fitness, based on the direct *economic benefits* that high social status may entail, or on a *direct* impact of status on fitness, because of some advantage in mating.

Our results show that social rewards have only a limited impact on behavior when preferences are determined by economic fitness. If preferences are restricted to only two types, social and asocial, then, under plausible conditions, it is only possible to regulate positive externalities. If one allows for preferences with any non negative weight on social rewards then, again, only two types of preferences can last, either social or asocial, depending on the nature of the interactions. However, the weight that social preferences put on social status adjusts to social rewards in such a way to make the equilibrium action *independent* of the social rewards. This evolutionarily stable outcome may be below or above the efficient outcome.

Alternative models of social status, such as Okuno-Fujiwara and Postlewaite (1995), do not rely on a direct effect of status on utility. Instead, the impact on behavior is derived from the reputation that one accumulates, which influences the behavior of others. In this framework, social status is quite potent and efficiency is attainable. These models rely on a positive probability of long life and perfect foresight, conditions which allow the application of the folk theorem. Our approach considers short lived agents with limited foresight, allowing the ‘‘blind’’ forces of evolution to select preferences. It is perhaps not surprising that under such circumstances, it is more difficult to achieve efficiency.

Acknowledgements

We would like to thank Giacomo Corneo, Eddie Dekel, Martin Dufwenberg, Peter Funk, Paul DiMaggio, Ariel Rubinstein, John Romer, Avner Shaked, Martin Weitzman, P. Young, the participants of the seminars at LSE, Tilburg, Oslo and Tel Aviv universities and the ISPE conference on ‘‘The Economics of Status’’ at Bonn University, and the Editors and two anonymous referees of this journal, for valuable comments and discussion.

Appendix A

Proof of proposition 1

Claim 1: Assume that $P(x_i,x_j)$ satisfies the condition $P_2(x_i,x_j)\ P_{12}(x_i,x_j) > 0$, then:

(i) For any q, if $\sigma < 0$ then, in any possible matching, a type 1 individual will have a *lower* equilibrium payoff than a type 0 individual. That is,

(ia) $P(x(1,0,q),x(0,1,q)) < P(x(0,0),x(0,0))$,

(ib) $P(x(1,1,q),x(1,1,q)) < P(x(0,1,q),x(1,0,q))$.

(ii) For any q there exists a positive σ_0 such that for all $0 < \sigma \leq \sigma_0$, a type 1 individual will have a *higher* equilibrium payoff than a type 0 individual, in any possible matching. That is,

(iia) $P(x(1,1,q),x(1,1,q)) > P(x(0,1,q),x(1,0,q))$,

(iib) $P(x(1,0,q),x(0,1,q)) > P(x(0,0),x(0,0))$.

Proof: When two individuals of type 0 are matched, the first order condition for player i is

$$E(x^{e})P_1(x_i,x_j) = 0. \tag{A1}$$

When player i is of type 1 and player j is of type 0, the first order conditions are

$$E(x^{e}P_1(x_i,x_j) + \sigma = 0, \tag{A2}$$

$$E(x^{e}P_1(x_j,x_i) = 0. \tag{A3}$$

When two individuals of type 1 meet, the first order condition for player i is

$$E(x^{e})P_1(x_i,x_j) + \sigma = 0. \tag{A4}$$

Let $R_\alpha(x)$, $\alpha = 0,1$, denotes the reaction function of a player of type α and define $\Psi_\alpha(x) \equiv P(x,R_\alpha(x))$, $\alpha = 0,1$, then differentiating $\Psi_\alpha(x)$, and using the first order conditions, imply that, in equilibrium:

(i) $\Psi'_0(x(0,0)) = P_2(x(0,0),x(0,0))R'_0(x(0,0))$, where, $R'_0(x(0,0)) = -P_{12}(x(0,0),x(0,0))/P_{11}(x(0,0),x(0,0))$.

(ii) $\Psi'_0(x(1,0,q)) = -\sigma/E(x^{e}) + P_2(x(1,0,q),x(0,1,q))R'_0(x(1,0,q))$, where, $R'_0(x(1,0,q)) = -P_{12}(x(0,1,q),x(1,0,q))/P_{11}(x(0,1,q),x(1,0,q))$.

