

# On the Evolution of Individualistic Preferences: Complete versus Incomplete Information Scenarios<sup>α</sup>

Efe A. Ok<sup>γ</sup>

Fernando Vega-Redondo<sup>z</sup>

March 1999

(this version: May 1999)

## Abstract

We study the evolution of preferences via payoff monotonic dynamics in strategic environments with and without complete information. It is shown that, with complete information and subgroup matching, empirically plausible interdependent preference relations may entail the local instability of individualistic preferences (which target directly the maximization of material payoffs/fitness). The said instability may even be global if the subgroup size is large enough. In contrast, under incomplete information (unobservability of preference types), we show that independent preferences are globally stable in a large set of environments, and locally stable in essentially any standard environment, provided that the number of subgroups that form in the society is large. Since these results are obtained within the context of a very general model, they may be thought of as providing an evolutionary rationale for the prevalence of individualistic preferences.

JEL Classification: C72, D62.

Keywords: Evolution, Preferences, Incomplete Information.

---

<sup>α</sup>We would like to thank Alberto Bisin, Kalyan Chatterjee, Levent Koçkesen, Larry Samuelson and Rajiv Sethi for useful comments and suggestions. The first author gratefully acknowledges financial support from the National Science Foundation through Grant No. 9808208, and the second author support by the Spanish Ministry of Education, CICYT Project no. 970131. Support from the C. V. Starr Center for Applied Economics at New York University is also acknowledged.

<sup>γ</sup>Department of Economics, New York University, 269 Mercer Street, New York NY 10003. E-mail: okefe@fasecon.econ.nyu.edu.

<sup>z</sup>Facultad de Económicas and Instituto Valenciano de Investigaciones Económicas, University of Alicante, Alicante, Spain. E-mail: fvega@merlin.fae.ua.es.

# 1 Introduction

Traditional economic analysis takes the individual preferences as an exogenous datum and does not consider their particular form as a relevant object of study. Moreover, there appears to be a consensus in the field about which type of preferences to allow in economic models. The working assumption of a vast majority of economists on this regard is that an individual's behavior is guided by the sole motive of the maximization of one's own material payoffs. However, this assumption has recently come under severe criticism in light of the evidence obtained from numerous experiments that involve strategic situations. In turn, this has led many economists to consider alternative (non-individualistic) preference structures that yield predictions which accord better with the experimental regularities.<sup>1</sup> It appears that time has come to critically examine the validity of exclusively modeling one's preferences as "individualistic" (i.e. independent of others' payoffs), and to ask deeper questions about the basis and plausibility of alternative preference structures.

One natural approach to addressing this issue is to adopt an evolutionary perspective and ask the following question: What type of preferences are evolutionarily stable, in the sense of inducing material payoffs at least as high as any alternative "mutant" in any given environment? If one subscribes to the widely held view that

(i) success (reproductive or otherwise) is an increasing function of material payoffs, and

(ii) individual preferences are inherited either by genetic transmission and/or imitation,

any clear-cut answer to this question will provide a useful "evolutionary rationalization" of certain preference structures (at least in some environments). This is precisely the approach adopted here.<sup>2</sup>

Since the goals of agents with individualistic preferences may be conceived identical to those of natural selection (i.e. the maximization of material resources), the evolutionary approach has been often used in economics to justify the assumption of material payoff maximization (cf. Friedman, 1953). However, it turns out that this line of argument is not true without some non-trivial qualifications, for it is possible that non-individualistic preferences are materially more rewarding than individualistic preferences in certain strategic environments. For instance, if the dynamics of

---

<sup>1</sup>See, among others, Bolton and Ockenfels (1998), Fehr and Schmidt (1998) and Levine (1998).

<sup>2</sup>An alternative (and by all means complementary) approach would be to model the formation of preferences via positing that preferences are transmitted through generations by means of the socialization actions of the parents. While mostly prominent in the field of cultural anthropology, this cultural evolutionary approach has received some attention from economists as well. For instance, the recent work by Bisin and Verdier (1998, 1999) focus on the implications of modeling the intergenerational transmission of cultural traits (and, in particular, preferences) as a result of the deliberate inculcation attempts of rational parents who evaluate the ex ante well-being of their children by using their own preferences. We refer the reader to Selten (1991) for a lively comparison of the genetic and cultural approaches.

evolution takes place through pairwise random matching and/or local interaction, then altruistic preferences may well be evolutionarily stable (Fershtman and Weiss, 1998, Bester and Güth, 1998, and Eshel et al., 1998). In a similar vein, Koçkesen et al. (1998, 1999) show that if all agents interact with every other agent simultaneously, then negatively interdependent (spiteful) preferences have a sharp evolutionary edge against individualistic (selfish) preferences in a large class of strategic models (that contain, for instance, the common pool resource and public good games). The evolutionary case for individualistic preferences is thus not at all straightforward.

In this paper, we aim to lay out an evolutionary foundation for individualistic preferences. We shall argue that the conclusions reached by the above cited studies depend crucially on the implicit assumption that preferences are common knowledge (i.e. the underlying game is one of complete information). Intuitively, under complete information, non-individualistic preferences serve as credible commitment devices, and hence it is not surprising that they may prove profitable in terms of the intrinsic (material) payoffs of a given game. However, under incomplete information (more precisely, under unobservability of preferences), they lose credibility, and as we intend to show, they may consequently be dominated by individualistic preferences in almost any environment.

To be more precise, consider a finite population in which each individual has either individualistic or an arbitrarily mixed type of non-individualistic preferences. Suppose as well that agents are randomly matched in subgroups to play a given symmetric game in strategic form, and assume that the two types of preferences under consideration have different equilibrium implications in this game (so that they are behaviorally distinguishable). Loosely stated, the question that we ask is this: How do the (expected) material payoffs of the individualistic and non-individualistic agents compare in equilibrium at various population compositions? As might be expected, it turns out that the answer depends crucially on the following two considerations: (i) the extent of information agents have on the "type" of their opponents; (ii) the number of subgroups in the society.

In what follows, we first take up the complete information scenario in which we assume that the types of each individual are perfectly observable. Building upon the earlier results of Koçkesen et al. (1999), we show that, in a variety of economic environments, individualistic preferences are locally unstable. Put more concretely, we provide examples of classes of games in which "spiteful" agents obtain more material payoffs than material-payoff maximizing individualistic agents in any equilibrium, at least when the former represent a small fraction of the whole population. We also show that, depending on the particulars of the environment, individualistic preferences may even be dominated globally, i.e. irrespectively of the frequencies of each type in the population.

Our main concern in this paper is, however, to understand the arguably more realistic scenario in which individuals do not observe their opponents' preferences but they are only informed of the

overall population frequencies. In such an incomplete-information setup, the relative magnitudes of the population and subgroup sizes become important since they determine the effective uncertainty experienced by any player in predicting her opponents' types. Thus, if the subgroups are relatively large (and the effective matching uncertainty is therefore small), the conclusions obtained under perfect observability are recovered, i.e. interdependent preferences prevail in a wide class of economic environments. On the other hand, if the subgroups are relatively small compared to the population, matching uncertainty does have "bite," and in this case, we reach to a very different conclusion: only individualistic preferences can survive in the long run under any material payoff-responsive evolutionary process. We view this result as a benchmark that provides a rigorous evolutionary rationale for individualistic preferences. It is worth stressing that, in contrast with the earlier results obtained in the literature, this observation remains valid in any game that satisfies certain standard convexity and continuity conditions, and what is more, it applies "globally" within a very large class of environments.<sup>3</sup>

The paper is organized as follows. In Section 2 we introduce a general model in which the stability of individualistic preferences is analyzed under the hypothesis of perfect observability of types. In this section we provide a number of examples that demonstrate how easily can individualistic preferences be "beaten" by others in their own turf (i.e. in terms of material payoffs). In Section 3, we prove two general limit results which establish the evolutionary superiority of individualistic preferences under the hypothesis of imperfect observability of types. A number of caveats and suggestions for future work are contained in the concluding Section 4.

## 2 Stability of Individualism under Complete Information

In this section we shall introduce an evolutionary framework in which one can study the stability properties of different types of preference structures, in particular, the stability of interdependent preferences. Our aim is to demonstrate that there is no reason to view such preferences as evolutionarily stable, provided that evolutionary dynamics takes place in a setting in which all individuals observe perfectly the types of every other agent.

By a pre-environment, we mean a 4-tuple  $(n; k; X; \frac{1}{4})$ , where

(i)  $n \geq 2$ ;  $f_2, 3, \dots, g$ ;

---

<sup>3</sup>A local version of this result is obtained by Dekel et al. (1998) who work with finite games and continuum populations (see also Ely and Yankaya, 1997, and Güth and Peleg, 1997). While the main emphasis of Dekel et al. (1998) is not on the evolution of preferences per se but rather on the implications of this notion for prevailing behavior, there are some similarities between the results obtained by these authors and those reported here. We shall thus elaborate on the relation between each of them in the sequel (cf. Remark 4 in Section 3).

- (ii)  $k \geq 2$ ;  $n \geq k$ ;  $n$  is a divisor of  $n$ ;
- (iii)  $X$  is any nonempty set,
- (iv)  $\varphi : X^k \rightarrow \mathbb{R}$  is any function.

We view  $N = \{f_1, \dots, f_n\}$  as a population of  $n$  individuals which is randomly matched into  $n/k$  many subgroups of size  $k$ : So if  $k = 2$ ; we talk of a pairwise random matching environment, and if  $k = n$ , we talk of a playing the ...eld environment. Once the matching takes place, the players of each subgroup  $f_{i_1}, \dots, f_{i_k}$  play a  $k$ -person game in which  $\varphi(x_{i_1}, \dots, x_{i_k})$  stands for the material payoffs that the player  $i_j$  obtains at the action profile  $x$ : The function  $\varphi$  is thus thought of as measuring the access of an individual to resources that may enhance reproduction. Alternatively, one may think of  $\varphi(x_{i_1}, \dots, x_{i_k})$  as a direct measure of the ...tness of player  $i_j$  at  $x$ :

We refer to an individual who targets solely the maximization of her material payoffs as an individualistic (or independent, or materialist) player. A player  $j \in \{i_1, \dots, i_k\}$ ; however, need not view the mapping  $x \mapsto \varphi(x_j, x_{-j})$  as her actual objective function. Instead, we posit that the payoff function of  $j$  may be represented by means of an alternative mapping  $x \mapsto \psi_j(x_j, x_{-j})$ , in which case we say that  $j$  is a non-individualistic player.

