



THE UNIVERSITY OF  
**SYDNEY**

Economics Working Paper Series

2014 - 14

Comment on Homo Moralis: when assortativity  
evolves

Jonathan Newton

September 2014

# Comment on Homo Moralis: when assortativity evolves

Jonathan Newton<sup>a,1</sup>

<sup>a</sup>*School of Economics, University of Sydney.*

---

## Abstract

Differing degrees of assortativity in matching can be expected to have both genetic and cultural determinants. When assortativity is subject to evolution, the main result of [Alger and Weibull \(2013\)](#) on the evolution of stable other-regarding preferences does not hold. Instead, both non-Nash and Pareto inefficient behavior are evolutionarily unstable.

*Keywords:* evolution, moral values, assortative matching

**JEL Classification Numbers:** C73.

---

## 1. Introduction

[Alger and Weibull \(2013\)](#) show that, under an exogenously given matching protocol, a population consisting of types whose behavior is determined by a particular utility function, *homo hamiltonensis*, is robust to invasion by types (genotypes) whose behavior (phenotype) differs from that of *homo hamiltonensis*. The level of other-regarding behavior exhibited by *homo hamiltonensis* depends directly on the level of assortativity in matching shown by small populations of invading mutants. Specifically, in a population of whom a proportion  $1 - \varepsilon$  are *homo hamiltonensis* and a proportion  $\varepsilon$  are some invading type  $\tau$ , the level of other-regarding behavior by *homo hamiltonensis* is given by  $\sigma := \lim_{\varepsilon \rightarrow 0} Pr[\tau|\tau, \varepsilon]$ , where  $Pr[\tau|\tau, \varepsilon]$  is the probability that an agent matches with a  $\tau$ -type given that he himself is of type  $\tau$  and that there are  $\varepsilon$   $\tau$ -types in

---

<sup>1</sup>Come rain or come shine, I can be reached at [jonathan.newton@sydney.edu.au](mailto:jonathan.newton@sydney.edu.au), telephone +61293514429. This work was completed while the author was supported by a Discovery Early Career Researcher Award funded by the Australian Research Council.

the population. That is, *homo hamiltonensis*' behavior depends directly on the behavior, as manifested via the degree of assortativity, of an invading type.

It is assumed by [Alger and Weibull \(2013\)](#) that the degree of assortativity is type independent. That is,  $\sigma$  does not depend on  $\tau$ . This is a very strong assumption, whether  $Pr[\tau|\tau, \varepsilon]$  is considered to be biologically or culturally determined. In fact, a deep and interesting literature exists that looks at the evolution of assortative behavior, in which mutants can exhibit higher or lower degrees of assortativity.<sup>2</sup> In addition, factors that indirectly lead to greater or lesser assortativity, such as the predilection to roam far from home or habitat location and size, are subject to evolutionary pressures.<sup>3</sup> Cultural determinants of assortativity also differ as social groups vary in degree of hostility to outsiders and openness to external influence.<sup>4</sup>

Considering the above, it is important to include assortativity in the possible behaviors determined by evolution. That is, to consider  $\sigma_\tau = \lim_{\varepsilon \rightarrow 0} Pr[\tau|\tau, \varepsilon]$ , where  $Pr[\tau|\tau, \varepsilon]$  depends on  $\tau$ . Following this change, the predictions of [Alger and Weibull \(2013\)](#) no longer hold. If there exists a *selfish rover* type,  $\tau_r$ , whose strategic behavior is determined solely by individual fitness considerations, and for whom  $\sigma_{\tau_r} = 0$ , then non-Nash equilibrium behavior is evolutionarily unstable. Moreover, if there exists a *Kantian parochial* type,  $\tau_p$ , whose strategic behavior maximizes fitness from symmetric strategy profiles, and for whom  $\sigma_{\tau_p} = 1$ , then Pareto inefficient behavior is also evolutionarily unstable.

## 2. Model and result

Consider a population whose individuals are randomly matched into pairs to engage in a symmetric interaction with the common strategy set  $X$ . An individual playing strategy  $x$  against an individual playing strategy  $y$  receives

---

<sup>2</sup>See, for example, [Cara et al. \(2008\)](#); [Dieckmann and Doebeli \(1999\)](#); [Matessi et al. \(2002\)](#); [Otto et al. \(2008\)](#); [Pennings et al. \(2008\)](#); [Servedio \(2010\)](#).

