

A heterogeneous evolutionary stable population under assortative matching: Exploring the diversity of preferences

Charles Ayoubi¹ and Boris Thurm²

¹ Chaire en économie et management de l'innovation, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland,

`charles.ayoubi@epfl.ch`,

² LEURE Laboratory, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland.

`boris.thurm@epfl.ch`

Abstract. Recent studies have shown that a population acting not only upon self-interest but also exhibiting some morality preference has an evolutionary advantage. Specifically, in the setting of a symmetric fitness game, a resident population is evolutionary stable against all types of mutants if it has the utility function of Homo-moralis, with a morality equal to the assortativity. In this paper, we extend the scope of analysis allowing for the presence of a diversity of preferences in the population. Establishing a Payoff Equality condition, we prove the possibility of co-existence of two residents in the population. We then introduce a tripartite assortment function and study the conditions for the evolutionary stability of this diverse population. In the case of a constant assortment function, we show the existence of an evolutionary stable and heterogeneous resident population.

Keywords: Evolutionary Game Theory, Homo-moralis, Fitness game, Evolutionary stability, Assortativity

1 Introduction

The choice of a self-centered selfish utility in economics has long been challenged and moderated by economists since the early history of the profession. Smith (1759) for instance, in his “Theory of moral sentiments” considered the idea of moral motivation in human decisions. More recently, Arrow (1973), Laffont (1975), Sen (1977), and Tabellini (2008) also included moral value in their consideration of economic behaviors. In a comprehensive analysis of such alternative preference models, Alger and Weibull (2013) introduce Homo-moralis, a new type of preferences which appears to be the most robust in terms of evolutionary stability. The model of Homo-moralis integrates the classical selfish “Homo-economicus” and adds to it a “moral term” which accounts for the preference of the right thing to do. The strength of the model of Alger and Weibull

holds not only in its accounting for morality but also in the fact that a population of agents can be evolutionary stable if it has a certain type of Homo-moralis preferences. The degree of morality of the only evolutionary stable agent depends on the assortativity that governs the interactions between agents.

The introduction of Homo-moralis elegantly tackles the limits of the use of selfishness-based rational preferences. However, by predicting the survival of one and only one type of preferences in the society it limits the possibilities to a unique uniform type of population with the same preferences and behaviors. As introduced by Alger and Weibull (2013), Homo-moralis implies the presence of a single type of agent in the evolutionary surviving population. This conclusion contradicts the observed heterogeneity of behaviors and preferences among agents. Agents exhibit a variety of behaviors and actions when confronted with similar choices. This paper aims at evaluating the possibility of including this heterogeneity in the underlying mathematical model supporting the superiority of Homo-moralis.

In a recent discussion on the evolutionary foundations and the policy implications of Homo-moralis preferences, Alger and Weibull (2016) suggest the presence in a society of a variety of individuals with different preferences. These preferences could be different types of Homo-moralis (i.e. with different morality coefficients) or even have other kinds of preferences (altruism, spitefulness, etc.). Being influenced by various cultural, geographical and societal factors, individuals might develop distinct preferences. This paper aims at evaluating the evolutionary stability of such a hypothesis and the conditions under which distinct preferences could steadily co-exist in an interactive society. Hence, we first extend the assortment function introduced by Bergstrom (2003) to three interacting populations. We then show that under few conditions it is possible to observe the coexistence of two types of resident populations consistently playing different strategies and achieving Payoff Equality. We also show that in the case of a constant assortment function there exist a heterogeneous evolutionary stable resident population.

We implement our analysis in the context of large populations with pairwise matched interactions. Following Güth and Yaari (1992) we build our model within an indirect evolutionary framework where the behavior of agents is driven by personal preferences but the evolutionary success is driven by some exogenous payoff (fitness) function. In order to avoid seeing the agents deviating from their objective payoff maximization (Robson (1990), Dekel, Ely and Yilankaya (2007)), we consider that the (subjective) personal preferences of agents are only known by the agent in question. However, even though the preferences are the agents' private information, building on Bergstrom (2003), we consider that the interactions between agents are not random. Specifically, the meeting probability between two individuals follows an assortative matching process rather than the more classical uniform random matching. This assortative matching process is exogenous and is defined by the *index of assortativity* which sets the likelihood for a mutant to interact with another mutant rather than with a resident when a mutation occurs. The fitness game we consider clearly applies to symmetric

interactions but also to asymmetric interactions with ex-ante symmetry. In other words, in each interaction individuals are as likely to be in one or the side of the interaction. For studying the evolutionary stability, we follow the standard definition of Maynard Smith and Price (1973), applying it to preference evolution rather than strategies. To do so we follow Dekel et al. (2007) and consider that in a given matched interaction agents play a Bayesian Nash equilibrium of the game under incomplete information.

Using this theoretical setting, we explore the overlooked aspect of population diversity and find that under an assortative matching of agents, it is possible to have a heterogeneous population exhibiting Payoff Equality and evolutionary stability to any mutant invasion. The organization of the rest of the paper is as follows: Section 2 sets up the model extending the assortment function to more than three populations, Section 3 introduces the Payoff Equality condition, Section 4 studies evolutionary stability in the case of a constant assortment function and Section 5 concludes.

2 Model and definitions

2.1 A Population with one resident

The first part of this section introduces the general setting in the classical case of a bilateral interaction between a unique resident and a mutant. Therefore we introduce some definitions and results useful for the rest of the paper.

As in Alger and Weibull (2013), we consider a population where individuals are randomly matched into pairs to engage in a symmetric interaction with the common strategy set X . We assume that X is a nonempty, compact and convex set. Individuals are utility maximizers, and their behavior depends on their type $\theta \in \Theta$, i.e. their preferences which are described by a continuous utility function $u_\theta : X^2 \rightarrow \mathbb{R}$. Individuals' success in the game is determined by the resulting payoffs: if $\theta \in \Theta$ plays strategy $x \in X$ when $\tau \in \Theta$ plays strategy $y \in X$, θ gets $\pi(x, y)$ where we assume $\pi : X^2 \rightarrow \mathbb{R}$ to be continuous.

We consider a population of two types $\theta, \tau \in \Theta$. The two types and their respective shares define a population state $s = (\theta, \tau, \varepsilon)$, where $\varepsilon \in (0, 1)$ is the population share of τ . If ε is small, we call θ the resident type and τ the mutant type (see Figure 1).

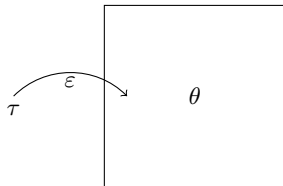


Fig. 1. A two-types population with θ the resident and τ the mutant

The matching process is random and exogenous, and it may be assortative. Let $\Pr[\tau|\theta, \varepsilon]$ be the probability that individual θ is matched with individual τ in the population state $s = (\theta, \tau, \varepsilon)$, we can define the assortment function and assortativity:

Definition 1 (Assortment function and assortativity). *In a population state $s = (\theta, \tau, \varepsilon)$ with $\varepsilon \in (0, 1)$, let $\phi(\varepsilon)$ be the the difference between the conditional probabilities for an individual to be matched with an individual with type θ , given that the individual herself has type θ or τ .*

We have: $\phi(\varepsilon) = \Pr[\theta|\theta, \varepsilon] - \Pr[\theta|\tau, \varepsilon]$, defining an assortment function $\phi : (0, 1) \rightarrow [-1, 1]$.

Assuming ϕ is continuous and converges as ε tends to zero, the assortativity $\sigma \in [0, 1]$ is the limit of ϕ in zero: $\lim_{\varepsilon \rightarrow 0} \phi(\varepsilon) = \sigma$

Individuals choose their strategy in order to maximize their utility. A Bayesian Nash Equilibrium (BNE) is a pair of strategies, one for each type, where each strategy is a best reply to the other in the given population state:

Definition 2 (Bayesian Nash Equilibrium). *In any state $s = (\theta, \tau, \varepsilon)$, a strategy pair $(x^*, y^*) \in X^2$ is a (Bayesian) Nash Equilibrium if:*

$$\begin{cases} x^* \in \operatorname{argmax}_{x \in X} & \Pr[\theta|\theta, \varepsilon] \cdot u_\theta(x, x^*) + \Pr[\tau|\theta, \varepsilon] \cdot u_\theta(x, y^*) \\ y^* \in \operatorname{argmax}_{y \in X} & \Pr[\theta|\tau, \varepsilon] \cdot u_\tau(y, x^*) + \Pr[\tau|\tau, \varepsilon] \cdot u_\tau(y, y^*) \end{cases}$$

The set of Bayesian Nash Equilibria in population state $s = (\theta, \tau, \varepsilon)$, i.e. all solutions (x^, y^*) , is called $B^{NE}(s) \subseteq X^2$.*

This definition allows us to introduce a useful result put forward by Alger and Weibull (2013):

Lemma 1 (Alger and Weibull Lemma 1). *$B^{NE}(s)$ is compact for each $(\theta, \tau, \varepsilon) \in \Theta^2 \times [0, 1)$. If u_θ and u_τ are concave in their first arguments then $B^{NE}(s) \neq \emptyset$. The correspondence $B^{NE}(\theta, \tau, \cdot) : [0, 1) \rightrightarrows X^2$ is upper hemi-continuous.*

Concerning the (objective) payoff (fitness) functions, we have, when θ plays $x \in X$ and τ plays $y \in X$, the following resulting payoffs:

$$\begin{aligned} \Pi_\theta(x, y, \varepsilon) &= \Pr[\theta|\theta, \varepsilon] \cdot \pi(x, x) + \Pr[\tau|\theta, \varepsilon] \cdot \pi(x, y) \\ \Pi_\tau(x, y, \varepsilon) &= \Pr[\theta|\tau, \varepsilon] \cdot \pi(y, x) + \Pr[\tau|\tau, \varepsilon] \cdot \pi(y, y) \end{aligned} \tag{1}$$

Hence, considering the case of a resident population of type θ , what happens when mutants of type τ "invade" the resident population? If the resident population θ withstands a small-scale invasion of τ , then θ is called evolutionary stable against τ . Or, in other words, if an individual θ makes a mistake when deciding on its strategy and plays like a type τ , the error will not propagate:

Definition 3 (Evolutionary-stability). A type $\theta \in \Theta$ is evolutionary stable against a type $\tau \in \Theta$ if there exists an $\bar{\varepsilon} > 0$ such that $\Pi_\theta(x^*, y^*, \varepsilon) > \Pi_\tau(x^*, y^*, \varepsilon)$ in all Nash equilibria (x^*, y^*) in all states $s = (\theta, \tau, \varepsilon)$ with $\varepsilon \in (0, \bar{\varepsilon})$.