Consequently, we define an environment as a 2-tuple  $E = (j; \psi_j)$ ; where  $j$  is a pre-environment, and  $\psi_j : X^k \rightarrow \mathbb{R}$  is any function. Since a stability analysis is meaningful only in the case of polymorphic populations, we define a population composition as a vector of types  $t = (t_1, \dots, t_n) \in \{0, 1\}^n$  such that  $t_1 = 1$  and  $t_n = 0$ : Thus, by definition, a population composition contains at least one individualistic and one non-individualistic agent. Our convention is to let individual 1 possess individualistic preferences while keeping individual  $n$  non-individualistic. Given our symmetric setting, this convention is without loss of generality.

Given an environment  $E$  and a population composition  $t$ ; the players of each subgroup  $f_{i_1}, \dots, f_{i_k}$  play the  $k$ -person game in strategic form  $G_{f_{i_1}, \dots, f_{i_k}}(t) = (X; u_j)_{j=i_1, \dots, i_k}$ , where

$$u_j(x) = \begin{cases} \varphi(x_j, x_{-j}); & \text{if } t_j = 1 \\ \psi_j(x_j, x_{-j}); & \text{if } t_j = 0 \end{cases} \quad (1)$$

for all  $x \in X^k$  and  $j = i_1, \dots, i_k$ : Notice that if  $t_j = 1$  for all  $j$ ; then this game is a symmetric game played by individualistic players who maximize their material payoffs. The set of all pure-strategy Nash equilibria of  $G_{f_{i_1}, \dots, f_{i_k}}(t)$  is denoted by  $N(G_{f_{i_1}, \dots, f_{i_k}}(t))$ :

Let  $A$  denote the  $k$ -element subsets of  $\{f_1, \dots, f_n\}$ ; and define

$$A_{1/k} = \{A \subseteq A : |A| = k\} \quad \text{and} \quad A_{n/k} = \{A \subseteq A : |A| = n/k\} \quad (2)$$

We next define the following set of equilibrium outcome vectors:

$$N_k(t) = \{f(x^A)_{A \in A_{1/k}} : x^A \in N(G_A(t))\}$$

A vector  $(x^A) \in N_k(t)$  simply specifies an equilibrium outcome in every possible game played by  $k$  individuals. Thus, the expected payoff of an individualistic player prior to the realization of the subgroup matching is:

$$\pi_{\frac{1}{2}}(t; (x^A)) = \frac{1}{|A_{\frac{1}{2}}|} \sum_{A \in A_{\frac{1}{2}}} \pi_A(x^A; x_{i-1}^A); \quad (3)$$

provided that the members of each subgroup  $A \in A_{\frac{1}{2}}$  (that contains at least one individualistic player) coordinate on the equilibrium  $x^A$  when playing  $G_A(t)$ : Similarly, conditional on an equilibrium selection  $(x^A) \in N_k(t)$ ; the expected payoff of a non-individualistic individual is

$$\pi_{\frac{1}{2}}(t; (x^A)) = \frac{1}{|A_{\frac{1}{2}}|} \sum_{A \in A_{\frac{1}{2}}} \pi_A(x_n^A; x_{i-n}^A) \quad (4)$$

for any population composition  $t$ .

We are now ready to state the following two stability concepts, both of which are rooted in the widely used material-payoff monotonicity property.

**Definition 1.** Let  $E = (i; \frac{1}{2})$  be any environment. The individualistic preferences are said to be **locally unstable** in  $E$ ; if

$$\pi_{\frac{1}{2}}(\frac{1}{4}; \dots; \frac{1}{4}; \frac{1}{2}; (x^A)) < \pi_{\frac{1}{2}}(\frac{1}{4}; \dots; \frac{1}{4}; \frac{1}{2}; (x^A)) \quad \text{for all } (x^A) \in N_k(\frac{1}{4}; \dots; \frac{1}{4}; \frac{1}{2});$$

We say that individualistic preferences are **exterminated** by  $\frac{1}{2}$ ; if

$$\pi_{\frac{1}{4}}(t; (x^A)) < \pi_{\frac{1}{2}}(t; (x^A))$$

for all  $(x^A) \in N_k(t)$  and all population compositions  $t$ :

In words, we say that individualistic preferences are locally unstable in  $E$  if a “mutant” non-individualistic agent obtains a higher material payoff than the average material payoff of the individualistic agents across all matching contingencies and for any equilibrium. The intuition behind this definition is that mutations are so rare that they occur in only a single individual at a time and that the number of offspring that each parent leaves behind is an increasing function of the material payoffs she earns through her adulthood. Thus, material payoffs (or more directly, fitness) can be considered as the “currency” of natural selection. The type that is on average more successful in terms of this currency is regarded more highly by the forces of evolution. It is then natural to focus on individualistic preferences in the theory of preference evolution, for the owners of such preferences are unique in that they measure success in terms of the “right” evolutionary currency.

Yet it is not difficult to see that individualistic preferences are locally unstable in a vast plethora of environments. In fact, in essentially any environment with pairwise random matching, there exists

an alternative preference relation that would yield a higher expected material payoff to its owner relative to individualistic preferences. This elementary fact is formalized in the next example.

**Example 1.** Consider any environment  $\gamma \in (n; 2; X; \frac{1}{4})$  (with pairwise random matching) such that the 2-person symmetric game  $(X; u_j)_{j=1,2}$ ; where  $u_j(x) = \frac{1}{4}(x_j; x_{-j})$ ;  $x \in X^2$ ; has infinitely many pure strategy equilibria.<sup>4</sup> Denote the highest material payoff that a player may receive in equilibrium of this game by  $u^*$ : For simplicity, assume next that the best response map  $B : X \rightarrow X$  of player 1 is single-valued:  $B(b) = \arg \max_{a \in X} \frac{1}{4}(a; b)$  for each  $b \in X$ : Clearly, we have

$$\frac{1}{4}(a; B(a)) \leq u^* \text{ for some } a \in X: \tag{5}$$

In what follows, we postulate that  $\frac{1}{4}$  has the following additional property:

$$\frac{1}{4}(a; B(a)) > u^* \text{ for some } a \in X: \tag{6}$$

We therefore restrict our attention to games in which a Stackelberg leader (in our case player 2) benefits strictly from being a first-mover in terms of material payoffs.<sup>5</sup>

Now define the function  $\frac{1}{2} : X^2 \rightarrow \mathbb{R}$  by

$$\frac{1}{2}(a; b) = \begin{cases} \frac{1}{4}(a; B(a)) & \text{if } a \in \arg \max_{c \in X} \frac{1}{4}(c; B(c)) \\ 0 & \text{otherwise,} \end{cases}$$

and consider  $E = (\gamma; \frac{1}{2})$  as an environment. By (6), for any  $(x^A) \in N_2(\frac{1}{4}; \dots; \frac{1}{4}; \frac{1}{2})$ ; we have

$$\mathbb{E}_{\frac{1}{4}}(\frac{1}{4}; \dots; \frac{1}{4}; \frac{1}{2}; (x^A)) = \frac{\mu}{n_i - 1} u^* + \frac{\mu}{n_i - 1} \frac{1}{4}(B(a^*); a^*) \text{ and } \mathbb{E}_{\frac{1}{2}}(\frac{1}{4}; \dots; \frac{1}{4}; \frac{1}{2}; (x^A)) = \frac{1}{4}(a^*; B(a^*))$$

for some  $a^* \in X$  with  $\frac{1}{4}(a^*; B(a^*)) > u^*$ : It readily follows that individualistic preferences are locally unstable in  $E$ :<sup>6</sup>  $\nexists$

There is a sense in which Example 1 is not convincing, however. Indeed, the non-individualistic preference relation considered in this example is completely game specific, and is clearly “tailored” to do materially better than the individualistic preferences in the equilibrium of 2-person games. In the context of preference evolution, it may not be suitable to focus on all preferences that are defined on the action space  $X$ ; for such preferences may be meaningless in a variety of environments with action

<sup>4</sup>More generally, we may assume that the set of equilibrium payoffs of this game is compact.

<sup>5</sup>This is of course a very mild requirement, and there is a sense in which it holds generically. In particular, this requirement disallows environments such as those with constant  $\frac{1}{4}$ :

<sup>6</sup>Notice that, by virtue of (5),  $\mathbb{E}_{\frac{1}{2}}(\frac{1}{4}; \dots; \frac{1}{4}; \frac{1}{2}) \leq \mathbb{E}_{\frac{1}{4}}(\frac{1}{4}; \dots; \frac{1}{4}; \frac{1}{2})$  holds even if the pre-environment does not satisfy (6). Thus, we may say that independent preferences fail to be locally stable in essentially any environment, while they may be stable in the sense of Liapunov in some environments.

spaces that differ from  $X$ . Instead, it may be argued that people are, say, either altruist or not; that is, they play a large class of games (if not all) with the same preference relation. This idea leads one to consider preference relations that depend on the outcomes only through the induced distribution of payoffs. Such interdependent preferences do not depend on the underlying environment and include the most standard formulations of altruistic, reciprocal and spiteful preferences (cf. Sethi and Somanathan, 1999).

The question then is if the instability of individualistic preferences can be obtained in an economically interesting class of environments that involve meaningful interdependent preferences. The answer turns out to be yes. Among others, the recent results due to Koçkesen et al. (1998, 1999) show that there is a large class of playing-the-...eld type environments in which quite reasonable (interdependent) preference structures can drive individualistic preferences even into extinction. The following example aims to demonstrate this point.

**Example 2.** Consider any pre-environment  $\bar{e} = (n; n; R_+; \frac{1}{4})$  such that

$$u_i(a_1; \dots; a_n) = a_i \cdot h\left(\sum_{i=1}^n a_i\right) \cdot \frac{a_i}{\sum_{i=1}^n a_i} \quad ; \quad a_i \geq 0; \quad i = 1; \dots; n;$$

for some differentiable  $h : R_+ \rightarrow R_+$ . The pre-environment  $\bar{e}$  can then be thought of as a situation in which the entire population interacts through a common-pool resource game in which individual  $j$  receives a share of the total output that is proportional to her share of aggregate extraction effort. We thus think of  $x_j$  and  $\sum_{i=1}^n x_i$  as the extraction effort exerted by individual  $j$  and the aggregate extraction effort, respectively, and interpret  $h(\sum_{i=1}^n x_i)$  as the average return to effort. Accordingly, we normalize the opportunity cost per unit of extractive effort to 1; and assume that  $h' < 0$  and  $h(0) > 1 > \lim_{s \rightarrow 1} h(s)$ : It is clear that these properties of  $h$  ensure the existence of a unique  $s^* > 0$  such that  $h(s) = 1$  if, and only if,  $s = s^*$ :

we define the function  $\frac{1}{2} : R_+^n \rightarrow R$  as

$$\frac{1}{2}(a_1; \dots; a_n) = \frac{1}{4}(a_1; a_{i-1}) \cdot \sum_{j \neq i} (a_j; a_{i-j})^2;$$

and consider  $E = (\bar{e}; \frac{1}{2})$  as an environment. This particular preference structure is a special instance of "negatively interdependent preferences," which reflect Duesenberry's relative income hypothesis. Hence, unlike the non-individualistic preferences considered in Example 1, we now consider a preference relation which belongs to a family that is widely used in many distinct areas of economics. In particular, our present formulation is nothing but a minor modification of Shubik's beat-the-average objective function, the analogues of which occupy a center stage in the theory of strategic delegation (cf. Vickers (1985) and Fershtman and Judd (1987)).

