Observational Learning and Parental Influence

by María Sáez
(2018)

James J. Heckman

THE UNIVERSITY OF CHICAGO

Econ 350, Winter 2021

1. Introduction

- Role models make others aware of alternatives which were unknown to them or thought to be beyond their capabilities.
- This informational aspect of role models, though neglected in the economics literature, has been emphasized in the psychological literature on social learning.
- In fact, family members, friends, teachers and other figures who serve as an inspiration are key elements in Bandura's (1977) social cognitive theory.
- According to this theory, learning from models or “observational learning” is the channel through which culture is transmitted across individuals and across generations (Rosenthal and Zimmerman (1978)).

- The objective of the paper is to analyze the dynamics of culture in an economy populated by families whose children see different role models, children can learn only what they see, and parents choose the extent of their influence vis-a-vis the other models.

- The main idea is that acquiring a cultural variant (or type) which is different from the one of the parents is like moving to a new location which has a different view of society.

- Children born in the new location see and consider alternatives which are different from those their parents could see and consider when they were children, and this can lead to their own moving when they become adults.

- The possibility of moving out of the family's cultural milieu is the engine of cultural change in the paper.

- The model builds on four basic tenets.

- First, the parents are a child's first and most important influence (Cavalli-Sforza and Feldman (1981), Heckmann, Stixrud, and Urzua (2006)) but not the only one (Harris (1998)).

- Second, in order to consider choosing an alternative or copying a behavior a decision maker has to be aware of it and pay attention to it (Simon (1957), Masatlioglu, Nakajima and Ozbay (2013)).

- Third, the models one pays attention to are more likely to belong to the group of people with whom one regularly associates (Bandura (1977)).

- Fourth, choices over different alternatives are best described by random choice functions (McFadden (2000), Rieskamp, Busemeyer and Mellers (2006), Gul, Natenzon and Pesendorfer (2012), Echenique, Saito and Tserenjigmid (2013) and Manzini and Mariotti (2013)).

- These four ideas are incorporated in an overlapping generations model of cultural transmission with stochastic choices.
- Assume that there is a finite number of mutually exclusive cultural variants or types and a large number of families.
- Each family consists of an adult (parent) and a child.
- The adult's type determines the environments where the child will grow and acquire its type through the observation of models.
- Families are heterogeneous and there are as many family types as types.
- A family is characterized by a set of role models, a probabilistic choice function determining the probabilities that the different role models are copied and a vector of parents' evaluations of the different types.

2. The model

- Consider a discrete time overlapping generations model with two-period-lived agents.

- The two periods correspond to childhood and adulthood.

- In every period, there is a large number of families consisting of an adult and his child.

- Each adult is of a given type $i \in T$, where $T = \{1,2, \ldots, n\}$ is the set of possible types and $n \geq 2$.

- Let $\mathcal{T}$ denote the set of all finite nonempty subsets of T.

- The adult's type is acquired during childhood through a process of observational learning.

- The type a child ends up being depends on the family it is born in.

- The family of an i-parent, is characterized by three elements:

  i. A set of role models, $M^i \in \mathcal{T}, i \in M^i$
  ii. A vector of parent's valuations $V_i = (V_i^1, \ldots V_i^n)$ where V_i^j indicates the value to an i-parent of a j-child, and
  iii. A probabilistic choice function $\rho_i: T \times \mathcal{T} \rightarrow [0,1]$ indicating the probability of acquiring each type.

- Assume that children copy only what they see and that all what is seen can be copied with positive probability:

Assumption 1. $\rho_i(j, M^i) > 0$ *for all* $j \in M^i$, $\rho_i(j, M^i) = 0$ *for all* $j \notin M^i$, *and* $\sum_{j \in M^i} \rho_i(j, M^i) = 1$.

- Assume that parents choose the extent of their influence vis-a-vis the world outside the family.

- Let $b_i \in [0,1]$ be the effort exerted by an i-parent to keep his child immune to outside influences so that with a probability equal to the effort b_i the child becomes of type i and with probability $(1 - b_i)$ the child copies a model from the set M_i according to the probabilistic choice function i.

