

Fertile or Influential? A Cultural Transmission Theory of Demographic Transition

Roman Zakharenko*

September 15, 2011

Abstract

I propose a theory of demographic transition that relates the phenomenon with changing patterns of cultural transmission. The objective of every individual is to be influential, i.e. maximize the extent to which one's views/cultural traits have been absorbed by other people. Acquisition of education makes individuals more influential but reduces the family size. In "traditional" societies that are culturally isolated from outside world, one's influence is inelastic with respect to education, hence low incentive to get educated and large family size. In modern societies with low costs of cultural transmission, one's influence is highly sensitive to own education, hence high educational effort and small family size. Empirical applications of the theory are also discussed.

Keywords: demographic transition, social influence, economic growth

JEL codes: J13, O15, Z13

*International College of Economics and Finance, State University Higher School of Economics, Moscow, Russia. Email: rzakharenko@hse.ru. Web: www.rzak.ru. I thank Maxim Arnold, Kenneth Binmore, Ronald Lee, Lesley Newson, seminar participants at NRU Higher School of Economics, participants of the Alpine Population and International Economic Association conferences, who's ideas have influenced this research.

1 Introduction

In the modern world, the *Homo sapiens* appears to be the only living species that voluntarily restricts own fertility and, for this reason, goes not maximize genetic success defined as the representation of own genes in subsequent generations. Moreover, it is currently widely held that such a voluntary restraint has emerged less than two hundred years ago, and only very recently has impacted the vast majority of the global population. For any biology-minded scholar, an explanation of the phenomenon must be related to characteristics and abilities that are unique to the *Homo sapiens*, and that began to play a prominent role in the recent century.

Beginning about thirty years ago, a number of influential works (Dawkins 1976, Cavalli-Sforza and Feldman 1981, Boyd and Richerson 1985) emphasized one such unique human ability – the ability for cultural transmission, i.e. transmission of knowledge, social norms, attitudes, and skills from one individual to another.

A number of recent studies have related the changes in the cultural transmission process to changes in fertility. Newson, Postmes, Lea, Webley, Richerson, and McElreath (2007) and Newson (2009) relate fertility decline with decreased interaction between relatives. If relatives encourage each other to have many children, high-fertility cultural traits dominate; but in the modern society, with few relatives and many strangers around, low-fertility traits evolve.

While it is theoretically plausible that an individual wants to encourage his/her relative to increase fertility (inclusive fitness hypothesis, Hamilton (1964)), it is less clear why does an individual needs to be encouraged by other relatives. Theoretical biology states that living species are genetically hard-wired to maximize genetic fitness regardless of the presence or absence of the relatives; the proposed theory does not offer a “rigorous” explanation of the deviation of modern humans from genetic-fitness-maximization paradigm.

The objective of this paper is to propose a theory of demographic transition that relates fertility reduction with the same exogenous change – transition from a society comprised mainly of relatives to the one with many non-relatives around, – but through a different mechanism which is detailed below. The proposed theory is similar in spirit to (but is not a copy of) the “memetic” theory of demographic transition proposed by Blackmore (2000).

I assume that the objective of every individual is to be *influential*, where influence is defined as the extent to which one’s cultural traits/views/memes have been absorbed by

other individuals in the next generation.¹ One can influence both own children (“vertical transmission” in the language of Cavalli-Sforza and Feldman (1981)), whose number is endogenous, and the children of others (“oblique transmission”). Likewise, children may be affected by both own parents as well as other individuals of the previous generation.

The main tradeoff faced by an individual is the tradeoff between the time spent on family and the time spent on education. The former positively affects the number of children and increases the number of individuals that one can influence,² while the latter makes one’s cultural trait more appealing (“high quality”) for a subset of individuals in the next generation.

The time allocated to education is spent on absorption of cultural traits of the past generation that are assumed to be necessary inputs for own cultural traits. Individuals optimally determine to what extent they absorb the cultural trait of each particular past individual, with more appealing cultural traits being absorbed more intensively.

In traditional societies with limited contact between non-relatives, children absorb cultural traits of their own parents regardless of the “quality” of these traits. This creates incentives for parents to save time on improvement of own traits and allocate as much as possible time on family, which results in large family size. In modern societies, learning from non-parents (and therefore influencing non-children) is nearly as easy as learning from parents, which makes demand for one’s traits more elastic with respect to their appeal and induces people to spend more time on education and thus have fewer children. In a version of the model, I show that fertility declines even when people cannot influence the outside world themselves, but their children are influenced by the outside world. Too fertile, and therefore too uneducated, parents will not be able to influence even their own children in the presence of ideological competitors from the outside world.

The rest of the paper is organized as follows. Section 1.1 discusses the mainstream theories of demographic transition. Section 1.2 discusses the empirical evidence in support of the proposed theory. Section 2 defines the general version of the model and conducts

¹Within this paper, I make an *ad hoc* assumption about the concern for influence and do not discuss the evolutionary forces that lead to emergence of such concern. Intuitively, in the evolutionary past when human psychology evolved, social influence could have had a positive effect on one’s genetic success, hence the preference for being influential. But in the modern society, the preference for influence could have turned against its original purpose of genetic success maximization.

²One could also assume that parents derive utility not only from influence on children but also from their number, which is consistent with the objectives assumed in theoretical biology. Such additional assumption is however redundant for the purposes of this paper and does not change the main results qualitatively.

preliminary analysis, while section 3 offers a detailed analysis of equilibria in various model settings. Section 4 concludes.

1.1 Mainstream “socioeconomic” theories of demographic transition

In the field of Economics, the existing literature associates the fertility decline with improving socioeconomic conditions of the households. The modern mainstream theories of demographic transition are build on contributions of Becker (1960) and Becker and Lewis (1973), the main idea of which is that children should be viewed as a consumption good which requires a time input to enjoy; fertility declines in response to increased earnings opportunities of household members (mainly women), who switch from enjoying the number of children to enjoying their “quality”, as well as consumption of other goods. A benchmark formal theory that relates fertility and income growth is Becker, Murphy, and Tamura (1990) who develop a multiple equilibria model. In this model, the low-income equilibrium is characterized by low return to human capital which, in turn, results in high fertility and zero per-capita income growth; a high-income equilibrium is associated with high return to human capital, low fertility, and rapid growth. More recent contributions in this stream of literature include Galor and Weil (1996) who argue that technological progress, which has diminished the role of muscle power and increased the role of brainpower, increased the economic opportunities of women and thus made them substitute children for consumption; Galor and Weil (2000) build a model of demographic transition in which fertility first rises and then declines in response to income growth, which is consistent with the European historic trend.

Another popular in Economics explanation of declining fertility is the “old age security hypothesis” by Caldwell (1976) who argues that when social security is poor, parents have many children, expecting that children will support them in their old age; improved social security reduces the need for children and therefore results in a lower fertility.

Finally, a third popular stream of literature attributes declining fertility to a declining child mortality rate: if parents are risk averse, they respond to better health services by sharply reducing the number of births, which results in a smaller number of surviving children. Recent examples of this school of thought include Kalemli-Ozcan (2003) and Tamura (2006).

It seems very plausible that all above mentioned factors of fertility decline – increased return to human capital (especially that of women), better social security, lower child mortality – are often associated with increased interaction of individuals with non-relatives, and thus to changing patterns of cultural transmission. Therefore, the three above factors may be correlated with unobserved social influence, rather than affect fertility *per se*. In the next section, I discuss empirical evidence that at least part of fertility decline should be attributed to the effect of outside cultural influence, rather than to the effect of socioeconomic change.