We claim that individualistic preferences are exterminated by  $\frac{1}{2}$ : To prove this claim, we shall show that, for any population composition  $t$ , player 1 must obtain a strictly lower material payoff than player  $n$  in any equilibrium of the game  $G_{f_1, \dots, n_g}(t) \in (fR_+; u_j g_{j=1, \dots, n})$ ; where  $u_j$  is defined as in (1) with  $\frac{1}{4}$  and  $\frac{1}{2}$  being as specified above. (Of course, in our symmetric framework, this means that any individualistic player obtains a strictly lower payoff than any individual of type  $\frac{1}{2}$  at any equilibrium.) Consider any population composition  $t$  and an arbitrary  $x^\pi \in N(G_{f_1, \dots, n_g}(t))$ : Relabelling if necessary, we assume that

$$t_j = \begin{cases} \frac{1}{4} & \text{if } j = 1; \dots; r; \\ \frac{1}{2} & \text{if } j = r + 1; \dots; n, \end{cases}$$

for some  $r \in \{2, \dots, n-1\}$ ; that is, we let the first  $r$  individuals to be individualistic. We shall first establish that  $\min_{x_1^\pi, x_n^\pi} > 0$ : To verify this claim, assume for the sake of contradiction that  $\frac{\partial}{\partial x_i^\pi} x_i^\pi > 0$ : Then,  $\frac{1}{4}(x_j^\pi; x_{i_j}^\pi) < 0$  for all  $j$ ; and hence it is immediate that  $x_j^\pi = 0$  must hold since  $x^\pi$  is an equilibrium. (In the case of interdependent agents, this follows from the fact that  $\frac{\partial}{\partial x_n^\pi}(x_n^\pi; x_{i_n}^\pi) = \frac{\partial}{\partial x_n} < 0$  whenever  $\frac{\partial}{\partial x_i^\pi} x_i^\pi > 0$ .) This, however, contradicts  $\frac{\partial}{\partial x_i^\pi} x_i^\pi > 0$ . If, on the other hand,  $\frac{\partial}{\partial x_i^\pi} x_i^\pi = 0$ ; then we must again have  $x_j^\pi = 0$  for all  $j = 1; \dots; r$ ; for otherwise we would have  $\frac{1}{4}(x_j^\pi; x_{i_j}^\pi) > 0 = \frac{1}{4}(0; x_{i_j}^\pi)$  for any  $i \in (0; x_j^\pi)$ ; which contradicts that  $x^\pi$  is an equilibrium. Similarly,  $x_j^\pi = 0$  must hold for all  $j = r + 1; \dots; n$ ; since if  $x_j^\pi > 0$  for a player  $j$  of type  $\frac{1}{2}$ ; the first order condition for  $j$  and the hypothesis that  $\frac{\partial}{\partial x_i^\pi} x_i^\pi = 0$  would imply that  $\frac{\partial}{\partial x_j^\pi}(x_j^\pi; x_{i_j}^\pi) = \frac{\partial}{\partial x_j} x_j^\pi h^0(0) = 0$ ; which contradicts  $x_j^\pi > 0$ : We may therefore conclude that  $\frac{\partial}{\partial x_i^\pi} x_i^\pi < 0$ . But then, for any  $i \in (0; x_i^\pi)$ ; we have  $\frac{1}{4}(i; x_{i_j}^\pi) > \frac{1}{4}(0; x_{i_j}^\pi)$  for all  $j = 1; \dots; r$  and  $\frac{1}{2}(i; x_{i_j}^\pi) > \frac{1}{2}(0; x_{i_j}^\pi)$  for all  $j = r + 1; \dots; n$ . Thus, we must have  $x_j^\pi > 0$  for all  $j$ ; the equilibrium  $x^\pi$  must be an interior one. Consequently,  $\frac{1}{4}(x_j^\pi; x_{i_j}^\pi) > 0$  for all  $j$ :

Given that  $\min_{x_1^\pi, x_n^\pi} > 0$ ; the first order conditions

$$\frac{\partial}{\partial x_1} \left( \frac{1}{4}(x_1^\pi; x_{i_1}^\pi) \right) = h \left( \frac{\partial}{\partial x_i^\pi} x_i^\pi \right)_{i=1} + x_1^\pi h^0 \left( \frac{\partial}{\partial x_i^\pi} x_i^\pi \right) = 0 \quad (7)$$

and

$$\frac{\partial}{\partial x_n} \left( \frac{1}{2}(x_n^\pi; x_{i_n}^\pi) \right) = \frac{\partial}{\partial x_n} \left( \frac{1}{4}(x_n^\pi; x_{i_n}^\pi) \right)_{i \in \{2, \dots, n\}} + \frac{\partial}{\partial x_n} \left( \frac{1}{4}(x_n^\pi; x_{i_n}^\pi) \right) = 0 \quad (8)$$

must be satisfied. Since, by definition of  $\frac{1}{4}$ ;  $\frac{\partial}{\partial x_n} \left( \frac{1}{4}(x_n^\pi; x_{i_n}^\pi) \right) < 0$  for all  $i \in \{2, \dots, n\}$  and  $\frac{1}{4}(x_n^\pi; x_{i_n}^\pi) > 0$  for all  $i$ ; it follows from (8) that  $\frac{\partial}{\partial x_n} \left( \frac{1}{2}(x_n^\pi; x_{i_n}^\pi) \right) < 0$ : Combining this with (7), we find that  $x_n^\pi > x_1^\pi$ . This, in turn, implies that  $\frac{1}{4}(x_n^\pi; x_{i_n}^\pi) > \frac{1}{4}(x_1^\pi; x_{i_1}^\pi)$  as we sought.  $\text{Q.E.D.}$

The example above relies heavily on the hypothesis that all individuals in the society interact with each other (intragroup selection). A natural question thus concerns the modification of the

conclusions in the case of subgroup matching scenarios. Indeed, an analysis akin to that of Banerjee and Weibull (1995) would readily entail that individualistic preferences need not be exterminated in environments with pairwise matching even though they may have strategic disadvantage against (interdependent) preferences of the sort considered in Example 2. Nevertheless, our next example shows that subgroup matching falls short as well of warranting the local stability of individualistic preferences.

**Example 3.** In this example we shall use the same pre-environment used in Example 2 except that here we shall focus on a  $k$ -wise random matching scenario. Thus, the pre-environment we shall study is  $\Gamma = (n; k; R_+; \mathcal{U})$  where

$$\mathcal{U}(a_1; \dots; a_k) = a_1 \cdot h \left( \prod_{i=1}^{k-1} \frac{a_i}{a_i + 1} \right); \quad a_i \geq 0; \quad i = 1; \dots; k;$$

for some twice continuously differentiable  $h : R_+ \rightarrow R_+$  with  $h' < 0$  and  $h(0) > 1 > \lim_{s \rightarrow 1^-} h(s)$ . Moreover, here we assume that  $h'' < 0$ ; which in turn guarantees the strictly submodularity of the subgroup game:  $\partial^2 \mathcal{U}(a) / \partial a_i \partial a_j < 0$  for all  $a \in R_+^k$  and  $i \neq j$ :

Now, for any  $\epsilon > 0$ ; define the function  $\mathcal{U}_\epsilon : R_+^k \rightarrow R$  as

$$\mathcal{U}_\epsilon(a_1; \dots; a_k) = \mathcal{U}(a_1; a_{i-1}) \cdot \left( \prod_{j \neq i} \mathcal{U}(a_j; a_{j-1}) \right)^\epsilon;$$

and consider  $E_\epsilon = (\Gamma; \mathcal{U}_\epsilon)$  as an environment. We claim that there exists an  $\epsilon$  such that individualistic preferences are locally unstable in  $E_\epsilon$ :

To see this, let us first note that, there exists an  $\epsilon > 0$  such that we have  $j \in N(G_A(\mathcal{U}; \dots; \mathcal{U}; \mathcal{U}_\epsilon)) \iff j = 1$  for all  $A \in A_{\mathcal{U}_\epsilon}$  and all  $\epsilon \in [0; \epsilon]$ .<sup>7</sup> Consequently, by restricting our attention to small  $\epsilon$ 's, we are able to avoid the equilibrium selection problem. In what follows, we shall show that, when  $\epsilon$  is small, an individualistic player (individual 1, in particular) makes strictly less material payoffs in equilibrium than the non-individualistic player  $n$ ; whether or not he is matched with player  $n$  in the same group. Put formally, we shall prove that there exists a small enough  $\epsilon > 0$  such that

$$\mathcal{U}(x_n^\epsilon; x_{i-1}^\epsilon) > \mathcal{U}(x_1^\epsilon; x_{i-1}^\epsilon) \tag{9}$$

and

$$\mathcal{U}(x_n^\epsilon; x_{i-1}^\epsilon) > \mathcal{U}(a^\epsilon; \dots; a^\epsilon) \tag{10}$$

where

$$f(x^\epsilon)g = N(G_{f_1; \dots; f_{k-1}; g}(\mathcal{U}; \dots; \mathcal{U}; \mathcal{U}_\epsilon)) \quad \text{and} \quad f(a^\epsilon)g = N(G_{f_1; \dots; f_k}(\mathcal{U}; \dots; \mathcal{U})).$$
<sup>8</sup>

<sup>7</sup>The proof of this claim is routine and is thus omitted.

<sup>8</sup>Notice that, since  $G_{f_1; \dots; f_k}(\mathcal{U}; \dots; \mathcal{U})$  is a symmetric game and has a unique equilibrium, this equilibrium must be symmetric.

