



**HAL**  
open science

# Rationalizability and Interactivity in Evolutionary OLG Models

Grégory Ponthière

► **To cite this version:**

Grégory Ponthière. Rationalizability and Interactivity in Evolutionary OLG Models. 2013. hal-00814699

**HAL Id: hal-00814699**

**<https://pjse.hal.science/hal-00814699>**

Preprint submitted on 17 Apr 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**PARIS SCHOOL OF ECONOMICS**  
ÉCOLE D'ÉCONOMIE DE PARIS

**WORKING PAPER N° 2013 – 09**

## **Rationalizability and Interactivity in Evolutionary OLG Models**

**Gregory Ponthiere**

**JEL Codes : D03, J10, O11, Z10**

**Keywords: Evolutionary OLG models; Survival; Extinction; Fitness ordering; Rationalization**



**PARIS-JOURDAN SCIENCES ÉCONOMIQUES**

48, Bd JOURDAN – E.N.S. – 75014 PARIS  
TÉL. : 33(0) 1 43 13 63 00 – FAX : 33 (0) 1 43 13 63 10  
[www.pse.ens.fr](http://www.pse.ens.fr)

# Rationalizability and Interactivity in Evolutionary OLG Models\*

Gregory Ponthiere<sup>†</sup>

April 15, 2013

## Abstract

We use the theory of rationalizable choices to study the survival and the extinction of types (or traits) in evolutionary OLG models. Two properties of evolutionary processes are introduced: rationalizability by a fitness ordering (i.e. only the most fit types survive) and interactivity (i.e. a withdrawal of types affects the survival of other types). Those properties are shown to be logically incompatible. We then examine whether the evolutionary processes at work in canonical evolutionary OLG models satisfy rationalizability or interactivity. We study  $n$ -types versions of the evolutionary OLG models of Galor and Moav (2002) and Bisin and Verdier (2001), and show that, while the evolutionary process at work in the former is generally rationalizable by a fitness ordering, the opposite is true for the latter, which exhibits, in general, interactivity.

*Keywords:* evolutionary OLG models; survival; extinction; fitness ordering; rationalization.

*JEL codes:* D03, J10, O11, Z10.

---

\*The author is grateful to Geir Asheim, Michael Caputo and to an anonymous referee for their helpful suggestions and comments on this paper.

<sup>†</sup>Paris School of Economics and Ecole Normale Supérieure, Paris. Address: ENS, Department of Social Sciences, 48 Boulevard Jourdan, building B, second floor, office B, 75014 Paris, France. E-mail: gregory.ponthiere@ens.fr Telephone: 0033-143136204. Fax: 0033-143136222.

# 1 Introduction

Whereas standard overlapping generations (OLG) models, as developed by Al-lais (1947), Samuelson (1958) and Diamond (1965), allow, by construction, for age-heterogeneity within the population alive at a given time, growth theory has recently paid particular attention to evolutionary OLG models, which exhibit not only inter-cohort heterogeneity, but also intra-cohort heterogeneity.

Evolutionary OLG models describe the dynamics of intra-cohort heterogeneity in terms of the selection mechanism. The precise form of that selection process depends on the economic environment considered, and, in particular, on how types or traits are transmitted across generations. In those models, types or traits are transmitted through various transmission mechanisms, such as (effortless) hereditary transmission (Galor and Moav 2002, 2005), or transmission through parental socialization effort (Bisin and Verdier 2000, 2001). Evolutionary OLG models allow, thanks to intra-cohort heterogeneity, for the replication of the complex dynamics exhibited by *aggregated* variables. Those models were applied to various issues, such as the shift of growth regime (Galor and Moav 2002), the demographic transition (Galor and Moav 2005), the dynamics of religions and marriage (Bisin *et al* 2004), and the evolution of globalization and trade (Olivier *et al* 2008).

Whereas evolutionary OLG models have become increasingly used in (long-run) macroeconomics, little attention has been paid so far to the formal properties of the evolutionary processes that are present in those models. The goal of this paper is precisely to propose an exploration of the properties characterizing the evolutionary mechanisms at work in evolutionary OLG models.

For that purpose, we develop a general framework to study the survival and the extinction of types (or traits) within a population. We consider an evolutionary process as a survival or selection process, that is, a process that maps finite sets of types into non-empty subsets of types.

Although considering an evolutionary process as a selection process involves a simplification, that reduction allows us to use the analytical tools of rationalizable choice theory to cast new light on evolutionary processes.<sup>1</sup> Once an evolutionary process is reduced to a selection out of a set of types, one can ask whether there exists or not a *fitness ordering* defined on types, which could rationalize the observed selection. An evolutionary process is rationalizable by a fitness ordering when the selection is such that only the most fit types survive.

Rationalizability is an appealing concept to economists, because rational choice theory is the basis of modern microeconomic theory. When considering the selection of consumption baskets by an agent, economists study the question of its rationalizability by a preference ordering, in line with the revealed

---

<sup>1</sup>The study of evolutionary processes as survival or selection processes constitutes a simplification, because it focuses on the identification of the surviving types, and thus ignores the proportions in which the different types survive. However, one can, following Parfit (1984, p. 453-454), argue that, from the point of view of history, the most relevant difference is not the one between a high or a low prevalence, but, rather, the difference between survival and extinction.

preferences approach pioneered by Samuelson (1938, 1948).<sup>2</sup> In a similar way, when considering evolutionary OLG models where some types or traits are selected over time, it is tempting to investigate whether the associated selection process is rationalizable by a fitness ordering or not. If an evolutionary process is rationalizable by a fitness ordering, then that economy can be interpreted as an environment where only the "most fit" types survive. If, on the contrary, the evolutionary process is not rationalizable, then it does not make sense to say that "only the most fit types survive" in that environment.

Although rationalizability is a plausible property to economists, the same is not true for biologists. Indeed, following Darwin's *The Origin of Species* (Darwin 1859), a strong emphasis has been laid on the existence of interactions between the different species or types.<sup>3</sup> Those interactions play a major role in the survival of some species, as well as in the extinction of other species. Hence biology hardly supports the idea of rationalizability of an evolutionary process by a fitness ordering; on the contrary, it points to another property of an evolutionary process: interactivity. Interactivity captures the fact that the set of surviving types is likely to vary with the composition of the initial population of types. The reason is that different population groups interact with each others in various ways, so that whether some type survives or becomes extinct may depend on whether some other types are present in the population or not.<sup>4</sup>

Those two properties - rationalizability and interactivity - are plausible to, respectively, economists and biologists. However, as we show in this paper, those two properties are logically incompatible. An evolutionary process must satisfy *either* rationalizability *or* interactivity, but cannot satisfy both properties. No fitness ordering can rationalize an interactive evolutionary process.<sup>5</sup>

That impossibility result allows us, in a second stage, to distinguish, among the existing evolutionary OLG models, between the ones satisfying rationalizability, and the ones satisfying interactivity. For that purpose, we review  $n$ -type versions of some canonical evolutionary OLG models, relying either on (effortless) hereditary transmission mechanisms (Galor Moav 2002), or, alternatively, on socialization mechanisms (Bisin Verdier 2001). We show that, while the evolutionary process at work in the former is generally rationalizable by a fitness ordering, the opposite is true for the latter, which exhibits, in general, interactivity (except when the stationary distribution is an interior distribution).

This paper is organized as follows. Section 2 presents the framework, and defines two properties of evolutionary processes: rationalizability by a fitness ordering and interactivity. Section 3 demonstrates the non-existence of a rationalizable interactive evolutionary process. Section 4 reviews  $n$ -types versions of Galor and Moav (2002) and Bisin and Verdier (2001), and examine whether

---

<sup>2</sup>For experimental applications, see Sippel (1997).

<sup>3</sup>See in particular Chapter 3 *The Struggle for Existence*, where Darwin highlights, by means of various examples, the existence of interspecies interactions in the struggle against extinction.

<sup>4</sup>Interactions between types can be of various kinds: positive interactions (e.g. some types helping other types to survive), or negative ones (e.g. struggle for survival between types).

<sup>5</sup>The non-existence of a fitness ordering was firstly studied by Cohen (1985), who applied Arrow's Impossibility Theorem (Arrow 1963) to the evolutionary context of species selection.

the associated evolutionary processes satisfy rationalizability or interactivity. Section 5 concludes.

## 2 A general framework

This section presents a theory of evolutionary processes as selection processes, through which types or traits survive or become extinct over time.

Let us denote by  $P$  a non-empty set of types (or traits).<sup>6</sup> Throughout the paper, we assume that  $P$  has a finite cardinality. We denote by  $\Omega$  the set of all non-empty subsets of  $P$ , i.e. the set of all possible non-empty groups of types:

$$\Omega = \{N \subseteq P : N \neq \emptyset\}$$

An evolutionary process is defined as follows.

**Definition 1** *An evolutionary process is a map  $S: \Omega \rightarrow \Omega$  such that for all  $N \in \Omega$ ,  $S(N) \subseteq N$ .*

An evolutionary process can be interpreted as the selection or choice, from a non-empty set of types, of a non-empty subset of types. For any set of types  $N \in \Omega$ , the set  $S(N)$ , which is a subset of  $N$ , can be interpreted as the set of *surviving* types. On the contrary, the set  $N \setminus S(N)$  is the set of *extinct* types.

Considering an evolutionary process as a selection allows us to apply the analytical tools of rationalizable choice theory to the study of evolutionary processes.<sup>7</sup> Actually, when considering an evolutionary process as a choice, a first question that can be raised is the one of the *rationalizability* of the evolutionary process by a fitness ordering defined on the set of types. In informal terms, the intuition behind the rationalizability of an evolutionary process by means of a fitness ordering consists of the simple idea that "only the most fit types survived to the evolutionary process". To define the rationalizability property, let us first define what we mean by a fitness ordering.

**Definition 2** *A fitness ordering  $F$  on a set of types  $N \in \Omega$  is a binary relation that is reflexive, transitive and complete.*

For any two types  $x, y \in N$ , the relation  $xFy$  states that "type  $x$  is at least as fit as type  $y$ ". Given the fitness ordering  $F$  and a set of types  $N$ , we can define the maximal set  $M(F, N)$ , which includes all elements of  $N$  that are not dominated, in terms of fitness, by other elements of  $N$ .

**Definition 3** *The maximal set  $M(F, N) = \{x \in N : \forall y \in N, xFy\}$*

Having defined the fitness ordering and the maximal set, we can clarify what we mean by an evolutionary process that is *rationalizable* by a fitness ordering.

<sup>6</sup>The elements of  $P$  can consist of various things, depending on the context. In an economic context,  $P$  can be regarded as a set of individual characteristics of agents (e.g. preferences, productivities, etc.). In a biological context,  $P$  can be regarded as a set of different species.

<sup>7</sup>See Austen-Smith and Banks (1998) and Nitzan (2010).

**Definition 4** An evolutionary process  $S(\cdot)$  is rationalizable (RAT) by a fitness ordering if and only if there exists a fitness ordering  $F$  such that, for any  $N \subseteq \Omega$ ,  $S(N) = M(F, N)$ .

The RAT property states that there is an identity between, on the one hand, the set of surviving types from  $N$ , and, on the other hand, the set of the "most fit" types within  $N$  on the basis of the fitness ordering  $F$  defined on  $N$ . If, for instance, a type  $x$  survives across epochs, whereas other types  $y$  and  $z$  do not survive, the RAT property states that type  $x$  is undominated by types  $y$  and  $z$  in terms of fitness. Moreover, if type  $x$  is undominated in terms of fitness by types  $y$  and  $z$ , then it must also be the case that type  $x$  survives.