- I shall refer to b_i as the family bias.

- Following Bisin and Verdier (2001), assume that an i-parents assign value $V_i^j \geq 0$ to a type-j child and that parents choose b_i to maximize the expected value of their child,

$$\max_{b_i \geq 0} b_i V_i^i + (1 - b_i) \sum_{j \in M^i} \rho_i(j, M) V_i^j - c(b_i). \qquad (1)$$

where $c : [0,1] \rightarrow R$ is an increasing and convex cost with $c(0) = c'(0)$ and $c'(1) \geq \bar{C}$ for some large $\bar{C} > 0$; the last assumption guarantees that parents cannot determine their children's type with certainty.

- It is also assumed that parents can only be models for their own type so that the only way to favor another type is to allowed the children to learn from someone else; i.e, choose $b_i = 0$.

- The optimal bias b_i^* solves

$$\max\{0, \Delta_i(M^i, V_i, \rho_i)\} = c'(b_i^*), \tag{2}$$

where

$$\Delta_i(M^i, V_i, \rho_i) = V_i^i - \sum_{j \in M^i} \rho_i(j, M^i) V_i^j. \tag{3}$$

Note that

$$b_i^* = b(M^i, V_i, \rho_i) = c'^{-1}(\max\{0, \Delta_i(M^i, V_i, \rho_i)\}). \tag{4}$$

- Parents who view their own type as worse than the expected role model will exert no effort.

- Parents who view their type as better will exert positive effort and this effort will be higher than the difference between the value of the family type and the expected type.

- In a Bisin and Verdier (2001) world, all parents will exert a positive effort since, by assumption, $V_i^i > V_i^i$ for all $j \neq i$.

3. The dynamics

- Let $\pi_i(j, M^i, V_i, \rho_i)$ denote the probability that a child from an i-family who observes set M_i and whose parents have valuations V_i becomes a j-adult,

$$\pi_i(j, M^i, V_i, \rho_i) = \begin{cases} b(M^i, V_i, \rho_i) + (1 - b(M^i, V_i, \rho_i))\rho_i(i, M^i) & j = i \\ (1 - b(M^i, V_i, \rho_i))\rho_i(j, M^i) & j \neq i. \end{cases} \quad (5)$$

- This process of observational learning yields a system of difference equations which describes the dynamics of the population shares of adult's types at the beginning of each period:

$$x_i(t+1) = \sum_{j=1}^{n} \pi_j(i, M^j, V_j, \rho_j) x_j(t) \;\; i = 1, 2, ..., n. \quad (6)$$

- I next analyze the long-run dynamics given by (6), but first a useful definition. Let $I : T \times \mathcal{T} \to \{0,1\}$ be the following indicator function,

$$I(j, M) = \begin{cases} 1 & j \in M \\ 0 & j \notin M, \end{cases} \quad (7)$$

and let M be the set of the sets of role models, $\mathcal{M} = \{M^1, M^2, \dots, M^n\}$.

Definition 1. (Reachability). *$M^j \in \mathcal{M}$ is reachable from $M^i \in \mathcal{M}$ if there exists at least one sequence $\{M_1, M_2, .., M_l\}$ of l elements of $\mathcal{M}$ and a mapping $m : \mathbb{N} \to T$ such that $M_k = M^{m(k)}$, $m(1) = i$, $m(l) = j$ and $\prod_{g=2}^{l} I(m(g), M^{m(g-1)}) = 1$.*

Assumption 2. *Each set in $\mathcal{M}$ is reachable from all other sets in $\mathcal{M}$.*

The set $\mathcal{M} = \{\{1,2\}, \{1,2\}, \{2,3\}\}$ does not satisfy the above assumptions. M^1 is reachable from M^3 through the sequence $\{M^3, M^2, M^1\}$ since $2 \in M^3$ and $1 \in M^2$, and hence $I(2, M^3)I(1, M^2) = 1$. M^3 is not reachable from M^1 because $I(3, M^1) = 0$ and, though $I(2, M^1) = 1$, $I(3, M^2) = 0$.

- The reachability condition together with assumption 1 guarantees that some descendants of i-parents will be j-adults, for any $i, j \in T$.