(iii) $\Psi'_1(x(0,1,q)) = P_2(x(0,1,q),x(1,0,q))R'_1(x(0,1,q))$, where, $R'_1(x(0,1,q)) = -P_{12}(x(1,0,q),x(0,1,q)/P_{11}(x(1,0,q),x(0,1,q))$.

(iv) $\Psi'_1(x(1,1,q)) = -\sigma/E(x^{e}) + P_2(x(1,1,q),x(1,1,q))R'_1(x(1,1,q))$, where, $R'_1(x(1,1,q)) = -P_{12}(x(1,1,q),x(1,1,q))/P_{11}(x(1,1,q),x(1,1,q))$.

By assumption, $P_{11}(.,.) < 0$, $P_2(.,.)$ and $P_{12}(.,.)$ do not change sign and $P_2(.,.)P_{12}(.,.) > 0$. It follows that $\Psi'_0(x(0,0))$ and $\Psi'_1(x(0,1,q))$ are positive for *all* σ, and that $\Psi'_0(x(1,0,q))$ and $\Psi'_1(x(1,1,q))$ are positive for $\sigma < 0$. If σ is positive, the terms $\Psi'_0(x(1,0,q))$ and $\Psi'_1(x(1,1,q))$ may be positive or negative. However, as σ approaches 0, the terms $P_2(.,.)R'_\alpha(.)$ all approach $P_2(x(0,0),x(0,0))R'_\alpha(x(0,0))$, which is strictly positive and the term $E(x^e)$ approaches $E(x(0,0))$ which is a positive number. Since $\Psi'_\alpha(.)$ are continuous functions of σ, there exist a σ_0 such that $\Psi'_0(x(1,0,q))$ and $\Psi'_1(x(1,1,q))$ are all positive for all σ such that $\sigma<\sigma_0$.

Using the results above and the assumption that, for $\alpha=0,1$, $\Psi_\alpha(x)\equiv P(x,R_\alpha(x_i))$ is single peaked, we can now prove the four parts of Claim 1.

To prove part (ia) of Claim 1, we use the fact that, $\Psi'_0(x(0,0))>0$ and that for a negative σ, $x(1,0,q)<x(0,0)$. To prove part (ib) of Claim 1, we use the fact that $\Psi'_1(x(0,1,q))>0$ and that for a negative σ, $x(1,1,q)<x(0,1,q)$.

To prove part (iia) of Claim 1 we use the fact that, for $0<\sigma\leq\sigma_0$, $\Psi'_1(x(1,1,q))>0$ and $x(1,1,q)>x(0,1,q)$. To prove part (iib) of Claim 1, we use the fact that, for $0<\sigma\leq\sigma_0$, $\Psi'_0(x(1,0,q))>0$ and $x(1,0,q)>x(0,0)$.

This concludes the proof of Claim 1. The proof of Proposition 1 follows directly from Claim 1 and the definitions of evolutionary stable society. ■

Appendix B

The Cournot duopoly

This appendix illustrates our results for the Cournot duopoly game. Our purpose is twofold: (i) to establish that all our assumptions are satisfied in a non-trivial economic example (ii) to provide a more precise analysis of the efficacy of social rewards.

In a duopoly with linear demand function, and where each firm has quadratic costs, the profit of each firm is a *quadratic* function of the outputs of the two firms. The same holds in a simple differentiated product model, in which the demand for each product is a linear function of the prices of the two producers, yielding profit functions of each player which are *quadratic* in prices. The only formal difference between the two cases is that in the quantity game the interaction is negative, while in the price game the interaction is positive. To incorporate externalites in the duopoly game we may consider an economy consisting of many doupolists, where the output in each industry causes negative (e.g., pollution) or positive (e.g., R&D spillovers) externalities which affect everyone. To represent these ideas in a simple parametric form, we let

$$P(x_i,x_j)=(1-\gamma)x_i+\gamma x_i x_j-x_i^2/2. \tag{B1}$$

where, $-1/2<\gamma<\frac{1}{2}$, and let

$$E(x^e)=(1+x^e)^\epsilon \tag{B2}$$

where, $-1<\varepsilon<1$.