A type $\theta \in \Theta$ is evolutionary stable if it is evolutionary stable against all types $\tau \neq \theta \in \Theta$.

Definition 4 (Evolutionary-unstability). A type $\theta \in \Theta$ is evolutionary unstable if there exists a type $\tau \in \Theta$ and an $\bar{\varepsilon} > 0$ such that $\Pi_\theta(x^*, y^*, \varepsilon) < \Pi_\tau(x^*, y^*, \varepsilon)$ in all Nash equilibria (x^*, y^*) in all states $s = (\theta, \tau, \varepsilon)$ with $\varepsilon \in (0, \bar{\varepsilon})$.

A peculiar type of preference in our context, introduced by Alger and Weibull (2013), is the preference of Homo-moralis. This type of preferences includes all linear convex combination of selfishness and morality as follows:

Definition 5 (Homo-moralis). An individual is a homo-moralis if her utility function is of the form:

$$u_\kappa(x, y) = (1 - \kappa) \cdot \pi(x, y) + \kappa \cdot \pi(x, x) \quad (2)$$

where $\kappa \in [0, 1]$ is her degree of morality.

If the degree of morality is equal to the assortativity $\kappa = \sigma$, then the individual is called homo hamiltonensis. If $\kappa = 0$, then the individual is a homo-oeconomicus (fully selfish). If $\kappa = 1$, then the individual is a homo-kantiensis (fully moral).

For each type $\theta \in \Theta$, β_θ denotes the best-reply correspondence:

$$\beta_\theta(y) = \operatorname{argmax}_{x \in X} u_\theta(x, y) \quad \forall y \in X$$

and X_θ the set of fixed-points under β_θ :

$$X_\theta = \{x \in X : x \in \beta_\theta(x)\}$$

For example, we call X_σ is the set of fixed-points for homo-hamiltonensis, i.e. the Hamiltonian strategies. Moreover, we define the concept of behavioral-alike:

Definition 6 (Behavioral-alike). Types θ and τ are behavioral alikes if they are behaviorally indistinguishable. Let θ be the resident, the set of types τ that are behaviorally alikes to θ is called Θ_θ

$$\Theta_\theta = \{\tau \in \Theta : \exists x \in X_\theta \text{ s.t. } (x, x) \in B^{NE}(\theta, \tau, 0)\}$$

Finally, the type set Θ is called rich if for each strategy $x \in X$, there exists some type $\theta \in \Theta$ for which this strategy is strictly dominant. In other words, Θ is rich if, for any strategy in the game, there exists a mutant to play it.

These definitions allow us to clearly state the main result of Alger and Weibull (2013):

Theorem 1 (Alger and Weibull Theorem 1). *If $\beta_\sigma(x)$ is a singleton for all $x \in X_\sigma$, then homo-hamiltonensis is evolutionary stable against all types $\tau \notin \Theta_\sigma$. If Θ is rich, $X_\theta \cap X_\sigma = \emptyset$, and X_θ is a singleton, then θ is evolutionary unstable.*

The theorem means that only a strategy played by homo-hamiltonensis $x_\sigma \in X_\sigma$ is evolutionary stable. Considering a population of homo-hamiltonensis and a mutant that wishes to enter the population, if the mutant is not behaviorally-alike to homo-hamiltonensis, he has no chance. If the mutant is a homo-moralis with a morality different from the assortativity ($\kappa \neq \sigma$), such that this homo-moralis and homo-hamiltonensis are not behaviorally-alike, then to enter the population, the morality of the homo-moralis should evolve in direction of the assortativity.

In Alger and Weibull (2013), the type of resident is unique living in a perfectly homogeneous population. But is this a required feature of the resident population? What happens when the population is more diverse? Is it possible to have a stable population of two residents? Our paper explores the case of a population where two types live together.

2.2 A Population with two residents

We expand the previous model by adding another type. Now the population is made of two residents θ_1 and θ_2 , and one mutant τ (see Figure 2). The three types and their respective shares define a population state $s = (\theta_1, \theta_2, \tau, (\lambda, \varepsilon))$, where $\varepsilon \in (0, 1)$ is the population share of τ and $\lambda \in (0, 1)$ ³ is the share of θ_2 with respect to θ_1 . Thus, the population share of θ_2 is $\lambda_2 = \lambda(1 - \varepsilon)$ and the population share of θ_1 is $\lambda_1 = (1 - \lambda)(1 - \varepsilon)$.

Individuals are still randomly matched into pairs to engage in a symmetric interaction with the common strategy set X . The matching process is random and exogenous, and it may be assortative.

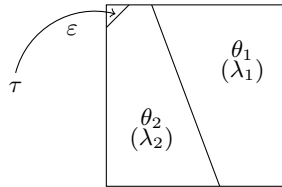


Fig. 2. A three-types population with θ_1 and θ_2 the residents and τ the mutant

For the sake of tractability, we introduce a new notation. The probability that individual θ_1 is matched with individual θ_2 is called $p_{21} = \Pr[\theta_2|\theta_1, (\lambda, \varepsilon)]$. Similarly, $p_{11} = \Pr[\theta_1|\theta_1, (\lambda, \varepsilon)]$ and $p_{31} = \Pr[\tau|\theta_1, (\lambda, \varepsilon)]$.

Extending the concept of (Bayesian) Nash Equilibrium seen above to the case of three types of agents $(\theta_1, \theta_2, \tau)$, we have:

³ When $\lambda = 0$ or $\lambda = 1$, there is only one resident in the population (Section 2.1).

Definition 7 (Bayesian Nash Equilibrium). *In a population state $s = (\theta_1, \theta_2, \tau, (\lambda, \varepsilon))$, $(x^1, x^2, y) \in X^3$ is a (Bayesian) Nash equilibrium if*

$$\begin{cases} x^1 \in \operatorname{argmax}_{x \in X} p_{11} \cdot u_{\theta_1}(x, x^1) + p_{21} \cdot u_{\theta_1}(x, x^2) + p_{31} \cdot u_{\theta_1}(x, y) \\ x^2 \in \operatorname{argmax}_{x \in X} p_{12} \cdot u_{\theta_2}(x, x^1) + p_{22} \cdot u_{\theta_2}(x, x^2) + p_{32} \cdot u_{\theta_2}(x, y) \\ y \in \operatorname{argmax}_{x \in X} p_{13} \cdot u_{\tau}(x, x^1) + p_{23} \cdot u_{\tau}(x, x^2) + p_{33} \cdot u_{\tau}(x, y) \end{cases} \quad (3)$$

Note that when τ is a mutant, i.e. when ε goes to zero, we have $p_{31} = p_{32} = 0$, and θ_1, θ_2 behave as if τ was not in the population. Bayesian Nash Equilibria are then the same as in Definition 2.

Matching probability and assortment

In this paragraph, building on Bergstrom (2003), we introduce a novel, tripartite assortment function allowing for matching and interactions between three distinct populations. In our case, the three populations are the two residents and the mutant.

We have three matching conditions on the probability:

$$\begin{cases} p_{11} + p_{21} + p_{31} = 1 \\ p_{12} + p_{22} + p_{32} = 1 \\ p_{13} + p_{23} + p_{33} = 1 \end{cases} \quad (4)$$

In other words, an individual is matched with an individual θ_1, θ_2 or τ with probability one.

There are also three balancing conditions:

$$\begin{cases} \lambda_2 \cdot p_{12} = \lambda_1 \cdot p_{21} \\ \varepsilon \cdot p_{13} = \lambda_1 \cdot p_{31} \\ \varepsilon \cdot p_{23} = \lambda_2 \cdot p_{32} \end{cases} \quad (5)$$

Balancing conditions insure the coherence of the assortative matching. In other words, they guarantee that the probability of the event "being of type θ_1 and being matched with an individual of type θ_2 " is the same as the probability of the event "being of type θ_2 and being matched with an individual of type θ_1 " so that no individual in the population is left without being matched. For instance:

$$\begin{aligned} \Pr[\theta = \theta_2] \cdot \Pr[\theta_1 | \theta_2, (\lambda, \varepsilon)] &= \Pr[\theta = \theta_1] \cdot \Pr[\theta_2 | \theta_1, (\lambda, \varepsilon)] \\ \lambda_2 \cdot p_{12} &= \lambda_1 \cdot p_{21} \end{aligned}$$

Now we need to redefine the notion of assortment function and assortativity for a population made of more than two types:

Definition 8 (Assortment function and assortativity). *Let θ be a member of a diverse population made of two types or more, and λ the share of θ in the*

population. The difference between the conditional probabilities for an individual to be matched with type θ , given that the individual herself has type θ or any other type ($\bar{\theta}$) is called $\phi(\lambda)$.