- Let the economy be described by the triplet $\varepsilon = \langle \mathcal{M}, V, \rho \rangle$ is the $V = \{V_1, \dots, V_n\}$ is the $(n \times n)$ matrix of parents' evaluations, $\rho = \{\rho_1, \dots, \rho_n\}$ is the n-vector of probabilistic choice functions and $\mathcal{M} = \{M^1, \dots, M^n\}$ is the set of sets of role models.

- The system of difference equations (6) can be written as

$$x(t+1) = x(t)\Gamma(\mathcal{E}), \tag{8}$$

where $\Gamma(\mathcal{E})$ is a $(n \times n)$ matrix with elements

$$\gamma_{ij}(\mathcal{E}) = \pi_i(j, M^i, V_i, \rho_i). \tag{9}$$

- Since Γ is a stochastic matrix (all entries are non-negative and all its row sums are unity) (8) is a Markov chain and I can apply standard results to characterize the long-run distribution of variants in the population.

- In particular, if Γ is such that the process is aperiodic and irreducible there exist a unique stationary distribution x^* so that,

$$x^* = x^*\Gamma(\mathcal{E}) = x(\mathcal{E}) \tag{10}$$

moreover,

$$\lim_{t\to\infty} x(t) = x^* \tag{11}$$

for any initial distribution.

- The process described by (8) is aperiodic if $\gamma_{ii} > 0$ for all i.

- A sufficient condition for this to hold is that children become like their parents with positive probability.

- This will always be the case if, as assumed, $i \in M^i$.

- Irreducibility means in our context that some descendants of i-individuals will be j-individuals after a finite number of generations, and this is true for any $i, j \in T$.

- This will be the case under assumption 1 if $\mathcal{M}$ is fully reachable.

- Proposition 1 below characterizes the unique long-run distribution of variants when the process is aperiodic.

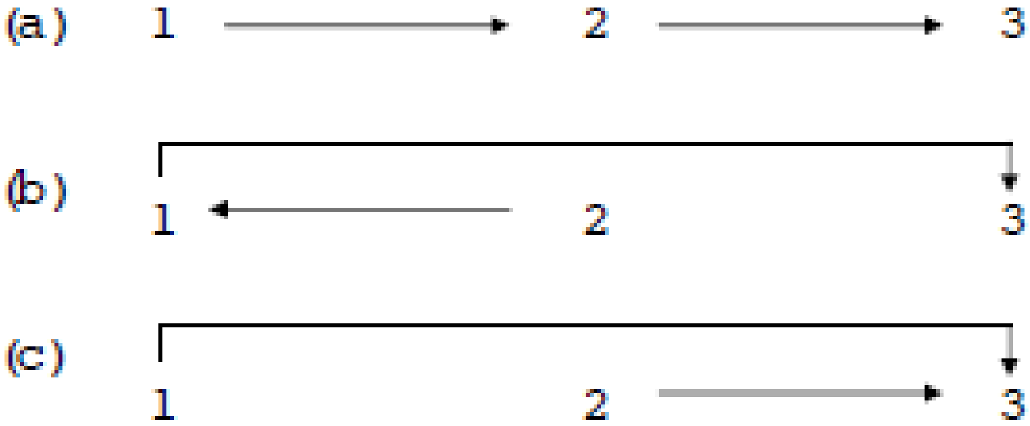
- The proposition is a new application of a result from Freidlin and Wentzel (1984) which uses a particular type of directed graphs, z-trees, to characterize the long-run distribution.

- Intuitively a-tree indicates how a state $z \in T$ of a finite Markov chain (a type in our application) can be reached from any other state without passing through any state more than once.

- Formally, a z-tree is a collection of arrows between elements of T such that i) every element $i \in T|\{z\}$ is the origin of one and only one arrow that leads to some other state $j \in T$, ii) there is a unique path starting in i that leads to z, and iii) there are no closed loops.

- Figures 1 shows all possible 3-trees for a three-type world.