Under specification Eq. (B1), $P(x_i,x_j)$ is strictly concave in x_i, $P_{11}(x_i,x_j)$ $P_{11}(x_j,x_i)>P_{12}(x_i,x_j)$ $P_{12}(x_j,x_i)$, and $P_2(x_i,x_j)$ and $P_{12}(x_i,x_j)$ are both positive (negative) if γ is positive (negative). Also, $P_1(0,0)=1-\gamma>0$. The implied reaction functions of types 0 and 1 are $R_0(x)=(1-\gamma)+\gamma x$ and $R_1(x)=(1-\gamma)+\gamma x+\sigma/E(x^e)$, respectively. Substituting these reaction functions, we see that $P(x,R_\alpha(x))$ is concave in x for $\alpha=0,1$, and thus single peaked. Under Eq. (B2), $E(x^e)$ is positive and is monotone increasing (decreasing) if ε is positive (negative). A positive (negative) ε indicates positive (negative) externalities.

Using Eq. (B1), it is easy to calculate the equilibrium actions and payoffs, for any given x^e. If two players of type 0 meet then, in equilibrium, $x(0,0)=1$ and

$$P(x(0,0),x(0,0)) = 1/2. \tag{B3}$$

If two players of type 1 meet then, in equilibrium, $x(1,1,q)=1+\sigma/(1-\gamma)E(x^e)$ and

$$P(x(1,1,q),x(1,1,q)) = 1/2 + \frac{\gamma}{1-\gamma}\left(\frac{\sigma}{E(x^e)}\right) + \frac{\gamma-1/2}{(1-\gamma)^2}\left(\frac{\sigma}{E(x^e)}\right)^2. \tag{B4}$$

If players type 0 and type 1 meet then, in equilibrium, $x(1,0,q)=1+\sigma/(1-\gamma)^2E(x^e)$ and $x(0,1,q)=1+\sigma\gamma/(1-\gamma)^2E(x^e)$ and the equilibrium payoffs are:

$$P(x(1,0,q),x(0,1,q) = 1/2 + \frac{\gamma^2}{1-\gamma^2}\left(\frac{\sigma}{E(x^e)}\right) + \frac{\gamma^2-1/2}{(1-\gamma^2)^2}\left(\frac{\sigma}{E(x^e)}\right)^2, \tag{B5}$$

$$P(x(0,1,q),x(1,0,q)) = 1/2 + \frac{\gamma}{1-\gamma^2}\left(\frac{\sigma}{E(x^e)}\right) + \frac{1}{2}\left(\frac{\gamma}{(1-\gamma)^2}\right)^2\left(\frac{\sigma}{E(x^e)}\right)^2. \tag{B6}$$

Using Eq. (B2), we can now demonstrate that the average action, x^e, is unique. Taking expectation over all possible pairs, Eq. (4) can be reduced to

$$x^e = 1 + \frac{q}{1-\gamma}\left(\frac{\sigma}{E(x^e)}\right). \tag{B7}$$

The uniqueness of x^e, given σ and q, follows from the fact that, under Eq. (B2), $E(x^e)(x^e-1)$ is monotone increasing in x^e . The monotonicity of $E(x^e)(x^e-1)$ implies that the average level of activity increases (decreases) with q, if σ is positive (negative), and for any positive q, an increase in σ raises the average activity level, irrespective of whether the externality effect is positive or negative. It can also be verified that the equilibrium activity level of any pair increases in x^e, whenever x^e satisfied Eq. (4) in the text.

We can now verify Proposition 1. From Eqs. (B3) and (B5), it is immediately seen that, for $\sigma<0$, $P(x(1,0,q),\ x(0,1,q))<P(x(0,0),\ x(0,0))$. Thus, type 0 is evolutionarily stable. By comparing Eqs. (B4) and (B6) it is evident that for $\sigma<0$,

$P(x(1,1,q), x(1,1,q)) < P(x(0,1), x(1,0))$. Thus, type 1 cannot be evolutionarily stable, as stated in part (i) of Proposition 1.