We have $\phi(\lambda) = \Pr[\theta|\theta, \lambda] - \Pr[\theta|\bar{\theta}, \lambda]$, defining an assortment function $\phi : (0, 1) \rightarrow [-1, 1]$.

Note that $\phi(\lambda) = \phi(1 - \lambda)$ because of balancing conditions and by symmetry since the assortment process is exogenous and does not depend on the type: $\Pr[\theta|\theta, \lambda] = 1 - \Pr[\bar{\theta}|\theta, \lambda]$ and $\Pr[\theta|\bar{\theta}, \lambda] = 1 - \Pr[\bar{\theta}|\bar{\theta}, \lambda]$.

Assuming ϕ is continuous and converges as λ tends to zero, the assortativity $\sigma \in [0, 1]$ is the limit of ϕ in zero: $\lim_{\lambda \rightarrow 0} \phi(\lambda) = \sigma$

For example, in a three-types population $s = (\theta_1, \theta_2, \tau, (\lambda, \varepsilon))$, for θ_1 we have:

$$\begin{aligned} \phi(\lambda_1) &= \Pr[\theta_1|\theta_1, (\lambda, \varepsilon)] - \Pr[\theta_1|\bar{\theta}_1, (\lambda, \varepsilon)] \\ &= \Pr[\theta_1|\theta_1, (\lambda, \varepsilon)] - \Pr[\theta_1|\theta_2 \cup \tau, (\lambda, \varepsilon)] \\ &= \Pr[\theta_1|\theta_1, (\lambda, \varepsilon)] - \frac{\lambda_2 \cdot \Pr[\theta_1|\theta_2, (\lambda, \varepsilon)] + \varepsilon \cdot \Pr[\theta_1|\tau, (\lambda, \varepsilon)]}{1 - \lambda_1} \\ &= p_{11} - \frac{\lambda_2 \cdot p_{12} + \varepsilon \cdot p_{13}}{1 - \lambda_1} \end{aligned}$$

The same applies for θ_2 and τ , giving us three assortment conditions:

$$\begin{cases} (1 - \lambda_1) \cdot \phi(\lambda_1) = (1 - \lambda_1) \cdot p_{11} - \lambda_2 \cdot p_{12} - \varepsilon \cdot p_{13} \\ (1 - \lambda_2) \cdot \phi(\lambda_2) = (1 - \lambda_2) \cdot p_{22} - \lambda_1 \cdot p_{21} - \varepsilon \cdot p_{23} \\ (1 - \varepsilon) \cdot \phi(\varepsilon) = (1 - \varepsilon) \cdot p_{33} - \lambda_1 \cdot p_{31} - \lambda_2 \cdot p_{32} \end{cases} \quad (6)$$

Solving the system of linear equations 4, 5 and 6, we can rewrite the matching probabilities in function of the share of each type in the population and the assortment. Let $\psi(\lambda) = \lambda \cdot (1 - \lambda) \cdot (1 - \phi(\lambda))$, we have:

$$\begin{aligned} p_{11} &= \lambda_1 + (1 - \lambda_1) \cdot \phi(\lambda_1) \\ p_{21} &= \frac{\psi(\lambda_1) + \psi(\lambda_2) - \psi(\varepsilon)}{2\lambda_1} \\ p_{31} &= \frac{\psi(\lambda_1) - \psi(\lambda_2) + \psi(\varepsilon)}{2\lambda_1} \\ p_{12} &= \frac{\psi(\lambda_1) + \psi(\lambda_2) - \psi(\varepsilon)}{2\lambda_2} \\ p_{22} &= \lambda_2 + (1 - \lambda_2) \cdot \phi(\lambda_2) \\ p_{32} &= \frac{-\psi(\lambda_1) + \psi(\lambda_2) + \psi(\varepsilon)}{2\lambda_2} \\ p_{13} &= \frac{\psi(\lambda_1) - \psi(\lambda_2) + \psi(\varepsilon)}{2\varepsilon} \\ p_{23} &= \frac{-\psi(\lambda_1) + \psi(\lambda_2) + \psi(\varepsilon)}{2\varepsilon} \\ p_{33} &= \varepsilon + (1 - \varepsilon) \cdot \phi(\varepsilon) \end{aligned} \quad (7)$$

We can calculate the limits when ε goes to zero, since τ is a mutant in this population. We have: $\lambda_1 \xrightarrow{\varepsilon \rightarrow 0} (1 - \lambda)$ and $\lambda_2 \xrightarrow{\varepsilon \rightarrow 0} \lambda$. For simplicity, we assume

$\phi(\cdot)$ to be differentiable, such that $\phi'(\lambda) = -\phi'(1-\lambda)$ because $\phi(\lambda) = \phi(1-\lambda)$.

$$\begin{aligned}
p_{11} &\xrightarrow{\varepsilon \rightarrow 0} (1-\lambda) + \lambda \cdot \phi(\lambda) \\
p_{21} &\xrightarrow{\varepsilon \rightarrow 0} \lambda \cdot (1 - \phi(\lambda)) \\
p_{31} &\xrightarrow{\varepsilon \rightarrow 0} 0 \\
p_{12} &\xrightarrow{\varepsilon \rightarrow 0} (1-\lambda) \cdot (1 - \phi(\lambda)) \\
p_{22} &\xrightarrow{\varepsilon \rightarrow 0} \lambda + (1-\lambda) \cdot \phi(\lambda) \\
p_{32} &\xrightarrow{\varepsilon \rightarrow 0} 0 \\
p_{13} &\xrightarrow{\varepsilon \rightarrow 0} -\frac{\lambda \cdot (1-\lambda) \cdot \phi'(\lambda)}{2} + (1-\lambda) \cdot (1 - \phi(\lambda)) + \frac{\phi(\lambda) - \sigma}{2} \\
p_{23} &\xrightarrow{\varepsilon \rightarrow 0} \frac{\lambda \cdot (1-\lambda) \cdot \phi'(\lambda)}{2} + \lambda \cdot (1 - \phi(\lambda)) + \frac{\phi(\lambda) - \sigma}{2} \\
p_{33} &\xrightarrow{\varepsilon \rightarrow 0} \sigma
\end{aligned} \tag{8}$$

We show the calculation to get p_{13} . We have:

$$\begin{aligned}
\psi(\lambda_1) &= (1-\lambda)(1-\varepsilon)(\lambda + \varepsilon(1-\lambda)) \cdot (1 - \phi(\lambda_1)) \\
\psi(\lambda_2) &= \lambda(1-\varepsilon)(1-\lambda + \varepsilon\lambda) \cdot (1 - \phi(\lambda_2))
\end{aligned}$$

So we can rewrite p_{13} :

$$p_{13} = (1-\varepsilon) \left(\frac{\lambda(1-\lambda) \cdot (\phi(\lambda_2) - \phi(\lambda_1))}{2\varepsilon} + \frac{(1-\lambda)^2(1 - \phi(\lambda_1)) - \lambda^2(1 - \phi(\lambda_2))}{2} + \frac{(1 - \phi(\varepsilon))}{2} \right)$$

And, using $\lim_{\varepsilon \rightarrow 0} \phi(\lambda_1) = \lim_{\varepsilon \rightarrow 0} \phi(\lambda_2) = \phi(\lambda)$, L'Hôpital's rule and $\phi'(\lambda) = -\phi'(1-\lambda)$, we have:

$$\begin{aligned}
\lim_{\varepsilon \rightarrow 0} p_{13} &= \frac{\lambda(1-\lambda)}{2} \cdot \lim_{\varepsilon \rightarrow 0} \left[\frac{\phi(\lambda_2) - \phi(\lambda_1)}{\varepsilon} \right] + \frac{(1-2\lambda)(1 - \phi(\lambda))}{2} + \frac{(1-\sigma)}{2} \\
&= \frac{\lambda(1-\lambda)}{2} \cdot \lim_{\varepsilon \rightarrow 0} \left[\frac{\phi(\lambda(1-\varepsilon)) - \phi((1-\lambda)(1-\varepsilon))}{\varepsilon} \right] + (1-\lambda)(1 - \phi(\lambda)) + \frac{\phi(\lambda) - \sigma}{2} \\
&= \frac{\lambda(1-\lambda) \cdot [-\lambda\phi'(\lambda) + (1-\lambda)\phi'(1-\lambda)]}{2} + (1-\lambda)(1 - \phi(\lambda)) + \frac{\phi(\lambda) - \sigma}{2} \\
&= -\frac{\lambda(1-\lambda)\phi'(\lambda)}{2} + (1-\lambda)(1 - \phi(\lambda)) + \frac{\phi(\lambda) - \sigma}{2}
\end{aligned}$$

We can obtain p_{23} from similar calculations. Note that it is not necessary (but convenient) to assume $\phi(\cdot)$ to be differentiable.