1.2 Empirical evidence

Religious isolation and fertility. A good example of cultural isolation and preserved high fertility is Amish community in the United States. They voluntarily abstain from modern communication devices such as phones and television, as well as from modern vehicles and travel opportunities. Such an abstention effectively limits the number of non-relatives with whom Amish community members interact. At the same time, Amish families are very large. Greksa (2002) estimates [marital] fertility among Old Order Amish at 7.7 children per woman, which makes Amish community one of the fastest-growing communities in the OECD countries. As citizens of the United States, they have access to all economic opportunities that their non-religious compatriots do; isolation from the outside cultural influence, rather than economic conditions, appears to be a factor of high Amish fertility.

The effect of television on fertility. A stark example of the proposed factor of fertility decline is a recent finding by La Ferrara, Chong, and Duryea (2008) who discover that the arrival of a TV channel broadcasting “soap operas” to a Brazilian municipality has a negative impact on fertility in that municipality. The content of the soap operas, arguably, does not have any “practical” knowledge that can be considered a contribution to the viewer’s human capital. A more plausible explanation of the phenomenon is that the new TV channel increases people’s exposure to the outside culture and thus changes the patterns of cultural transmission.

The noxious influence of the West. A well-known stylized fact is that fertility in the Eastern European countries declined sharply in the late 1980s and early 1990s, following political and economic liberalization. For example, fertility (births per woman) in Russia fell from 2.22 in 1987 to 1.23 in 1997, almost a twofold decline in one decade. At the same

time, these countries have experienced a large cultural change following the collapse of the Iron Curtain: previously unavailable views and fashions, labeled by the Soviet propaganda as the “noxious influence of the West”, became easily available, which has induced young people to spend time on acquisition of newly available knowledge and fashion and left less time for family. Indeed, Eastern European countries have experienced not only cultural but also economic transformations (e.g. inflation, rising unemployment, etc.) but, first, these changes began few years later than the sharp change in the fertility trend and, second, there is no clear theory of why macroeconomic instability should be related to declining fertility. Socioeconomic theories usually relate reduced fertility with improved, not deteriorated, economic conditions. Hence, newly available culture seems to be a more plausible factor of fertility change than the changing economic conditions.

2 Model

2.1 Basics

This is an overlapping-generations model, where each generation lives for two periods. Each generation is labeled by its period of birth, $t \in \{0, 1, \dots\}$. In each generation, there is a continuum G_t of individuals. It is assumed that individuals make all their decisions when young, and enjoy their utility when old. For simplicity, there is no time discounting.

Individuals of a given generation are distributed on a circle, with unitary density per unit of arc length; the size of the circle may change over generations as the population size changes. Members of two consecutive generations are linked to each other by parent-child relationship: each child has one parent; the number of children of a parent is endogenous. Formally, I define a parental operator $C : G_{t-1} \rightarrow G_t$ such that for every $i \in G_{t-1}$, $C(i)$ is the set of all children of i .

An individual $j \in G_t$ when young is influenced by (“influence” is rigorously specified in section 2.2) his parent $C^{-1}(j) \in G_{t-1}$, as well as by an exogenously defined subset $N(j) \subset G_{t-1}$ of parent’s contemporaries. Consequently, the same individual when old may influence his own children $C(j) \in G_{t+1}$ and an exogenously defined subset of children’s contemporaries. Figure 1 illustrates the geography of influence.

The rest of the model is presented in a somewhat unusual order: I first define individual constraints and then the utility function.

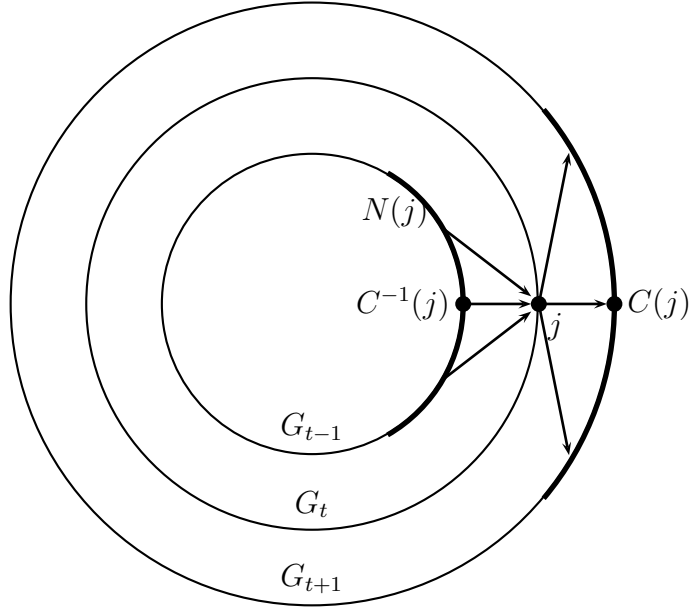


Figure 1: An illustration of the geography of influence

2.2 Constraints

Each individual $j \in G_t, \forall t$ is endowed with L units of time that can be utilized in two activities: learning and production of children. I assume that raising each child takes ν units of parent's time, so an individual j who has $n(j)$ children and spends $x(j)$ units of time learning has the following budget constraint:

$$x(j) + \nu n(j) \leq L \tag{1}$$

Individuals are allowed to choose non-integer number of children, which can be interpreted as follows. While the choice of a parent $n(j)$ may be non-integer, the actual number of children is a multinomial random variable with an expectation of $n(j)$. As defined below, parents are risk neutral with respect to the number of children, and thus the variance in the number of children is immaterial: any mean-preserving spread in the distribution of the number of children does not change parents' decisions. Also, I assume that the time burden of raising children depends on expected value $n(j)$, rather than on actual realization.

By assumption, there exists a learning lower bound: $x(j) \geq \underline{L} > 0$. This assumption reflects the fact early in their life, individuals can only learn and cannot reproduce.

Each individual $j \in G_t$ is characterized by the endogenous quality $q(j)$ of his *idea*, which

represents all knowledge possessed by the individual, and which can be viewed as an analog of human capital in conventional models of demographic transition. To develop an idea, j has to absorb, or get *influenced* by, ideas developed by the previous generation. This assumption relies on the fact that knowledge accumulation is a social activity; the most successful ideas “stand on the shoulders of giants” rather than are developed in isolation. The production function of j ’s idea is the following:

$$q(j) = \left([q(C^{-1}(j))y(C^{-1}(j), j)]^{\frac{\sigma-1}{\sigma}} + \int_{i \in G_{t-1} \setminus C^{-1}(j)} [q(i)y(i, j)]^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}} \quad (2)$$

Here $y(i, j) \geq 0$ is the influence of i on j ; $\sigma > 1$ is the elasticity of substitution between ideas.³ While past ideas $q(i)$ are given for j , the extent to which he is influenced by them, $y(i, j)$, is his control parameter, and is chosen subject to a time constraint outlined below. Also note that in (2), j ’s parent is the mass point in the continuous menu of past ideas available to j . This assumption is made to reflect the idea that parent’s influence is of the same order of magnitude as influence of everyone else.

Absorption of ideas incurs a time cost per unit of influence. I assume that j ’s time cost of learning from own parent is normalized to unity, while the cost of learning from $i \in G_{t-1} \setminus C^{-1}(j)$, denoted $\tau(i, j)$, is exogenous and may range from unity to infinity. The cost of learning from non-relatives is the key parameter of interest in this paper: its decrease implies increased outside influence, and affects fertility and accumulation of knowledge.