Examining Eqs. (B3)–(B6), it can be verified that for $0 < \sigma < \gamma^2 E(x^e)$, $P(x(1,1,q),x(1,1,q)) > P(x(0,1,q),x(1,0,q))$ and $P(x(1,0,q),x(0,1,q)) > P(x(0,0), x(0,0))$. Therefore, type 1 is evolutionarily stable for $0 < \sigma < \gamma^2 E(x(1,1,1)$, while type 0 is not evolutionarily stable for $0 < \sigma < \gamma^2 E(x(0,0))$. Since $x(0,0) = 1$ and is independent of σ, it follows immediately that a population consisting only of type 0 individuals is evolutionary unstable, for *all* σ such that $0 < \sigma < \gamma^2 2^{\epsilon}$. Since $x(1,1,1)$ depends on σ, it is more difficult to obtain an explicit expression (in terms of the parameters of Eqs. (B1) and (B2)) for the condition that σ is sufficiently small for status to have a positive effect on the fitness of type 1. Here we need to separate two cases, positive externalities, $\epsilon > 0$, and negative externalities, $\epsilon < 0$. In the case of positive externalities, it is easy to show that type 1 is evolutionarily stable for *all* σ such that $0 < \sigma < \gamma^2 2^{\epsilon}$. This follows from the observation that for a positive σ, $x(1,1,1) > x(0,0) = 1$ and, therefore, if $\epsilon > 0$ then $E(x(1,1,1)) > E(x(0,0)) = 2^{\epsilon}$. For the case of negative externalities, we can only say that the sufficient condition $0 < \sigma < \gamma^2 E(x(1,1,1)$ is satisfied for all $\sigma < \sigma_0$, where σ_0 is some critical value satisfying $0 < \sigma_0 < \gamma^2 2^{\epsilon}$. To see that such a value exists, consider the difference $\gamma^2 E(x(1,1,1)) - \sigma$ as a function of σ. Observe that, for $\epsilon < 0$, this function is monotone decreasing, positive at $\sigma = 0$ and negative at $\sigma = 2^{\epsilon}$. We have thus verified the existence of a positive critical value for σ, such that for all positive σ which are less then this critical value, type 1 is evolutionarily stable, while type 0 is not evolutionarily stable, as stated in part (ii) of Proposition 1.

We can use the example to describe the conditions for under and over provision. Define $f(x) = E(x)P(x,x)$ and assume the specifications Eqs. (B1) and (B2). Then,

$$f'(x) = (1+x)^{\epsilon-1}[x(\epsilon(1-\gamma)+\gamma) + x^2(\gamma - 1/2)(\epsilon+2) + 1 - \gamma], \tag{B8}$$

$$f''(x) = \frac{(\epsilon-1)f'(x)}{1+x} + (1+x)^{\epsilon-1}[(\epsilon(1-\gamma)+\gamma) + 2x(\gamma-1/2)(\epsilon+2)]. \tag{B9}$$

There is a unique positive maximizer of $f(x)$, x^{m}, because at the point where $f'(x) = 0$ and $x > 0$, $f''(x) < 0$. Recall that when all individuals don't care about social status they will all choose $x = 1$. From Eq. (B8), it is evident that the sign of $f'(1)$ is determined by the sign of $\epsilon/2 + 2\gamma$. Thus, in the absence of social rewards, the equilibrium is efficient if $\epsilon/2 + 2\gamma = 0$, because the externalities within and across pairs cancel each other. However, a positive (negative) $\epsilon/2 + 2\gamma$ implies under (over) provision and there is a potential corrective role for social rewards.

Consider now the determination of the pair (α^*, x^*) which solves Eqs. (5) and (6) in the text. Defining, $c_t = \sigma\alpha_t / E(x^e)$, we obtain that the following solutions for c^* and x^* satisfy Eqs. (5) and (6):

$$c^* = \frac{\gamma^2(1-\gamma)}{1-\gamma-\gamma^2}, \tag{B10}$$

$$x^* = \frac{(1-\gamma)}{1-\gamma-\gamma^2}. \tag{B11}$$

Note that these quantities depend only on the interaction parameter γ. For a positive σ, we get α^* from the definition $\alpha^* = c^*x^*/\sigma$. However, since Eq. (B10) implies a positive c^*, there is no admissible solution for α if $\sigma \leq 0$. In this case, the only evolutionarily stable equilibrium is such that no one cares about status, i.e., $\alpha^* = 0$, as stated in proposition 1.