Figure 1: 3-trees when $T = \{1,2,3\}$



- Freidlin and Wentzel (1984) associate to the arrow linking i with j, to be denoted $(i \to j)$, the probability γ_{ij} that the transition occurs, to each z-tree the product of the probabilities of all its arrows and to each state $z \in T$ the sum of all the numbers assigned to all its z-trees.

- Let q_z be the resulting number,

$$q_z = \sum_{h \in W_z} \prod_{i \to j \in h} \gamma_{ij}, \tag{12}$$

where W_z is the set of all z-trees and h is a z-tree. Substituting (5) in (12) and multiplying and dividing by $(1 - b_z^*)$ leads to,

$$q_z = \frac{Q(z, \mathcal{M}, \rho)}{1 - b_z^*} \prod_{i \in T} (1 - b_i^*), \tag{13}$$

where

$$Q(z, \mathcal{M}, \rho) = \sum_{h \in W_z} \prod_{k \to j \in h} \rho_k(j, M^k). \tag{14}$$

- Note that (13) will be positive if assumption 1 holds and the reachability condition is met.

- For instance, when $T = \{1,2,3\}$ and $\mathcal{M} = \{\{1,2,3\}, \{1,2\}, \{1,3\}\}$, the only 3-tree (see Figure 1) with a positive value is (b) since both $\rho_2(1, M^2)$ and $\rho_1(3, M^1)$ are positive.

- The other two trees, (a) and (c), have an arrow linking 2 and 3, but $\rho_2(3, M^2) = 0$ because $3 \notin M^2$.

- Aperiodicity and irreducibility guarantee the existence of a unique stationary distribution x^*.

- Let q be the n-dimensional vector which has q_i as its i-th element.

- Freidlin and Wentzel (1984, Lemma 3.1) show that if (8) is aperiodic,

$$x_i^* = \frac{q_i}{\sum_{k \in T} q_k}, \qquad i = 1, 2, ..., n. \tag{15}$$

PROPOSITION 1. *Suppose that assumptions 1 and 2 hold. Then,*

$$x_i^* = x_i(\mathcal{E}) = \frac{Q(i, \mathcal{M}, \rho)(1 - b_i^*)^{-1}}{\sum_{j \in T} Q(j, \mathcal{M}, \rho)(1 - b_j^*)^{-1}} \qquad i = 1, 2, ..., n., \tag{16}$$

where $b_i^* = b(M^i, V_i, \rho_i)$ *is given by (4).*

- It is clear from (16) that increases in b_i^* increase i's long share, provided all other families' biases remain unchanged.

- Interventions that change the importance of parents relative to other role models, like changing the age requirement for compulsory school, will have long-run effects only if they affect different family types to different extents.

- Note that $Q(i, \mathcal{M}, \rho)$ depends neither on M^i nor on V_i or ρ_i because these only determine the outflow from i whereas the i-tree considers only inflows to i.

- Note also that $b(M^i, V_i, \rho_i)$ only depend on variables related to family i.

- This implies that including some new element j in M_i without changing any other set of role models will influence b_i^* but no effect on $Q(i, \mathcal{M}, \rho)$ while changing some other family's set of role models will affect $Q(i, \mathcal{M}, \rho)$ but not b_i.

Example 1. *Assume that* $T = \{1,2,3\}$, $V_1 = (V,1,1)$, $V_2 = V_3 = (1,1,1)$ *with* $V \in (0, 5/2)$, $c(b) = b^2/2$, *and that children copy each variant in their set of role models with the same probability. Consider two alternative sets of sets of role models which differ only in* M^1: $\bar{\mathcal{M}} = \{\{1,2\},\{1,2,3\},\{2,3\}\}$ *and* $\hat{\mathcal{M}} = \{\{1,2,3\},\{1,2,3\},\{2,3\}\}$. *Type 2 and 3 parents exert no effort because all variants are similar from the parents' point of view. Type 1 parents prefer their own type* ($V > 1$) *and will choose a positive bias* ($b_1^* > 0$). *The respective long run shares of variant 1 are given by:*

$$\bar{x}_1^* = x_1(\bar{\mathcal{E}}) = \frac{2(1-\max\{0,(V-1)/2\})^{-1}}{2(1-\max\{0,(V-1)/2\})^{-1}+5}, \tag{17}$$

and

$$\hat{x}_1^* = x_1(\hat{\mathcal{E}}) = \frac{3(1-\max\{0,2(V-1)/3\})^{-1}}{3(1-\max\{0,2(V-1)/3\})^{-1}+10}. \tag{18}$$

$x_1(\bar{\mathcal{E}}) \gtreqless x_1(\hat{\mathcal{E}})$ *whenever* $V \lesseqgtr 13/7$.