Clearly, this will establish that individualistic preferences are locally unstable in  $E_\theta$  for small enough  $\theta > 0$ :

We begin by observing that (9) holds for all  $\theta > 0$ : The proof of this is identical to that given in Example 2 for the case  $n = k$  and  $\theta = 1$ ; and does not require any submodularity arguments. Here we shall thus concentrate only on establishing (10). To this end, let  $B : \mathbb{R}_+^{k-1} \rightarrow \mathbb{R}_+$  stand for the best response mapping of an individualistic player. (The single-valuedness of this map is ensured by the assumptions  $h^l < 0$  and  $h^l = 0$ .) We next define the function  $\gamma : [0, \theta] \rightarrow [0, \theta]$  by

$$\gamma(q) = B(\gamma(q); \dots; \gamma(q); q); \quad 0 \leq q \leq \theta.$$

By using the standard arguments based on the Brouwer's fixed point theorem and the implicit function theorem, we can show that  $\gamma$  is well-defined, and  $B$  and  $\gamma$  are  $C^1$  functions on  $(0, \theta)$ : Two further observations about the function  $\gamma$  are in order. First, note that the definition of  $\gamma$  readily entails that

$$(\gamma(q_\theta); \dots; \gamma(q_\theta); q_\theta) = x_n^{(\theta)} \quad \text{whenever} \quad q_\theta \in \arg \max_{q \in [0, \theta]} \frac{1}{2} \gamma(q; \gamma(q_\theta); \dots; \gamma(q_\theta)) \quad (11)$$

for any  $\theta \in [0, \theta]$ : Second, we may use again the implicit function theorem to establish that  $\gamma$  is a strictly decreasing function. In particular,

$$\gamma^l(a^n) = \frac{\frac{\partial B(a^n; \dots; a^n)_{k-1}}{\partial x_{k-1}}}{1 - \sum_{i=1}^{k-2} \frac{\partial B(a^n; \dots; a^n)_{x_i}}{\partial x_i}} < 0; \quad (12)$$

where the inequality is an immediate consequence of the strict submodularity of  $\gamma$ : We are now ready to prove the following

Claim 1. There exists a  $\bar{q} > a^n$  such that  $\gamma(q; \gamma(q); \dots; \gamma(q)) > \gamma(a^n; \dots; a^n)$  for all  $q \in (a^n; \bar{q}]$ :

Proof of Claim 1. Define  $\alpha(q) = \gamma(q; \gamma(q); \dots; \gamma(q))$  for all  $q \in [0, \theta]$ ; and observe that

$$\alpha^l(a^n) = \frac{\frac{\partial \gamma(a^n; \dots; a^n)}{\partial x_1}}{\frac{\partial \gamma(a^n; \dots; a^n)}{\partial x_1}} + \sum_{j=2}^k \frac{\frac{\partial \gamma(a^n; \dots; a^n)}{\partial x_j}}{\frac{\partial \gamma(a^n; \dots; a^n)}{\partial x_j}} \gamma^l(a^n) = (k-1) a^n h^l(k a^n) \gamma^l(a^n) > 0$$

in view of the fact that  $\frac{\partial \gamma(a^n; \dots; a^n)}{\partial x_1} = 0$  and  $\gamma^l(a^n) < 0$ : The claim then readily follows from the continuity of  $\alpha : k$

Claim 2.  $x_n^{(\theta)} > a^n$  for all  $\theta \in [0, \theta]$ :

Proof of Claim 2. Fix any  $\theta \in [0, \theta]$ : The claim follows from the fact that  $x_1^{(\theta)} = \dots = x_{k-1}^{(\theta)}$  and (9).  $k$

Now, take any strictly decreasing sequence  $\theta^m \in (0, \theta)$  such that  $\lim_{m \rightarrow \infty} \theta^m = 0$ : Since  $x_n^{(\theta^m)} \in [0, \theta]$  (as is shown in Example 2), there must exist a convergent subsequence of  $x_n^{(\theta^m)}$ ; which we

again denote by  $x_n^{\alpha}(\epsilon^m)$  for simplicity. Let  $x_n^{\alpha}(\epsilon^m) \neq x_n^{\alpha}$  as  $m \rightarrow 1$ : But by (11) and continuity of  $\pi$ ; we have

$$N(G_{f_1, \dots, k_i}(\frac{1}{4}, \dots, \frac{1}{4}; \frac{1}{2\epsilon^m})) \ni (x_n^{\alpha}(\epsilon^m); \dots; x_n^{\alpha}(\epsilon^m); x_n^{\alpha}(\epsilon^m)) \neq (x_n^{\alpha}; \dots; x_n^{\alpha}; x_n^{\alpha})$$

as  $m \rightarrow 1$ : Thus, since the Nash equilibrium correspondence has a closed graph and  $\lim_{\epsilon^m \rightarrow 0} \epsilon^m = 0$ ; we find  $(x_n^{\alpha}; \dots; x_n^{\alpha}; x_n^{\alpha}) \in N(G_{f_1, \dots, k_i}(\frac{1}{4}, \dots, \frac{1}{4}; \frac{1}{4}))$ ; which in turn implies that  $x_n^{\alpha} = \pi(x_n^{\alpha}) = a^{\alpha}$ : Therefore, by Claim 2, there must exist an integer  $M > 0$  such that  $m \geq M$  implies that  $\pi(x_n^{\alpha}(\epsilon^m)) > a^{\alpha}$ ; where  $\epsilon$  is chosen as in Claim 1. Consequently, by (11) and Claim 1, we obtain

$$\pi(x_n^{\alpha}(\epsilon^m); x_n^{\alpha}(\epsilon^m)) = \pi(x_n^{\alpha}(\epsilon^m); \pi(x_n^{\alpha}(\epsilon^m)); \dots; \pi(x_n^{\alpha}(\epsilon^m))) > \pi(a^{\alpha}; \dots; a^{\alpha})$$

for all  $m \geq M$ : This proves (10), and hence as noted earlier, completes the proof of the fact that individualistic preferences are locally unstable in  $E_{\epsilon}$  for small  $\epsilon$ :  $\square$

The upshot of the above examples is that individualistic preferences do not possess evolutionary stability properties in arbitrary environments under complete information. In particular, due to their strategic advantage, negatively interdependent preferences (that preconditions a player to derive utility from being materially better off than others) appear to invade a monomorphic population composed only of individualistic agents in many interesting environments.<sup>9</sup> In fact, depending on the environment (e.g. one with pairwise matching), even altruist preferences would constitute successful mutations as shown by Fershtman and Weiss (1998) and Bester and Güth (1998). Similar results are reported by Sethi and Somanathan (1999) in the case of reciprocal preferences (of Levine, 1998) that allow for both altruism and spite.

In view of the preceding discussion, one might be tempted to conclude that, while commonly used in economic theory, the case for individualistic preferences is hardly compelling from an evolutionary viewpoint. Yet, in the rest of this paper, we shall argue that it has been the implicit assumption of complete information that is mostly responsible for the apparent evolutionary weakness of individualistic preferences. As we presently show, relaxing this assumption may lead us to a very different setting in which such preferences do emerge as truly focal.

---

<sup>9</sup>This observation parallels the well-known spiteful effect of Hamilton (1970): a finite monomorphic population adopting a Nash equilibrium strategy may be vulnerable to invasion by a 'spiteful' mutant adopting a strategy that is a nonoptimal response to the action profile of the incumbents, but which reduces the average payoff of the incumbents so greatly that it falls below the payoff of the mutant (cf. Rhode and Stegeman 1996, Palomino 1996, and Vega-Redondo, 1997). However, notice that here we are considering the harder problem of preference evolution as opposed to behavior evolution, and hence attribute a trait like "spite" to preferences as opposed to behavior. Consequently, we assume that all players behave rationally given their preferences, and then investigate the properties of the equilibria conditional on the underlying preference distribution.

### 3 Stability of Individualism under Incomplete Information

The previous analysis demonstrates that individualistic preferences should not be viewed as preeminent from an evolutionary perspective, provided that preferences evolve according to payoff monotonic dynamics and the economic environment (within which evolution takes place) displays complete information. Depending on the structure of the game that is played by subgroups, certain plausible preference relations may have a definitive strategic advantage over individualistic preferences. One way of thinking about this observation is to interpret a non-individualistic player as an agent who commits herself in an observable manner to play the game the way some such (genuine) non-individualistic person would play. Observability of this commitment usually guarantees an advantageous outcome to the committing party. For instance, a player who commits herself to play “hawk” in the usual hawk-dove game against an individualistic player secures herself the best outcome, since an individualistic player would rationally play “dove” given that she observes the commitment of his opponent. But what happens if we dispense with the “observability” assumption in this scenario? Or, more precisely, what happens if we assume that each player has only incomplete information about the type of her opponents in the subgroup? This gives rise to a strategic context where players can no longer tailor their behavior to the type profile of the subgroup they happen to lie in. Hence, each player must choose a “flat” strategy that, in expected terms, provides her with maximal payoffs given her own preferences (individualistic or not) and the strategies chosen by the opponents. In this section, our aim is to show that this leads individualistic preferences to be essentially the only type of preferences that enjoy natural selection privileges in a very large class of environments, provided that the number of subgroups in the population is large enough.

Before introducing formally the incomplete information model that we shall investigate here, we first turn to one of the examples discussed in the previous section, and see how the main conclusion obtained there fails if we relax the complete information assumption. The following example will also provide some intuition for the main results of this paper.

**Example 4.** Consider the setting described in Example 1, and assume again that the population composition is  $t = (\frac{1}{4}; \dots; \frac{1}{4}; \frac{1}{2})$ : Suppose next that a player cannot exactly identify the type of her opponent but rather assigns a probability to her opponent being individualistic. One natural way of choosing the prior beliefs of such a player is by linking them to the population shares of each type. Thus, let us assume that each player assigns probability  $(n_i - 2)/(n_i - 1)$  to the event that her opponent (in the two-person subgroup) has individualistic preferences. On the other hand, the beliefs of the only  $\frac{1}{2}$ -type player are of course degenerate on her opponent being a  $\frac{1}{4}$ -type.

Given these beliefs, we may model the game played within each subgroup as a Bayesian game in

the standard manner. Parametrizing matters by  $n$  (the population size), let  $\frac{3}{4}_n : \mathbb{R}^2 \rightarrow \mathbb{R}$  be any map such that

$$\frac{3}{4}_n(\frac{1}{4}) \geq \arg \max_{a \in X} \left[ \frac{n-2}{n-1} \frac{1}{4}(a; \frac{3}{4}_n(\frac{1}{4})) + \frac{1}{n-1} \frac{1}{4}(a; \frac{3}{4}_n(\frac{1}{2})) \right]$$

and

$$\frac{3}{4}_n(\frac{1}{2}) \geq \arg \max_{a \in X} \frac{1}{2}(a; \frac{3}{4}_n(\frac{1}{4})).$$

Clearly,  $\frac{3}{4}_n$  can be considered as a Bayes-Nash equilibrium for the subgroup Bayesian game described above.