The total amount of time spent on learning from all sources must add up to $x(j)$:

$$y(C^{-1}(j), j) + \int_{i \in G_{t-1} \setminus C^{-1}(j)} y(i, j)\tau(i, j)di = x(j) \quad (3)$$

2.3 Objective function

Individuals derive utility from own influence on the next generation, defined as the share of own idea among all ideas absorbed by the next generation; they place a special emphasis on

³The constant elasticity of substitution function of ideas production was chosen to reflect the fact that individuals are influenced more by better ideas, and respond smoothly to changes in the quality of ideas.

their own children. Formally,

$$U_0(j) = \sum_{k \in C(j)} \frac{y(j, k)}{y(j, k) + \int_{j' \in G_t \setminus j} y(j', k) dj'} + \int_{k \in G_{t+1} \setminus C(j)} \frac{y(j, k)}{y(C^{-1}(k), k) + \int_{j' \in G_t \setminus C^{-1}(k)} y(j', k) dj'} dk \quad (4)$$

In other words, own children are a mass point in the continuous set of future individuals that may be influenced. By construction of the model, all children of a given parent face identical conditions and thus make identical choices, which enables us to write the expected (before the actual number of children is known) utility of an individual as

$$U(j) = EU_0(j) = \frac{L - x(j)}{\nu} \frac{y(j, C(j))}{y(j, C(j)) + \int_{j' \in G_t \setminus j} y(j', C(j)) dj'} + \int_{j' \in G_t \setminus C(j)} \frac{L - x(j')}{\nu} \frac{y(j, C(j'))}{y(j', C(j')) + \int_{j'' \in G_t \setminus j'} y(j'', C(j')) dj''} dj' \quad (5)$$

where $\frac{L-x(j)}{\nu}$ is the expected number of children of j and $y(j_1, C(j_2))$ is the influence of j_1 on a representative child of j_2 . Throughout the rest of the paper, we refer to expected utility $U(j)$ as “utility”.

2.4 Initial conditions

It is assumed that every individual in the initial generation G_0 has n_0 children; the qualities of ideas $q(i), i \in G_0$ may be heterogenous and drawn from a known distribution.

2.5 Equilibrium concept

The strategy of individual j is to choose educational length $x(j)$ and the learning intensity from every past individual $y(i, j), i \in G_{t-1}$ such that own utility is maximized. The strategies of j 's contemporaries are taken by j as given. The strategies of the next generation are viewed by j as a function of his quality of idea: $y(j, k) = y_{j,k}(q(j))$ and, whenever $k \in C(j)$, $x(k) = x_k(q(j))$.⁴ In other words, members of a generation t act as Stackelberg leaders with respect to members of the generation $t + 1$.

⁴The learning choice of k , $x(k)$, may depend on $q(j)$ only if k is a child of j , because otherwise j has an infinitesimal impact on k 's quality of idea and thus on the choice of education length.

2.6 Preliminary analysis and refinement

An individual choice of j 's learning intensity $y(i, j)$ depends on j 's beliefs about how his quality of idea, $q(j)$, affects his influence on the next generation, $y(j, k)$. While it is quite natural to assume that individuals with better ideas are more influential, other equilibria may arise. For example, if $j \in G_t$ believes that the next generation will learn more from those with the *worst* ideas in G_t , j will minimize $q(j)$ by choosing the minimal educational time \underline{L} and by learning the worst ideas in G_{t-1} . To refine equilibria, we characterize a rather mild sufficient condition for the “more is better” equilibrium, defined in Lemma 1. The Lemma makes use of the following notations:

$$\bar{q}_j(x) \equiv E(j)^{\frac{1}{\sigma-1}} x \quad (6)$$

$$\underline{q}_j(x) \equiv \min_{i \in G_{t-1}} \left(\frac{q(i)}{\tau_{i,j}} \right) x \quad (7)$$

$$E(j) \equiv q(C^{-1}(j))^{\sigma-1} + \int_{i \in N(j)} \left(\frac{q(i)}{\tau} \right)^{\sigma-1} di \quad (8)$$

Formula (6) can be shown to be the maximal attainable quality of idea, i.e. the result of maximization of (2) over $y(i, j)$ subject to (3). Likewise, (7) is the minimal attainable quality of idea. The parameter $E(j)$ is referred to as the *learning environment* of individual j : it shows how knowledgeable j 's potential teachers are, and how easy it is to absorb their ideas. By definition, $E(j)$ does not depend on j 's decisions and is taken by him as given.

Lemma 1 *Suppose that for some $j \in G_t$ and for every $k \in G_{t+1}$, j 's influence on k as a function of j 's quality of idea, $y(j, k) = y_{j,k}(q(j))$ is such that $\frac{\partial y_{j,k}(z)}{\partial z} \geq 0$ for $z \in [0, \bar{q}_j(\underline{L})]$, with strict inequality for own children and/or a positive-mass subset of others in G_{t+1} . Then, j 's demand for past ideas is increasing in their quality: $\frac{\partial y_{i,j}(z)}{\partial z} \geq 0$ for $z \in [0, \infty]$ and for every $i \in G_{t-1}$, with strict inequality whenever the learning cost $\tau(i, j)$ is finite.*

Proof. By definition (5), utility of j is decreasing with own education $x(j)$. By assumption of the Lemma, own quality of idea $q(j)$ positively affects influence on own children and/or a positive mass of others if $q(j) \in [0, \bar{q}_j(\underline{L})]$, thus utility increases with $q(j)$ in that region. Therefore, the indifference curves in the $\{x(j), q(j)\}$ space are upward sloping when $q(j) \in [0, \bar{q}_j(\underline{L})]$, with higher utility in the upper-left corner, as shown on Figure 2. The boundaries of the feasibility set in the same space are determined by maximizing (minimizing) (2) with

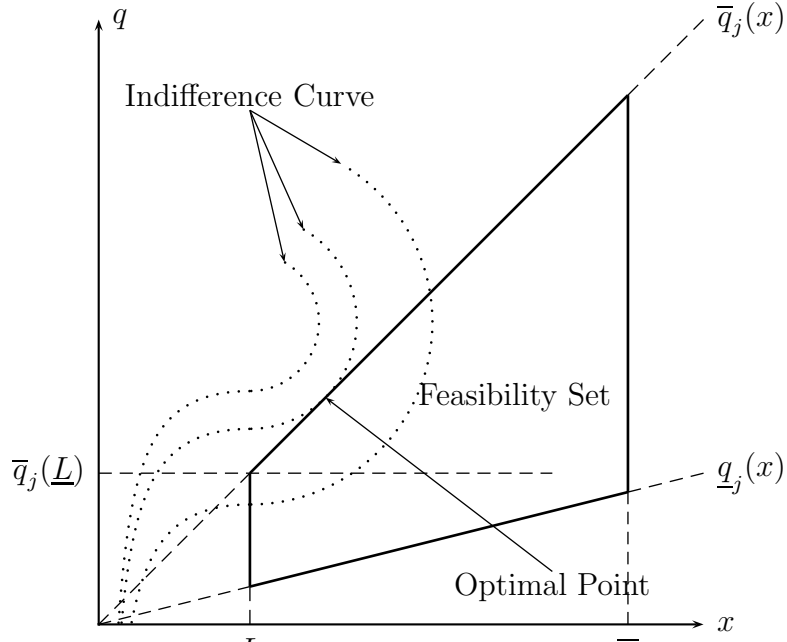


Figure 2: Utility maximization in the $\{x, q\}$ space

respect to $y(i, j)$ subject to (3), thus the set of feasible points is contained between the rays defined by (6) and (7), with $x \in [\underline{L}, \bar{L}]$, and is illustrated by bold lines on Figure 2. Maximization of utility then implies that the optimal point must be on the $\bar{q}_j(x)$ line, which is attained by maximizing (2) with respect to $y(i, j)$ subject to (3), which results in the following optimal learning intensities:

$$y(i, j) = \frac{x(j)}{E(j)} q(i)^{\sigma-1} \tau(i, j)^{-\sigma} \quad (9)$$

which implies that j 's demand for past ideas is strictly increasing in their quality as long as $\tau(i, j) < \infty$. ■

The Lemma entails the following

Corollary 1 *If the assumption of Lemma 1 is true for every individual j from a given generation t , the statement of the Lemma is true for every individual of every generation preceding t .*

The proof is done by inverse induction: if the assumption of the Lemma holds for those in G_t , they are influenced more by those in G_{t-1} with better ideas, which ensures that the assumption of the Lemma also holds for those in G_{t-1} .