- Figure 1 shows x_1^* (solid), b_1^* (dashed) and $\pi_1(1,.)$ (dotted) for the example above.

- The thick lines correspond to $\bar{\varepsilon}$ and the thin ones to $\hat{\varepsilon}$.

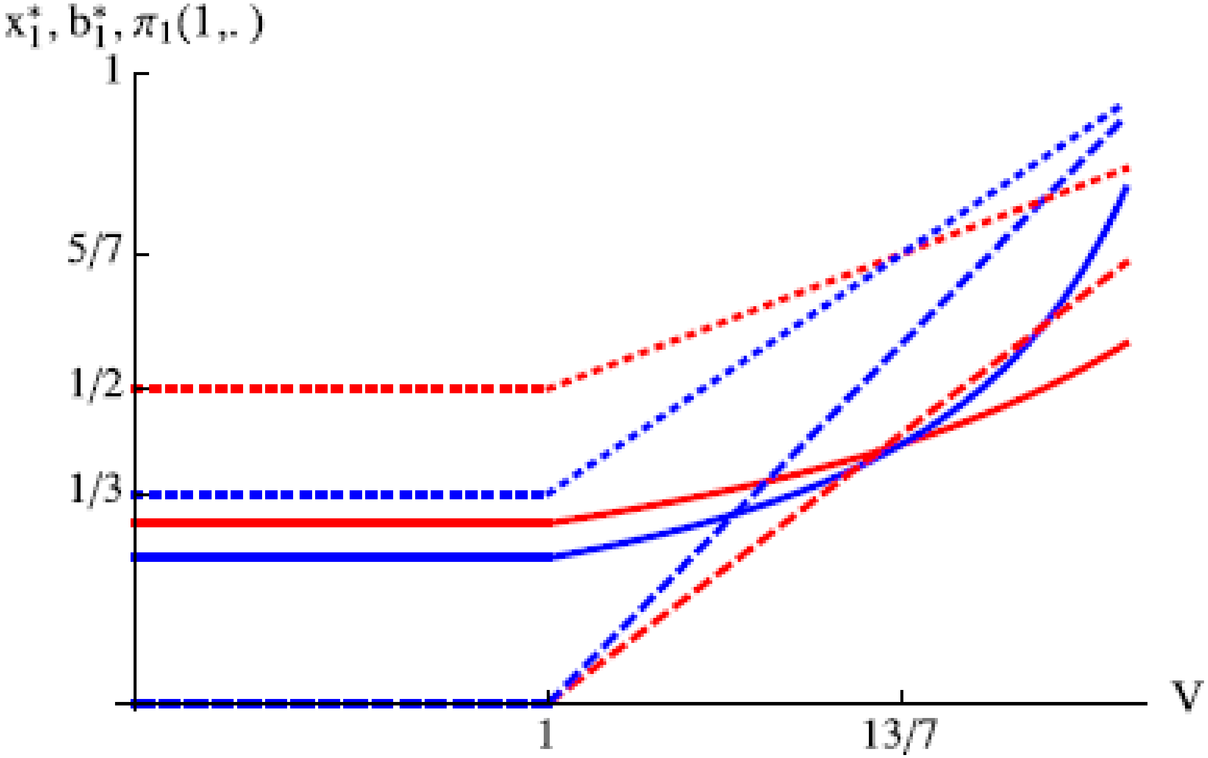
- Note that when $V \leq, b_1^* = 0$ for both, $\hat{\varepsilon}$ and $\bar{\varepsilon}$, and x_1^* is independent of V in both cases.

- As V increases beyond 1, type 1 parents start exerting effort.