In what follows, we assume that the above equilibrium  $\frac{3}{4}_n$  obtains in all of the subgroups formed in this environment. Thus, exactly  $(n-2)$  many individualistic players end up with a payoff of  $\frac{1}{4}(\frac{3}{4}_n(\frac{1}{4}); \frac{3}{4}_n(\frac{1}{4}))$  in this equilibrium. On the other hand, the material payoffs of the only  $\frac{1}{2}$ -type individual is  $\frac{1}{4}(\frac{3}{4}_n(\frac{1}{2}); \frac{3}{4}_n(\frac{1}{4}))$ ; while the payoffs of her single individualistic opponent is  $\frac{1}{4}(\frac{3}{4}_n(\frac{1}{4}); \frac{3}{4}_n(\frac{1}{2}))$ : Consequently, there would be good reason to conclude that individualistic preferences are locally stable (especially if the equilibrium  $\frac{3}{4}_n$  is unique) if the following inequality applied:

$$\frac{1}{4}(\frac{3}{4}_n(\frac{1}{2}); \frac{3}{4}_n(\frac{1}{4})) < \frac{n-2}{n-1} \frac{1}{4}(\frac{3}{4}_n(\frac{1}{4}); \frac{3}{4}_n(\frac{1}{4})) + \frac{1}{n-1} \frac{1}{4}(\frac{3}{4}_n(\frac{1}{4}); \frac{3}{4}_n(\frac{1}{2})). \quad (13)$$

To see that this inequality may well be expected to hold, assume that  $X$  is a compact and convex subset of  $\mathbb{R}$ ; and let  $\frac{1}{4}$  be continuous and strictly quasi-concave in its first argument. For the sake of contradiction, assume that (13) fails for infinitely many  $n$ : Then, by passing to a subsequence of  $(\frac{3}{4}_n(\frac{1}{2}); \frac{3}{4}_n(\frac{1}{4}))$  that converges to a strategy profile  $(\frac{3}{4}_1(\frac{1}{2}); \frac{3}{4}_1(\frac{1}{4}))$  for which (13) fails, we find that  $\frac{1}{4}(\frac{3}{4}_1(\frac{1}{2}); \frac{3}{4}_1(\frac{1}{4})) \geq \frac{1}{4}(\frac{3}{4}_1(\frac{1}{4}); \frac{3}{4}_1(\frac{1}{4}))$ : But by the closed graph property of the equilibrium correspondence, we must have  $\frac{3}{4}_1(\frac{1}{4}) \in B(\frac{3}{4}_1(\frac{1}{4}))$ ; where  $B$  is the best response correspondence. By strict quasi-concavity, however, this can hold only if  $\frac{3}{4}_1(\frac{1}{2}) = \frac{3}{4}_1(\frac{1}{4})$ : Hence, in any environment where a  $\frac{1}{2}$ -type player and a  $\frac{1}{4}$ -type player would play the game differently when they assign probability one to facing a  $\frac{1}{4}$ -type, (13) must hold. In any such environment, therefore, we may conclude that individualistic preferences are locally stable.

Before concluding, let us further illustrate the point by providing a concrete example with a unique equilibrium that verifies (13). Consider the environment  $((n; 2; [0; 1]; \frac{1}{4}; \frac{1}{2})$  where  $\frac{1}{4}(a; b) = a(2 - a - b)$  for all  $a; b \in [0; 1]$  and  $\frac{1}{2}$  is as defined in Example 1.<sup>10</sup> It is readily verified that (6) holds and  $\frac{3}{4}_n(\frac{1}{2}) = 1$  for all  $n$ : In turn, through simple calculus, we find  $\frac{3}{4}_n(\frac{1}{4}) = (2n - 3)/(3n - 4)$  which converges to  $2/3$

<sup>10</sup>The subgroup game is thus a symmetric Cournot duopoly game with a linear demand and cost structure. Given this interpretation, the  $\frac{1}{2}$ -type player can be thought of as a manager who will choose the Stackelberg output (due to his contract, say) irrespective of her opponent's play.

as  $n \rightarrow \infty$ : By the previous observation, therefore, (13) holds in this setting for all large enough  $n$ :<sup>11</sup>

We now wish to improve upon the observation noted in the above example by focusing on the global stability properties of individualistic preferences in a large class of environments. To this end, let us consider an arbitrary environment  $E = (i; \frac{1}{2})$ . As noted earlier, given any prevailing population composition  $t = (t_1; \dots; t_n) \in \mathcal{F}(\frac{1}{4}; \frac{1}{2})^n$ ; the key implicit assumption underlying the complete-information context studied in Section 2 is that  $t$  is common knowledge. As in Example 4, we shall instead explore here the alternative assumption that, beyond her own type, each player is informed only of the frequency

$$\bar{t} = \frac{1}{n} \sum_{i \in N} t_i = \frac{1}{n} \sum_{j \in N} t_j$$

of individualistic players (and, therefore, of non-individualistic players as well) present in the whole population. The strategic situation encountered by any individual  $j$  placed in a subgroup  $i_1; \dots; i_k$  can then be modeled as a  $k$ -person Bayesian game. Formally, we may define this game as

$$G_{i_1, \dots, i_k}^B = (\mathcal{F}(\frac{1}{4}; \frac{1}{2}); \bar{t}; u_j)_{j = i_1, \dots, i_k}$$

where  $u_j$  is as defined in (1). Here the beliefs of the players are such that a  $\frac{1}{4}$ -type player assigns probability  $\frac{n^1(t) - 1}{n - 1}$  to the event that an arbitrary opponent is  $\frac{1}{4}$ -type, while a  $\frac{1}{2}$ -type assigns probability  $\frac{n^1(t)}{n - 1}$  to the same event. (Notice again that we are letting individuals to condition their beliefs on their own types.) In what follows, we shall refer to an environment with this information structure as an environment under incomplete information.

As in Dekel et al. (1998) and mainly for the sake of notational simplicity, we shall restrict our attention to a symmetric context where the strategy of player  $j$  does not depend on the types of the individuals gathered with her in the subgroup. Thus, her strategy is given by some function

$$\sigma^j : \mathcal{F}(\frac{1}{4}; \frac{1}{2}) \rightarrow [0; 1] \times X$$

that specifies the action chosen  $\sigma^j(t_j; \bar{t}) \in X$  depending both on player  $j$ 's private information – her type  $t_j \in \mathcal{F}(\frac{1}{4}; \frac{1}{2})$  – and the commonly available public information, i.e. the prevailing frequency of individualistic players. Since we choose to focus on symmetric strategy profiles,  $\sigma^j$  is in fact independent of  $j$ . Thus, we let  $\sigma^j = \sigma$  for all  $j = 1; \dots; n$ ; and identify the notions of “strategy” and “strategy profile” in what follows.

<sup>11</sup>In fact, it turns out that (13) holds in this particular example for all  $n = 4; 6; \dots$  which is verified by routine calculation. It is important to note, however, that  $\frac{1}{4}(\frac{3}{4}_n(\frac{1}{2}); \frac{3}{4}_n(\frac{1}{4})) > \frac{1}{4}(\frac{3}{4}_n(\frac{1}{4}); \frac{3}{4}_n(\frac{1}{2}))$  holds for all  $n$ ; and hence intragroup selection favors  $\frac{1}{4}$ -type players (as would be expected from Example 1). Yet this effect is overcome in this example by the effect of intergroup selection which favors the  $\frac{1}{4}$ -types, thereby illustrating the combined power of random matching and incomplete information.

Given a population composition  $t$ ; the expected payoffs earned by an individualistic player when she plays  $x \in X$  and the rest of the players play the strategy  $\frac{3}{4}$  is:

$$v_{\frac{3}{4}}(x; \frac{3}{4}; t) = \sum_{s=0}^k \mu_{k-i} \prod_{r=1}^s \frac{\prod_{q=0}^{k_i s_i - 2} (n_i - 1 - t_i - r)}{(n_i - 1)(n_i - 2) \dots (n_i - k + 1)} \mu_{\frac{3}{4}}(x; h_{\frac{3}{4}}(\frac{3}{4}; t))_{i_s}; h_{\frac{3}{4}}(\frac{3}{4}; t))_{k_i s_i - 1} \quad (2)$$

where  $h_{\mu} i_r$  stands for the  $r$ -fold replica of an object  $\mu$  for any positive integer  $r$  and, by convention,  $i_{r=1}^0(\mu) = 1$ : Similarly, the expected payoff of a  $\frac{1}{2}$ -type player when she plays  $x \in X$  and the rest of the players play  $\frac{3}{4}$  is:

$$v_{\frac{1}{2}}(x; \frac{3}{4}; t) = \sum_{s=0}^k \mu_{k-i} \prod_{r=0}^{k_i s_i - 2} \frac{\prod_{q=1}^{n_i - 1 - t_i - q} (n_i - 1 - t_i - r)}{(n_i - 1)(n_i - 2) \dots (n_i - k + 1)} \mu_{\frac{1}{2}}(x; h_{\frac{3}{4}}(\frac{3}{4}; t))_{k_i s_i - 1}; h_{\frac{3}{4}}(\frac{3}{4}; t))_{i_s} \quad (3)$$

We are now ready to introduce the (symmetric) Bayesian equilibrium concept that will be used in the present incomplete-information setup.

**Definition 2.** Let  $E = (j; \frac{1}{2})$  be an environment under incomplete information. A strategy  $\frac{3}{4}$  is called a **Symmetric Bayes-Nash Equilibrium (SBNE)** if, for all population compositions  $t$ ,

$$v_{\frac{3}{4}}(\frac{3}{4}(\frac{3}{4}; t); \frac{3}{4}; t) \geq v_{\frac{3}{4}}(x; \frac{3}{4}; t) \quad \text{for all } x \in X$$

and

$$v_{\frac{1}{2}}(\frac{3}{4}(\frac{1}{2}; t); \frac{3}{4}; t) \geq v_{\frac{1}{2}}(x; \frac{3}{4}; t) \quad \text{for all } x \in X:$$

If a strategy  $\frac{3}{4}$  is an SBNE for  $E$ , then the symmetric profile  $h_{\frac{3}{4}}(t; \frac{3}{4}; t)_{i_k}$  is simply a standard Bayesian equilibrium for the  $k$ -person game played by each subgroup, when players are randomly chosen from a population with an overall composition given by  $t$ . Therefore, SBNE is an equilibrium concept that embodies (symmetric) Bayesian equilibrium behavior under all possible population compositions. Since we shall provide a global stability analysis below, such a formulation is in the nature of things.

We next define the notion of global stability that will be invoked in the subsequent analysis.