3 Payoff Equality

The core contribution of this paper resides in the consideration of the cohabitation of more than one type in the resident population. What happens then? If one type dominates the other, for example if θ_1 gets a higher payoff than θ_2 , it seems unlikely that θ_2 would survive. Thus, we need to impose a condition for the mutual coexistence of two distinct types without one overcoming (or invading) the other. We call it the Payoff Equality condition. This section evaluates the conditions under which this Payoff Equality is met.

3.1 Payoff Equality condition

Let's assume $(x^1, x^2) \in X^2$ is a BNE in the population state $s = (\theta_1, \theta_2, \lambda)$, with $x^1 \neq x^2$ and $\lambda \in (0, 1)$. The Payoff Equality condition is met when:

$$\Pi_{\theta_1}(x^1, x^2, \lambda) = \Pi_{\theta_2}(x^1, x^2, \lambda) \quad (9)$$

We can write the payoffs of θ_1 and θ_2 (Equation (1)) using the matching probabilities with only two residents derived in (8) and the notation $\pi^{ij} = \pi(x^i, x^j)$:

$$\begin{aligned} \Pi_{\theta_1}(x^1, x^2, \lambda) &= (1 - \lambda + \lambda\phi(\lambda)) \cdot \pi^{11} + \lambda(1 - \phi(\lambda)) \cdot \pi^{12} \\ \Pi_{\theta_2}(x^1, x^2, \lambda) &= (1 - \lambda)(1 - \phi(\lambda)) \cdot \pi^{21} + (\lambda + (1 - \lambda)\phi(\lambda)) \cdot \pi^{22} \end{aligned} \quad (10)$$

Under what conditions on the payoffs $(\pi^{11}, \pi^{12}, \pi^{21}, \pi^{22})$ and on $\lambda \in (0, 1)$ are total payoffs are equal?

We simply equate the expressions of the payoff functions of θ_1 and θ_2 (equation (10)), which gives, after rearranging, the following condition:

$$\lambda(1 - \phi(\lambda))(\pi^{11} + \pi^{22} - \pi^{12} - \pi^{21}) = \pi^{11} - \pi^{21} - \phi(\lambda)(\pi^{22} - \pi^{21})$$

This equation can be rewritten by reordering the terms to put forward the share of θ_1 in the population $((1 - \lambda))$, as follows:

$$(1 - \lambda)(1 - \phi(\lambda))(\pi^{11} + \pi^{22} - \pi^{12} - \pi^{21}) = \pi^{22} - \pi^{12} - \phi(\lambda)(\pi^{11} - \pi^{12})$$

Then using the notations:

$$\begin{aligned} Q_\pi &= \pi^{11} - \pi^{21} - \phi(\lambda)(\pi^{22} - \pi^{21}) \\ R_\pi &= \pi^{22} - \pi^{12} - \phi(\lambda)(\pi^{11} - \pi^{12}) \\ S_\pi &= (\pi^{11} + \pi^{22} - \pi^{12} - \pi^{21}) \end{aligned} \quad (11)$$

Note that, combining the equations in (11), we have: $Q_\pi + R_\pi = (1 - \phi(\lambda))S_\pi$.

We can rewrite the Payoff Equality condition to obtain two equivalent equations, one for λ and the other for $(1 - \lambda)$:

$$\lambda(1 - \phi(\lambda))S_\pi = Q_\pi \quad (12)$$

$$(1 - \lambda)(1 - \phi(\lambda))S_\pi = R_\pi \quad (13)$$

We have the following proposition:

Proposition 1 (Payoff Equality). *Let $(x^1, x^2) \in X^2$ be a BNE in the population state $s = (\theta_1, \theta_2, \lambda)$, with $x^1 \neq x^2$ and $\lambda \in (0, 1)$. The Payoff Equality condition $\Pi_{\theta_1}(x^1, x^2, \lambda) = \Pi_{\theta_2}(x^1, x^2, \lambda)$ is satisfied if and only if:*

1. $Q_\pi = 0$ and $S_\pi = 0$, or
2. $Q_\pi \neq 0$ and $R_\pi \neq 0$ are of the same sign, and $\lambda(1 - \phi(\lambda)) = Q_\pi/S_\pi$, or
3. $Q_\pi = 0, R_\pi = 0, S_\pi \neq 0$ and $\phi(\lambda) = 1$.

Note that for all other possible cases, Payoff Equality is not satisfied.

Proof. First, let's assume that Payoff Equality is satisfied. We distinguish two cases:

When $\phi(\lambda) = 1$:

Since we have Payoff Equality, equations (12) and (13) are satisfied and we necessarily have that $Q_\pi = R_\pi = 0$ (since $(1 - \phi(\lambda)) = 0$). Then either $S_\pi = 0$ and we are in case 1. of the proposition, or $S_\pi \neq 0$ and we are in case 3. of the proposition.

When $\phi(\lambda) \neq 1$:

Payoff equality (equations (12) and (13)) implies only two possibilities.

Either $Q_\pi = 0$ and then $S_\pi = 0$ and we are in the case 1. of the proposition.

Or $Q_\pi \neq 0$ and then since $(1 - \phi(\lambda)) > 0$ and $\lambda > 0$, Q_π and S_π are of the same sign. Moreover, since $\lambda < 1$, R_π and S_π have the same sign. Finally, dividing (12) by $S_\pi \neq 0$ we have: $\lambda(1 - \phi(\lambda)) = Q_\pi/S_\pi$ and we are in case 2. of the proposition.

For the converse, using similar arguments, if one of the three cases (1.,2.,3. of Proposition 1) is true then the condition stated in equation (12) is satisfied and we have Payoff Equality. ■

Here are a few examples for each case of the proposition:

1. (a) If $\pi^{12} = \pi^{11} = \pi^{22} = \pi^{21}$. No matter the share of each type, they will get the same payoff.
- (b) If $\pi^{21} = \pi^{11} \neq \pi^{22} = \pi^{12}$, and if $\phi(\lambda) = 0$. Without assortment, the matching probability are equal to the share in the population.
2. (a) If $\{\pi^{11} = \pi^{22}, \pi^{11} > \pi^{12}, \pi^{11} > \pi^{21}\}$ or if $\{\pi^{11} = \pi^{22}, \pi^{11} < \pi^{12}, \pi^{11} < \pi^{21}\}$, and if $\lambda = (\pi^{11} - \pi^{21})/S_\pi$ and $\phi(\lambda) \neq 1$. In the special case where $\pi^{12} = \pi^{21}$, then λ should be equal to 1/2.
- (b) If $\{\pi^{12} = \pi^{21} = \pi^{11} \neq \pi^{22}\}$ or if $\{\pi^{11} \neq \pi^{22} = \pi^{12} = \pi^{21}\}$, and if $\lambda = -\phi(\lambda)/(1 - \phi(\lambda))$. Note that in this case the assortment is negative: $\phi(\lambda) < 0$.
3. An assortativity equal to 1 means that θ_1 only meets θ_1 , getting π^{11} , while θ_2 only meets θ_2 , getting π^{22} .
4. If $\pi^{12} = \pi^{11} \neq \pi^{22} = \pi^{21}$, then there is no λ that satisfies Payoff Equality. No matter the share in the population, one player will always get more than the other.

3.2 Case of constant assortment

Having established the conditions for an equality of the payoffs among two types of resident population, we now consider the case of a constant assortment function and show the peculiarity of homo-hamiltonensis introduced by Alger and Weibull (2013) in this special case. Homo-hamiltonensis' strategies are closely related to the Payoff Equality condition when the assortment function is constant. We recall that X_σ is the set of fixed-points for homo-hamiltonensis, i.e.

the strategies that homo-hamiltonensis could play when she is the only resident in the population. We define: $X_1 \times X_2 = \{(x^1, x^2) \subseteq B^{NE}(s) \text{ such that } x^1 \neq x^2\}$. $X_1 \times X_2$ is the set of solutions of the Bayesian Nash Equilibrium problem such that θ_1 and θ_2 do not play the same strategy. In the following, we assume that the assortment is constant, equal to the assortativity (by continuity): $\forall \lambda \in [0, 1], \phi(\lambda) = \sigma \in [0, 1]$. We have the following theorem:

Theorem 2 (Payoff Equality and Hamiltonian strategies). *When the assortment is constant, suppose that X_σ is not a singleton and let $(\theta_1, \theta_2, \lambda)$ be a population where θ_1 plays x^1 , θ_2 plays x^2 , such that $(x^1, x^2) \in X_1 \times X_2 \subset X_\sigma^2$, and $\lambda = Q_\pi / (1 - \sigma)S_\pi$ the share of θ_2 in the population.*

If $\beta_\sigma(x)$ is a singleton for all $x \in X_\sigma$, then the population $(\theta_1, \theta_2, \lambda)$ satisfies the Payoff Equality condition.