<sup>3</sup>[Bearhop et al. \(2005\)](#); [Dyson-Hudson and Smith \(1978\)](#); [López-Sepulcre and Kokko \(2005\)](#).

<sup>4</sup>[Cashdan \(2001\)](#); [Choi and Bowles \(2007\)](#); [Fry and Söderberg \(2013\)](#).

a payoff, representing biological fitness,  $\pi(x, y)$ , where  $\pi : X^2 \rightarrow \mathbb{R}$ . The pair  $\langle X, \pi \rangle$  is the *fitness game*.  $X$  is a nonempty, compact and convex set in a topological vector space and  $\pi$  is continuous. Each individual is characterized by a type  $\theta \in \Theta$  which defines a continuous utility function  $u_\theta : X^2 \rightarrow \mathbb{R}$  and an index of assortativity  $\sigma_\theta \in [0, 1]$ . An individual's type is his private information.

Consider a population with two types present and define a *population state*  $s = (\theta, \tau, \varepsilon)$ , where  $\theta, \tau \in \Theta$  are the two types and  $\varepsilon \in (0, 1)$  is the population share of type  $\tau$ . The random matching process is such that, with probability  $Pr[\tau|\theta, \varepsilon]$  a given individual of type  $\theta$  is matched with an individual of type  $\tau$ , and with probability  $Pr[\theta|\tau, \varepsilon]$  a given individual of type  $\tau$  is matched with an individual of type  $\theta$ . Let  $Pr[\tau|\tau, \varepsilon]$  satisfy  $\lim_{\varepsilon \rightarrow 0} Pr[\tau|\tau, \varepsilon] \rightarrow \sigma_\tau$ . The balancing condition for heterogeneous matchings then implies that  $\lim_{\varepsilon \rightarrow 0} Pr[\theta|\theta, \varepsilon] \rightarrow 1$ .

For a state  $s = (\theta, \tau, \varepsilon)$ , strategies  $x \in X$  used by type  $\theta$  and  $y \in X$  used by type  $\tau$ , the average fitness of each type is

$$\Pi_\theta(x, y, \varepsilon) = Pr[\theta|\theta, \varepsilon] \cdot \pi(x, x) + Pr[\tau|\theta, \varepsilon] \cdot \pi(x, y), \quad (1)$$

$$\Pi_\tau(x, y, \varepsilon) = Pr[\theta|\tau, \varepsilon] \cdot \pi(y, x) + Pr[\tau|\tau, \varepsilon] \cdot \pi(y, y). \quad (2)$$

It is assumed that the strategies chosen by individuals of both types are a (Bayesian) Nash equilibrium.

**Definition 2.1.** In any state  $s = (\theta, \tau, \varepsilon)$ , a strategy pair  $(x^*, y^*) \in X^2$  is a (Bayesian) Nash Equilibrium,  $(x^*, y^*) \in B^{\text{NE}}(s)$ , if

$$\begin{cases} x^* \in \arg \max_{x \in X} Pr[\theta|\theta, \varepsilon] \cdot u_\theta(x, x^*) + Pr[\tau|\theta, \varepsilon] \cdot u_\theta(x, y^*), \\ y^* \in \arg \max_{y \in X} Pr[\theta|\tau, \varepsilon] \cdot u_\tau(y, x^*) + Pr[\tau|\tau, \varepsilon] \cdot u_\tau(y, y^*). \end{cases} \quad (3)$$

This definition defines, for fixed types  $\theta, \tau$ , an equilibrium correspondence  $B^{\text{NE}}(\theta, \tau, \cdot) : (0, 1) \rightrightarrows X^2$  that maps mutant population shares to equilibria. Letting  $Pr[\cdot|\cdot, 0] = \lim_{\varepsilon \rightarrow 0} Pr[\cdot|\cdot, \varepsilon]$ , the domain of  $B^{\text{NE}}(\theta, \tau, \cdot)$  can be extended to  $[0, 1)$ .

The same definition of evolutionary instability as [Alger and Weibull \(2013\)](#) is used.