Rationalizability captures the intuition according to which "only the most fit types survive". That intuition is appealing to economists. According to the RAT property, the survival or extinction of a type would *reveal* the existence of a fitness ordering between types, exactly as individual choice processes can reveal the existence of a preference ordering in microeconomic theory.

Besides the RAT property, other properties can also be expected from evolutionary processes. In particular, biologists have, since the early stages of the discipline, emphasized that the set of types or traits that survive or become extinct over time is likely to vary with the presence or the absence of other types or traits in the population. That property can be called interactivity.

**Definition 5** An evolutionary process  $S(\cdot)$  is interactive (I) if and only if there exists  $x \in N \subseteq P$  such that  $S(N) \cap N \setminus \{x\} \neq \emptyset$  and  $S(N \setminus \{x\}) \neq S(N) \cap N \setminus \{x\}$ .

Interactivity states that, within the set of surviving types, there exists a type that would not have survived in the hypothetical case where another type had been absent, or, alternatively, there exists, in the set of non-surviving types, a type that would have survived if another type had been absent.

The plausibility of the interactivity property is related to likelihood of *interactions* between different types. Those interactions can be of various kinds, since these can favour or disfavor the survival of other types. Negative interactions occur when the withdrawal of some type allows the survival of other types that would have otherwise not survived. Take the example of three species  $\{x, y, z\}$  sharing the same piece of land. Suppose that only the predator species  $x$  survives, so that  $S(\{x, y, z\}) = \{x\}$ . Let us now suppose that species  $x$  is withdrawn from the piece of land. If species  $y$  and/or  $z$  survive in that new context, we have that  $S(\{y, z\}) \neq S(\{x, y, z\}) \cap \{y, z\}$ , which is an occurrence of interactivity due to *negative* interactions. But interactivity can also be caused by *positive* interactions among types. In that case, the withdrawal of some type from the population leads to the extinction of other types that would have otherwise survived thanks to positive interactions.

### 3 An impossibility result

Rationalizability (RAT) and interactivity (I) seem, at first sight, to be quite plausible properties of evolutionary processes. It makes sense to assume that

the surviving types are the "most fit" types, i.e. the undominated types in terms of fitness. It is also reasonable to suppose that an evolutionary process is not independent from the composition of the population in terms of types, but, rather, exhibits interactivity.

However, as shown in this section, properties RAT and I are logically incompatible. To show that incompatibility, we need first to prove the following lemma. That lemma is imported from the theory of rationalizable choices (see Austen-Smith and Banks 1998, Nitzan 2010), since rationalizing an evolutionary process by means of a fitness ordering on types is formally similar to rationalizing a choice by means of a preference ordering on alternatives. It states two conditions that are necessary and sufficient for an evolutionary process to be rationalizable by a fitness ordering. Those conditions are named Properties  $\alpha$  and  $\beta$ , exactly as these are referred to in the theory of rationalizable choices.

**Lemma 1** *An evolutionary process  $S(\cdot)$  is rationalizable if and only if it satisfies Properties  $\alpha$  and  $\beta$ , defined as follows:*

- Property  $\alpha$ :  $\forall O, N \subseteq \Omega, O \subseteq N \implies S(N) \cap O \subseteq S(O)$ .
- Property  $\beta$ :  $\forall O, N \subseteq \Omega, O \subseteq N, S(O) \cap S(N) \neq \emptyset \implies S(O) \subseteq S(N)$ .

**Proof.** See Nitzan (2010), Theorem 3.2. ■

Property  $\alpha$  states that the withdrawal of some types cannot prevent the survival of the types that would have survived from the initial set. Thus the evolutionary process must be consistent in contraction, since contracting the set of types cannot prevent those who survived initially from surviving in the contracted set. Property  $\beta$  states that expanding the set of types must not prevent the types that survived initially from surviving under the expanded set of types. Thus an evolutionary process must be consistent in expansion.

Thanks to Lemma 1, it can be shown that rationalizability and interactivity are logically incompatible. No evolutionary process can satisfy both properties. Moreover, any evolutionary process must satisfy either RAT or I.

**Proposition 1** *There exists no evolutionary process  $S(\cdot)$  satisfying both RAT and I. There exists no evolutionary process  $S(\cdot)$  satisfying neither RAT nor I.*

**Proof.** See the Appendix. ■

The intuition behind that result can be explained as follows. As stated in Lemma 1, rationalizability implies Properties  $\alpha$  and  $\beta$ , which can be interpreted as, respectively, a contraction-consistency property and an expansion-consistency property. Interactivity must, by the variation of the set of surviving types due to a contraction of the set of types, violate either contraction-consistency or expansion-consistency.

To see why interactivity is incompatible with Property  $\alpha$ , take the following example with  $P = \{x, y, z\}$ . Suppose that a type  $x$  survives only because another type, let us say type  $y$ , is present. We have, for instance,  $S(\{x, y, z\}) = \{x, y, z\}$  and  $S(\{x, z\}) = \{z\}$ . It is easy to see that  $S(\cdot)$  does not satisfy Property  $\alpha$ , since  $S(\{x, y, z\}) \cap \{x, z\} \not\subseteq S(\{x, z\}) = \{z\}$ , contrary to what Property  $\alpha$

requires. As a consequence, there does not exist a fitness ordering rationalizing the evolutionary process  $S(\cdot)$ .<sup>8</sup>

To illustrate the incompatibility of Property  $\beta$  with interactivity, take the following example. Suppose that a type  $x$  does not survive because of the presence of a predatory type  $y$ , but would have survived in the absence of  $y$ . We have  $S(\{x, y, z\}) = \{y, z\}$  and  $S(\{x, z\}) = \{x, z\}$ . It is straightforward to see that  $S(\cdot)$  does not satisfy Property  $\beta$ . Indeed  $S(\{x, z\}) \cap S(\{x, y, z\}) = \{z\} \neq \emptyset$  but  $S(\{x, z\}) = \{x, z\} \not\subseteq S(\{x, y, z\}) = \{y, z\}$ . Hence, here again, there does not exist any fitness ordering  $F$  rationalizing  $S(\cdot)$ .<sup>9</sup> Therefore there exists no evolutionary process that is both rationalizable and interactive.

Those examples can also be used to explain the second part of Proposition 1, that is, any evolutionary process must be either rationalizable or interactive. To see why any violation of Property  $\alpha$  or  $\beta$  implies interactivity, let us consider the example with  $P = \{x, y, z\}$ , with  $S(\{x, y, z\}) = \{y, z\}$ . A violation of, for instance, Property  $\alpha$ , must be an occurrence of interactivity. The reason is that a violation of Property  $\alpha$  is a lack of contraction-consistency. In our example, it occurs when either  $S(\{y, z\}) = \{y\}$  or  $\{z\}$ , so that  $S(\{x, y, z\}) \cap \{y, z\} \not\subseteq S(\{y, z\})$ . But that situation implies  $S(\{y, z\}) \neq S(\{x, y, z\}) \cap \{y, z\}$ , that is, interactivity. Similarly, a violation of Property  $\beta$  is equivalent to expansion-inconsistency, and implies also interactivity.

Therefore an obvious corollary of Proposition 1 is that any evolutionary process must satisfy *one and only one* property: either rationalizability or interactivity. Whether an evolutionary process satisfies rationalizability or interactivity depends on its precise form. However, when the set of types is a pair, we know that the RAT property is necessarily satisfied.

**Lemma 2** *Take a pair of types  $P = \{x, y\}$ . Then any evolutionary process  $S(P)$  satisfies RAT and violates I.*

**Proof.** See the Appendix. ■

As a consequence, studying whether evolutionary processes satisfy rationalizability or interactivity is only relevant for sets of types  $P$  with  $|P| > 2$ .

Finally, to better illustrate the incompatibility between rationalizability and interactivity in the general case where  $|P| > 2$ , it may be worth introducing the following property, which is equivalent to Properties  $\alpha$  and  $\beta$ . That property can be called the Weak Axiom of Revealed Fitness, as it is the equivalent, in our context, of the Weak Axiom of Revealed Preferences (WARP) in consumption theory (see Samuelson 1938, 1948).

**Definition 6** *An evolutionary process  $S(\cdot)$  satisfies the Weak Axiom of Revealed Fitness (WARF) if and only if,  $\forall O, N \subseteq \Omega$ ,  $x \in S(N)$ ,  $y \in N \setminus S(N)$  and  $y \in S(O) \implies x \notin O$ .*

<sup>8</sup>Indeed, if such a fitness ordering  $F$  existed, we would have, from  $S(\{x, y, z\}) = \{x, y, z\}$ , that type  $x$  is undominated by type  $z$  in fitness terms:  $xFz$ . But from  $S(\{x, z\}) = \{z\}$ , we have that type  $x$  is dominated by type  $z$  in fitness terms, so that  $\neg xFz$ . It is thus impossible to find a fitness ordering  $F$  such that  $xFz$  and  $\neg xFz$ .

<sup>9</sup>Indeed, from  $S(\{x, y, z\}) = \{y, z\}$ , we have  $\neg xFz$ , but from  $S(\{x, z\}) = \{x, z\}$ , we have  $xFz$ , which contradicts  $\neg xFz$ .

The WARF states that, if type  $y$  does not survive from a set of types  $N$ , whereas another type  $x$  survives from  $N$ , then if type  $y$  survives from another set of types  $O$ , it must be the case that type  $x$  does not belong to  $O$ .<sup>10</sup>

From the theory of rationalizable choices, we know that the WARF is equivalent to Properties  $\alpha$  and  $\beta$ .<sup>11</sup> Hence, in our context, a rationalizable evolutionary process must satisfy the WARF property, and an interactive evolutionary process must violate the WARF property. To see this, take the example of a three-type population, with  $S(\{x, y, z\}) = \{y, z\}$  and  $S(\{x, z\}) = \{x, z\}$ . That evolutionary process is interactive: type  $x$  does not survive in the presence of type  $z$  when type  $y$  is there, but survives in the presence of type  $z$  when type  $y$  is absent. That evolutionary process also violates WARF:  $S(\{x, y, z\}) = \{y, z\}$  reveals that type  $z$  dominates, in fitness terms, type  $x$ , whereas  $S(\{x, z\}) = \{x, z\}$  reveals that type  $x$  is not dominated, in fitness terms, by type  $z$ . Thus no fitness ordering can be revealed from  $S(\{x, y, z\}) = \{y, z\}$  and  $S(\{x, z\}) = \{x, z\}$ .

## 4 Evolutionary OLG models

Section 3 showed that an evolutionary process must be *either* rationalizable by a fitness ordering *or* interactive, but cannot satisfy both properties. The logical impossibility to have an interactive rationalizable evolutionary process raises the question of the property satisfied by existing evolutionary OLG models.

To answer that question, this section reviews some canonical evolutionary OLG models. In order to study the rationalizability or interactivity of the associated evolutionary processes, we proceed in two stages. First, we focus on models of (effortless) hereditary transmission, in which types are transmitted through biological reproduction. For that purpose, we pay a particular attention to the evolutionary models developed by Galor and Moav in the context of unified growth theory (see Galor and Moav, 2002, 2005). Then, in a second stage, we study models of cultural transmission through (costly) socialization. That second stage is carried out by focusing on models of socialization and cultural transmission by Bisin and Verdier (2000, 2001, 2011).