Throughout the rest of the paper, we assume the “more is better” equilibrium and the demand for past ideas determined by (9).

Using (6), we can rewrite (9) and (8):

$$y(i, j) = \frac{x(j)}{E(j)} E(i) x(i)^{\sigma-1} \tau(i, j)^{-\sigma} \quad (10)$$

$$E(j) = E(C^{-1}(j)) x(C^{-1}(j))^{\sigma-1} + \int_{i \in N(j)} E(i) \left(\frac{x(i)}{\tau} \right)^{\sigma-1} di \quad (11)$$

Using (10) extrapolated one generation forward, $y(j, k) = \frac{x(k)}{E(k)} E(j) x(j)^{\sigma-1} \tau(j, k)^{-\sigma}$, and canceling out $\frac{x(k)}{E(k)}$ from the numerators and denominators of both components, we can rewrite the utility (5) as follows:

$$U(j) = \frac{L - x(j)}{\nu} \frac{E(j) x(j)^{\sigma-1}}{E(j) x(j)^{\sigma-1} + \int_{j' \in N(C(j))} E(j') x(j')^{\sigma-1} \tau(j', C(j))^{-\sigma} dj'} + \int_{j': j \in N(C(j'))} \frac{L - x(j')}{\nu} \frac{E(j) x(j)^{\sigma-1} \tau(j, C(j'))^{-\sigma}}{E(j') x(j')^{\sigma-1} + \int_{j'' \in N(C(j'))} E(j'') x(j'')^{\sigma-1} \tau(j'', C(j'))^{-\sigma} dj''} \quad (12)$$

3 Results

3.1 A primer: isolated families

Consider an extreme case in which children can only learn from their parents, that is $\tau(j, k) = \infty$ whenever $k \notin C(j)$. Then, the utility (12) boils down to $U(j) = \frac{L - x(j)}{\nu}$, i.e. is equal to the family size. When families cannot communicate with one another, we end up with a standard Darwinian/Malthusian equilibrium in which parents maximize the number of children and have the minimal level of education, $x(j) = \underline{L}$.

3.2 Outside influence

3.2.1 Specification

This section targets the phenomenon of reduced fertility due to increased outside cultural influence, as discussed in the introduction of the paper. We model “outside” influence in the following way. All members of every generation t are randomly partitioned into *artists*, with probability α , and *commoners*, with probability $1 - \alpha$. In math, $G_t = G_t^A \cup G_t^C$, such

that $|G_t^A| = \alpha|G_t|$ and $|G_t^C| = (1 - \alpha)|G_t|$. The main focus of this section is fertility of commoners; for simplicity, we assume that artists do not have family and spend all available time of education: $x(j) = L, \forall j \in G_t^A, \forall t$.

Every individual $k \in G_{t+1}$, when young, is influenced by own parent and by a subset $N(k) \subset G_t^A$ of artists of the previous generation. The set $N(k)$ is such that its mass is fixed, $|N(k)| = N, \forall k \in G_{t+1}, \forall t$, and $\Pr(j \in N(k)) = \frac{N}{|G_t^A|}, \forall j \in G_t^A$. The cost of influence of $j \in N(k)$ on k is $\tau(j, k) = \tau \in [0, \infty)$, while influence of $j \notin N(k)$ on k is prohibitively (infinitely) costly. For convenience, we assume that children of the same parent are influenced by the same set of artists: $N(k) = N(k')$ if $k, k' \in C(j)$ for some $j \in G_t$; we denote such set of artists by $N(C(j))$.

The above assumptions imply that commoners cannot influence anyone except own children: $\tau(j, k) = \infty, \forall j \in G_t^C, \forall k \notin C(j), \forall t$. Note that the case $\alpha = 0$ (that is, everyone is a commoner) is identical to the case considered in section 3.1.

In this setup, the utility of a commoner $j \in G_t^C$ is

$$U(j) = \frac{L - x(j)}{\nu} \frac{E(j)x(j)^{\sigma-1}}{E(j)x(j)^{\sigma-1} + \int_{j' \in N(C(j))} E(j')L^{\sigma-1}\tau^{-\sigma}dj'} = \frac{L - x(j)}{\nu} \frac{x(j)^{\sigma-1}}{x(j)^{\sigma-1} + R(j)} \quad (13)$$

where $R(j) \equiv \frac{\int_{j' \in N(C(j))} E(j')L^{\sigma-1}\tau^{-\sigma}dj'}{E(j)}$ is the *relative outside influence* of generation t artists on the children of j . The outside influence increases with lowered costs of cultural transmission τ and with increased span of influence of artists N . Since artists are drawn randomly from the population, their average learning environment is equal to that of the entire population, and the relative outside influence can be redefined as

$$R(j) \equiv L^{\sigma-1}\tau^{-\sigma}N \frac{\bar{E}_t}{E(j)} \quad (14)$$

where $\bar{E}_t = \frac{1}{|G_t|} \int_{j \in G_t} E(j)dj$ is the average learning environment.

Define $B(x, R) \equiv \frac{x^{\sigma-1}}{x^{\sigma-1} + R}$; it is equal to relative influence of an individual with education x on his children who face relative outside influence of R . $B(x, R)$ has the following properties:

$$B_x(x, R) = \frac{\sigma - 1}{x} B(x, R)(1 - B(x, R)) \quad (15)$$

$$B_R(x, R) = -\frac{1}{R} B(x, R)(1 - B(x, R)) \quad (16)$$

The utility (13) can be then presented as

$$U(j) = \frac{L - x(j)}{\nu} B(x(j), R(j)) \quad (17)$$

3.2.2 Effect of outside influence on fertility

The utility (17) can be shown to have a unique local maximum with respect to $x(j)$. The optimal educational length $x(R(j))$, is determined by maximization of (17) subject to $x(j) \geq \underline{L}$, which results in the following first-order condition:

$$\frac{1}{\nu} B(x(R(j)), R(j)) F(x(R(j)), R(j)) \begin{cases} \leq 0 & x(R(j)) = \underline{L} \\ = 0 & x(R(j)) > \underline{L} \end{cases} \quad (18)$$

where

$$F(x, R) \equiv (\sigma - 1) \frac{L - x}{x} (1 - B(x, R)) - 1 \quad (19)$$

Given that $B(x, R) > 0$ for every interior point x , the first-order condition of optimal education boils down to

$$F(x(R(j)), R(j)) \begin{cases} \leq 0 & x(R(j)) = \underline{L} \\ = 0 & x(R(j)) > \underline{L} \end{cases} \quad (20)$$

The function F has the following properties:

$$\begin{aligned} F_x &= -(\sigma - 1)(1 - B(x, R)) \left(\frac{L}{x^2} + \frac{L - x}{x} \frac{\sigma - 1}{x} B(x, R) \right) < 0 \\ F_R &= (\sigma - 1)(1 - B(x, R)) \frac{L - x}{x} \frac{1}{R} B(x, R) > 0 \\ F(0, R) &= \infty, F(L, R) = -1 \end{aligned}$$

We can now establish the main result of this section.