Proof. Since θ_1 plays $x^1 \in X_\sigma$, θ_2 plays $x^2 \in X_\sigma$ with $x^1 \neq x^2$, and $\beta_\sigma(x)$ is a singleton for all $x \in X_\sigma$, we have:

$$\begin{cases} x^1 \in \operatorname{argmax}_{x \in X} u_\sigma(x, x^1) & \Rightarrow \quad \forall x \neq x^1 \in X, \pi(x^1, x^1) > (1 - \sigma) \cdot \pi(x, x^1) + \sigma \cdot \pi(x, x) \\ x^2 \in \operatorname{argmax}_{y \in X} u_\sigma(y, x^2) & \Rightarrow \quad \forall y \neq x^2 \in X, \pi(x^2, x^2) > (1 - \sigma) \cdot \pi(y, x^2) + \sigma \cdot \pi(y, y) \end{cases}$$

In particular, for $x = x^2$ and $y = x^1$, we have:

$$\begin{cases} \pi^{11} > (1 - \sigma) \cdot \pi^{21} + \sigma \cdot \pi^{22} & \Rightarrow \quad Q_\pi > 0 \\ \pi^{22} > (1 - \sigma) \cdot \pi^{12} + \sigma \cdot \pi^{11} & \Rightarrow \quad R_\pi > 0 \end{cases}$$

Consequently, since $\lambda = Q_\pi / ((1 - \sigma)S_\pi)$ by assumption, we are in case 2. of Proposition 1 and the Payoff Equality condition is satisfied. ■

3.3 Some examples on finite games

We now look at several examples in two-strategies games to illustrate the Payoff Equality condition. We use the notation defined above $\pi^{ij} = \pi(x^i, x^j)$ to describe the payoff obtained by a player playing x^i when her opponent plays x^j . Moreover, we denote by π_{ij} the payoff when pure strategy i is played against pure strategy j . Also, we study in the following examples the case of a constant assortment function.

Example 1 (Coordination game).

In the coordination game, players get reward only if they play the same strategy. Suppose that the θ_1 type plays strategy A while the θ_2 type plays strategy B. For a θ_1 agent, when she is matched with a θ_1 agent, she gets the payoff $\pi^{AA} = \pi_{AA}$, and when she is matched with a θ_2 agent, she gets $\pi^{AB} = \pi_{AB} = 0$.

Thus, $S_\pi = \pi_{AA} + \pi_{BB} > 0$, $Q_\pi = \pi_{AA} - \sigma\pi_{BB}$ and $R_\pi = \pi_{BB} - \sigma\pi_{AA}$. If $\pi_{AA} = \pi_{BB}$, then we must have $\sigma = 1$ to satisfy Payoff Equality (we are then in

case 3. of Proposition 1). If $\pi_{AA} > \pi_{BB}$, then $\lambda = (\pi_{AA} - \sigma\pi_{BB})/[(1 - \sigma)(\pi_{AA} + \pi_{BB})]$ satisfies Payoff Equality if $\sigma < \pi_{BB}/\pi_{AA}$ (case 2. of Proposition 1). Finally, if $\pi_{AA} < \pi_{BB}$, then $\lambda = (\pi_{AA} - \sigma\pi_{BB})/[(1 - \sigma)(\pi_{AA} + \pi_{BB})]$ satisfies Payoff Equality if $\sigma < \pi_{AA}/\pi_{BB}$ (case 2. of Proposition 1).

Table 1. Coordination game ($\pi_{AA} > 0$; $\pi_{BB} > 0$)

	A	B
A	(π_{AA}, π_{AA})	$(0, 0)$
B	$(0, 0)$	(π_{BB}, π_{BB})

For example, when $\pi_{AA} = 2$ and $\pi_{BB} = 1$, $\lambda = (2 - \sigma)/[3(1 - \sigma)]$ satisfies Payoff Equality when $\sigma < 1/2$. With $\sigma = 1/5$, $\lambda = 3/4$, and θ_1 and θ_2 get the same total payoff $\Pi = 0.8$. Under uniform random matching, $\sigma = 0$, we get the mixed Nash equilibrium with $\lambda = 2/3$.

Table 2. Coordination game example

	A	B
A	$(2, 2)$	$(0, 0)$
B	$(0, 0)$	$(1, 1)$

Example 2 (Prisoner's dilemma).

In the prisoner's dilemma, players can either cooperate or defect, getting $\pi_{CD} = 0 < \pi_{DD} < \pi_{CC} < \pi_{DC}$. Suppose a homo-kantiensis (θ_1 , playing C) plays with a homo-economicus (θ_2 , playing D).

Table 3. Prisoner's dilemma

	C	D
C	(π_{CC}, π_{CC})	(π_{CD}, π_{DC})
D	(π_{DC}, π_{CD})	(π_{DD}, π_{DD})

(a) As a first example we consider the case when $\pi_{CD} = 0$, $\pi_{DD} = 1$, $\pi_{CC} = 4$ and $\pi_{DC} = 6$. We then have $S_\pi = -1 < 0$, $Q_\pi = -2 + 5\sigma$ and $R_\pi = 1 - 4\sigma$. Thus, there is a solution when $1/4 < \sigma < 2/5$. With $\sigma = 1/3$, then $\lambda = 1/2$ and homo-kantiensis and homo-oeconomicus co-exist and get the same payoff equal to $\Pi = 9/3$, which would be impossible without assortativity where only homo-oeconomicus survives.

(b) Now let $\pi_{CD} = 0$, $\pi_{DD} = 1$, $\pi_{CC} = 4$ and $\pi_{DC} = 5$, we have $S_\pi = 0$, $Q_\pi = -1 + 4\sigma$ and $R_\pi = 1 - 4\sigma$. Thus, we need $\sigma = 1/4$ to satisfy Payoff Equality.

(c) Finally, let $\pi_{CD} = 0$, $\pi_{DD} = 1$, $\pi_{CC} = 4$ and $\pi_{DC} = 4.5$, we have $S_\pi = 0.5 > 0$, $Q_\pi = -0.5 + 3.5\sigma$ and $R_\pi = 1 - 4\sigma$. Thus, there is a solution when $1/7 < \sigma < 1/4$. For example, when $\sigma = 1/5$, then $\lambda = 1/2$ and homo-kantiensis and homo-oeconomicus live together and get the same payoff equal to $\Pi = 12/5$.

The three cases above are summarized in Table 4 below.

Table 4. Prisoner's dilemma examples

	C	D		C	D		C	D
(a)	C (4, 4)	D (0, 6)	(b)	C (4, 4)	D (0, 5)	(c)	C (4, 4)	D (0, 4.5)
	D (6, 0)	D (1, 1)		D (5, 0)	D (1, 1)		D (4.5, 0)	D (1, 1)

4 Evolutionary stability

After having assessed the conditions under which the coexistence of two distinct resident populations is possible in Section 3, we study the evolutionary stability of the resulting diverse resident population against mutant invasions. In all this section, we consider the case of a constant assortment function and give the conditions for evolutionary stability in this case. Note that, by continuity, the assortment function is equal to the assortativity for all $\lambda \in [0, 1]$:

$$\forall \lambda \in [0, 1], \phi(\lambda) = \sigma \in [0, 1]$$

As we did for BNE we need to extend the definition of evolutionary stability to a resident population having more than one type:

Definition 9 (Evolutionary-stability). *In a mixed population made of two distinct types $\theta_1, \theta_2 \in \Theta$, with λ the share of θ_2 , the population $(\theta_1, \theta_2, \lambda)$ is evolutionary stable against a type $\tau \neq \theta_1, \theta_2 \in \Theta$ if:*

1. θ_1 and θ_2 satisfy the Payoff Equality condition when the mutant is absent: $\Pi_{\theta_1}(x^1, x^2, \lambda) = \Pi_{\theta_2}(x^1, x^2, \lambda)$ in all (Bayesian) Nash equilibria (x^1, x^2) in the population state $s = (\theta_1, \theta_2, \lambda)$;
2. There exists an $\bar{\varepsilon} > 0$ such that $\Pi_{\theta_1}(x^1, x^2, y, (\lambda, \varepsilon)) > \Pi_\tau(x^1, x^2, y, (\lambda, \varepsilon))$ and $\Pi_{\theta_2}(x^1, x^2, y, (\lambda, \varepsilon)) > \Pi_\tau(x^1, x^2, y, (\lambda, \varepsilon))$ in all Nash equilibria (x^1, x^2, y) in all states $s = (\theta_1, \theta_2, \tau, (\lambda, \varepsilon))$ with $\varepsilon \in (0, \bar{\varepsilon})$.

Moreover, θ_1 and θ_2 are evolutionary stable if they are evolutionary stable against all types $\tau \neq \theta_1, \theta_2 \in \Theta$.

We consider the case of finite symmetric 2×2 fitness games. Let A be the matrix of the payoffs in this game, with π_{ij} is the payoff when pure strategy i is played against pure strategy j . The payoff obtained by an individual playing strategy x when matched with an individual playing y is then:

$$\pi(x, y) = xAy \quad (14)$$

Let a population $s = (\theta_1, \theta_2, \tau, (\lambda, \epsilon))$, where θ_1 and θ_2 are the residents playing $x^1 \neq x^2$, and τ a mutant playing y such that $(x^1, x^2, y) \in X^3$ is a (Bayesian) Nash equilibrium, i.e. (3) is satisfied. Since we are in a two-strategies game, we can express the strategy y in function of the strategies x^1 and x^2 . For this purpose, let $\alpha_1, \alpha_2, \in [0, 1]$ be the probabilities that θ_1, θ_2 attach to the first pure strategy: $x^1 = (\alpha_1, 1 - \alpha_1)$ and $x^2 = (\alpha_2, 1 - \alpha_2)$. Then, there exists $\gamma \in \mathbb{R}$ such that $y = \gamma x^1 + (1 - \gamma)x^2 = (\gamma\alpha_1 + (1 - \gamma)\alpha_2, 1 - \gamma\alpha_1 - (1 - \gamma)\alpha_2)$ ⁴.