In order to examine the evolutionary process at work in those frameworks, we must, given Lemma 2, consider versions of those models where the set of types  $P$  has a cardinality strictly larger than 2. Throughout this section, we will examine  $n$ -types versions of evolutionary OLG models, in order to keep the analysis as general as possible. Focusing on  $n$ -types models allows us to be certain that rationalizability, if it holds, is not due to a low postulated number of types, but is a general property of the evolutionary process under study.

<sup>10</sup>It is easy to see that, if one focuses on evolutionary processes involving only two types, the WARF property is necessarily satisfied.

<sup>11</sup>On that equivalence, see Nitzan (2010).

## 4.1 Models of effortless transmission: Galor Moav (2002)

Galor and Moav's (2002) 2-period OLG model provides the foundations of an evolutionary growth theory, since this shows that the take-off from a period of stagnation to sustained growth can be explained by the evolutionary advantage associated to human traits that are complementary to the growth process.

That model is not the only one linking economic growth with the dynamics of heterogeneity. More recently, Galor and Moav (2005) developed another 2-period OLG model with intracohort heterogeneity, whose goal is to explain the demographic transition.<sup>12</sup> In both models, the stylized fact under study - i.e. economic or demographic transition - is explained by means of a shift of evolutionary advantage, from "low quality" agents towards "high quality" agents. The main difference between those models lies in how "quality" is modeled. In Galor Moav (2002), "quality" is the level of children's education, whereas, in Galor Moav (2005), "quality" is the level of a genetically predetermined somatic investment, which coincides both with the unitary child cost and with the degree of child's robustness when facing epidemics. Those two models being close to each other, we focus here only on Galor Moav (2002).

The adult population at time  $t$ , denoted by  $L_t$ , is divided into  $n$  types of agents  $i \in \{1, 2, \dots, n\}$ , who differ in their preferences with respect to the "quality" or the education of their children. The intensity of parental taste for children's education is denoted by the parameter  $\beta^i$ :

$$\beta^1 > \beta^2 > \dots > \beta^n \quad (1)$$

As we will see below, the preference parameter  $\beta^i$  is a key determinant of how agents solve the trade-off between the quantity and the quality of their children.

The adult population  $L_t$  includes  $L_t^i$  agents of type  $i \in \{1, 2, \dots, n\}$ :

$$L_t = \sum_{i=1}^n L_t^i = L_t \sum_{i=1}^n q_t^i \quad (2)$$

where  $q_t^i \equiv L_t^i/L_t$  is the share of type  $i \in \{1, 2, \dots, n\}$  in the adult population. We obviously have:  $\sum_{i=1}^n q_t^i = 1$ .

### 4.1.1 Microfoundations

A single good is produced at every period by means of the technology:

$$Y_t = H_t^{1-\alpha} (A_t X)^\alpha \quad (3)$$

where  $Y_t$  is total output,  $H_t$  is aggregate quantity of efficiency units of labour,  $X$  is the (fixed) land, and  $A_t$  is the level of technology.

The supply of labour depends on agents' decisions on the number of children and on the time investment in the education or "quality" of each child. An

---

<sup>12</sup>The demographic transition is defined as the shift from a demographic equilibrium with high fertility and high mortality to an equilibrium with low fertility and low mortality.

agent of type  $i \in \{1, 2, \dots, n\}$  of generation  $t$  is endowed with  $h_t^i$  efficiency units of labour at time  $t$ . The aggregate supply of efficiency units  $H_t$  is:

$$H_t = L_t^1 f_t^1 h_t^1 + L_t^2 f_t^2 h_t^2 + \dots + L_t^n f_t^n h_t^n \quad (4)$$

where  $f_t^i$  is the fraction of time devoted to labour by an agent of type  $i$ .

In the absence of property rights, the return to land is zero, and the wage per efficiency unit of labour  $w_t$  is equal to the output per efficiency unit of labour:

$$w_t = x_t^\alpha \quad (5)$$

where  $x_t = A_t X / H_t$  denotes effective resources per efficiency unit of labour.

The preferences of a young adult agent of type  $i \in \{1, 2, \dots, n\}$  are represented by a log-linear utility function as follows:

$$u_t^i = (1 - \gamma) \ln(c_t^i) + \gamma [\ln(n_t^i) + \beta^i \ln(h_{t+1}^i)] \quad (6)$$

where  $u_t^i$  is the utility of the agent,  $\gamma$  is a time preference parameter ( $0 < \gamma < 1$ ),  $c_t^i$  is the consumption,  $n_t^i$  is the number of children, while  $h_{t+1}^i$  is the human capital of the child, all expressed for an agent of type  $i$  at time  $t$ . The parameter  $\beta^i$  reflects parental taste for children's education ( $0 < \beta^i \leq 1$ ). This is the source of intracohort heterogeneity in our economy.

The transmission of the preference parameter  $\beta^i$  across generations takes place through biological reproduction: preferences are hereditary, and are transmitted from generations to generations within a dynasty.

The human capital stock of children  $h_{t+1}^i$  depends on the parental time investment in education  $e_{t+1}^i$  and on some rate of technological progress  $g_{t+1} = \frac{A_{t+1} - A_t}{A_t}$ , by means of the following function:

$$h_{t+1}^i = h(e_{t+1}^i, g_{t+1}) \quad (7)$$

which is increasing concave in  $e_{t+1}^i$  and decreasing convex in  $g_{t+1}$ , while the cross derivative is strictly positive.<sup>13</sup>

The endowment of parental time, valued as  $w_t h_t^i$  in terms of consumption, can be spent either on consumption  $c_t^i$ , or on producing children. Each child has a fixed time cost  $\tau$ , which is assumed to be sufficiently small so as to allow for a positive population growth ( $\tau < \gamma$ ), as well as a cost in terms of education,  $e_{t+1}^i$ . Hence the budget constraint of an agent of type  $i$  is:

$$c_t^i \leq w_t h_t^i (1 - \tau n_t^i - e_{t+1}^i n_t^i) \quad (8)$$

Parents of type  $i \in \{1, 2, \dots, n\}$  choose the number of children  $n_t^i$  and time education investment  $e_{t+1}^i$  in such a way as to maximize their lifetime welfare:

$$\begin{aligned} & \max_{n_t^i, e_{t+1}^i} (1 - \gamma) \ln(w_t h_t^i (1 - \tau n_t^i - e_{t+1}^i n_t^i)) + \gamma [\ln(n_t^i) + \beta^i \ln(h(e_{t+1}^i, g_{t+1}))] \\ & \text{s.t. } w_t h_t^i (1 - \tau n_t^i - e_{t+1}^i n_t^i) \geq \tilde{c} \\ & \text{s.t. } (n_t^i, e_{t+1}^i) \geq 0 \end{aligned}$$

<sup>13</sup>The intuition is that the time required for learning a new technology diminishes with the level of education and increases with the rate of technological progress.

where  $\tilde{c}$  is the subsistence consumption necessary for having children.

The optimal number of children for an adult of type  $i \in \{1, 2, \dots, n\}$  satisfies:

$$\begin{aligned} n_t^i(\tau + e_{t+1}^i) &= \gamma && \text{if } x_t^\alpha h_t^i \geq \tilde{c}/(1 - \gamma) \\ n_t^i(\tau + e_{t+1}^i) &= 1 - \tilde{c}/x_t^\alpha h_t^i && \text{if } x_t^\alpha h_t^i \leq \tilde{c}/(1 - \gamma) \\ n_t^i &= 0 && \text{if } x_t^\alpha h_t^i \leq \tilde{c} \end{aligned} \quad (9)$$

Three distinct cases can arise. If the potential income  $x_t^\alpha h_t^i$  is lower than the threshold  $\tilde{c}$ , agents of type  $i$  have no children at all. If the potential income  $x_t^\alpha h_t^i$  is higher than  $\tilde{c}$  but below  $\tilde{c}/(1 - \gamma)$ , agents of type  $i$  choose a strictly positive number of children, which is increasing in the parental potential income. In that case, the fraction of time devoted for child rearing is below  $\gamma$ , and is increasing in the potential income. If the potential income  $x_t^\alpha h_t^i$  is higher than  $\tilde{c}/(1 - \gamma)$ , the fraction of time devoted for child rearing is equal to  $\gamma$ , and is invariant to the level of the parental potential income, *ceteris paribus*.

Regarding the optimal education investment, we have:

$$\begin{aligned} \beta^i h_e(e_{t+1}^i, g_{t+1}) - \frac{h(e_{t+1}^i, g_{t+1})}{(\tau + e_{t+1}^i)} &= 0 \text{ if } e_{t+1}^i > 0 \\ &\leq 0 \text{ if } e_{t+1}^i = 0 \end{aligned} \quad (10)$$

It is only under a sufficiently high level of  $\beta^i$  that there is an interior optimum level of  $e_{t+1}^i$ ; otherwise, for low levels of  $\beta^i$ , parents choose zero education:  $e_{t+1}^i = 0$ .<sup>14</sup> Galor and Moav show that  $e_{t+1}^i$  is increasing in  $g_{t+1}$  and in  $\beta^i$ . On the contrary, technological progress decreases the number of children of an individual of type  $i$ , by raising the investment in the "quality" of children  $e_{t+1}^i$ .

Hence, agents can belong to one of the following three groups. First, agents with potential incomes lower than the subsistence threshold (i.e.  $x_t^\alpha h_t^i \leq \tilde{c}$ ) have no children at all, and thus do not invest in education ( $n_t^i = e_{t+1}^i = 0$ ). Agents of that type become extinct, due to the absence of descendants. Second, agents with potential incomes  $\tilde{c} \leq x_t^\alpha h_t^i \leq \tilde{c}/(1 - \gamma)$  have a strictly positive number of children, but do not necessarily invest in their children's education, depending on how large their taste for children's education  $\beta^i$  is ( $n_t^i > 0$ ,  $e_{t+1}^i \geq 0$ ). The fraction of time dedicated to rearing children is, for those agents, increasing with the potential income. Hence, among agents of that group for which  $e_{t+1}^i = 0$ , those who enjoy a higher potential income have also more children. Third, agents with a potential income larger than  $\tilde{c}/(1 - \gamma)$  have a strictly positive number of children. Here again, whether those agents spend on education or not depends on how large  $\beta^i$  is. In that case, we thus have also  $n_t^i > 0$ ,  $e_{t+1}^i \geq 0$ . However, a major difference with respect to the second group lies in that fact that those agents spend a fraction  $\gamma$  of their time on child rearing, which is invariant to the level of parental potential income.

<sup>14</sup>To guarantee that *at least* some individuals will invest in education, it is assumed, following Galor and Moav, that, for individuals with the highest valuation of quality (i.e.  $\beta^1$ ), there is a positive investment in the quality of children.

At any point in time, whether an individual has some children or not depends on the level of his potential income  $x_t^\alpha h_t^i$ , and, thus, depends not only on the level of his human capital  $h_t^i$ , but, also, on how large the level of technology  $A_t$  and the surface of land  $X$  are in comparison to the aggregate supply of efficiency units of labor  $H_t$ . If  $H_t$  is low in comparison to the level of technology and the surface of land, then all individuals have potential incomes that exceed the threshold  $\tilde{c}$ , and, hence, have some children. However, if  $H_t$  is high in comparison to the level of technology and the surface of land, then individuals with low human capital have a potential income smaller than  $\tilde{c}$ , and, thus, cannot have children.