Proposition 1 *An individual acquires more education and thus has a smaller family in response to increased relative outside influence on her children.*

Proof. The effect of increasing relative outside influence $R(j)$, either due to increased mass of observed artists or due to lowered costs of learning from artists, on educational choice $x(j)$ may be found from the implicit function theorem:

$$\frac{dx(R)}{dR} = - \frac{F_R(x(R), R)}{F_x(x(R), R)} = \frac{1}{R} \frac{x(R)(L - x(R))B(x(R), R)}{L + (\sigma - 1)(L - x(R))B(x(R), R)} > 0 \quad (21)$$

for R such that $x(R) > \underline{L}$; the effect is zero for R such that $x(R) < \underline{L}$.

The family size $\frac{L-x(j)}{\nu}$ is therefore reduced by increasing $R(j)$. ■

The intuitive explanation of the result is as follows. As a community of commoners becomes more exposed to outside influence from artists, young community members become increasingly inclined to pick up cultural traits from outside of the community, which makes every particular old community member less influential. Young individuals who anticipate to become old themselves in the future respond to the shock by acquiring more knowledge, which makes them more influential, at the cost of reduced family size.

3.2.3 Evolution of learning environments

We next investigate the evolution of individual learning environments over generations. Specifically, we are interested whether within-generation heterogeneity of learning environments diminishes over generations.

Define by \mathfrak{E} the operator that maps the learning environment of a parent $E(j)$ into the learning environment of her children $E(C(j))$, given the average learning environment \bar{E}_t ; it is defined by (cf.(11))

$$\mathfrak{E}(E) \equiv Ex(R(E))^{\sigma-1} + L^{\sigma-1}\tau^{1-\sigma}N\bar{E} \quad (22)$$

where (cf.14) $R(E) = L^{\sigma-1}\tau^{-\sigma}N\frac{\bar{E}}{E}$.

The operator has the following useful properties.

Lemma 2 *Suppose the distribution of learning environments of a given generation is such that the average learning environment \bar{E} is positive and finite. Then,*

1. $\mathfrak{E}(E)$ increases with E ;
2. $\frac{\mathfrak{E}(E)}{E}$ decreases with E .

Proof.

1. Observe that $\mathfrak{E}(E)$ can be presented as a function of the relative outside influence:

$$\mathfrak{E}(E) = L^{\sigma-1}\tau^{-\sigma}N\frac{\bar{E}}{R(E)}x(R(E))^{\sigma-1} + L^{\sigma-1}\tau^{1-\sigma}N\bar{E}$$

Since $R(E)$ is inversely related to E , to prove the claim it is sufficient to show that $\frac{x(R)^{\sigma-1}}{R}$ is decreasing in its argument. Indeed,

$$\begin{aligned} \frac{d}{dR} \frac{x(R)^{\sigma-1}}{R} &= \frac{x(R)^{\sigma-1}}{R} \left(\frac{dx(R)}{dR} \frac{\sigma-1}{x(R)} - \frac{1}{R} \right) \\ &\stackrel{\text{cf. (21)}}{=} - \frac{x(R)^{\sigma-1}}{R^2} \frac{L}{L + (\sigma-1)(L-x(R))B(x(R), R)} < 0 \end{aligned} \quad (23)$$

2. Observe that

$$\frac{\mathfrak{E}(E)}{E} = x(R(E))^{\sigma-1} + \tau R(E) \quad (24)$$

An increase in E causes a decrease in R , which causes both components of (24) to decrease.

■

A straightforward corollary of Lemma 2 is that, whenever learning environments of generation t are heterogenous,

1. parents with better learning environments have larger families, and their children have better learning environments too;
2. the heterogeneity diminishes over time: dynasties that start with lower learning environments catch up over time.

3.3 Symmetric influence: overview

We next focus on the case in which all individuals have an equal opportunity to influence the next generation. In this setting, the educational choice of an individual depends on educational choices of contemporaries, hence we analyze the Nash equilibrium and its properties.

We assume that every individual $j \in G_t$ when young is influenced by, besides own parent $C^{-1}(j) \in G_{t-1}$, by a randomly selected subset $N(j) \subset G_{t-1} \setminus C^{-1}(j)$; all those in $G_{t-1} \setminus C^{-1}(j)$ are selected into $N(j)$ with equal probabilities. As in section 3.2, children of the same parent are influenced by the same set of parent's contemporaries.⁵ The cost of influence for non-relatives is $\tau(N(j), j) = \tau \geq 1$. By assumption, the mass of those in $N(j)$ is fixed at

⁵This assumption is not essential for analysis, but allows to save on notation.

$|N(j)| = N$. Therefore, the mass of those that $i \in G_{t-1}$ may influence when old is $N \frac{|G_t|}{|G_{t-1}|}$, i.e. N multiplied by the average family size.

The analysis of the general case, with an arbitrary distribution of learning environments of the initial generation, is tedious; we constrain ourselves to a simpler case when all members of a generation have the same learning environment. We later show that within-generation equality of learning environments is preserved over generations.

With the assumed symmetry of learning environments, the utility (12) of individual j , as a function of own learning choice $x(j)$, can be simplified as follows:

$$\begin{aligned}
U_j(x) &= \frac{L-x}{\nu} \frac{x^{\sigma-1}}{x^{\sigma-1} + \int_{j' \in N(C(j))} x(j')^{\sigma-1} \tau^{-\sigma} dj'} \\
&+ \int_{j': j \in N(C(j'))} \frac{L-x(j')}{\nu} \frac{x^{\sigma-1} \tau^{-\sigma}}{x(j')^{\sigma-1} + \int_{j'' \in N(C(j'))} x(j'')^{\sigma-1} \tau^{-\sigma} dj''} dj' \\
&= \frac{L-x}{\nu} B(x, R(j)) + x^{\sigma-1} \int_{j': j \in N(C(j'))} \frac{L-x(j')}{\nu} \frac{\tau^{-\sigma}}{x(j')^{\sigma-1} + R(j')} dj' \quad (25)
\end{aligned}$$

where $R(j) \equiv \int_{j' \in N(C(j))} x(j')^{\sigma-1} \tau^{-\sigma} dj'$ is the outside influence on j 's children, and $B(x, R) \equiv \frac{x^{\sigma-1}}{x^{\sigma-1} + R}$ is one's influence on own children.

The results of optimization with respect to own learning choice, $x(j)$, are qualitatively different under different values of σ , which is detailed in the following Lemma.

Lemma 3 1. *The utility $U_j(x)$ defined by (25) has at most two local maxima on $(0, L]$.*

2. *$U_j(x)$ has at most one interior local maximum (and therefore the other one, if exists, is at $x = L$).*

3. *If*

$$B(L, R(j)) \leq \frac{\sigma}{2(\sigma-1)} \quad (26)$$

the utility $U_j(x)$ has only one local maximum on $x \in (0, L]$.

Note that when $\sigma < 2$, the condition (26) holds for any values of $R(j)$, because $B(L, R(j)) \leq 1$ by definition of B .

The proof of Lemma 3 is contained in Appendix A. Lemma 3 entails the following

Corollary 2 *Any point $x^* \in (0, L)$ of utility local maximum is also the point of global maximum if and only if $U(x^*) \geq U(L)$.*

Since the utility maximization outcomes are qualitatively different at different values of σ , we analyze all possible cases in separate sections.