- This effort is larger under $\hat{\varepsilon}$ than under $\bar{\varepsilon}$ because their children are less likely to copy the family trait in the former than in the latter (1/3 < 1/2).

- Note that when V = 13=7 the probability that children become like the parents, $\pi_1(1,.)$, is the same under the two information structures, though parental effort is different, and the long run equilibrium share is the same.

Figure 2: Example 1



- In this example the presence of the new alternative in M_1 reduces the quality (from the parents' viewpoint) of the set of role models when $V > 1$.
- As a result, type 1 parents exert more effort to make more likely that children acquire the family's preferred type.
- Whether this higher effort is enough to compensate the negative effect the new model has on the children depends on V.
- For small V's the change leads to a decrease in x_1^*.
- The effect is the opposite for large V's.

- This result can explain why some parents find it optimal to limit the contact with other groups or are so careful in the choice of peers.

- A well-known example of the former are the Amish of North America who advocate a simple lifestyle with very strict codes of behavior.

- For instance, private use of cars is mostly banned because it would, according to Amish Country News, "quicken the pace of their life, erase geographical limits, weaken social control, and eventually ruin their community."

4. Comparative statics

- Focus on choice probabilities that satisfy the axiom of independence of irrelevant alternatives (IIA) and on sets with mutual observation (MO).

- Mutual observation requires that if a mathematician's child is (is not) aware of the existence of farmers, then a farmer's child is (is not) aware of the existence of mathematicians.

- More formally,

Definition 2. (Mutual observability (MO)). *Observation is mutual between M^i and M^j if $I(j, M^i) = I(i, M^j)$.*

Assumption 3. *All pair of sets in $\mathcal{M}$ satisfy mutual observability.*

- MO is not satisfied in $\mathcal{M} = \{\{1,2,3\},\{1,2,3\},\{2,3\}\}$ since $I(3, M^1) \neq I(1, M^3)$.

- Mutual observability rules out several configurations as well as the comparative statics we can perform.

- For instance, if $T = \{1,2,3\}$ either $M^i = T$ for all i or one type observes all other types and the other two observe the former, and if we add, as we did in example 1, 3 to M^1, we should also add 1 to M^3.

- Mutual observation will hold if we think of types as geographical locations and assume that one cannot see without being seen.

- The axiom of independence of irrelevant alternatives requires that the relative probabilities of choosing one alternative over another does not depend on the choice set (constant ratio rule), provided that both alternatives are included in the set.

- More formally,

Definition 3. (Independence of irrelevant alternatives (IIA)). *The probabilistic choice function* $\rho_i : T \times \mathcal{T} \to [0,1]$ *satisfies the axiom of independence of irrelevant alternatives if for any pair* $j, k \in T$ *and any two sets* $X, Y \in \mathcal{T}$ *containing* j *and* k*,*

$$\frac{\rho_i(j,X)}{\rho_i(k,X)} = \frac{\rho_i(j,Y)}{\rho_i(k,Y)}. \tag{19}$$

- The axiom requires that the comparison between any two alternatives is not affected by the presence of other options.

- To see the implications of IIA assume that $T = \{1,2,3\}$ and that in any set with only two elements the decision maker choses each with probability 1=2.

- If IIA holds, the same decision maker should choose each alternative with probability 1/3 when all three alternatives are in the set.

- IIA hold, for instance, in example 1.

- IIA will be violated if two of the options are "similar" to each other and distinct from the third (Debreu (1960)), for instance if 2 and 3 are two types of engineers and 1 is a singer.

- In this case we should have $\rho_i(1, T) = 1/2$ and $\rho_i(2, T) = \rho_i(3, T) = 1/4$.

- IIA will also fail when there are two "extreme" options and a "compromise" one (Simonson and Tversky (1992)), for instance if 1 is a mathematician, 2 is an economist and 3 is a historian.

- If economics is viewed as a good compromise between mathematics and history, we should have $\rho_i(2, T) = 1/2$ and $\rho_i(1, T) = \rho_i(3, T) = 1/4$.
- In what follows I analyze the implications of IIA and MO for the long-run distribution.
- It is important to remark that proposition 1 only requires reachability and can accommodate all the violations mentioned above as well as non mutual observation.