**Definition 2.2.** A type  $\theta \in \Theta$  is *evolutionarily unstable* if there exists a type  $\tau \in \Theta$  and  $\bar{\varepsilon} > 0$  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  $\varepsilon \in (0, \bar{\varepsilon})$ .

Two types are now defined, the first of which will guarantee that non-Nash behavior is unstable, the second of which will guarantee that Pareto inefficient behavior is unstable. Note that, replacing a universal value of  $\sigma$  by type specific  $\sigma_\tau$ , both of these types are varieties of *homo hamiltonensis* (Alger and Weibull, 2013).

**Definition 2.3.**

The *selfish rover* type  $\tau_r$  satisfies  $u_{\tau_r}(x, y) = \pi(x, y)$ ;  $\sigma_{\tau_r} = 0$ .

The *Kantian parochial* type  $\tau_p$  satisfies  $u_{\tau_p}(x, y) = \pi(x, x)$ ;  $\sigma_{\tau_p} = 1$ .

For each type  $\theta \in \Theta$ , let  $\beta_\theta : X \rightrightarrows X$  denote the best response correspondence,  $\beta_\theta(y) = \arg \max_{x \in X} u_\theta(x, y) \forall y \in X$ , and  $X_\theta \subseteq X$  the set of fixed points under  $\beta_\theta$ ,

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

Note that  $X_{\tau_r}$  corresponds to the set of symmetric Nash equilibria when selfish individuals maximize their own fitness. In contrast,  $X_{\tau_p}$  corresponds to the set of Pareto efficient symmetric strategy profiles.

**Theorem 2.1.** *If  $X_\theta \cap X_{\tau_r} = \emptyset$  and  $\tau_r \in \Theta$ , then  $\theta$  is evolutionarily unstable.*

*If  $X_\theta \cap X_{\tau_p} = \emptyset$  and  $\tau_p \in \Theta$ , then  $\theta$  is evolutionarily unstable.*

*Proof.* Let  $s = (\theta, \tau_r, \varepsilon)$ ,  $(x^*, y^*) \in B^{\text{NE}}(\theta, \tau_r, 0)$ . Note that  $Pr[\theta|\theta, 0] = 1$ ,  $Pr[\tau_r|\tau_r, 0] = 0$ . It follows that

$$x^* \in \arg \max_{x \in X} u_\theta(x, x^*) \tag{4}$$

which implies  $x^* \in X_\theta$ . Also,

$$y^* \in \arg \max_{y \in X} u_{\tau_r}(y, x^*) = \arg \max_{y \in X} \pi(y, x^*). \tag{5}$$

If  $x^* \in \arg \max_{y \in X} \pi(y, x^*)$ , then  $x^* \in X_{\tau_r}$ , contradicting  $X_\theta \cap X_{\tau_r} = \emptyset$ . Therefore,  $x^* \notin \arg \max_{y \in X} \pi(y, x^*)$ , which implies that  $\pi(y^*, x^*) > \pi(x^*, x^*)$ , hence

$\Pi_{\tau_r}(x^*, y^*, 0) > \Pi_{\theta}(x^*, y^*, 0)$ , and the first part of the theorem follows from the continuity argument of [Alger and Weibull \(2013, Proof of Theorem 1\)](#).

Now, let  $s = (\theta, \tau_p, \varepsilon)$ ,  $(x^*, y^*) \in B^{\text{NE}}(\theta, \tau_p, 0)$ . Note that  $Pr[\theta|\theta, 0] = 1$ ,  $Pr[\tau_p|\tau_p, 0] = 1$ . (4) continues to hold so  $x^* \in X_{\theta}$ . Now,

$$y^* \in \arg \max_{y \in X} u_{\tau_r}(y, y^*) = \arg \max_{y \in X} \pi(y, y^*). \quad (6)$$

If  $x^* \in \arg \max_{y \in X} \pi(y, y)$ , then  $x^* \in X_{\tau_p}$ , contradicting  $X_{\theta} \cap X_{\tau_p} = \emptyset$ . Therefore,  $x^* \notin \arg \max_{y \in X} \pi(y, y)$ , which implies that  $\pi(y^*, y^*) > \pi(x^*, x^*)$ , hence  $\Pi_{\tau_p}(x^*, y^*, 0) > \Pi_{\theta}(x^*, y^*, 0)$ , and the second part of the theorem follows from the continuity argument of [Alger and Weibull \(2013, Proof of Theorem 1\)](#).  $\square$