Therefore, when the aggregate supply of efficiency units of labor  $H_t$  is high in comparison to the level of technology and the surface of land, the economy is characterized by a "struggle for life", in the sense that only the sufficiently gifted agents can have descendants, and thus only the most gifted dynasties, in terms of human capital, can survive over time. The existence of such a struggle for life constitutes, together with heterogeneity, the two necessary conditions for the existence of natural selection mechanisms as described by Darwin (1859). The next section examines the evolutionary process at work in that economy.

#### 4.1.2 The evolutionary process

Let us now study the dynamics of heterogeneity. For that purpose, we assume, without loss of generality, that, at the initial stage of development, type-specific human capital levels satisfy:

$$h_0^1 > h_0^2 > \dots > h_0^{n-1} > h_0^n \quad (11)$$

Thus higher quality types have also a higher initial human capital.

Within Galor and Moav's model, whether one type of agents tends to become, over time, more or less widespread than another type depends only on fertility differentials across types.

To see this, note that the number of agents of type  $i \in \{1, 2, \dots, n\}$  in a cohort follows the law:

$$L_{t+1}^i = n_t^i L_t^i \quad (12)$$

Hence, given that the total adult population can be written as:  $L_{t+1} = n_t^1 L_t^1 + n_t^2 L_t^2 + \dots + n_t^n L_t^n$ , we know that  $q_t^i$  follows the dynamic law:

$$q_{t+1}^i = \frac{n_t^i}{n_t} q_t^i \quad (13)$$

where  $n_t \equiv \frac{n_t^1 L_t^1 + n_t^2 L_t^2 + \dots + n_t^n L_t^n}{L_t}$  denotes the average fertility rate, while  $q_t^i \equiv L_t^i / L_t$  is the share of type  $i \in \{1, 2, \dots, n\}$  in the adult population.

Whether the fraction of the population having a particular type  $i$  rises or falls over time depends only on whether the fertility rate associated with that type is higher or lower than the average fertility rate:

$$q_{t+1}^i \leq q_t^i \iff n_t^i \leq n_t$$

It is straightforward to see that, if some type  $i$  has a potential income lower than  $\tilde{c}$ , then that type becomes extinct, since  $n_t^i = 0$  implies  $q_{t+1}^i = 0$ . Thus extinction of a type occurs in Galor Moav (2002) when that type has too low income potential. For higher income levels, the initial associated fertility rates are positive, and whether one type will tend to expand or not in the population depends on fertility differentials across types. As shown above, type-specific fertility rates depend strongly on the chosen level of education. Parents with a high  $\beta^i$  choose a higher amount of education per child, and thus make fewer children. On the contrary, agents with a lower  $\beta^i$  choose a lower amount of education per child, and invest more in the quantity of children.

In the light of those observations, we can distinguish between  $n$  distinct cases regarding the evolution of the composition of the population over time, depending on which type  $i \in \{1, \dots, n\}$  is the "marginal" one, i.e. the lowest "quality" type that does not become extinct initially. For each of those cases, we need to distinguish between: (1) the composition of the population after a *finite* number of periods; (2) the composition of the population after an *infinite* number of periods.<sup>15</sup> For the sake of presentation, we will present all those cases by referring to a Case  $j$ , where type  $j \in \{1, 2, \dots, n\}$  is the marginal type.

Suppose that the initial potential incomes  $x_0^\alpha h_0^i$  of all types  $i$  with  $i > j$  are so low that  $n_0^i = 0$ . As a result of poverty and stagnation, types  $\{j + 1, j + 2, \dots, n\}$  become extinct. The only surviving types are types  $\{1, 2, \dots, j\}$ , who tend to favor the "quality" of children over the "quantity" of children. Hence, they exhibit low fertility and high investment in education. However, among those surviving "quality" types, one type, type  $j$ , has a larger fertility than other types, because of a lower concern for education (i.e. a lower  $\beta^i$ ). As a consequence, we have:  $n_t^j > \dots > n_t^1$  and type  $j$ , the most "quantity" type of the "quality" types, has the evolutionary advantage over other surviving "quality" types. Hence we have:

$$\begin{aligned} S(\{1, 2, \dots, n\}) &= \{1, 2, \dots, j\} \text{ (finite horizon)} \\ S(\{1, 2, \dots, n\}) &= \{j\} \text{ (infinite horizon)} \end{aligned}$$

The dynamics of heterogeneity must belong to one of those  $n$  cases, depending on the type  $j \in \{1, 2, \dots, n\}$  with the lowest potential income that can escape from initial extinction. That "marginal" type will turn out, in the future, to dominate the other types, because of its larger fertility. That dominance will be compatible with the survival of other types under a finite time horizon, but consists of a full generalization of the population under an infinite time horizon.

In the light of that result, we can now investigate whether the evolutionary process at work in Galor and Moav's framework satisfies rationalizability (RAT) or interactivity (I). The following proposition summarizes our results.

---

<sup>15</sup>The reason is that, if we have  $n_t^i > n_t^j > 0$ , type  $i$  has an evolutionary advantage over type  $j$ , but the latter type will not be extinct after a finite number of time periods (since  $n_t^j > 0$ ). However, from an asymptotic point of view, the inequality  $n_t^i > n_t^j$  implies that type  $j$  tends to be extinct when  $t \rightarrow \infty$ .

**Proposition 2** Consider the model of Galor and Moav (2002) with  $P = \{1, 2, \dots, n\}$ , after a finite or an infinite number of time periods.

- if  $(A_0X/H_0)^\alpha h_0^i \geq \tilde{c}$  for all  $i \in \{1, 2, \dots, n\}$ , then the evolutionary process satisfies RAT and not I;
- if  $(A_0X/H_0)^\alpha h_0^i \geq \tilde{c}$  for all  $i \in \{1, 2, \dots, j\}$  and if  $(A_0X/H_0)^\alpha h_0^i < \tilde{c}$  for all  $i \in \{j+1, j+2, \dots, n\}$ , then the evolutionary process satisfies RAT and not I; except if  $(A_0X/H_0)^\alpha h_0^{j+1} \geq \tilde{c}$  in the absence of some type  $i \in \{1, 2, \dots, n\}$ , in which case it satisfies I and not RAT.

**Proof.** See the Appendix. ■

Whether the evolutionary process satisfies the rationalizability or interactivity depends on initial conditions  $\{A_0, H_0, h_0^i\}$  and on structural parameters  $\{X, \alpha, \tilde{c}\}$ . When there is a large amount of land (i.e. high  $X$ ), a quite productive labour (i.e. a high  $\alpha$ ) and a low subsistence consumption  $\tilde{c}$ , all types will, for a large interval of values for initial conditions  $\{A_0, H_0, h_0^i\}$ , survive in the long-run, and the evolutionary process satisfies rationalizability. However, when the conditions are less favorable, so that at least one type becomes extinct, the evolutionary process may exhibit interactivity. Violations of the RAT property occur when the withdrawal of a type from the population allows the survival of (at least) one other type that would not have survived otherwise. Such a situation is possible, since the withdrawal of a type reduces  $H_0$ , and, thus, raises the potential parental income  $x_0^\alpha h_0^i$  of agents of other types.

In sum, when the living conditions are favorable (i.e. high  $X$ , high  $\alpha$  and low  $\tilde{c}$ ) and / or when initial conditions are sufficiently good (i.e. high values for  $A_0, H_0, h_0^i$ ), the evolutionary process satisfies rationalizability. However, when living conditions are harsher, interactivity may arise. Those results are robust to whether we consider a finite or an infinite time horizon.

## 4.2 Models of costly transmission: Bisin Verdier (2001)

Following the pioneer works of Cavalli-Sforza and Feldman (1981) and Boyd and Richerson (1985) applying models of evolutionary biology to the transmission of cultural types, Bisin and Verdier (2000, 2001) developed microeconomic models of socialization choices. In those models, parental socialization efforts determine the probability that children adopt the type of parents, and, indirectly, the dynamics of the transmission of traits over time.<sup>16</sup>

Bisin and Verdier (2001) consider a 2-period OLG model, where each parent has one child. Children are born without any type, and they acquire their type before becoming an adult, in the way that will be specified below. The adult population is divided in several types. We will focus here on a  $n$ -type economy:  $P = \{1, 2, \dots, n\}$ . As usual,  $q_t^i$  denotes the proportion of the population with type  $i \in \{1, \dots, n\}$  in the adult population at  $t$ .

<sup>16</sup>Those socialization models have become increasingly used, to explain phenomena such as the dynamics of religions and marriage (Bisin *et al* 2004), globalization and trade (Olivier *et al* 2008), or the dynamics of life expectancy (Ponthiere 2010). See Bisin and Verdier (2011).

As this is usually assumed in that literature (see Bisin *et al* 2009 and Montgomery 2010), we assume that the  $n$  different types differ regarding their degree of cultural intolerance, that is, regarding the extent to which they value the fact of having a child with their *own* trait. For simplicity, that degree of intolerance is assumed to be *symmetric* for any given type  $i$ , that is, the utility gain for a parent of type  $i$  from having a child of the same type  $i$  rather than a child of type  $j$  is the same as the utility gain for a parent of type  $i$  from having a child of type  $i$  rather than a child of type  $k$ . Denoting by  $\Delta V^i$  the utility gain of a parent of type  $i$  having a child of type  $i$  rather than of another type, we can rank the  $n$  types on the basis of their degree of intolerance as follows:

$$\Delta V^1 > \Delta V^2 > \Delta V^3 > \dots > \Delta V^n \quad (14)$$

Thus heterogeneity lies here on the side of socialization gains. One could, alternatively, have heterogeneity in the costs of socialization.

#### 4.2.1 Microfoundations

Cultural transmission can be either the result of direct vertical (parental) socialization, or the result of horizontal / oblique socialization:

- direct vertical socialization to the parent trait  $i \in \{1, 2, \dots, n\}$  occurs with a probability  $d_t^i$ ;
- if the child born in a family of type  $i \in \{1, 2, \dots, n\}$  is not directly socialized, which happens with a probability  $1 - d_t^i$ , the child will be socialized by picking up the trait of a role model chosen randomly in the population.<sup>17</sup>

Hence, the probability that a child born in a family of type  $i \in \{1, 2, \dots, n\}$  takes the types  $i$  and  $j$ , with  $i \neq j$ , are given respectively by:

$$P_{t+1}^{ii} = d_t^i + (1 - d_t^i)q_t^i \quad (15)$$

$$P_{t+1}^{ij} = (1 - d_t^i)q_t^j \quad (16)$$

The probabilities of direct socialization  $d_t^i$  are determined by parental socialization decisions. Parents have imperfect empathy: they care about their children, but only through their own preferences (and not the ones of their children). Let us denote by  $V^{ij}$  the utility to a type- $i$  parent of a type- $j$  child. We assume that  $V^{ii}$  and  $V^{ij}$  take constant values, and that  $V^{ii} > V^{ij}$ , so that parents have an incentive to socialize their children to their type.

The socialization takes place through a parental socialization effort, denoted by  $\tau_t^i \in [0, 1]$ .<sup>18</sup> That socialization effort has a welfare cost, which is usually assumed to be quadratic in the effort level:

$$C(\tau_t^i) = \frac{(\tau_t^i)^2}{2} \quad (17)$$

<sup>17</sup>The child takes the trait  $i$  with a probability  $q_t^i$  and the trait  $j \neq i$  with a probability  $q_t^j$ .