3.4 Symmetric influence: low σ

With $\sigma \leq 2$, there is a unique local maximum of (25) and hence, in equilibrium, all players of a given generation t play the same pure strategy with respect to own learning choice, denoted by z . An individual $j \in G_t$ who considers deviation from z then has to maximize the following utility with respect to $x \in [\underline{L}, L]$:

$$U(x) = \frac{L-x}{\nu} \frac{x^{\sigma-1}}{x^{\sigma-1} + Nz^{\sigma-1}\tau^{-\sigma}} + N \frac{L-z}{\nu} \frac{x^{\sigma-1}\tau^{-\sigma}}{z^{\sigma-1}(1+N\tau^{-\sigma})} \quad (27)$$

Denote by $\bar{N} \equiv N\tau^{-\sigma}$ a measure of outside influence, by $B(x, z, \bar{N}) \equiv \frac{x^{\sigma-1}}{x^{\sigma-1} + \bar{N}z^{\sigma-1}}$ a measure of one's influence on own children, and by $H(x, z, \bar{N}) \equiv \left(\frac{x}{z}\right)^{\sigma-1} \frac{\bar{N}}{1+\bar{N}}$ a measure of one's influence on other children; the utility (27) can then be further simplified to

$$\nu U(x, z, \bar{N}) = (L-x)B(x, z, \bar{N}) + (L-z)H(x, z, \bar{N}) \quad (28)$$

Using the fact that $B_x(x, z, \bar{N}) = \frac{\sigma-1}{x}B(1-B)$ and $H_x(x, z, \bar{N}) = \frac{\sigma-1}{x}H$, we can write the first order condition of optimal learning choice x^* as follows (cf.(18)):⁶

$$\nu \frac{dU(x, z, \bar{N})}{dx} = B(x, z, \bar{N})F(x, z, \bar{N}) + (L-z) \frac{\sigma-1}{x} H(x, z, \bar{N}) = 0 \quad (29)$$

where $F(x, z, \bar{N}) \equiv (L-x) \frac{\sigma-1}{x} (1-B(x, z, \bar{N})) - 1$ is identical to that defined by (19).

The equation (29) has a unique zero which determines the best-response learning choice $x(z, \bar{N})$. The equilibrium learning choice $z(\bar{N})$ is such that

$$z(\bar{N}) \equiv x(z(\bar{N}), \bar{N})$$

which results in the following closed-form solution for equilibrium learning intensity:

$$z(\bar{N}) = L \frac{\bar{N}(2+\bar{N})(\sigma-1)}{\bar{N}(2+\bar{N})(\sigma-1) + (1+\bar{N})} \quad (30)$$

⁶For simplicity, we assume the lower bound \underline{L} of educational choice is small enough and does not bind; the case $x = L$, i.e. zero family size, is also ignored as impossible in symmetric equilibrium.

which is increasing in \bar{N} . That is, a greater outside influence increases individual learning choice and reduces family size. The utility at the equilibrium point is

$$U(z(\bar{N}), z(\bar{N}), \bar{N}) = \frac{L - z(\bar{N})}{\nu} \quad (31)$$

i.e. is equal to the family size of a representative individual.

Next, we analyze equilibrium stability.

3.4.1 Within-generation stability of equilibrium

Lemma 4 *The Nash equilibrium described by (30) is stable, i.e. $\frac{dx(z, \bar{N})}{dz} < 1$ at the equilibrium point.*

Proof. By implicit function theorem, the requisite derivative can be found from (29) as $\frac{dx(z, \bar{N})}{dz} = -\frac{\frac{d^2U(x, z, \bar{N})}{dx dz}}{\frac{d^2U(x, z, \bar{N})}{dx^2}}$. Given the fact that $\frac{d^2U(x, z, \bar{N})}{dx^2} < 0$ at the equilibrium point, to prove the Lemma it is sufficient to show $\frac{d^2U(x, z, \bar{N})}{dx dz} + \frac{d^2U(x, z, \bar{N})}{dx^2} < 0$ at the same point. From (29),⁷

$$\begin{aligned} \nu \frac{d^2U}{dx dz} &= B_z F + B F_z - \frac{\sigma - 1}{x} H - (L - z) \frac{(\sigma - 1)^2}{xz} H \\ \nu \frac{d^2U}{dx^2} &= B_x F + B F_x + (L - z) \frac{(\sigma - 1)(\sigma - 2)}{x^2} H \end{aligned} \quad (32)$$

The sum $\frac{d^2U(x, z, \bar{N})}{dx dz} + \frac{d^2U(x, z, \bar{N})}{dx^2}$ can then be presented as a sum of four components:

$$\begin{aligned} \nu \left(\frac{d^2U(x, z, \bar{N})}{dx dz} + \frac{d^2U(x, z, \bar{N})}{dx^2} \right) &= (B_x + B_z) F + B(F_x + F_z) \\ &+ (L - z) H \frac{\sigma - 1}{x} \left(\frac{\sigma - 2}{x} - \frac{\sigma - 1}{z} \right) - \frac{\sigma - 1}{x} H \end{aligned} \quad (33)$$

In equilibrium, we have that $x = z$ and hence $B_x + B_z = \frac{\sigma - 1}{x} B(1 - B) - \frac{\sigma - 1}{z} B(1 - B) = 0$, which cancels out the first component of (33). From the definition of F , we have that $F_x = -(\sigma - 1) \frac{L}{x^2} (1 - B) - \left(\frac{\sigma - 1}{x} \right)^2 (L - x) B(1 - B)$ while $F_z = \frac{(\sigma - 1)^2}{xz} (L - x) B(1 - B)$, which implies, under $x = z$, that the second component $B(F_x + F_z) = -(\sigma - 1) \frac{L}{x^2} B(1 - B)$ is negative. It is straightforward to see that the third and fourth components are negative too, which proves the Lemma. ■

⁷Function arguments are omitted for brevity.

3.4.2 Intergenerational stability of the equilibrium

In this section, we verify whether within-generation equality of learning environments is preserved over generations after a small shock. Specifically, we assume that the learning environment of one individual⁸ deviates from that of contemporaries, and verify whether the learning environments of the deviant's successors converge to that of their contemporaries.

Suppose that generation t is partitioned into *typical individuals* and a single *deviant*. The learning environment of a typical individual is E_t , while that of a deviant is $E(j) \neq E_t$. By $e = \frac{E(j)}{E_t}$ we define the deviant's learning environment relative to that of contemporary typical individuals.

The children of generation- t typical individuals become generation= $t + 1$ typical individuals; their learning environments are then (cf.(11))

$$E_{t+1} = E_t z^{\sigma-1} (1 + N \tau^{-(\sigma-1)}) \quad (34)$$

where $z = z(\bar{N})$ is the equilibrium learning choice of a typical individual defined by (30). To calculate the learning environment of children of the deviant, we need to account for the fact that the deviant's learning choice, denoted x , differs from that of typical individuals, z .

The utility of the deviant is (cf.(28))⁹

$$\nu U(x, e) = (L - x) \frac{ex^{\sigma-1}}{ex^{\sigma-1} + \bar{N}z^{\sigma-1}} + (L - z) \frac{e\bar{N}x^{\sigma-1}}{z^{\sigma-1}(1 + \bar{N})} = (L - x)B(x, e) + (L - z)H(x, e)$$

where $B(x, e) \equiv \frac{ex^{\sigma-1}}{ex^{\sigma-1} + \bar{N}z^{\sigma-1}}$ and $H(x, e) \equiv \frac{e\bar{N}x^{\sigma-1}}{z^{\sigma-1}(1 + \bar{N})}$. The first-order condition for optimal learning choice x is (cf.(29))

$$B(x, e)F(x, e) + (L - z) \frac{\sigma - 1}{x} H(x, e) = 0 \quad (35)$$

where $F(x, e) = (L - x) \frac{\sigma-1}{x} (1 - B(x, e))$ is identical to that of (19). The first order condition uniquely defines the deviant's optimal learning choice, denoted $x(e)$. Note that, by definition of z , $x(1) = z$.

⁸alternatively, learning environments of a zero-mass subset of individuals

⁹We ignore arguments z and \bar{N} as they are held fixed in this section.