**Definition 3.** Let  $E = (j; \frac{1}{2})$  be an environment under incomplete information. Individualistic preferences  $\frac{1}{4}$  are said to **exterminate** alternative preferences  $\frac{1}{2}$  if, for any SBNE strategy  $\frac{3}{4}$  and any population composition  $t$ :

$$\frac{1}{jA_{\frac{1}{4}}} \sum_{A \in A_{\frac{1}{4}}} \mu_{\frac{1}{4}}(\frac{1}{4}(t; \frac{1}{2}; t); (\frac{3}{4}(t; \frac{1}{2}; t)))_{j \in 2A_{\frac{1}{4}}} > \frac{1}{jA_{\frac{1}{2}}} \sum_{A \in A_{\frac{1}{2}}} \mu_{\frac{1}{2}}(\frac{1}{4}(t; \frac{1}{2}; t); (\frac{3}{4}(t; \frac{1}{2}; t)))_{j \in 2A_{\frac{1}{2}}}$$

where  $A_{\frac{1}{4}}$  and  $A_{\frac{1}{2}}$  are defined as in (2). If this inequality holds for  $t = (h_{\frac{1}{4}} i_{n_i - 1}; \frac{1}{2})$ ; we say that individualistic preferences are **locally stable** in  $E$ .

The notion of global evolutionary dominance contemplated in this definition reflects a situation in which individuals displaying individualistic preferences enjoy a larger material payoff than those with non-individualistic preferences under any population configuration. In this sense, therefore, the sort of payoff monotonicity displayed here is entirely analogous to that already explained in motivating Definition 1.

In the present setting, our goal is to establish that individualistic preferences exterminate essentially any non-individualistic preferences if there is a sufficiently large number of subgroups in the society. We shall need a number of structural assumptions in formalizing this claim, which is obviously in sharp contrast with the examples considered in Section 2. To state these assumptions, however, we need to endow the action spaces of the environments that we will study with some topological and linear structure. In what follows, therefore, we shall confine our attention to environments with an action space  $X$  that is a compact and convex subset of an arbitrary metrizable topological vector space.<sup>12</sup> Furthermore, we will make use of the following assumptions that are posited on the primitives of an environment,  $(n; k; X; \mu; \nu)$ :

**Continuity (C).** The functions  $\mu; \nu : X^k \rightarrow \mathbb{R}$  are continuous.

**Strict Concavity (SC).** Given any  $x_i \in X^{k_i - 1}$ ; the function  $\mu(t; x_i) : X \rightarrow \mathbb{R}$  is strictly concave.

**No Trivial Equilibrium (NTE).** An SBNE  $\mu$  exists: Moreover, given  $\mu; \nu$ ; and  $k$ ; there exists some associated environment  $E$  such that  $\mu(\mu; \nu; t) \neq \mu(\nu; \nu; t)$  for every SBNE  $\mu$  and some population composition  $t$ .

With the only exception of (NTE), these properties are standard and there is no need to elaborate on them here. (However, we shall consider below a substantial weakening of (SC) that still has interesting implications.) On the other hand, (NTE) embodies a combination of two distinct postulates. The first is technical and tackles the problem of existence of an SBNE in a trivial manner: the underlying  $\mu; \nu$ ; and  $k$  are assumed to be such that an SBNE exists for every associated environment (i.e. every possible  $n$ ): Since our interest is presently on the properties of (existing) equilibria, we view this assumption only little more than a simplifying one. Moreover, it is not difficult to provide conditions on the fundamentals of the problem that would guarantee the existence of an SBNE (see Remark 1).

The second part of (NTE) is, on the other hand, critical for our main results. It allows us to

---

<sup>12</sup>While this is a minor point, we note that the metrizable condition may be replaced here with the weaker property of first countability. Since a compact and first countable topological space is sequentially compact, the subsequent analysis goes through without alteration with this weakening.

study environments in which individualistic and non-individualistic preferences are behaviorally distinguishable for at least one population composition and some environment. It must be obvious that such a postulate is really unexceptionable in that if two preferences are, at equilibrium, fully equivalent in behavioral terms, neither we (as observers) nor any evolutionary system may discriminate between them.<sup>13</sup>

We are now ready to prove the ...rst main result of this paper.

**Theorem 1.** Assume (C), (SC) and (NTE). Then, there exists some  $M > 0$  such that, if  $n \geq M$ ; individualistic preferences  $\mu_1$  exterminate non-individualistic preferences  $\mu_2$  in any incomplete-information environment  $E = ((n; k; X; \mu_1); \mu_2)$ :

**Proof.** Consider any  $\mu_1; \mu_2$  and  $k$  that satisfy (C), (SC) and (NTE). Let  $F_{\mu_1}(x; y; z; \pi)$  and  $F_{\mu_2}(x; y; z; \pi)$  stand, respectively, for the expected payoff of an individualistic and non-individualistic type when her strategy is  $x$ ; all (other) individuals of an individualistic type choose  $y$ ; all (other) non-individualistic individuals chooses  $z$ ; and there is prior probability  $\pi$  that a randomly chosen individual be of an individualistic type. By denoting the  $r$ -fold replica of an object  $\mu$  by  $\mu^r$  for any integer  $r$ ; we define the functions  $F_{\mu_1}$  and  $F_{\mu_2}$  on  $X^3 \in [0; 1]$  as

$$F_{\mu_1}(x; y; z; \pi) = \sum_{s=0}^{k-1} \binom{k-1}{s} \pi^s (1-\pi)^{k-1-s} \mu_1^s(x; \mu_1^{s+1}(y); \mu_2^{k-1-s}(z)) \quad (14)$$

and

$$F_{\mu_2}(x; y; z; \pi) = \sum_{s=0}^{k-1} \binom{k-1}{s} \pi^s (1-\pi)^{k-1-s} \mu_2^s(x; \mu_1^{s+1}(y); \mu_2^{k-1-s}(z)) \quad (15)$$

respectively. Here, we adopt the convention that  $F_{\mu_1}(x; y; z; 1) = \mu_1(x; \mu_1^{k-1}(y); \mu_2^0(z))$  and  $F_{\mu_1}(x; y; z; 0) = \mu_1(x; \mu_2^{k-1}(z))$ ; with a similar convention applying to  $F_{\mu_2}$  as well.

Some useful properties of these functions are established in the next three claims.

**Claim 1.** The equilibrium correspondence  $\alpha : [0; 1] \rightarrow X^2$  given by

$$\alpha(\pi) = (y; z) : y \in \arg \max_{x \in X} F_{\mu_1}(x; y; z; \pi); \quad z \in \arg \max_{x \in X} F_{\mu_2}(x; y; z; \pi) \quad (16)$$

has a closed graph.

**Proof of Claim 1.** Even though the proof is routine, we include it for completeness. Consider a sequence  $(\pi_m; y_m; z_m)$  in  $[0; 1] \in X^2$  such that  $\lim_{m \rightarrow \infty} (\pi_m; y_m; z_m) = (\pi; y; z)$  and  $(y_m; z_m) \in \alpha(\pi_m)$  for each  $m$ . Fix a  $x \in X$  and observe that the latter condition implies that  $F_{\mu_1}(y_m; y_m; z_m; \pi_m) \geq F_{\mu_1}(x; y_m; z_m; \pi_m)$  for all  $m$ . So, by continuity of  $F_{\mu_1}$ ; we have  $F_{\mu_1}(y; y; z; \pi) \geq F_{\mu_1}(x; y; z; \pi)$ . Since  $x$  is arbitrary here, and since the analogous reasoning applies to  $F_{\mu_2}$  as well, we obtain  $(y; z) \in \alpha(\pi)$ .  $\square$

<sup>13</sup>Analogous caveat is found in Ely and Y-lankaya (1997).

Claim 2. For any  $(\theta; y; z) \in X^2$  such that

$$y \in \arg \max_{x \in X} F_{\frac{1}{2}}(x; y; z; \theta) \quad \text{and} \quad z \in \arg \max_{x \in X} F_{\frac{1}{2}}(x; y; z; \theta); \quad (16)$$

we have  $y = z$ :

Proof of Claim 2. Assume that the claim is false, and take a  $(\theta; y; z)$  that satisfies (16) but  $y \neq z$ . Then, the common action  $y = z$  has to be an optimal response for both individualistic and non-individualistic types alike, given any population profile, as long as everyone else also adopts that same action. Formally, we find that

$$\begin{aligned} y &\in \arg \max_{x \in X} F_{\frac{1}{2}}(x; y; z; \theta) \\ &= \arg \max_{x \in X} \frac{1}{2} \sum_{i=1}^n x_i \sum_{s=0}^{k_i-1} \mu_{k_i-1}^s (1 - \theta)^{k_i - s} \\ &= \arg \max_{x \in X} \frac{1}{2} \sum_{i=1}^n x_i \sum_{s=0}^{k_i-1} \mu_{k_i-1}^s \frac{\prod_{r=1}^s (n_i - r) \theta^{k_i - s} (1 - \theta)^{s-1}}{(n_i - 1)(n_i - 2) \dots (n_i - k + 1)} \end{aligned}$$

for all  $\theta \in [0, 1]$  since

$$\sum_{s=0}^{k_i-1} \mu_{k_i-1}^s (1 - \theta)^{k_i - s} = 1 = \sum_{s=0}^{k_i-1} \mu_{k_i-1}^s \frac{\prod_{r=1}^s (n_i - r) \theta^{k_i - s} (1 - \theta)^{s-1}}{(n_i - 1)(n_i - 2) \dots (n_i - k + 1)}$$

Similarly, we also have

$$y \in \arg \max_{x \in X} \frac{1}{2} \sum_{i=1}^n x_i \sum_{s=0}^{k_i-1} \mu_{k_i-1}^s \frac{\prod_{r=1}^s (n_i - r) \theta^{k_i - s} (1 - \theta)^{s-1}}{(n_i - 1)(n_i - 2) \dots (n_i - k + 1)}$$

Thus, if we denote by  $\mathcal{A}(\theta)$  a strategy that satisfies

$$\mathcal{A}(\theta)_i = \mathcal{A}(\theta)_i \quad (\theta \in [0, 1]; i = 1, \dots, n);$$

we find that  $y \in \arg \max_{x \in X} \frac{1}{2} \sum_{i=1}^n x_i \mathcal{A}(\theta)_i$  and  $y \in \arg \max_{x \in X} \frac{1}{2} \sum_{i=1}^n x_i \mathcal{A}(\theta)_i$  for any population composition  $\theta$ . Therefore, we conclude that  $\mathcal{A}$  is a "trivial" SBNE for  $E$ ; which contradicts (NTE).  $\square$

Claim 3. There exists an  $\epsilon > 0$  such that

$$|F_{\frac{1}{2}}(y; y; z; \theta) - F_{\frac{1}{2}}(z; y; z; \theta)| \leq \epsilon$$

for any  $(\theta; y; z) \in X^2$  that satisfies (16).