<sup>18</sup>One can, alternatively, make the probability of direct vertical socialization depend on a segregation effort, whose goal is to minimize the influence of external sources on the child.

Bisin and Verdier consider a wide set of socialization mechanisms, which relate the probability of direct vertical socialization  $d_t^i$  to the parental socialization effort  $\tau_t^i$ . One well-know socialization mechanism is entitled "It's the family". Under that mechanism, the probability of direct vertical socialization to trait  $i$  is merely dependent on the parental socialization effort, as follows:

$$d_t^i = \tau_t^i \quad (18)$$

Hence the probabilities of keeping the parental type  $i \in \{1, 2, \dots, n\}$  or adopting another type  $j \neq i$  are:

$$P_{t+1}^{ii} = \tau_t^i + (1 - \tau_t^i)q_t^i \quad (19)$$

$$P_{t+1}^{ij} = (1 - \tau_t^i)q_t^j \quad (20)$$

Let us now consider the parental socialization decision. The socialization choice for a parent of type  $i \in \{1, 2, \dots, n\}$  can be written as:

$$\begin{aligned} \max_{\tau_t^i \in [0,1]} & -C(\tau_t^i) + P_{t+1}^{ii}V^{ii} + \sum_{j \neq i} P_{t+1}^{ij}V^{ij} \\ \text{s.t. } & P_{t+1}^{ii} = \tau_t^i + (1 - \tau_t^i)q_t^i \\ \text{s.t. } & P_{t+1}^{ij} = (1 - \tau_t^i)q_t^j \end{aligned}$$

The first-order condition can be rewritten as:

$$\tau_t^i = (V^{ii} - V^{ij}) \sum_{j \neq i} q_t^j > 0 \quad (21)$$

$V^{ii} - V^{ij} \equiv \Delta V^{ij} > 0$  is the relative value of child with the same type as the parent of type  $i$  in comparison with that child taking trait  $j \neq i$ . Following Bisin and Verdier (2011), we assume that the degrees of cultural intolerance  $\Delta V^{ij}$  are symmetric, that is,  $\Delta V^{ij} = \Delta V^{ik}$  for all  $j \neq k$ . We will thus denote  $\Delta V^{ij}$  by  $\Delta V^i$ . That assumption is known as the "symmetric intolerances" assumption.

The parental socialization effort is decreasing in the fraction of the population sharing the parent's type,  $q_t^i$ .<sup>19</sup> This phenomenon is called "cultural substitution": the larger the proportion of the population sharing the parental type, the lower is the socialization effort chosen by those parents, since they can rely on the society as a whole for the socialization of their child to their type.

#### 4.2.2 The evolutionary process

The adult population with the type  $i$  at time  $t + 1$ , denoted by  $q_{t+1}^i$ , is:

$$q_{t+1}^i = q_t^i [d_t^i + (1 - d_t^i)q_t^i] + \sum_{j \neq i} q_t^j [(1 - d_t^j)(q_t^i)] \quad (22)$$

---

<sup>19</sup>Indeed  $q_t^i = 1 - \sum_{j \neq i} q_t^j$ .

The first term is the number of children born in a family of type  $i$  and who kept the type of their parent, whereas the second term is the number of children born in a family of type  $j \neq i$  and who took the type  $i$ .

Substituting for  $\tau_t^i = (1 - q_t^i)\Delta V^i$  and  $\tau_t^j = (1 - q_t^j)\Delta V^j$ , we get:

$$q_{t+1}^i = q_t^i + q_t^i \left[ \sum_{j \neq i} q_t^j (1 - q_t^i) \Delta V^i - \sum_{j \neq i} q_t^j (1 - q_t^j) \Delta V^j \right] \quad (23)$$

Hence, in the presence of  $n$  types, the dynamics of population is summarized by an  $n$ -dimensional dynamic system. As all  $q_t^i$  sum up to 1, that system can be reduced to an  $(n - 1)$ -dimensional dynamic system. Obviously, the study of the dynamics of heterogeneity within such a framework is far from trivial. That study has been recently carried out by Bisin *et al* (2009), but in a continuous time setting. Given that the precise purpose of the present paper lies in the study of rationalizability, we will, throughout the rest of this section, shift to a continuous time model, and use Bisin *et al* (2009)'s findings to explore the properties of the associated evolutionary process.<sup>20</sup>

In a continuous time setting, the proportion of the adult population with type  $i$  follows the dynamic law:

$$\dot{q}^i = q^i \left[ (1 - q^i) \Delta V^i - \sum_{j \neq i} q^j (1 - q^j) \Delta V^j \right] \quad (24)$$

That expression is supposed to hold for all  $i \in \{1, 2, \dots, n\}$ .

Bisin *et al* (2009) study the dynamics of heterogeneity in that economy. They particularly focus on the existence, uniqueness and stability of a stationary distribution of the population across the  $n$  possible types, that is, a distribution that can replicate itself over time. Their main results are the following:<sup>21</sup>

- Any degenerate distribution (i.e. where  $q^i = 1$  for some type  $j$  and  $q^i = 0$  for types  $i \neq j$ ) is a locally unstable stationary distribution.
- A stationary distribution with  $k$  surviving types, denoted by  $F^k$ , exists if and only if, for any type  $i \in F^k$ , we have:

$$\Delta V^i > (k - 1)G^k$$

where  $\frac{1}{G^k} = \sum_{j \in F^k} \frac{1}{\Delta V^j}$ .

- The stationary distribution is defined by:

$$q^i = 1 - \frac{(k - 1)}{\Delta V^i} G^k$$

for  $i \in F^k$  and  $q^j = 0$  for  $j \notin F^k$ .

<sup>20</sup>Note that the focus on the continuous time model involves some simplification, since the equivalent discrete time model may exhibit a significantly more complex dynamics.

<sup>21</sup>See Bisin *et al* (2009), propositions 1 to 4.

- There exists a type  $k^* \geq 2$  such that a unique stationary distribution  $F^{k^*}$  exists. That stationary distribution is locally stable.
- $F^{k^*}$  contains the types with the highest degree of cultural intolerance, that is, all traits  $i$  such that  $i \geq k^*$ .
- There exist no stationary distribution  $F^k$  with  $k > k^*$ .
- Stationary distributions  $F^k$  with  $k < k^*$  are locally unstable.
- If  $\sum_{j \in \{1, \dots, n\}} \frac{1}{\Delta V^j} > \frac{n-1}{\min_{i \in \{1, \dots, n\}} \Delta V^i}$ , then there exists a unique interior locally stable stationary distribution, with, for all  $i \in \{1, 2, \dots, n\}$ :

$$q^i = 1 - \frac{(n-1)}{\Delta V^i} \left[ \sum_{j \in \{1, \dots, n\}} \frac{1}{\Delta V^j} \right]^{-1}$$

This is the special case where the margin type  $k^*$  coincides with the least intolerable type, i.e. type  $n$ .

Those results can be summarized as follows. Take the general case where the initial composition of the population is not a stationary distribution of types. Then, two cases can arise. First, if  $\sum_{i \in \{1, \dots, n\}} \frac{1}{\Delta V^i} > \frac{n-1}{\min_{i \in \{1, \dots, n\}} \Delta V^i}$ , all types  $\{1, \dots, n\}$  survive in the long-run. Second, if  $\sum_{i \in \{1, \dots, n\}} \frac{1}{\Delta V^i} \leq \frac{n-1}{\min_{i \in \{1, \dots, n\}} \Delta V^i}$ , only a subset of types, denoted by  $\{1, \dots, k^*\}$  survives in the long-run, whereas types  $\{k^* + 1, \dots, n\}$  become extinct. In that second case, agents with a degree of cultural intolerance higher or equal to  $\Delta V^{k^*}$  will survive in the long-run, whereas types with a low degree of intolerance will become extinct.

Therefore, and focusing on the general case where the initial distribution is not a stationary distribution, we have either:

$$S(\{1, 2, \dots, n\}) = \{1, 2, \dots, n\}$$

or

$$S(\{1, 2, \dots, n\}) = \{1, \dots, k^*\}$$

We are now able to discuss whether the evolutionary process satisfies rationalizability or interactivity.

**Proposition 3** *Consider the model of Bisin and Verdier (2001) with  $P = \{1, \dots, n\}$  under the "It's the family" socialization mechanism.*

- If  $\sum_{i \in \{1, \dots, n\}} \frac{1}{\Delta V^i} > \frac{n-1}{\min_{i \in \{1, \dots, n\}} \Delta V^i}$ , then the evolutionary process satisfies RAT and not I;
- If  $\sum_{i \in \{1, \dots, n\}} \frac{1}{\Delta V^i} \leq \frac{n-1}{\min_{i \in \{1, \dots, n\}} \Delta V^i}$ , then the evolutionary process satisfies I and not RAT.

**Proof.** See the Appendix. ■

The first case is the situation where the stationary distribution is an interior distribution, that is, a situation where all types survive. In that case, any contraction of the set of types will preserve the initial set of survivors, so that the evolutionary process is rationalizable by a fitness ordering. However, in the second case, where only a subset of types survives, the evolutionary process is not rationalizable. As shown in the Appendix, the reason why the evolutionary process present in Bisin Verdier does not satisfy rationalizability lies in a violation of Property  $\beta$ . As we show, the withdrawal of some initially surviving type can make an initially non-surviving type survive, against Property  $\beta$ . It follows from this that, in the case where some types do not initially survive, the evolutionary process is interactive.

In sum, the evolutionary process present in the Bisin Verdier model, which relies on a microfounded socialization mechanism, exhibits rationalizability by a fitness ordering or interactivity, depending on whether the equilibrium stationary distribution of types is interior or not. The interiority of the stationary distribution of types depends on the shape of the distribution of cultural intolerance in the population. A central aspect of that distribution is the extent to which the least intolerant type is quite intolerant or not, i.e. whether  $\min \Delta V^i$  is more or less large. When the least intolerant type exhibits a high degree of intolerance, the stationary distribution of types is an interior equilibrium. In that case, Property  $\beta$  is necessarily satisfied, as there could hardly be a "new" surviving type when some other type is withdrawn from the population. Hence the evolutionary process is rationalizable. However, when the least intolerant type exhibits a low degree of cultural intolerance, the stationary distribution of types is not an interior equilibrium, as some types become extinct. In that case, Property  $\beta$  is not satisfied, and the evolutionary process is not rationalizable by a fitness ordering.

## 5 Conclusions

Evolutionary OLG models have become widespread in growth theory, since these allow for a modelling of the dynamics of heterogeneity, which has been shown to be crucial for the understanding various phenomena, such as the birth of economic growth (Galor and Moav 2002), the demographic transition (Galor and Moav 2005), and the dynamics of religions and marriage (Bisin *et al* 2004).

The goal of this paper was to propose a framework to study the properties of the evolutionary processes at work in those models. In particular, we focused on the properties of those evolutionary processes in terms of survival and extinction of types or traits in the long-run.

For that purpose, we first introduced, within a general finite-population model, two properties of an evolutionary process. On the one hand, the rationalization by a fitness ordering, in the sense that the selection must have been such that "only the most fit ones have survived". On the other hand, the interactivity of the evolutionary process, in the sense that the set of selected types

is sensitive to the presence of some other types. Those two formal properties were shown to be logically incompatible.