Finally, the learning environment of the deviant's children is

$$E(C(j)) = E(j)x(e)^{\sigma-1} + E_t z^{\sigma-1} N \tau^{-(\sigma-1)}$$

and, in relative terms,

$$\mathfrak{R}(e) \equiv \frac{E(C(j))}{E_{t+1}} = \frac{ex(e)^{\sigma-1} + z^{\sigma-1} N \tau^{-(\sigma-1)}}{z^{\sigma-1} (1 + N \tau^{-(\sigma-1)})} \equiv \frac{ex(e)^{\sigma-1} + z^{\sigma-1} \bar{N} \tau}{z^{\sigma-1} (1 + \bar{N} \tau)} \quad (36)$$

The system analyzed in Section 3.4 is said to be *intergenerationally stable* if deviant's children learning environment relative to that of their contemporaries, $\frac{E(C(j))}{E_{t+1}}$, is closer to unity than the deviant's learning environment relative to that of her contemporaries, $\frac{E(j)}{E_t}$. In math, the operator \mathfrak{R} is such that $\mathfrak{R}(1) = 1$ and $-1 < \mathfrak{R}'(1) < 1$.

Lemma 5 *The system analyzed in Section 3.4 is intergenerationally stable.*

Proof. The fact that $\mathfrak{R}(1) = 1$ is trivially verified, recalling that $x(1) = z$. We focus on the analysis of the first derivative.

$$\begin{aligned} \mathfrak{R}'(e) &= \frac{x(e)^{\sigma-1} + \frac{\sigma-1}{x(e)} e \frac{dx(e)}{de} x(e)^{\sigma-1}}{z^{\sigma-1} (1 + \bar{N} \tau)} \\ \mathfrak{R}'(1) &= \frac{1 + \frac{\sigma-1}{x(1)} \frac{dx(1)}{de}}{1 + \bar{N} \tau} \end{aligned} \quad (37)$$

From (35) and from the implicit function theorem, $\frac{dx(1)}{de} = -\frac{U_{xe}(z,1)}{U_{xx}(z,1)}$; we analyze the numerator and the denominator of the last expression in detail. Using the definitions of B and H , as well as (30), we obtain

$$U_{xe}(z, 1) = B_e(z, 1)F(z, 1) + B F_e(z, 1) + (L - z) \frac{\sigma - 1}{z} H_e(z, 1) = \frac{B^2(1 - B)}{1 + B} \quad (38)$$

$$U_{xx}(z, 1) = \frac{\sigma - 1}{z} \left[\frac{B^2(1 - B)}{1 + B} - B \left((1 - B) + \frac{1}{\sigma - 1} \right) \right] \quad (39)$$

First, observe that $U_{xe} - \frac{z}{\sigma-1} U_{xx} > 0$, which implies $\mathfrak{R}'(1) > 0$. Proving $\mathfrak{R}'(1) < 1$ is equivalent to proving $U_{xe} + \frac{z}{\sigma-1} U_{xx} \bar{N} \tau < 0$, which we do below. Observe that, by definition

of B , $\bar{N} = \frac{1-B(y,1)}{B(y,1)}$, and that $\tau \geq 1$.

$$\begin{aligned}
U_{xe} + \frac{z}{\sigma-1} U_{xx} \frac{1-B}{B} \tau &\stackrel{\underbrace{\leq}_{U_{xx}<0}}{\leq} U_{xe} + \frac{z}{\sigma-1} U_{xx} \frac{1-B}{B} = \frac{B(1-B)}{1+B} - (1-B)^2 - \frac{1-B}{\sigma-1} \\
&\stackrel{\underbrace{\leq}_{\sigma \leq 2}}{\leq} \frac{B(1-B)}{1+B} - (1-B)^2 - (1-B) < 0
\end{aligned}$$

■

3.5 Symmetric influence: high σ

When $\sigma > 2$, the second component of utility (25) is convex, and, according to Lemma 3, two local maxima are possible: one in the interior and one at $x = L$. There are then three qualitatively different cases.

First, when all individuals choose the interior argmaximum attained at $z(\bar{N})$ characterized by (30), and the utility of a representative individual at $x = z(\bar{N})$ is higher or equal than that at $x = L$: $U(z(\bar{N}), z(\bar{N}), \bar{N}) \geq U(L, z(\bar{N}), \bar{N})$, which is equivalent to

$$\left(\frac{\bar{N}(2 + \bar{N})(\sigma - 1)}{\bar{N}(2 + \bar{N})(\sigma - 1) + (1 + \bar{N})} \right)^{\sigma-1} \geq \frac{\bar{N}}{1 + \bar{N}}$$

In this case, the equilibrium and its properties are identical to that described in section 3.4.

Another scenario is when the utility of a representative individual is maximized at $x = L$. This case however implies that everyone chooses to have zero family size and there is no one to influence in the next generation; such outcome clearly cannot be sustained in equilibrium as everyone has an incentive to deviate and have own family.

Finally, the mixed strategy equilibrium is attained when the representative individual is indifferent between the interior and the corner utility argmaxima, and randomizes between the two. The utility of the representative individual, as a function of own learning choice x , interior learning choice of others z , probability p of the corner learning choice, and outside influence \bar{N} , can be presented as (cf.(25))

$$\nu U(x, z, p, \bar{N}) = (L-x) \frac{x^{\sigma-1}}{x^{\sigma-1} + \bar{N} [(1-p)z^{\sigma-1} + pL^{\sigma-1}]} + (1-p)(L-z) \frac{\bar{N} x^{\sigma-1}}{z^{\sigma-1} + \bar{N} [(1-p)z^{\sigma-1} + pL^{\sigma-1}]} \quad (40)$$

A detailed analysis of the mixed strategy equilibrium is tedious; we constrain ourselves to

description of the moments that are used to calculate the equilibrium values of p and $x = z$. The first moment is the first-order condition of the utility interior maximum, analogous to (29). After some simplification, it can be shown to be

$$\frac{\sigma - 1}{z}(L - z) \frac{\bar{N} [(1 - p)z^{\sigma-1} + pL^{\sigma-1}]}{z^{\sigma-1} + \bar{N} [(1 - p)z^{\sigma-1} + pL^{\sigma-1}]} - 1 + \frac{\sigma - 1}{z}(1 - p)(L - z)\bar{N} = 0$$

The second moment is equality of utility at the two local maxima, $U(z, z, p, \bar{N}) = U(L, z, p, \bar{N})$, which can be reduced to

$$\left(\frac{z}{L}\right)^{\sigma-1} - \frac{(1 - p)\bar{N}}{1 + (1 - p)\bar{N}} = 0$$

4 Conclusion

This paper develops an “influential” theory of demographic transition, according to which the decline of fertility over time occurs due to increased outside cultural influence rather than due to increased economic opportunities of parents. The objective of parents is to maximize own cultural influence on own children and on other people. With increased outside influence, children’s demand for parents’ ideas becomes more sensitive to their quality, which induces parents to spend more time on education and reduce the amount of time spent on family. It is difficult to imagine a mother of seven children attend a rock concert or play in a rock band; rock music is a high-influence entity. Access of people to this kind of activities should have an effect on fertility choices.

The model developed in this paper, for simplicity, assumes that ideas are unidimensional; in fact, there may exist a vast array of different types of ideas. Different societies may be susceptible to different types of ideas, and it may be hard to tell in advance which ideas will affect a particular society. La Ferrara, Chong, and Duryea (2008) discover that fertility in Brazil declines after the arrival of a particular television channel broadcasting particular types of shows; the same channel is likely to have little effect on African households. Sudden openness of the Soviet bloc to the Western culture in the 1980s seems to have had a sharp effect on fertility in Eastern Europe (e.g. Russia and Ukraine) but not on that in Central Asia (e.g. Uzbekistan and Tajikistan) which had experienced only a gradual fertility reduction, with no apparent change in the trend around mid-1980s. A future extension of the model might consider different types of ideas (“religions”), with high or infinite elasticity of

substitution between ideas of different religions. Then, children would be inclined to absorb the religion of parents (from both parents and other individuals of the same religion), and would have little or no interest to ideas from different religions.