Restricting our attention to a positive σ, we may ask what is the relation between x^* and the efficient level x^m. The function $f'(x)$ which determines x^m changes sign only once, being first positive and then negative. It is, therefore, sufficient to evaluate $f'(x^*)$ in order to determine the sign of $x^m - x^*$. Simple calculations show that

$$f'(x^*) = (1+x^*)^{\epsilon-1}\left(\frac{(1-\gamma)}{1-\gamma-\gamma^2}\right)^2\left[\left(2\gamma+\frac{\epsilon}{2}\right) - 2\gamma^2 - \gamma^3 - \epsilon\gamma^2\right]. \tag{B12}$$

It is seen from equation (B12) that if $\epsilon/2 + 2\gamma = 0$, so that the equilibrium with no social rewards is efficient, then $f'(x^*) < 0$, implying that $x^m < x^*$. (This can be verified directly by noting that, by Eq. (B11), $x^* > 1$ and that $x^m = 1$ when $\epsilon/2 + 2\gamma = 0$.) However, for γ close to $\frac{1}{2}$ and ϵ close to 1, $f'(x^*) > 0$, implying that $x^m > x^*$. Thus, the evolutionarily stable equilibrium is generally inefficient and, depending on the parameters, there may be either under or over provision.

References

Akerlof, G., 1980. A theory of social custom, of which unemployment may be one consequence. Quarterly Journal of Economics 94, 749–775.

Arrow, K.J., 1971. Political and economic evaluation of social effects and externalities. In: Intriligator, M. (Ed.), Frontier of Quantitative Economics. North Holland, Amsterdam.

Basu, K., 1995. Civil institution and evolution: concepts, critique and models. Journal of Development Economics 46, 19–33.

Becker, G., 1992. Habits, addictions, and traditions. Kyklos 45, 327–345.

Becker, G., Mulligan, C., 1997. On the endogenous determination of time preference. Quarterly Journal of Economics 112, 729–758.

Bester, H., Guth, W., 1997. Is altruism evolutionary stable? Journal of Economic Behavior and Organization, forthcoming.

Dekel, E., Scotchmer, S., 1994. On the evolution of attitudes towards risk in winner-take-all games. Mimeo.

Elster, J., 1989. Social norms and economic theory. Journal of Economic Perspectives 4, 99–117.

Fershtman, C., Judd, K., 1987. Incentive equilibrium in oligopoly. American Economic Review 77, 927–940.
Fershtman, C., Weiss, Y., 1997. Why do we care what others think about us. In: Ben Ner, A., Putterman, L. (Eds.), Economics, Values and Organization. Cambridge University Press, Cambridge.
Fershtman, C., Murphy, K.M., Weiss, Y., 1996. Social status, education and growth. Journal of Political Economy 104, 108–132.
Fudenberg, D., Tirole, J., 1992. Game Theory. MIT Press, Cambridge MA.
Hammerstein, P., Selten, R., 1994. Game theory and evolutionary biology. In: Aumann, R.J., Hart, S., Handbook of Game Theory with Economic Applications, vol 2, chapter 28. Elsevier, Amsterdam.
Hirshleifer, J., 1980. Privacy: Its origin, function, and future. Journal of Legal Studies 9, 649–664.
Lazear, E.P., Rosen, S., 1981. Rank-order tournaments as optimum labor contracts. Journal of Political Economy 89, 841–864.
Maynard Smith, J., 1982. Evolution and the Theory of Games. Cambridge University Press, Cambridge.
Okuno-Fujiwara, M., Postlewaite, A., 1995. Social norms and random matching games. Games and Economic Behavior 9, 79–109.
Robson, J.A., 1996. A biological basis for expected and non-expected utility. Journal of Economic Theory 68, 397–424.
Rogers, A.R., 1994. Evolution of time preference by natural selection. American Economic Review 84, 460–481.
Weibull, J.W., 1995. Evolutionary Game Theory. MIT press, Cambridge MA.