Suppose that the population $s = (\theta_1, \theta_2, \lambda)$ satisfies the Payoff Equality condition: $\Pi_{\theta_1}(x^1, x^2, \lambda) = \Pi_{\theta_2}(x^1, x^2, \lambda) = \Pi_{\theta}$. Then, we can write the differences in payoff between the residents and the mutant at the limit:

Lemma 2 (Difference in payoffs). *For a population $s = (\theta_1, \theta_2, \lambda)$ engaged in a symmetric 2×2 fitness game and satisfying the Payoff Equality condition, the difference in payoffs between the residents and a mutant τ at the limit (when ϵ converges to zero) is:*

$$\Pi_{\theta} - \Pi_{\tau} = \sigma\gamma(1 - \gamma)(\alpha_1 - \alpha_2)^2 (\pi_{11} + \pi_{22} - \pi_{12} - \pi_{21})$$

Proof. We can write the payoffs obtained by each type. Using the notation $\pi(y, y) = \pi^{33}$:

$$\begin{cases} \Pi_{\theta_1}(x^1, x^2, y, (\lambda, 0)) = \Pi_{\theta_1} = (1 - \lambda + \lambda\sigma) \cdot \pi^{11} + \lambda(1 - \sigma) \cdot \pi^{12} \\ \Pi_{\theta_2}(x^1, x^2, y, (\lambda, 0)) = \Pi_{\theta_2} = (1 - \lambda)(1 - \sigma) \cdot \pi^{21} + (\lambda + (1 - \lambda)\sigma) \cdot \pi^{22} \\ \Pi_{\tau}(x^1, x^2, y, (\lambda, 0)) = \Pi_{\tau} = (1 - \lambda)(1 - \sigma) \cdot \pi^{31} + \lambda(1 - \sigma) \cdot \pi^{32} + \sigma \cdot \pi^{33} \end{cases}$$

Since we are in a symmetric and finite fitness game, we can rewrite the payoffs in function of the matrix payoff A using (14):

$$\begin{cases} \Pi_{\theta_1} = x^1 [(1 - \lambda)(1 - \sigma)Ax^1 + \lambda(1 - \sigma)Ax^2] + \sigma x^1 Ax^1 \\ \Pi_{\theta_2} = x^2 [(1 - \lambda)(1 - \sigma)Ax^1 + \lambda(1 - \sigma)Ax^2] + \sigma x^2 Ax^2 \\ \Pi_{\tau} = y [(1 - \lambda)(1 - \sigma)Ax^1 + \lambda(1 - \sigma)Ax^2] + \sigma y Ay \end{cases}$$

From Payoff Equality, we know that $\Pi_{\theta_1} = \Pi_{\theta_2} = \Pi_{\theta}$. Thus, $\gamma\Pi_{\theta_1} + (1 - \gamma)\Pi_{\theta_2} = \Pi_{\theta}$ with $\gamma \in \mathbb{R}$ such that $y = \gamma x^1 + (1 - \gamma)x^2$. We can then write the difference between the payoff of the residents and the payoff of the mutant as follows:

$$\begin{aligned} \Pi_{\theta} - \Pi_{\tau} &= \gamma\Pi_{\theta_1} + (1 - \gamma)\Pi_{\theta_2} - \Pi_{\tau} \\ &= \sigma (\gamma x^1 Ax^1 + (1 - \gamma)x^2 Ax^2 - yAy) \\ &= \sigma\gamma(1 - \gamma) (x^1 Ax^1 + x^2 Ax^2 - x^1 Ax^2 - x^2 Ax^1) \end{aligned}$$

⁴ Note that the only case where this decomposition does not exist is when $\alpha_1 = \alpha_2$, and this case is not possible in our analysis since we have $x^1 \neq x^2$.

We can further develop this expression, using the pure-strategies payoffs:

$$\begin{cases} x^1 Ax^1 = \alpha_1^2 \pi_{11} + \alpha_1(1 - \alpha_1)(\pi_{21} + \pi_{12}) + (1 - \alpha_1)^2 \pi_{22} \\ x^2 Ax^2 = \alpha_2^2 \pi_{11} + \alpha_2(1 - \alpha_2)(\pi_{21} + \pi_{12}) + (1 - \alpha_2)^2 \pi_{22} \\ x^1 Ax^2 = \alpha_1 \alpha_2 \pi_{11} + \alpha_1(1 - \alpha_2)\pi_{12} + (1 - \alpha_1)\alpha_2 \pi_{21} + (1 - \alpha_1)(1 - \alpha_2)\pi_{22} \\ x^2 Ax^1 = \alpha_1 \alpha_2 \pi_{11} + \alpha_2(1 - \alpha_1)\pi_{12} + (1 - \alpha_2)\alpha_1 \pi_{21} + (1 - \alpha_2)(1 - \alpha_1)\pi_{22} \end{cases} \quad (15)$$

Rearranging, the difference in payoff is:

$$\Pi_\theta - \Pi_\tau = \sigma\gamma(1 - \gamma)(\alpha_1 - \alpha_2)^2 (\pi_{11} + \pi_{22} - \pi_{12} - \pi_{21})$$

■

When studying the sign of this difference in payoffs, it is useful to understand what is the sign of $\gamma(1 - \gamma)$. Without loss of generality and by symmetry we can assume $\alpha_2 < \alpha_1$, i.e. θ_1 plays the pure strategy 1 with a greater probability than θ_2 . Let η be the probability that τ attaches to pure strategy 1, $\eta = \gamma\alpha_1 + (1 - \gamma)\alpha_2$. If $\eta \in (\alpha_2, \alpha_1)$, then $\gamma \in (0, 1)$ and $\gamma(1 - \gamma) > 0$. If $\eta = \alpha_1$ or $\eta = \alpha_2$, then $\gamma(1 - \gamma) = 0$. Else $\gamma(1 - \gamma) < 0$ (see Fig. 3).

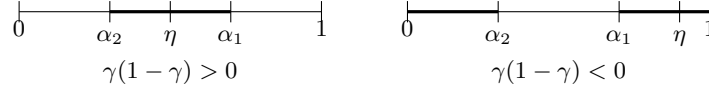


Fig. 3. Sign of $\gamma(1 - \gamma)$ depending on probabilities attached to the first pure strategy

Before studying when a population of two residents is evolutionary stable, we need to introduce one of the results of Alger and Weibull (2013) for symmetric 2×2 fitness game.

Lemma 3 (Proposition 2 of Alger and Weibull (2013)). *Let*

$$\hat{x}(\sigma) = \min \left\{ 1, \frac{\pi_{12} + \sigma\pi_{21} - (1 + \sigma)\pi_{22}}{(1 + \sigma)(\pi_{12} + \pi_{21} - \pi_{11} - \pi_{22})} \right\}$$

When $\sigma > 0$,

1. If $\pi_{11} + \pi_{22} - \pi_{12} - \pi_{21} > 0$, then $X_\sigma \subseteq \{0, 1\}$.
2. If $\pi_{11} + \pi_{22} - \pi_{12} - \pi_{21} = 0$, then

$$X_\sigma = \begin{cases} \{0\}, & \text{if } \pi_{12} + \sigma\pi_{21} - (1 + \sigma)\pi_{22} < 0 \\ [0, 1], & \text{if } \pi_{12} + \sigma\pi_{21} - (1 + \sigma)\pi_{22} = 0 \\ \{1\}, & \text{if } \pi_{12} + \sigma\pi_{21} - (1 + \sigma)\pi_{22} > 0 \end{cases}$$

3. If $\pi_{11} + \pi_{22} - \pi_{12} - \pi_{21} < 0$, then

$$X_\sigma = \begin{cases} \{0\}, & \text{if } \pi_{12} + \sigma\pi_{21} - (1 + \sigma)\pi_{22} \leq 0 \\ \{\hat{x}(\sigma)\}, & \text{if } \pi_{12} + \sigma\pi_{21} - (1 + \sigma)\pi_{22} > 0 \end{cases}$$

We omit in what follows the case where $\sigma = 0$ (uniform random matching) well-studied in the literature. The following proposition exhibits the link between Hamiltonian strategies and evolutionary stability of a population of two residents:

Proposition 2 (Evolutionary stability in a symmetric 2×2 fitness game).

In a symmetric 2×2 fitness game where the assortment is constant and strictly positive, let $(\theta_1, \theta_2, \lambda)$ be a population where θ_1 plays x^1 , θ_2 plays x^2 , such that $(x^1, x^2) \in X_1 \times X_2$.

If X_σ is a singleton and if Θ is rich, then there does not exist any evolutionary stable population made of two residents.