Proof of Claim 3. Define  $E$  to be the set of all  $(\theta; y; z) \in X^2$  such that (16) is satisfied. The set  $E$  is closed by Claim 1. On the other hand, note that by Claim 2, (SC), and the definition of  $F_{\frac{1}{2}}$ ; we must have  $F_{\frac{1}{2}}(y; y; z; \theta) > F_{\frac{1}{2}}(z; y; z; \theta)$  for all  $(\theta; y; z) \in E$ : Thus, if the claim were false, we could find a sequence  $(\theta_m; y_m; z_m)$  in  $E$  such that

$$\lim_{m \rightarrow \infty} (F_{\frac{1}{2}}(y_m; y_m; z_m; \theta_m) - F_{\frac{1}{2}}(y_m; y_m; z_m; \theta_m)) = 0; \quad (17)$$



requirements can be dispensed with. On the one hand, the obvious need for (NTE) was already discussed above. The requirement that the population size  $n$  be large enough relative to  $k$  is also crucial. Only if  $n=k$  is large will there be a significant "matching uncertainty" and the postulated incomplete information may have genuine "biting power." Thus, only in this case will type unobservability may overcome the complete-information effects emphasized in Section 2.

As an extreme illustration of this point, suppose that we were to force  $k = n$ ; even allowing  $n$  to become arbitrarily large. This is the underlying context of Example 2, the only difference being that the discussion there was carried out under the assumption of complete information. However, it is clear that the alternative assumption of incomplete information would leave the conclusions completely unaffected in this case because of the symmetry of the game – even if individuals might be uncertain about the type of each particular player, this is an irrelevant piece of information as long as the corresponding population frequencies (which are now identical to group frequencies) are accurately known. Hence, even under incomplete information, one can obtain the extermination of individualistic preferences in the context of Example 2, in contrast with the conclusion reached by Theorem 1.

We should note that assumption (SC), as used in Theorem 1, may be thought of as unusually strong for a game theoretical model. It is then of interest to see how much of Theorem 1 can be salvaged if (SC) is replaced by the following weaker requirement.

**Strict Quasi-Concavity (SQC).** Given any  $x_i \in X^{k_i - 1}$ ; the function  $\%(\xi; x_i) : X \rightarrow \mathbb{R}$  is strictly quasi-concave.

This question is answered by our second main result:

**Theorem 2.** Assume (C), (SQC) and (NTE). Then, there exists some  $M > 0$  such that, if  $n \geq M$ ; individualistic preferences  $\%$  are locally stable in any incomplete-information environment  $E = ((n; k; X; \%); \%$ ):

**Proof.** The proof is a straightforward modification of the proof of Theorem 1. Observe first that Claim 2 of that proof is valid in the present setting. Claim 3, on the other hand, needs to be reformulated as follows:

Claim 3<sup>st</sup>. There exist an  $\epsilon > 0$  and a  $\delta \in (0; 1)$  such that

$$F_{\%}(y; y; z; \delta) \geq F_{\%}(z; y; z; \delta) - \epsilon$$

holds for all  $(y; z) \in [0; 1] \times X^2$  that satisfies (16).

**Proof of Claim 3<sup>st</sup>.** Observe that

$$\lim_{\delta \rightarrow 1} F_{\%}(x; a; b; \delta) = \%(x; a; b); \quad x; a; b \in X;$$

Therefore by continuity of  $F_{\frac{1}{n}}$  in  $\theta$  and by (SQC), there must exist a  $\theta^* \in (0; 1)$  such that  $F_{\frac{1}{n}}(\theta^*; a; b; \theta^*)$  is strictly quasi-concave on  $X$  for all  $a; b \in X$  and all  $\theta^* \in [\theta^*; 1]$ : Given this observation, the argument used in proving Claim 3 of the proof of Theorem 1 applies here without modification to yield the sought result.  $\square$

In view of Claim 3<sup>a</sup>; if we let  $\theta_n = 1 - \frac{1}{n}$ ;  $n = k; 2k; \dots$ ; the proof here may be completed by reproducing the line of argument used for Theorem 1.  $\square$

Theorem 2 is a local counterpart of Theorem 1. It deals with a larger set of environments than Theorem 1, but it delivers only local stability of individualistic preferences. (It is, however, less “local” than it might first appear, as explained in Remarks 2 and 3 below.)

Both of our main results underscore the same point: the evolutionary rationale for individualistic preferences stands strong in incomplete-information environments if there is a large number of subgroups involved. Combining this point with our former discussion of Example 3, one arrives at the summarizing conclusion that, among the factors that underlie the evolutionary stability of “materialist behavior,” the information structure plays a central role.

We conclude this section with further remarks that elaborate on different aspects of our analysis.

**Remark 1. (Existence of equilibrium)** As noted earlier, the assumption of (NTE) is used in Theorems 1 and 2 to guarantee trivially the existence of an SBNE for the environments under consideration. We may of course solve the existence problem in a less trivial manner. For instance, if  $\frac{1}{2}$  is concave in the first argument and (SC) holds, then it can be shown that an SBNE must exist. This observation can be used to settle the existence issue that is relevant for Theorem 1. (In particular, an SBNE exists for the environment described in Example 3 for sufficiently small  $\theta$ .) On the other hand, if  $\frac{1}{2}$  is strictly quasi-concave in the first argument and (SQC) holds, then again an SBNE exists, provided that the number of  $\frac{1}{2}$ -types are less than a fixed number and  $n$  is large. This observation applies directly to the setting covered by Theorem 2.<sup>14</sup>  $\square$

**Remark 2. (Multiple mutants)** The local stability concept contemplated in Definition 3 is based on the premise that mutations occur very infrequently and in an uncorrelated fashion so that only a single mutant may exist at any given point in time. Theorem 2 is thus silent with respect to the evolutionary stability of individualistic preferences against not-so-infrequent or correlated

---

<sup>14</sup>Both of these claims can be proved by using a suitable fixed point theorem in the standard manner. After all, as long as we can ensure the quasi-concavity of  $\frac{1}{n}$  and  $\frac{1}{2}$  in the first argument, there is no problem with invoking a standard existence theorem. (Notice that quasi-concavity of  $\frac{1}{2}$  and (SQC) are not sufficient for this since a convex combination of quasi-concave functions need not be quasi-concave.) Alternatively, one may use more sophisticated (pure strategy) equilibrium existence theorems that are based on weaker convexity requirements (such as diagonal quasi-concavity); see, for instance, Baye et al. (1993).

mutations. Yet, it turns out that individualistic preferences perform successfully against multiple mutations as well. Consider first the possibility that several homogeneous mutants may appear simultaneously – see Remark 3 for the case of heterogeneous mutants. Given any positive integer  $m$ ; say that individualistic preferences are  $m$ -locally stable in an environment  $E$ , if the inequality posited in Definition 3 holds for all population compositions that consist of at most  $m$ -many mutants. It is easy to see that we may replace the term “locally stable” with “ $m$ -locally stable” for any  $m$  (which is independent of  $n$ ) without introducing any other modification in Theorem 2.<sup>15</sup> Thus, this result does not really embody a knife-edge conclusion: it convincingly carries the message that individualistic preferences are in general powerful enough to expel any “relatively small” mutation in a vast set of environments with incomplete information.  $\forall$

**Remark 3. (Multiple non-individualistic types)** Concerning mutation, another interesting issue pertains to the possibility that several kinds of mutants may be in simultaneous co-existence with the individualistic type. In that case, of course, the approach must be adapted accordingly since our framework allows only for dichotomic configurations (i.e. individualistic preferences and one alternative type). However, along the lines explained in Remark 2, it is straightforward to observe that as long as the total number of (possibly heterogeneous) mutants  $m$  remains small relative to  $n$ ; no significant changes have to be introduced, either in the modeling of the problem or in the required arguments.

If several non-individualistic types may be simultaneously present, a more difficult issue pertains to whether one can still obtain the global-dominance conclusion established by Theorem 1. In this respect, a crucial point concerns how to extend Assumption (NTE) to that more general context. Suppose that we have a finite set of possible (non-individualistic) preference types  $T = \{ \frac{1}{2}_1; \dots; \frac{1}{2}_m \}$ . Then, a frequency distribution may be identified with a vector in the  $m$ -dimensional simplex  $\Phi^m$ ,  $\mathbf{1} = (1_0; 1_1; \dots; 1_m)$ ; where  $1_s$  stands for the frequency of type  $\frac{1}{2}_s$  and  $1_0$  is that of  $\frac{1}{4}$ . A generalization of (NTE) that is sufficient for our purposes can be formulated as follows.<sup>16</sup>

---

<sup>15</sup>For example, in the proof of an extended Theorem 2, one would simply have to allow for sequences  $\mathbf{1}_n$  where  $1_n \geq 1_j \quad m=n$  for all  $n$ .

<sup>16</sup>Alternatively, one could have considered the following weaker generalization of (NTE):  
 (NTE<sup>00</sup>) An SBNE  $\frac{3}{4}$  exists: Moreover, given  $k; \frac{1}{4}$ ; and any  $\frac{1}{2}_s \in T$ , there exists some associated environment  $E$  such that  $\frac{3}{4}(\frac{1}{4}; \mathbf{1}(t)) \notin \frac{3}{4}(\frac{1}{2}_s; \mathbf{1}(t))$  for every SBNE  $\frac{3}{4}$  and some population composition  $t$ :

However, under this alternative assumption, one cannot exploit the line of argument that established Theorem 1. The key point here concerns the proof of Claim 2, where it could not be argued that if alternative types choose the same equilibrium action at some population configuration, they must also do so everywhere.