### 3. Discussion

Theorem 2.1 shows that when both assortativity and preferences evolve, neither non-Nash behavior nor inefficient behavior can persist indefinitely in the long run. This is not particularly troublesome: the world, after all, is a dynamic and changing place. However, it is relevant to ask under what conditions the results of [Alger and Weibull \(2013\)](#) will pertain. It is clear that a necessary condition is that changes in assortativity take place on a much longer timescale than changes in preferences.<sup>5</sup> Unfortunately, such situations are difficult to conceive and would not seem to be common. The examples in the paper under discussion do not help here. The example of “Kin” (p.2286, op.cit.) is uncontroversial but not germane as past behavior of close relatives would likely be observable, from which their (pheno)type could be inferred. Besides, in very small groups, there is a significant chance of a mutation attaining fixation via genetic drift, regardless of the direction of selection. The example of “Geography, Homophily and Business Partnerships” (p.2287, op.cit.) is one where assortativity (as determined by the probability,  $p(n, N)$ , of matching within the same group, p.2288,

---

<sup>5</sup>This is not required for similar work ([Wilson and Dugatkin, 1997](#)) where behavioral type is observed and therefore individuals can intentionally assort by type.

op.cit.) would be expected to vary and be subject to selection on a similar timescale to preferences.

Finally, note that even if assortativity and preferences are determined at different genetic loci and simultaneous mutation is rare, the implications of Theorem 2.1 still hold. Given any fixed choice behavior in a population, there is no selection for or against different degrees of assortativity, so genetic drift will create clusters of rovers and clusters of parochials, thus providing hospitable environments for the invasion of selfish or Kantian behavior.

Alger, I., Weibull, J.W., 2013. Homo moralis—preference evolution under incomplete information and assortative matching. *Econometrica* 81, 2269–2302.

Bearhop, S., Fiedler, W., Furness, R.W., Votier, S.C., Waldron, S., Newton, J., Bowen, G.J., Berthold, P., Farnsworth, K., 2005. Assortative mating as a mechanism for rapid evolution of a migratory divide. *Science* 310, 502–504.

Cara, M.A.R.d., Barton, N.H., Kirkpatrick, M., 2008. A model for the evolution of assortative mating. *The American Naturalist* 171, pp. 580–596.

Cashdan, E., 2001. Ethnocentrism and xenophobia: A crosscultural study. *Current Anthropology* 42, pp. 760–765.

Choi, J.K., Bowles, S., 2007. The coevolution of parochial altruism and war. *Science* 318, 636–640.

Dieckmann, U., Doebeli, M., 1999. On the origin of species by sympatric speciation. *Nature* 400, 354–357.

Dyson-Hudson, R., Smith, E.A., 1978. Human territoriality: An ecological reassessment. *American Anthropologist* 80, pp. 21–41.

Fry, D.P., Söderberg, P., 2013. Lethal aggression in mobile forager bands and implications for the origins of war. *Science* 341, 270–273.

López-Sepulcre, A., Kokko, H., 2005. Territorial defense, territory size, and population regulation. *The American Naturalist* 166, pp. 317–325.

- Matessi, C., Gimelfarb, A., Gavrilets, S., 2002. Long-term buildup of reproductive isolation promoted by disruptive selection: how far does it go? *Selection* 2, 41–64.
- Otto, S.P., Servedio, M.R., Nuismer, S.L., 2008. Frequency-dependent selection and the evolution of assortative mating. *Genetics* 179, 2091–2112.
- Pennings, P.S., Kopp, M., Meszéna, G., Dieckmann, U., Hermisson, J., 2008. An analytically tractable model for competitive speciation. *The American Naturalist* 171, E44–E71.
- Servedio, M.R., 2010. Limits to the evolution of assortative mating by female choice under restricted gene flow. *Proceedings of the Royal Society B: Biological Sciences* .
- Wilson, D.S., Dugatkin, L.A., 1997. Group selection and assortative interactions. *American Naturalist* , 336–351.