Then, in a second stage, we considered  $n$ -types versions of several canonical evolutionary OLG models, such as the models of hereditary transmission by Galor and Moav (2002), and the models of socialization by Bisin and Verdier (2001). We showed that, while the evolutionary process at work in Galor-Moav is generally rationalizable by a fitness ordering, the opposite is true for the evolutionary process in Bisin-Verdier, which exhibits, in general, interactivity.

In sum, the two general properties of evolutionary processes introduced in this paper (rationalizability by a fitness ordering and interactivity) allow us to clearly distinguish between evolutionary OLG models that could, at first glance, be regarded as very close. Distinguishing between, on the one hand, rationalizable evolutionary processes, and, on the other hand, interactive evolutionary processes, matters for the study of heterogeneity, since these two kinds of evolutionary processes differ on a particular aspect that is most relevant for studying situations where heterogeneity is important: the capacity of the initial structure of heterogeneity to affect, under the prevailing evolutionary process, the set of types that will survive in the long-run. Under rationalizable evolutionary processes, the initial structure of heterogeneity does not affect the set of surviving types, whereas the opposite result holds under interactive processes, where the set of surviving types is sensitive to the initial structure of heterogeneity. Therefore the partition between rationalizable and interactive evolutionary processes casts some original light on the OLG literature aimed at describing the dynamics of heterogeneity.

## 6 References

- Allais, M., 1947. *Economie et Intérêt*. Paris: Imprimerie Nationale.
- Arrow, K.J., 1963. *Social Choice and Individual Values*, 2nd edition, Yale University Press, New Haven.
- Austin-Smith, D., Banks, J., 1998. *Positive Political Theory I: Collective Preferences*, University of Michigan Press, Michigan.
- Bisin, A., Topa, G., & Verdier, T., 2004. Religious intermarriage and socialization in the United States. *Journal of Political Economy*, 112, 615-665.
- Bisin, A., Topa, G., & Verdier, T., 2009. Cultural transmission, socialization and the population dynamics of multiple state traits distributions. *International Journal of Economic Theory*, 5(1), 139-154.
- Bisin, A. & Verdier, T., 2000. Beyond the melting pot: cultural transmission, marriage and the evolution of ethnic and religious traits. *Quarterly Journal of Economics*, 115, 955-988.
- Bisin, A., & Verdier, T., 2001. The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory*, 97, 298-319.
- Bisin, A., & Verdier, T., 2011. The economics of cultural transmission and socialization. *Handbook of Social Economics*, forthcoming.
- Boyd, R. & Richerson, P., 1985. *Culture and the Evolutionary Process*. Chicago, IL, University of Chicago Press.

Cavalli Sforza, L. & Feldman, M., 1981. Cultural Transmission and Evolution. A Qualitative Approach. Princeton, NJ, Princeton University Press.

Cohen, J., 1985. Can fitness be aggregated?, *The American Naturalist*, 125(5), 717-729.

Darwin, C., 1859. *On the Origin of Species By Means of Natural Selection, or The Preservation of Favoured Races In The Struggle For Life*. London, John Murray Albemarle Street.

Diamond, P., 1965. National debt in a neoclassical growth model. *American Economic Review*, 55(5), 1126-1150.

Galor, O., Moav, O., 2002. Natural selection and the origin of economic growth. *Quarterly Journal of Economics*, 117(4), 1133-1191.

Galor, O., Moav, O., 2005. Natural selection and the evolution of life expectancy. CEPR Discussion Papers, 5373.

Montgomery, J., 2010. Intergenerational cultural transmission as an evolutionary game. *American Economic Journal: Microeconomics*, 2(4), 115-136.

Nitzan, S., 2010. *Collective Preferences and Choice*, Cambridge University Press, Cambridge.

Olivier, J., Thoenig, M., & Verdier, T., 2008. Globalization and the dynamics of cultural identity. *Journal of International Economics*, 76(2), 356-370.

Parfit, D. *Reasons and Persons*, Oxford University Press, New-York.

Ponthiere, G., 2010. Unequal longevities and lifestyles transmission. *Journal of Public Economic Theory*, 12(1), 93-126.

Samuelson, P., 1938. A note on the pure theory of consumer's behaviour. *Economica*, 5, 61-71.

Samuelson, P., 1948. Consumption theory in terms of revealed preferences. *Economica*, 15, 243-253.

Samuelson, P., 1958. An exact consumption-loan model of interest with or without the social contrivance of money. *Journal of Political Economy*, 66(6), 467-482.

Sippel, R., 1997. An experiment on the pure theory of consumer's behaviour. *Economic Journal*, 107, 1431-1444.

## 7 Appendix

### 7.1 Proof of Proposition 1

Let us first show the logical incompatibility between RAT and I.

Suppose that  $S(\cdot)$  satisfies I. Then, by definition, there exists  $x \in N \subseteq P$  such that  $S(N) \cap N \setminus \{x\} \neq \emptyset$  and  $S(N \setminus \{x\}) \neq S(N) \cap N \setminus \{x\}$ .  $S(N \setminus \{x\}) \neq S(N) \cap N \setminus \{x\}$  implies either  $S(N \setminus \{x\}) \not\subseteq S(N) \cap N \setminus \{x\}$  or  $S(N \setminus \{x\}) \not\supseteq S(N) \cap N \setminus \{x\}$  or both.

Let us first assume that  $S(N \setminus \{x\}) \not\supseteq S(N) \cap N \setminus \{x\}$ . Denoting  $N \setminus \{x\}$  by  $O$ , this is equivalent to  $S(O) \not\supseteq S(N) \cap O$ , in contradiction with Property  $\alpha$ . This implies, given Lemma 1, that RAT is violated. Let us now assume that  $S(N \setminus \{x\}) \not\subseteq S(N) \cap N \setminus \{x\}$ . Denoting  $N \setminus \{x\}$  by  $O$ , we have  $O \subseteq N$  and  $S(O) \not\subseteq S(N) \cap O$ . This clearly violates  $S(O) \subseteq S(N)$ , in contradiction with Property  $\beta$ . This implies, given Lemma 1, that RAT is violated.

Let us now show that any evolutionary process must satisfy either RAT or I. That result can be proved by *reductio ad absurdum*. Let us suppose the existence of an evolutionary process  $S(\cdot)$  that satisfies neither RAT nor I. The non-satisfaction of RAT implies, by Lemma 1, that Property  $\alpha$  and/or Property  $\beta$  is violated.

Assume first that Property  $\alpha$  is violated and Property  $\beta$  is satisfied. This implies that  $\exists O, N \subseteq \Omega$ , with  $O \subseteq N$ , and such that  $S(N) \cap O \not\subseteq S(O)$ . Therefore the withdrawal of some type from  $N$  prevents the survival of at least one type that used to survive from  $N$ . Let us denote the withdrawn type as  $x$ . We have thus:  $N \setminus \{x\} \subseteq N$  and  $S(N) \cap N \setminus \{x\} \not\subseteq S(N \setminus \{x\})$ , which implies that  $S(N \setminus \{x\}) \neq S(N) \cap N \setminus \{x\}$ . Hence Property I is necessarily satisfied. Thus a contradiction is reached. An evolutionary process that violates Property  $\alpha$  must satisfy I, and thus cannot violate both RAT and I.

Assume now that Property  $\beta$  is violated and Property  $\alpha$  is satisfied. This implies that  $\exists O, N \subseteq \Omega$ , with  $O \subseteq N$ , and such that  $S(O) \cap S(N) \neq \emptyset$  and  $S(O) \not\subseteq S(N)$ . In other words, the addition of some type, let us say  $y$ , to the initial set of types, let us say set  $O$ , prevents the survival of some initially surviving type. We have thus:  $O \subseteq O \cup \{y\}$  and  $S(O \cup \{y\}) \cap O \not\subseteq S(O)$ , which implies that  $S(O) \neq S(O \cup \{y\}) \cap O$ . Hence Property I is necessarily satisfied. Thus a contradiction is reached. An evolutionary process that violates Property  $\beta$  must satisfy I, and thus cannot violate both RAT and I.

Finally, when an evolutionary process violates both Properties  $\alpha$  and  $\beta$ , the two previous arguments hold, and the evolutionary process must satisfy I. Hence it is impossible to violate both RAT and I. One - and only one - of those properties must be satisfied.

## 7.2 Proof of Lemma 2

Take  $P = \{a, b\}$ . From the non-emptiness of the set of surviving types, we know that  $S(\{a, b\}) = \{a\}$  or  $\{b\}$  or  $\{a, b\}$ , and that  $S(\{a\}) = \{a\}$  and  $S(\{b\}) = \{b\}$ .

Let us show that Property  $\alpha$  is always valid. Property  $\alpha$  requires:  $O, N \subseteq \Omega$ ,  $O \subseteq N \implies S(N) \cap O \subseteq S(O)$ . Take the case where  $S(\{a, b\}) = \{a\}$ . We have  $\{a\} \subseteq \{a, b\}$  and  $S(\{a, b\}) \cap \{a\} = \{a\} \subseteq S(\{a\}) = \{a\}$ , in conformity with Property  $\alpha$ . We have also  $\{b\} \subseteq \{a, b\}$  and  $S(\{a, b\}) \cap \{b\} = \emptyset \subseteq S(\{b\}) = \{b\}$ , in conformity with Property  $\alpha$ . The same rationale holds when  $S(\{a, b\}) = \{b\}$ . Take now the case where  $S(\{a, b\}) = \{a, b\}$ . We have  $\{a\} \subseteq \{a, b\}$  and  $S(\{a, b\}) \cap \{a\} = \{a\} \subseteq S(\{a\}) = \{a\}$ , in conformity with Property  $\alpha$ . We have also  $\{b\} \subseteq \{a, b\}$  and  $S(\{a, b\}) \cap \{b\} = \{b\} \subseteq S(\{b\}) = \{b\}$ . Hence Property  $\alpha$  is satisfied.

Let us now show that Property  $\beta$  is also valid. Property  $\beta$  requires:  $O, N \subseteq \Omega$ ,  $O \subseteq N$  and  $S(O) \cap S(N) \neq \emptyset \implies S(O) \subseteq S(N)$ . Take the case where  $S(\{a, b\}) = \{a\}$ . We have  $\{a\} \subseteq \{a, b\}$  and  $S(\{a, b\}) \cap S(\{a\}) = \{a\} \neq \emptyset$  and  $S(\{a\}) = \{a\} \subseteq S(\{a, b\}) = \{a\}$ , in conformity with Property  $\beta$ . We have also  $\{b\} \subseteq \{a, b\}$  and  $S(\{a, b\}) \cap S(\{b\}) = \emptyset$ , so that Property  $\beta$  is satisfied. The same rationale holds when  $S(\{a, b\}) = \{b\}$ . Take now the case where  $S(\{a, b\}) = \{a, b\}$ . We have  $\{a\} \subseteq \{a, b\}$  and  $S(\{a, b\}) \cap S(\{a\}) = \{a\} \neq \emptyset$  and

$S(\{a\}) = \{a\} \subseteq \{a, b\} = \{a, b\}$ , in conformity with Property  $\beta$ . We have also  $\{b\} \subseteq \{a, b\}$  and  $S(\{a, b\}) \cap S(\{b\}) = \{b\} \neq \emptyset$  and  $S(\{b\}) = \{b\} \subseteq S(\{a, b\}) = \{a, b\}$ . Hence Property  $\beta$  is satisfied.

### 7.3 Proof of Proposition 2

**Finite time horizon** Consider first the case where  $(A_0X/H_0)^\alpha h_0^i \geq \tilde{c}$  for all  $i \leq n$ . In that case, we have  $S(\{1, \dots, n\}) = \{1, \dots, n\}$ .