Proof. First, note that if the population does not respect the Payoff Equality condition, it is not evolutionary stable. Thus, we consider next a population that respects the Payoff Equality condition (Proposition 1). We know from Lemma 2 that the difference in payoffs between the residents and a mutant τ at the limit satisfies:

$$\Pi_\theta - \Pi_\tau = \sigma\gamma(1 - \gamma)(\alpha_1 - \alpha_2)^2 (\pi_{11} + \pi_{22} - \pi_{12} - \pi_{21})$$

We can show using (15) that $S_\pi = (\alpha_1 - \alpha_2)^2 (\pi_{11} + \pi_{22} - \pi_{12} - \pi_{21})$. Hence, rewriting the expression above, we have:

$$\Pi_\theta - \Pi_\tau = \sigma\gamma(1 - \gamma)S_\pi$$

We consider the three different cases in which X_σ is a singleton using Lemma 3:

1. If $\pi_{11} + \pi_{22} - \pi_{12} - \pi_{21} > 0$, $X_\sigma = \{0\}$ or $X_\sigma = \{1\}$,

Since Θ is rich, if θ_1 or θ_2 does not play pure strategies, it is always possible to find a mutant such that $\gamma(1 - \gamma) < 0$ (discussion of Fig.3). In this case the difference between the two payoff above is negative and the mutant earns more than the residents at the limit since $\sigma > 0$.

Else, if θ_1 and θ_2 both play pure strategies, then one of them is playing the Hamiltonian strategy. Without loss of generality and by symmetry, suppose θ_1 is playing this Hamiltonian strategy, and that $X_\sigma = \{1\}$ i.e. homo-hamiltonensis and θ_1 play the first pure strategy while θ_2 plays the second pure strategy. We are then in case 2. or 3. of Proposition 1 and we have $Q_\pi, R_\pi \geq 0$ and $S_\pi > 0$. Let $x \in X$, such that $x \neq x^2$, i.e. $x = (\rho, 1 - \rho)$ with $\rho \in (0, 1]$. Then:

$$\begin{aligned} (1 - \sigma)\pi(x, x^2) + \sigma\pi(x, x) &= \pi_{22} - \rho R_\pi - \sigma\rho(1 - \rho)S_\pi \\ &< \pi_{22} \end{aligned}$$

Thus, for all x in X such that $x \neq x^2$, $u_\sigma(x, x^2) < u_\sigma(x^2, x^2)$. This means that the strategy played by θ_2 , i.e. the second pure strategy is also a Hamiltonian strategy and thus $X_\sigma = \{0, 1\}$ which contradicts the fact that X_σ is a singleton. Hence there is no diverse evolutionary stable resident population in this case.

2. If $\pi_{11} + \pi_{22} - \pi_{12} - \pi_{21} = 0$. Hence, we have $S_\pi = 0$ ($\alpha_1 \neq \alpha_2$ else the residents play the same strategy). Thus, from Proposition 1, we also have $Q_\pi = R_\pi = 0$. Subtracting, the expression $Q_\pi - S_\pi$, using (15), we find:

$$Q_\pi - S_\pi = (\alpha_1 - \alpha_2)[\alpha_2(1 + \sigma)(\pi_{11} + \pi_{22} - \pi_{12} - \pi_{21}) + (\pi_{12} + \sigma\pi_{21} - (1 + \sigma)\pi_{22})]$$

Hence, we have $\pi_{12} + \sigma\pi_{21} - (1 + \sigma)\pi_{22} = 0$. Therefore, case 2. of Lemma 3 implies that $X_\sigma = [0, 1]$ and is not a singleton.

3. If $\pi_{11} + \pi_{22} - \pi_{12} - \pi_{21} < 0$, then since Θ is rich, it is always possible to find a mutant such that $\gamma(1 - \gamma) > 0$ so that the mutant earns more than the residents at the limit since $\sigma > 0$.

So in the different cases when X_σ is a singleton and Θ is rich, we have shown that there exists a mutant that earns more than the residents at the limit. Using Lemma 1 we can say that by continuity of the payoffs this strict inequality holds in a neighborhood. ■

The proposition specifies cases when it is not possible to have a population of two residents that is not evolutionary stable. Is there a case where evolutionary stability for a population of two residents is actually possible? The answer is yes, when $X_\sigma = \{0, 1\}$ ⁵. Let Θ_i be the set of mutants τ that are behaviorally indistinguishable from a resident θ_i . We have the following theorem:

Theorem 3 (Evolutionary stability in a symmetric 2×2 fitness game).

In a symmetric 2×2 fitness game where the assortment is constant and strictly positive. Suppose $X_\sigma = \{0, 1\}$, and let $(\theta_1, \theta_2, \lambda)$ be a population where θ_1 plays x^1 , θ_2 plays x^2 , such that $(x^1, x^2) \in X_1 \times X_2 \subset X_\sigma^2$, ($x^1 \neq x^2$) and $\lambda = Q_\pi / ((1 - \sigma)S_\pi)$ the share of θ_2 in the population.

If $\beta_\sigma(x)$ is a singleton for all $x \in X_\sigma$, then the population $(\theta_1, \theta_2, \lambda)$ is evolutionary stable against all types $\tau \notin \Theta_1 \cup \Theta_2$.

Proof. First, the population $(\theta_1, \theta_2, \lambda)$ satisfies the Payoff Equality condition when the mutant is absent from Theorem 2 and we also have $S_\pi > 0$.

Then, from Lemma 2, we know that the difference in payoffs between the residents and a mutant τ at the limit is:

$$\Pi_\theta - \Pi_\tau = \sigma\gamma(1 - \gamma)(\alpha_1 - \alpha_2)^2 (\pi_{11} + \pi_{22} - \pi_{12} - \pi_{21})$$

Second, since $(x^1, x^2) \in X_1 \times X_2 \subset X_\sigma^2$ and $X_\sigma = \{0, 1\}$, it means that θ_1 and θ_2 will play the two pure strategies. Without loss of generality and by symmetry, we can assume that θ_1 plays the pure strategy 1 ($\alpha_1 = 1$), and that θ_2 plays the pure strategy 2 ($\alpha_2 = 0$). Thus, γ is actually the probability that τ attaches to the pure strategy 1. Moreover, since $\tau \notin \Theta_1 \cup \Theta_2$, the mutant

⁵ In the last possible case, $X_\sigma = [0, 1]$, any strategy gives the same payoff meaning that there does not exist a evolutionary stable population, even for single-type population

cannot play a pure strategy and $\gamma \in (0, 1)$ i.e. $\gamma(1 - \gamma) > 0$. We also have $S_\pi = \pi_{11} + \pi_{22} - \pi_{12} - \pi_{21} > 0$, and $\sigma > 0$.

Consequently, the difference in payoff at the limit is:

$$\Pi_\theta - \Pi_\tau = \sigma\gamma(1 - \gamma)S_\pi > 0$$

We have shown:

$$\begin{aligned} \Pi_{\theta_1}(x_1, x_2, y, (\lambda, 0)) &> \Pi_\tau(x_1, x_2, y, (\lambda, 0)) \\ \text{and } \Pi_{\theta_1}(x_1, x_2, y, (\lambda, 0)) &> \Pi_\tau(x_1, x_2, y, (\lambda, 0)) \end{aligned}$$

for all $(x_1, x_2) \in X_1 \times X_2$ in the population state $s = (\theta_1, \theta_2, \lambda)$ and for any $\tau \notin \Theta_1 \cup \Theta_2$.

We can then use Lemma 1, to conclude the proof by continuity. ■

This last theorem proves the existence of a heterogeneous and evolutionary stable population in the particular case of a constant assortment function. In the following section, we expose some examples and discuss the main implications.

5 Discussion and Examples

In this section, we first explore the question of evolutionary stability in the examples we studied in Section 3. Then we discuss the differences between a heterogeneous population and a population constituted by a single type of resident. And we finally discuss the possibility of testing empirically this diversity as well as potential future research strands.

5.1 Examples

Example 3 (Coordination game).

In the coordination game we considered in Section 3, we had $\pi_{AA} = 2$ and $\pi_{BB} = 1$. We saw that we had Payoff Equality when $\phi(\lambda) < 1/2$.

In the case of a constant assortment: $\phi(\lambda) = \sigma = 1/5 \forall \lambda \in [0, 1]$, $\lambda = 3/4$, and θ_1 and θ_2 get the same total payoff $\Pi = 0.8$. In this case we have $S_\pi = \pi_{AA} + \pi_{BB} = 3 > 0$. We therefore know that the difference in payoffs between the residents and the mutant is: $\Pi_\theta - \Pi_\tau = \sigma\gamma(1 - \gamma)S_\pi > 0 \forall \gamma \in (0, 1)$, and we can conclude using Theorem 3 that this population is evolutionary stable.

Table 5. Coordination game example

	A	B
A	(2, 2)	(0, 0)
B	(0, 0)	(1, 1)

Example 4 (Prisoner's dilemma).

In our examples of the prisoner's dilemma, we study the evolutionary stability of a population made of a homo-oeconomicus that always defects and a homo-kantiensis that always cooperates.

(a) First example when $\pi_{CD} = 0$, $\pi_{DD} = 1$, $\pi_{CC} = 4$ and $\pi_{DC} = 6$. With a constant assortment $\sigma = 1/3$, then with $\lambda = 1/2$ the populations satisfies the Payoff Equality condition. However, we have $S_\pi = -1 < 0$, and since the difference in payoffs between the residents and the mutant is: $\Pi_\theta - \Pi_\tau = \sigma\gamma(1 - \gamma)S_\pi$, any mutant would earn more than the residents! The population is not evolutionary stable.