(NTE<sup>0</sup>) An SBNE  $\sigma$  exists: Moreover, given  $k; \frac{1}{2}$ ; any  $\frac{1}{2} \in T$ , and any frequency vector  $\frac{1}{2} \in \Phi^m$ ; there exists an associated environment  $E$  such that  $\sigma(\frac{1}{2}; \frac{1}{2}(t)) \in \sigma(\frac{1}{2}; \frac{1}{2}(t))$  for every SBNE  $\sigma$  and some population composition  $t$  such that  $\frac{1}{2}_r(t) = \frac{1}{2}_r$  for all  $0 \in r \in s$ :

Clearly, (NTE<sup>0</sup>) boils down to (NTE) when there is just one alternative non-individualistic type. On the other hand, when there is more than one such kind of alternative preferences, (NTE<sup>0</sup>) simply requires that, for any non-individualistic type, there must always exist some reassignment of its relative frequencies vis a vis the individualistic type where each of them (the individualistic and non-individualistic types) differ in the equilibrium-induced behavior. Under this assumption (together with (C) and (SC)), it is not difficult to see that one can establish a direct “polymorphic” counterpart of Theorem 1. That is, individualistic preferences still exterminate all non-individualistic types, no matter how many of them there are and what are their initial frequencies.  $\forall$

Remark 4. (Related literature) As noted earlier, there are some recent papers that have tackled the study of preference evolution along lines similar to those pursued here. Particularly relevant to the present study are Ely and Y-lankaya (1997), Güth and Peleg (1997), and Dekel et al. (1998), which are henceforth referred to as EY, GP, and DEY, respectively. DEY show that, in the case of continuum populations, if players observe their opponents’ preferences, every evolutionarily stable profile must induce an efficient pattern of play in terms of material payoffs. Instead, if the types are unobservable, DEY establish that only a situation in which individuals’ preferences are consistent with material payoff maximization may be evolutionarily stable. EY and GP report analogous results applicable to the case of unobservable types. Apart from certain specific modeling details, there are three important aspects in which the present work differs from these papers. We take up each of these issues in turn.

(i) The cardinality of the population: In the three papers indicated (as well as in most of the related literature<sup>17</sup>), the underlying assumption concerning the population consists of a continuum of infinitesimal players (this is explicitly assumed in EY and DEY, and implicitly in GP). However, this assumption leaves little room for some of the strategically-relevant considerations that are well known to play a significant role in many evolutionary models, as discussed in detail by Hamilton (1970) and Schaefer (1989). For this reason, our present framework involves instead a finite (but typically large) population, a feature which allows for the possibility that negatively interdependent preferences (e.g. “spiteful” or envious) may play a significant role in the analysis. When the population cardinality is the continuum, this possibility is much curtailed, even if actual play takes place within small (finite) groups. For, in this respect, the analysis of continuum models may embody sharp discontinuities

---

<sup>17</sup>See, for example, the evolutionary explanations of time preference by Rogers (1994), risk aversion by Robson (1996), or altruism by Bester and Güth (1998).

with large-but-finite population models. Our discussion of Example 3 is a good case in point. No matter how large the population is in that context, any single agent displaying spiteful preferences will see her offspring eventually dominate a population originally made of agents with individualistic preferences.<sup>18</sup> However, if we had decided to model directly a large population through the continuum assumption, no (infinitesimal) agent would have had any possibility of affecting the population state and, therefore, if endowed with non-materialist preferences, she would have been unable to survive in that scenario.

(ii) The range of the stability analysis: To the best of our knowledge, almost every paper in this literature who has addressed the issue of preference evolution has relied on local stability concepts.<sup>19</sup> For instance, while GP and DEY use stability concepts that carry an equilibrium flavor (like the traditional ESS), the paper by EY focuses on a dynamic scenario by invoking the usual notion of asymptotic stability. In contrast, in the present paper we also identify reasonable conditions under which one can provide global answers to the problem at hand (see Example 2 and Theorem 1) along with local answers (Examples 1, 3 and 4, and Theorem 2).

(iii) The matching scenario: Most of the related literature has focused on simple matching contexts where the population is matched in pairs to play a bilateral game. Here, we have extended this schematic framework to allow for groups of varying size – including, as a special case, the scenario in which the entire population is involved in a playing-the-field game. Leaving aside its wider generality, the main advantage of this approach is that it allows us to focus on the relative group size as a relevant (in fact, key) consideration. Its main role in the analysis is to parametrize in a transparent manner the degree of “informativeness” allowed by our incomplete-information framework. A related idea is modeled explicitly in DEY by positing an exogenous probability of true observation of the opponent’s preferences. Here, we maintain the pure dichotomy between perfect and imperfect observation (as in EY and GP) but introduce a varying extent of effective unobservability in terms of relative group size. ¥

---

<sup>18</sup>This means that, in a variety of interesting finite-population scenarios, the assumption of perfect observability of preferences need not lead to an efficient (or Nash) pattern of behavior in terms of fitness, which is in contrast with DEY. Thus, our results underscore the fact that population finiteness will generally have important implications on the issue of preference evolution.

<sup>19</sup>An exception is the paper by Koçkesen et al. (1999) which, as explained, addresses only a scenario with complete information.

## 4 Conclusion

In this paper, we have shown that results pertaining to the evolution of preferences under random matching are bound to depend crucially on the information content of the postulated strategic interaction. If interaction takes place under perfect observability of preferences, a monomorphic population composed only of individualistic (sel...sh) agents may well be vulnerable to invasion by “mutant” preferences that involve traits like spite or altruism. This observation is not at all surprising and is very much in line with the earlier results established in the literature. Unfortunately, one ...nds that, with complete information, little can be said about stable population compositions in general. The long-run predictions of the evolutionary dynamics in this case happen to depend very much on the speci...c details of the strategic environments under consideration.

At the other polar extreme, we have a scenario in which preferences are unobservable. Among the various ways in which this context could be modeled, we have chosen to posit that individuals are aware of the overall population frequencies of types but not of those corresponding to the particular opponents with whom they are matched. This assumption conveniently allows us to analyze the situation as a standard Bayesian game with common priors. Interestingly, we have found that this approach yields a sharp limit result which (save for standard technical assumptions) is independent of the particulars of the games under consideration: individualistic preferences are globally stable in a vast set of environments, provided there is a large enough number of subgroups in the population.

In our view, the aforementioned result provides a fundamental evolutionary rationale for individualistic preferences. However, it is also telling that the presence of incomplete information is not enough to ensure the stability of individualistic preferences when the size of the population is small relative to the size of the subgroups. It would thus be a mistake to conclude on the basis of the present ...ndings that evolution favors unequivocally the individualistic preferences. For it turns out that one needs to know more about the structure of the matching mechanism to identify the precise implications of evolutionary forces on the selection of preferences. In particular, individualistic agents need not thrive in situations in which intragroup selection is more powerful than intergroup selection, even if types are unobservable.

To conclude, let us stress that, while we view the present work as a promising start, it does not say much about alternative evolutionary scenarios that may also involve incomplete information. In particular, further development of the theory must include an assessment of how robust is our analysis to the consideration of environments involving alternative matching processes (such as local interaction or assortative matching), partial observability of types, and socialization e...orts of parents (such as the “imperfect empathy” model), among other issues that we did not consider here.



## References

- Banerjee, A. and Weibull, J. W. (1995). "Evolutionary Selection and Rational Behavior" in *Learning and Rationality in Economics* (Kirman, A. P., and Salmon, M., Eds.). Oxford: Blackwell.
- Baye, M., Tian, G., and Zhou J. (1993). "Characterization of the Existence of Equilibria in Games with Discontinuous and Non-quasiconcave Payoffs," *Review of Economic Studies* 60, 935-948.
- Bester, H. and Güth W. (1998). "Is Altruism Evolutionarily Stable?" *Journal of Economic Behavior and Organization* 34, 193-209.
- Bisin, A., and Verdier T. (1998). "On the Cultural Transmission of Preferences for Social Status," *Journal of Public Economics* 70, 75-98.
- Bisin, A., and Verdier T. (1999). "The Economics of Cultural Transmission and the Dynamics of Preferences," mimeo, NYU.
- Bolton, G., and Ockenfels A. (1998). "A Theory of Equity, Reciprocity and Competition," mimeo, Penn State University.
- Dekel, E., Ely, J. and Y-lankaya, O. (1998). "Evolution of Preferences," mimeo, Northwestern University.
- Ely, J. and Y-lankaya, O. (1997). "Evolution of Preferences and Nash Equilibrium," mimeo, Northwestern University.
- Eshel, I., Samuelson, L., and Shaked, A. (1998). "Altruists, Egoists and Hooligans in a Local Interaction Model," *American Economic Review* 88, 157-179.
- Fehr, E., Schmidt, K. (1999). "A Theory of fairness, Competition and Cooperation," *Quarterly Journal of Economics*, forthcoming.
- Fershtman, C. and Judd, K. L. (1987). "Incentive Equilibrium in Oligopoly," *American Economic Review* 77, 927-40.
- Fershtman, C. and Weiss, Y. (1998). "Social Rewards, Externalities and Stable Preferences," *Journal of Public Economics* 70, 53-74.
- Friedman, M. (1953). "The Methodology of Positive Economics," in *Essays in Positive Economics*, Chicago: University of Chicago Press.
- Güth, W., and Yaari, M. (1992). "Explaining Reciprocal Behavior in Simple Strategic Games: An Evolutionary Approach," in *Explaining Forces and Change: Approaches to Evolutionary Economics* (Witt, U., ed.) Ann Arbor: University of Michigan Press.
- Güth, W., and Peleg, B. (1997). "When will the Fittest Survive? An Indirect Evolutionary Analysis," mimeo, University of Berlin.
- Hamilton, W.D. (1970). "Selfish and Spiteful Behavior in an Evolutionary Model," *Nature*, 228, 1218-20.

- Koçkesen, L., Ok, E. A., and Sethi, R. (1997). "On the Strategic Advantage of Negatively Interdependent Preferences," C.V. Starr Center Working Paper 97-34, NYU.
- Koçkesen, L., Ok, E. A., and Sethi, R. (1999). "Evolution of Interdependent Preferences in Aggregative Games," *Games and Economic Behavior*, forthcoming.
- Levine, D. (1998). "Modeling Altruism and Spitefulness in Experiments," *Review of Economic Dynamics* 1, 593-622.
- Robson, A. (1996): "The Evolution of Attitudes to Risk: Lottery Tickets and Relative Wealth," *Games and Economic Behavior* 14, 190-207.
- Rogers, A.R. (1994): "Evolution of Time Preference by Natural Selection," *American Economic Review*, 84, 460-82.
- Palomino, F. (1996). "Noise Trading in Small Markets," *Journal of Finance* 51, 1537-50.
- Rhode, P. and M. Stegeman (1996). "Learning, Mutation, and Long-Run Equilibria in Games: A Comment," *Econometrica* 64, 443-49.
- Schaefer, M.E. (1989). "Are Profit Maximizers the Best Survivors?: A Darwinian Model of Economic Natural Selection," *Journal of Economic Behavior and Organization* 12, 29-45.
- Selten, R. (1991). "Evolution, Learning, and Economic Behavior," *Games and Economic Behavior*, 3, 3-24.
- Sethi, R. and Somanathan, E. (1999). "Preference Evolution and Reciprocity," mimeo, Barnard College, Columbia University.
- Vega-Redondo, F. (1997): "The Evolution of Walrasian Behavior," *Econometrica* 65, 375-84.
- Vickers, J. (1985). "Delegation and the Theory of the Firm," *Economic Journal*, Supplement 95, 138-147.