The model developed in this paper predicts that high outside influence, besides reducing fertility, makes people worse off. However, reduced fertility may have a positive externality on other families by increasing the amount of resources per capita, the effect not considered in the paper. Therefore, the aggregate effect of increased outside influence on well-being may be positive.

A Proof of Lemma 3

Define a function $W_j(x)$ such that

$$\frac{dW_j(x)}{dx} \equiv \nu x^{-(\sigma-2)} \frac{dU_j(x)}{dx} = x^{-(\sigma-2)} B(x, R(j)) F(x, R(j)) + K(j) \quad (41)$$

where $K(j)$ is a positive constant and $F(x, R)$ is defined by (19). By definition, $\frac{dW_j(x)}{dx}$ has the same sign as $\frac{dU_j(x)}{dx}$ for all $x > 0$, and therefore a point $x > 0$ is a local maximum (minimum) of $W_j(x)$ if and only if it is the point of a local maximum (minimum) of $U_j(x)$. We focus on the properties of $W_j(x)$ in the following analysis.

By definition of B and F , $W_j(x)$ is strictly increasing on $(0, x^*]$, where x^* is such that $F(x^*) = 0$. Hence, we focus on studying W_j on $x \in (z^*, L]$. We start by investigating the sign of the second derivative of W_j . In equations that follow, all function arguments and subscripts are suppressed for brevity.

$$\begin{aligned} \frac{d^2W(x)}{dx^2} &= -(\sigma - 2)x^{-(\sigma-1)}BF + (\sigma - 1)x^{-(\sigma-1)}B(1 - B)F \\ &\quad - x^{-(\sigma-1)}B(\sigma - 1)\frac{L}{x}(1 - B) - x^{-(\sigma-1)}B(\sigma - 1)\left(\frac{L}{x} - 1\right)(\sigma - 1)B(1 - B) \end{aligned} \quad (42)$$

Division of (42) by $Bx^{-(\sigma-1)}$ will change neither its sign nor its zeros. We obtain

$$\begin{aligned} B^{-1}x^{\sigma-1}\frac{d^2W(x)}{dx^2} &= -(\sigma - 2)F + (\sigma - 1)(1 - B)F \\ &\quad - (\sigma - 1)\frac{L}{x}(1 - B) - (\sigma - 1)\left(\frac{L}{x} - 1\right)(\sigma - 1)B(1 - B) \end{aligned} \quad (43)$$

By adding $(\sigma - 1)B + (\sigma - 1)(1 - B) - (\sigma - 1)$, which equals zero, to the right-hand side of

(43), we obtain

$$\begin{aligned}
B^{-1}x^{\sigma-1}\frac{d^2W(x)}{dx^2} &= -(\sigma-2)F + (\sigma-1)(1-B)F - (\sigma-1)\left(\frac{L}{x}-1\right)(1-B) \\
&- (\sigma-1)B\left((\sigma-1)\left(\frac{L}{x}-1\right)(1-B)-1\right) - (\sigma-1) \\
&= -(\sigma-2)F + (\sigma-1)(1-B)F - (F+1) - (\sigma-1)BF - (\sigma-1) \\
&= -2(\sigma-1)BF - \sigma
\end{aligned} \tag{44}$$

By definition of x^* , $F(x, R(j))$ is strictly negative on $x \in (x^*, L]$ and is decreasing from 0 to -1 , therefore (44) is strictly increasing from $-\sigma$ to $2(\sigma-1)B(L, R(j)) - \sigma$ on $(z^*, L]$. Whenever (26) holds, $\frac{d^2W(x)}{dx^2}$ is strictly negative for any value of $x \in (x^*, L)$; negative second derivative implies concavity of W and uniqueness of argmaximum, which proves the third statement of Lemma 3. When (26) does not hold, there exists $x^{**} \in (x^*, L)$ such that $\frac{d^2W(x)}{dx^2} < 0$ for $x \in (x^*, x^{**})$ and $\frac{d^2W(x)}{dx^2} > 0$ for $x \in (x^{**}, L)$. Then, there may be at most one local maximum of W on $x \in (x^*, x^{**})$; another possible maximum is at $x = L$. This proves the first and second statements of Lemma 3. ■

References

- BECKER, G. S. (1960): “An Economic Analysis of Fertility,” in *Demographic and Economic Change in Developed Countries*, pp. 225 – 256. UMI.
- BECKER, G. S., AND H. G. LEWIS (1973): “On the Interaction between the Quantity and Quality of Children,” *The Journal of Political Economy*, 81(2), S279–S288.
- BECKER, G. S., K. M. MURPHY, AND R. TAMURA (1990): “Human Capital, Fertility, and Economic Growth,” *The Journal of Political Economy*, 98(5), S12–S37.
- BLACKMORE, S. (2000): *The meme machine*. Oxford University Press.
- BOYD, R., AND P. J. RICHERSON (1985): *Culture and the Evolutionary Process*. The University of Chicago Press.
- CALDWELL, J. C. (1976): “Toward A Restatement of Demographic Transition Theory,” *Population and Development Review*, 2(3/4), 321–366.
- CAVALLI-SFORZA, L. L., AND M. W. FELDMAN (1981): *Cultural Transmission and Evolution*. Princeton University Press.
- DAWKINS, R. (1976): *The Selfish Gene*. Oxford University Press.
- FAN, C. S., AND O. STARK (August 2008): “Looking at the Population Problem Through the Prizm of Heterogeneity: Welfare and Policy Analyses,” *International Economic Review*, 49, 799–835(37).
- GALOR, O., AND O. MOAV (2002): “Natural Selection and the Origin of Economic Growth,” *Quarterly Journal of Economics*, 117(4), 1133 – 1191.
- GALOR, O., AND D. N. WEIL (1996): “The Gender Gap, Fertility, and Growth,” *The American Economic Review*, 86(3), 374–387.
- (2000): “Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and beyond,” *The American Economic Review*, 90(4), 806–828.
- GREKSA, L. P. (2002): “Population growth and fertility patterns in an Old Order Amish settlement,” *Annals of Human Biology*, 29(2), 192 – 201.

- HAMILTON, W. D. (1964): “The genetical evolution of social behaviour, I,” *Journal of Theoretical Biology*, 7(1), 1 – 16.
- KALEMLI-OZCAN, S. (2003): “A stochastic model of mortality, fertility, and human capital investment,” *Journal of Development Economics*, 70, 103–118(16).
- KOHLER, H.-P., F. C. BILLARI, AND J. A. ORTEGA (2002): “The Emergence of Lowest-Low Fertility in Europe during the 1990s,” *Population and Development Review*, 28(4), 641–680.
- LA FERRARA, E., A. CHONG, AND S. DURYEA (2008): “Soap Operas and Fertility: Evidence from Brazil,” RES Working Papers 4573, Inter-American Development Bank, Research Department.
- MUNSHI, K., AND J. MYAUX (2006): “Social norms and the fertility transition,” *Journal of Development Economics*, 80(1), 1 – 38.
- NEWSON, L. (2009): “Cultural versus reproductive success: Why does economic development bring new tradeoffs?” *American Journal of Human Biology*, 21(4), 464–471.
- NEWSON, L., T. POSTMES, S. LEA, P. WEBLEY, P. J. RICHERSON, AND R. MCELREATH (2007): “Influences on communication about reproduction: the cultural evolution of low fertility,” *Evolution and Human Behavior*, 28(3), 199 – 210.
- TAMURA, R. (2006): “Human capital and economic development,” *Journal of Development Economics*, 79(1), 26 – 72.