(b) Now let $\pi_{CD} = 0$, $\pi_{DD} = 1$, $\pi_{CC} = 4$ and $\pi_{DC} = 5$, we have $S_\pi = 0$, $Q_\pi = -1 + 4\phi(\lambda)$ and $R_\pi = 1 - 4\phi(\lambda)$. With a constant assortment $\sigma = 1/4$, any $\lambda \in (0, 1)$ satisfies Payoff Equality. On the other hand, any mutant would also earn the same payoff. Actually this case contradicts the assumption that β_σ is a singleton since any strategy is a best reply.

(c) Finally, let $\pi_{CD} = 0$, $\pi_{DD} = 1$, $\pi_{CC} = 4$ and $\pi_{DC} = 4.5$. With a constant assortment $\sigma = 1/5$, then with $\lambda = 1/2$ the populations satisfies the Payoff Equality condition and $\Pi_\theta = 12/5$. Moreover, we have $S_\pi = 0.5 > 0$, and the difference in payoffs between the residents and the mutant is: $\Pi_\theta - \Pi_\tau = \sigma\gamma(1 - \gamma)S_\pi > 0 \forall \gamma \in (0, 1)$, i.e. this population is evolutionary stable.

Table 6. Prisoner's dilemma examples

(a)	<table style="border-collapse: collapse; text-align: center;"> <tr><td style="padding: 0 5px;"><i>C</i></td><td style="padding: 0 5px;"><i>D</i></td></tr> <tr><td style="border: 1px solid black; padding: 2px;">(4, 4)</td><td style="border: 1px solid black; padding: 2px;">(0, 6)</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">(6, 0)</td><td style="border: 1px solid black; padding: 2px;">(1, 1)</td></tr> </table>	<i>C</i>	<i>D</i>	(4, 4)	(0, 6)	(6, 0)	(1, 1)	(b)	<table style="border-collapse: collapse; text-align: center;"> <tr><td style="padding: 0 5px;"><i>C</i></td><td style="padding: 0 5px;"><i>D</i></td></tr> <tr><td style="border: 1px solid black; padding: 2px;">(4, 4)</td><td style="border: 1px solid black; padding: 2px;">(0, 5)</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">(5, 0)</td><td style="border: 1px solid black; padding: 2px;">(1, 1)</td></tr> </table>	<i>C</i>	<i>D</i>	(4, 4)	(0, 5)	(5, 0)	(1, 1)	(c)	<table style="border-collapse: collapse; text-align: center;"> <tr><td style="padding: 0 5px;"><i>C</i></td><td style="padding: 0 5px;"><i>D</i></td></tr> <tr><td style="border: 1px solid black; padding: 2px;">(4, 4)</td><td style="border: 1px solid black; padding: 2px;">(0, 4.5)</td></tr> <tr><td style="border: 1px solid black; padding: 2px;">(4.5, 0)</td><td style="border: 1px solid black; padding: 2px;">(1, 1)</td></tr> </table>	<i>C</i>	<i>D</i>	(4, 4)	(0, 4.5)	(4.5, 0)	(1, 1)
<i>C</i>	<i>D</i>																						
(4, 4)	(0, 6)																						
(6, 0)	(1, 1)																						
<i>C</i>	<i>D</i>																						
(4, 4)	(0, 5)																						
(5, 0)	(1, 1)																						
<i>C</i>	<i>D</i>																						
(4, 4)	(0, 4.5)																						
(4.5, 0)	(1, 1)																						

In the general case, there will exist a diverse evolutionary-stable population in the prisoner dilemma if $\pi_{CC} + \pi_{DD} - \pi_{DC} - \pi_{CD} > 0$. If this condition is met, then the evolutionary-stable population will be a mix of homo-oeconomicus and homo-kantiensis. Note that this result is not the same as in Alger and Weibull (2013). Indeed, in this case ($\pi_{CC} + \pi_{DD} - \pi_{DC} - \pi_{CD} > 0$), homo-hamiltonensis would either cooperate or defect, i.e. everybody behaves as a homo-oeconomicus and defects, or everybody behaves as a homo-kantiensis and cooperates. On the other hand, we found a diverse evolutionary-stable population with a share behaving as homo-oeconomicus and the other as homo-kantiensis.

5.2 Heterogenous population vs single-type Homo-hamiltonensis resident

Using the frame of evolutionary stability formally established by Maynard Smith and Price (1973), Alger and Weibull proved the evolutionary stability of a particular type of preference, homo-hamiltonensis. Although our paper also proves the evolutionary superiority of Hamiltonian strategies in the case of a constant

assortment function, it is important to note that our result has different implications. While, in the case of a single resident population, all the population necessarily plays the same strategy thus imposing a unique expected played strategy, we show that it is possible for a population to exhibit different behaviors and different types of strategies played without infringing the evolutionary stability. It is worthwhile insisting here on the importance of assortative matching in making this diversity possible. As can be seen in example 4 above, it is because the assortment function is strictly positive that homo-kantiensis can survive without being eliminated by the homo-oeconomicus present in the resident population.

As first expectation about a diverse population exhibiting an evolutionary stability against mutants, we could have hypothesized the fact that the population would "on average" have a Hamiltonian preference in order to be coherent with the main result of Alger and Weibull (2013). In other words, an apparently good candidate diverse population would have been a population composed by fully-selfish and fully-moral individuals with a share σ of fully moral individuals in order to be "coherent" with the Hamiltonian preference. However, as we showed in Section 4 this case cannot be evolutionary stable whereas some other can be. Consequently, our case of a diverse resident population is also different from a homo-hamiltonensis playing a mixed strategy. In fact, in his thesis John Nash exposed a second interpretation of mixed Nash equilibria (also called mass-action), which would consist in large populations playing the pure strategies composing the mixed equilibrium rather than a single player randomizing his play (Leonard (1994)). The heterogeneous evolutionary stable population we find above could have been an application of this case if we had a diversity that mirrors a mixed Hamiltonian strategy. However, we showed in Section 4 that it is possible to have a diversity of resident populations even when homo-hamiltonensis plays only pure strategies.

5.3 Empirical testing and further research

While empirical behavioral research often aims at finding the parameters of the preferences of individuals, it would be an interesting challenge to try to estimate the diversity of preference in a population considering the possibility of the presence of several types of preferences in the analysis. We expect allowing for diversity in the preferences of individuals to give a more precise estimation of what is driving the strategic behavior of agents in assortative symmetric (fitness) games. This paper aims at opening the way for analyses accounting for a potential diversity of preferences. Many extensions and improvements can be undertaken to deepen the understanding of heterogeneous populations. First, exploring the possibility of a non-constant assortment function will offer a better understanding of the role assortment plays in allowing for the diversity of preferences and might imply the presence of evolutionary stable preference not playing Hamiltonian strategies. Second, the assortative matching could be extended to more than three agents and it would be interesting identifies some patterns allowing for a generalization of assortative games. Finally, the case of

evolutionary stability of diverse populations could be extended to finite games with more than two pure strategies and to infinite games.

6 Conclusion

Building on the analysis of Alger and Weibull (2013) which introduced the Homo-moralis preference type and proved its evolutionary stability, this paper extends the scope of analysis of evolutionary stability in the context of assortative interaction with imperfect information. Introducing a tripartite assortment function and the Payoff Equality condition in the case of two competing resident populations, we prove the existence of diverse and evolutionary stable equilibria. Using the case of constant assortment function we prove the possibility of co-existence of two different types in the population being together evolutionary stable. This paper aims at opening the way towards better consideration of the diversity of preferences moving away from the more classical use of representative agents and homogeneous populations.

References

1. Alger, Ingela and Weibull, Jörgen W: "Homo-moralis preference evolution under incomplete information and assortative matching." *Econometrica* 81 (6), 2269–2302 (2013)
2. Alger, Ingela and Weibull, Jörgen W: "Evolution and Kantian morality." *Games and Economic Behavior* 98 (2016): 56-67.
3. Arrow, K. (1973): "Social Responsibility and Economic Efficiency," *Public Policy*,
4. Bergstrom, T. (2003): "The Algebra of Assortative Encounters and the Evolution of Cooperation," *International Game Theory Review*, 5, 211-228.
5. Dekel, E., J. C. Ely, and O. Yilankaya (2007): "Evolution of Preferences," *Review of Economic Studies*, 74, 685-704.
6. Güth, W, and M. Yaari (1992): "An Evolutionary Approach to Explain Reciprocal Behavior in a Simple Strategic Game," in *Explaining Process and Change—Approaches to Evolutionary Economics*, ed. by U. Witt. Ann Arbor: University of Michigan Press.
7. Laffont, J.-J. (1975): "Macroeconomic Constraints, Economic Efficiency and Ethics: An introduction to Kantian Economics," *Económica*, 42, 430-437.
8. Leonard. R. 1994, Reading Cournot, reading Nash: The creation and stabilization of the Nash equilibrium. *Economic Journal* 104: 492 - 511
9. Maynard Smith, J., and G. R. Price (1973): "The Logic of Animal Conflict," *Nature*, 246, 15-18.
10. Robson, A. J. (1990): "Efficiency in Evolutionary Games: Darwin, Nash and the Secret Hand shake," *Journal of Theoretical Biology*, 144, 379-396.
11. Smith, A. (1759): *The Theory of Moral Sentiments*, ed. by D. D. Raphael and A. L. Macfie. Oxford: Clarendon Press; New York: Oxford University Press (1976).
12. Sen, A. K. (1977): "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory," *Philosophy & Public Affairs*, 6, 317-344.
13. Tabellini, G. (2008): "Institutions and Culture," *Journal of the European* 6, 255-294.