Let us now check whether the model still satisfies RAT. If we withdraw type  $j$ , we obtain the subset  $\{1, \dots, j-1, j+1, \dots, n\} \subseteq \{1, \dots, j-1, j, j+1, \dots, n\}$ . Note that it must be the case that  $S(\{1, \dots, j-1, j+1, \dots, n\}) = \{1, \dots, j-1, j+1, \dots, n\}$ , since, if agents of type  $j$  disappear, then  $H_t$  goes down, and thus the potential income of other types goes up, so that this cannot push these towards extinction. Hence, given  $S(\{1, \dots, j-1, j+1, \dots, n\}) = \{1, \dots, j-1, j+1, \dots, n\}$ , it is surely true that  $S(\{1, \dots, j-1, j, j+1, \dots, n\}) \cap \{1, \dots, j-1, j+1, \dots, n\} = \{1, \dots, j-1, j+1, \dots, n\} \subseteq S(\{1, \dots, j-1, j+1, \dots, n\}) = \{1, \dots, j-1, j+1, \dots, n\}$ , in conformity with Property  $\alpha$ .

Regarding Property  $\beta$ , we have, when taking  $\{1, \dots, j-1, j+1, \dots, n\} \subseteq \{1, \dots, j-1, j, j+1, \dots, n\}$  and  $S(\{1, \dots, j-1, j+1, \dots, n\}) \cap S(\{1, \dots, j-1, j, j+1, \dots, n\}) \neq \emptyset$ , that  $S(\{1, \dots, j-1, j+1, \dots, n\}) = \{1, \dots, j-1, j+1, \dots, n\} \subseteq S(\{1, \dots, j-1, j, j+1, \dots, n\})$ , in conformity with Property  $\beta$ .

Hence, by Lemma 1, RAT is necessarily satisfied in that case.

Consider now the case where  $(A_0X/H_0)^\alpha h_0^i \geq \tilde{c}$  for all  $i \leq j$  and  $(A_0X/H_0)^\alpha h_0^i < \tilde{c}$  for all  $i > j$ .

In that case, we have  $S(\{1, \dots, n\}) = \{1, \dots, j\}$ .

Let us now check whether the model still satisfies RAT. If we withdraw type  $i \leq j$ , we obtain the subset  $\{1, \dots, i-1, i+1, \dots, n\} \subseteq \{1, \dots, i-1, i, i+1, \dots, n\}$ .

Two cases can arise.

If  $(A_0X/H_0)^\alpha h_0^{j+1} < \tilde{c}$ , then  $S(\{1, \dots, i-1, i+1, \dots, n\}) = \{1, \dots, i-1, i+1, \dots, j\}$ . Hence, given  $S(\{1, \dots, i-1, i+1, \dots, n\}) = \{1, \dots, i-1, i+1, \dots, j\}$ , it is surely true that  $S(\{1, \dots, i-1, i, i+1, \dots, n\}) \cap \{1, \dots, i-1, i+1, \dots, n\} = \{1, \dots, i-1, i+1, \dots, j\} \subseteq S(\{1, \dots, i-1, i+1, \dots, n\}) = \{1, \dots, i-1, i+1, \dots, j\}$ , in conformity with Property  $\alpha$ . Moreover, regarding Property  $\beta$ , we have, when taking  $\{1, \dots, i-1, i+1, \dots, n\} \subseteq \{1, \dots, i-1, i, i+1, \dots, n\}$  and  $S(\{1, \dots, i-1, i+1, \dots, n\}) \cap S(\{1, \dots, i-1, i, i+1, \dots, n\}) \neq \emptyset$ , that  $S(\{1, \dots, i-1, i+1, \dots, n\}) = \{1, \dots, i-1, i+1, \dots, j\} \subseteq S(\{1, \dots, i-1, i, i+1, \dots, n\})$ , in conformity with Property  $\beta$ .

Hence, by Lemma 1, RAT is necessarily satisfied in that case.

If, however,  $(A_0X/H_0)^\alpha h_0^{j+1} \geq \tilde{c}$ , then  $S(\{1, \dots, i-1, i+1, \dots, n\}) = \{1, \dots, i-1, i+1, \dots, j+1\}$ . Hence, given  $S(\{1, \dots, i-1, i+1, \dots, n\}) = \{1, \dots, i-1, i+1, \dots, j+1\}$ , it is surely true that  $S(\{1, \dots, i-1, i, i+1, \dots, n\}) \cap \{1, \dots, i-1, i+1, \dots, n\} = \{1, \dots, i-1, i+1, \dots, j+1\} \subseteq S(\{1, \dots, i-1, i+1, \dots, n\}) = \{1, \dots, i-1, i+1, \dots, j+1\}$ , in conformity with Property  $\alpha$ . Hence Property  $\alpha$  remains valid. However, regarding Property  $\beta$ , we have, when taking  $\{1, \dots, i-1, i+1, \dots, n\} \subseteq \{1, \dots, i-1, i, i+1, \dots, n\}$  and  $S(\{1, \dots, i-1, i+1, \dots, n\}) \cap S(\{1, \dots, i-1, i, i+1, \dots, n\}) \neq \emptyset$ , that  $S(\{1, \dots, i-1, i+1, \dots, n\}) = \{1, \dots, i-1, i+1, \dots, j+1\} \not\subseteq S(\{1, \dots, i-1, i, i+1, \dots, n\})$ , against Property  $\beta$ .

Hence, by Lemma 1, RAT is not satisfied in that case.

**Infinite time horizon** Consider first the case where  $(A_0X/H_0)^\alpha h_0^i \geq \tilde{c}$  for all  $i \leq n$ . In that case, we have  $S(\{1, \dots, n\}) = \{n\}$ .

Let us now check whether the model still satisfies RAT. If we withdraw type  $j$ , we obtain the subset  $\{1, \dots, j-1, j+1, \dots, n\} \subseteq \{1, \dots, j-1, j, j+1, \dots, n\}$ . Note that it must be the case that  $S(\{1, \dots, j-1, j+1, \dots, n\}) = \{n\}$ , since, if agents of type  $j$  disappear, then  $H_t$  goes down, but does not cause a change in the evolutionary advantage of type  $n$ . Hence, given  $S(\{1, \dots, j-1, j+1, \dots, n\}) = \{n\}$ , it is surely true that  $S(\{1, \dots, j-1, j, j+1, \dots, n\}) \cap \{1, \dots, j-1, j+1, \dots, n\} = \{n\} \subseteq S(\{1, \dots, j-1, j+1, \dots, n\}) = \{n\}$ , in conformity with Property  $\alpha$ .

Regarding Property  $\beta$ , we have, when taking  $\{1, \dots, j-1, j+1, \dots, n\} \subseteq \{1, \dots, j-1, j, j+1, \dots, n\}$  and  $S(\{1, \dots, j-1, j+1, \dots, n\}) \cap S(\{1, \dots, j-1, j, j+1, \dots, n\}) \neq \emptyset$ , that  $S(\{1, \dots, j-1, j+1, \dots, n\}) = \{n\} \subseteq S(\{1, \dots, j-1, j, j+1, \dots, n\})$ , in conformity with Property  $\beta$ .

Hence, by Lemma 1, RAT is necessarily satisfied in that case.

Consider now the case where  $(A_0X/H_0)^\alpha h_0^i \geq \tilde{c}$  for all  $i \leq j$  and  $(A_0X/H_0)^\alpha h_0^i < \tilde{c}$  for all  $i > j$ . In that case, we have  $S(\{1, \dots, n\}) = \{j\}$ .

Let us now check whether the model still satisfies RAT. If we withdraw type  $i \leq j$ , we obtain the subset  $\{1, \dots, i-1, i+1, \dots, n\} \subseteq \{1, \dots, i-1, i, i+1, \dots, n\}$ .

Two cases can arise.

If  $(A_0X/H_0)^\alpha h_0^{j+1} < \tilde{c}$ , then  $S(\{1, \dots, i-1, i+1, \dots, n\}) = \{j\}$ . Hence, given  $S(\{1, \dots, i-1, i+1, \dots, n\}) = \{j\}$ , it is surely true that  $S(\{1, \dots, i-1, i, i+1, \dots, n\}) \cap \{1, \dots, i-1, i+1, \dots, n\} = \{j\} \subseteq S(\{1, \dots, i-1, i+1, \dots, n\}) = \{j\}$ , in conformity with Property  $\alpha$ . Moreover, regarding Property  $\beta$ , we have, when taking  $\{1, \dots, i-1, i+1, \dots, n\} \subseteq \{1, \dots, i-1, i, i+1, \dots, n\}$  and  $S(\{1, \dots, i-1, i+1, \dots, n\}) \cap S(\{1, \dots, i-1, i, i+1, \dots, n\}) \neq \emptyset$ , that  $S(\{1, \dots, i-1, i+1, \dots, n\}) = \{j\} \subseteq S(\{1, \dots, i-1, i, i+1, \dots, n\})$ , in conformity with Property  $\beta$ . Hence, by Lemma 1, RAT is necessarily satisfied in that case.

If, however,  $(A_0X/H_0)^\alpha h_0^{j+1} \geq \tilde{c}$ , then  $S(\{1, \dots, i-1, i+1, \dots, n\}) = \{j+1\}$ . Hence, given  $S(\{1, \dots, i-1, i+1, \dots, n\}) = \{j+1\}$ , it is clear that  $S(\{1, \dots, i-1, i, i+1, \dots, n\}) \cap \{1, \dots, i-1, i+1, \dots, n\} = \{j\} \not\subseteq S(\{1, \dots, i-1, i+1, \dots, n\}) = \{j+1\}$ . Hence Property  $\alpha$  is no longer valid here. Moreover, regarding Property  $\beta$ , we have, when taking  $\{1, \dots, i-1, i+1, \dots, n\} \subseteq \{1, \dots, i-1, i, i+1, \dots, n\}$ , we have  $S(\{1, \dots, i-1, i+1, \dots, n\}) \cap S(\{1, \dots, i-1, i, i+1, \dots, n\}) \neq \emptyset$ , so that Property  $\beta$  is trivially satisfied. But as Property  $\alpha$  is not verified, it follows from Lemma 1 that RAT is violated.

## 7.4 Proof of Proposition 3

Two general cases must be considered: (1) the survival of all types; (2) the survival of some subset of types.

**Case (1): interior stationary distribution** We know from Bisin *et al.* (2009) that, when all types survive, it must be the case that:

$$\sum_{i \in \{1, \dots, n\}} \frac{1}{\Delta V^i} > \frac{n-1}{\min_{i \in \{1, \dots, n\}} \Delta V^i}$$

Let us check whether the associated evolutionary process satisfies RAT.

To check that property, let us withdraw some type, let us say type  $j$ , from the population.

The LHS of the above condition is reduced by  $\frac{1}{\Delta V^j}$ , while the RHS is reduced by  $\frac{1}{\min_{i \in \{1, \dots, n\} \setminus \{j\}} \Delta V^i}$ . Given that  $\frac{1}{\Delta V^j} \leq \frac{1}{\min_{i \in \{1, \dots, n\} \setminus \{j\}} \Delta V^i}$ , the initial inequality is preserved. Hence we have:

$$\sum_{i \in \{1, \dots, n\} \setminus \{j\}} \frac{1}{\Delta V^i} > \frac{n-2}{\min_{i \in \{1, \dots, n\} \setminus \{j\}} \Delta V^i}$$

This means that all types  $\{1, \dots, n\} \setminus \{j\}$  still survive. Denoting  $\{1, \dots, n\} \setminus \{j\}$  by  $O$ , and  $\{1, \dots, n\}$  by  $N$ , we have:  $O \subseteq N$  and  $S(N) \cap O \subseteq S(O)$ , in conformity with Property  $\alpha$ .

Consider now Property  $\beta$ . Clearly  $S(O) \cap S(N) \neq \emptyset$  and  $S(O) \subseteq S(N)$ , in conformity with Property  $\beta$ .

Hence, by Lemma 1, the evolutionary process satisfies RAT.

**Case (2): non-interior stationary distribution** Consider now the case where types with a degree of cultural intolerance lower than the one of type  $k^*$  do not survive, whereas types with a degree of cultural intolerance higher than the one of type  $k^*$  survive.

This means that, initially, we have:

$$\sum_{i \in \{1, \dots, n\}} \frac{1}{\Delta V^i} \leq \frac{n-1}{\min_{i \in \{1, \dots, n\}} \Delta V^i}$$

We know that, for the surviving types  $i \in \{1, \dots, k^*\}$ , the following condition holds (see Section 4.2):

$$\Delta V^i > \frac{k^* - 1}{\sum_{i \in \{1, \dots, k^*\}} \frac{1}{\Delta V^i}}$$

whereas, for the non-surviving types  $i \in \{k^* + 1, \dots, n\}$ , we have:

$$\Delta V^i \leq \frac{k^* - 1}{\sum_{i \in \{1, \dots, k^*\}} \frac{1}{\Delta V^i}}$$

To check whether Properties  $\alpha$  and  $\beta$  are valid or not in that case, we will consider that some type, let us say type  $j$ , is withdrawn from the population. Two cases can arise: either the withdrawn type did not survive initially (i.e.  $j > k^*$ ), or the withdrawn type did survive initially (i.e.  $j \leq k^*$ ). We will consider those two cases successively.

Consider first Property  $\alpha$ .

Consider first when the withdrawn type did not initially survive (i.e.  $j > k^*$ ).

Property  $\alpha$  requires that those who were initially surviving and remain in the population after the withdrawal of some type must still survive after that

withdrawal. In this context, Property  $\alpha$  requires that the new marginal type, i.e. the least intolerant surviving type after type  $j$  has been withdrawn, denoted by  $K^*$ , should be such that  $K^* \geq k^*$ . To show this, let us proceed by *reductio ad absurdum*. Suppose that  $K^* < k^*$ . Assume, for instance, that  $K^* = k^* - 1$ . This means that, after the withdrawal of type  $j$ , type  $k^*$  does not survive any more. From the condition for survival of type  $k^*$  *before* the withdrawal, we have:

$$\Delta V^{k^*} \left[ \sum_{i \in \{1, \dots, k^*\}} \frac{1}{\Delta V^i} \right] > k^* - 1 \iff \Delta V^{k^*} \left[ \sum_{i \in \{1, \dots, K^*\}} \frac{1}{\Delta V^i} \right] > k^* - 2$$

If type  $k^*$  does not survive *after* the withdrawal of type  $j$ , we have:

$$\Delta V^{k^*} \left[ \sum_{i \in \{1, \dots, K^*\}} \frac{1}{\Delta V^i} \right] \leq K^* - 1 \iff \Delta V^{k^*} \left[ \sum_{i \in \{1, \dots, K^*\}} \frac{1}{\Delta V^i} \right] \leq k^* - 2$$

The LHS is the same in the two conditions. Hence those two conditions cannot hold simultaneously, and a contradiction is reached. Therefore, it cannot be the case that  $K^* = k^* - 1$  and that type  $k^*$  does not survive. The same argument can be used to show that it cannot be the case that  $K^* = k^* - x$  for any  $x > 1$  and types  $k^*$  to  $k^* - x$  do not survive. Hence it cannot be the case that  $K^* < k^*$ . If  $K^* < k^*$ , type  $k^*$  must survive, and all types  $i < k^*$  as well, so that a contradiction is reached, since the survival of all initially surviving types contradicts  $K^* < k^*$ . Hence  $K^* < k^*$  cannot occur. Therefore we have  $K^* \geq k^*$ , in conformity with Property  $\alpha$ .

Take now the case where the withdrawn type used to survive:  $j \leq k^*$ .

Property  $\alpha$  requires that those who were initially surviving and remain in the population after the withdrawal of some type must still survive after that withdrawal. In technical terms, Property  $\alpha$  requires that the number of surviving types after the withdrawal, denoted by  $K^*$ , is at least equal to  $k^* - 1$ , that is,  $K^* \geq k^* - 1$ . To proof this, let us proceed by contradiction, and suppose that  $K^* < k^* - 1$ . Assume, for instance, that  $K^* = k^* - 2$ . This means that, after the withdrawal of type  $j$ , type  $k^*$  does not survive any more.

From the condition for survival of type  $k^*$  *before* the withdrawal, we have:

$$\Delta V^{k^*} \left[ \sum_{i \in \{1, \dots, k^*\}} \frac{1}{\Delta V^i} \right] > k^* - 1 \iff \Delta V^{k^*} \left[ \sum_{i \in \{1, \dots, K^*\} \setminus \{j\}} \frac{1}{\Delta V^i} \right] > k^* - 2 - \frac{\Delta V^{k^*}}{\Delta V^j}$$

If type  $k^*$  does not survive *after* the withdrawal of type  $j$ , we have:

$$\Delta V^{k^*} \left[ \sum_{i \in \{1, \dots, K^*\} \setminus \{j\}} \frac{1}{\Delta V^i} \right] \leq K^* - 1 \iff \Delta V^{k^*} \left[ \sum_{i \in \{1, \dots, K^*\} \setminus \{j\}} \frac{1}{\Delta V^i} \right] \leq k^* - 3$$

The LHS is the same in the two conditions. These two conditions are satisfied when  $1 < \frac{\Delta V^{k^*}}{\Delta V^j}$ , which is never true. Hence a contradiction is reached. Indeed,

it cannot be the case that  $K^* = k^* - 2$  and that type  $k^*$  does not survive. But if type  $k^*$  does survive, then it is also the case for type  $k^* - x$  for  $x > 0$ , so that  $K^* < k^* - 1$  is not possible. Hence the case where  $K^* < k^* - 1$  cannot occur. Therefore we have  $K^* \geq k^* - 1$ , in conformity with Property  $\alpha$ . Thus we have:  $\{1, \dots, n\} \setminus \{j\} \subseteq \{1, \dots, n\} \implies S(\{1, \dots, n\}) \cap \{1, \dots, n\} \setminus \{j\} \subseteq S(\{1, \dots, n\} \setminus \{j\})$ .

Consider now Property  $\beta$ .

Consider first when the withdrawn type did not initially survive (i.e.  $j > k^*$ ).

Property  $\beta$  requires that the set of surviving types after the withdrawal of some type cannot exceed the initial set of surviving types, but must be included in it. This means that there cannot be any new survivor due to the withdrawal. Property  $\beta$  requires that the new marginal type, i.e. the least intolerant surviving type after type  $j$  has been withdrawn, denoted by  $K^*$ , should be such that  $K^* \leq k^*$ . To show this, let us proceed by *reductio ad absurdum*, and assume that:  $K^* > k^*$ . Take first the case where  $K^* = k^* + 1$ . Given that the *ex ante* condition for no survival for type  $k^* + 1$  is:

$$\Delta V^{k^*+1} \leq \frac{k^* - 1}{\sum_{i \in \{1, \dots, k^*\}} \frac{1}{\Delta V^i}} \iff \Delta V^{k^*+1} \left[ \sum_{i \in \{1, \dots, K^*\}} \frac{1}{\Delta V^i} \right] \leq k^*$$

while the survival of type  $k^* + 1$  *after* the withdrawal requires:

$$\Delta V^{k^*+1} \left[ \sum_{i \in \{1, \dots, K^*\}} \frac{1}{\Delta V^i} \right] > K^* - 1 \iff \Delta V^{k^*+1} \left[ \sum_{i \in \{1, \dots, K^*\}} \frac{1}{\Delta V^i} \right] > k^*$$

The two conditions:

$$\Delta V^{k^*+1} \left[ \sum_{i \in \{1, \dots, K^*\}} \frac{1}{\Delta V^i} \right] \leq k^* \text{ and } \Delta V^{k^*+1} \left[ \sum_{i \in \{1, \dots, K^*\}} \frac{1}{\Delta V^i} \right] > k^*$$

cannot be both satisfied. A contradiction is reached. It cannot be the case that  $K^* = k^* + 1$  and type  $k^* + 1$  now survives. But the same is also true for types  $k^* + x$  for any  $x > 1$ . Hence  $K^* > k^*$  cannot occur. Therefore we have  $K^* \leq k^*$ , in conformity with Property  $\beta$ .

Take now the case where the withdrawn type used to survive:  $j \leq k^*$ .

Property  $\beta$  requires that the set of surviving types after the withdrawal of some type cannot exceed the initial set of surviving types, but must be included in it. This means that there cannot be any new survivor due to the withdrawal. In this context, Property  $\beta$  requires that the number of surviving types after the withdrawal, denoted by  $K^*$ , is at most equal to  $k^* - 1$ .

To see why that property is not satisfied when considering the withdrawal of an initially surviving type, let us show that it is possible that type  $k^* + 1$ , which did not survive initially, survives after the withdrawal. The *ex ante* condition

for no survival for type  $k^* + 1$  is:

$$\Delta V^{k^*+1} \leq \frac{k^* - 1}{\sum_{i \in \{1, \dots, k^*\}} \frac{1}{\Delta V^i}} \iff \Delta V^{k^*+1} \left[ \sum_{i \in \{1, \dots, k^*\}} \frac{1}{\Delta V^i} \right] \leq k^* - 1$$

That condition can be rewritten as:

$$\Delta V^{k^*+1} \left[ \sum_{i \in \{1, \dots, k^*+1\} \setminus \{j\}} \frac{1}{\Delta V^i} \right] \leq k^* - \frac{\Delta V^{k^*+1}}{\Delta V^j}$$

Note that the condition for the survival of type  $k^* + 1$  *after* the withdrawal is:

$$\Delta V^{k^*+1} \left[ \sum_{i \in \{1, \dots, k^*+1\} \setminus \{j\}} \frac{1}{\Delta V^i} \right] > K^* - 1$$

where the new number of surviving types  $K^*$  is equal to  $k^*$  as a result of type  $k^* + 1$ 's survival.

Those two conditions are compatible, since  $\frac{\Delta V^{k^*+1}}{\Delta V^j} \leq 1$ . Hence the non-survival of type  $k^* + 1$  before type  $j$  is withdrawn from the population is fully compatible with the survival of type  $k^* + 1$  after type  $j$  is withdrawn.

Hence we have a violation of Property  $\beta$  when an initially surviving type is withdrawn from the population.

We thus have, when  $j \leq k^*$ ,  $S(\{1, \dots, n\}) \cap S(\{1, \dots, n\} \setminus \{j\}) \neq \emptyset \implies S(\{1, \dots, n\} \setminus \{j\}) \not\subseteq S(\{1, \dots, n\})$ , against Property  $\beta$ .

By Lemma 1, it follows that the evolutionary process is not rationalizable.