Luce (1959) shows that the axiom of independence of irrelevant alternatives implies the so called Luce rule: For every alternative $j \in T$ there exist a strictly positive (Luce) value v_j such that the probability that i is chosen when the choice set is $M \in \mathcal{T}$ is equal to

$$\rho^L(i, M, v) = \frac{I(i, M)v_i}{\sum_{j \in M} v_j}, \tag{20}$$

where $v = (v_1, .., v_n)$.

Assumption 4. $\rho_i(j, M) = \rho^L(j, M, v)$ *where* ρ^L *is given by (20).*

- Luce rule (20) has had a long tradition in psychology and economics and is object of revived interest (see Gul, Natenzon and Pesendorfer (2012) and Echenique, Saito and Tserenjigmid (2013) for recent extensions of the rule).
- The Luce values are usually interpreted as the variants' attractiveness since variants with higher values are selected, under the rule, with higher probability (see Gul, Natenzon and Pesendorfer (2012)).
- Hereafter an economy will be characterized by the triplet $\varepsilon = \langle \mathcal{M}, V, v \rangle$, where $\mathcal{M} = \{M^1, \ldots, M^n\}$ are the sets of role models, $V = (V_1, \ldots, V_n)$ are the parents' evaluations and $v = (v_1, \ldots, v_n)$ are the Luce values.

PROPOSITION 2. *Assume that the economy $\mathcal{E} = \langle \mathcal{M}, \mathbf{V}, v \rangle$ satisfies assumptions 1,2,3 and 4. Then,*

$$x_i^* = x_i(\mathcal{E}) = \frac{v_i(1-b_i^*)^{-1}\sum_{k\in M^i} v_k}{\sum_{j=1}^n v_j(1-b_j^*)^{-1}\sum_{k\in M^j} v_k} \qquad i = 1, 2, ..., n. \tag{21}$$

where $b_i^* = b(M^i, V_i, \rho^L)$ is given by (4) and ρ^L by (20).

Proof. See Appendix.

Proposition 2 states that the long-run share of each variant i is proportional to the ratio,

$$\frac{v_i \sum_{k\in M^i} v_k}{1-b_i^*}, \tag{22}$$

so that variants with higher Luce values, sets of role models with variants with higher aggregate Luce value and larger family biases will have larger long-run shares.

- Note that Proposition 2 allows for any configuration of parent's evaluations V.

- I next focus on three simple cases which are easy to analyze: unbiased, positively biased and negatively biased parents.

- I say that an i-parent is unbiased if $V_i^j = V \geq 0$ for all j and denote the vector of such valuations by V^u.

- Unbiased parents consider all variants as similar and decide not to interfere in their children's choices.

- An i-parent is positively biased towards variant k if he considers variant k as superior to all other variants.

- This is captured by assuming that $V_i^k = V > 1$ and $V_i^j = 1$ for all $j \neq k$ and denote the vector of such valuation by V^{k+}. A special case of a positive bias is the family bias. In this case the positive bias is towards the own type ($k = i$).

- This is the case considered by the literature on cultural transmission in the tradition of Bisin and Verdier (2002).

- Finally, an i-parent is negatively biased towards variant k if he considers type k as inferior to all other types.

- I capture this by assuming that $V_i^k = 1$ and $V_i^j = V > 1$ for all $j \neq k$ and denote the vector of such valuation by V^{k-}.

LEMMA 1. *Assume that* $V_i \in \{V^u, V^{1+}, .., V^{n+}, V^{1-}, .., V^{n-}\}$. *Then,*

i) $b(M^i, V_i, \rho_i) > 0$ *when*

a) $V_i = V^{i+}$ *or*

b) $V_i = V^{k-}$, $k \in M^i$ *and* $k \neq i$.

ii) $b(M^i, V_i, \rho_i) = 0$, *otherwise.*

- According to the previous lemma only family-biased parents and parents who are negatively biased towards a variant which is not the family's will to exert positive effort to transmit the family trait.

- All other biases lead to zero effort because parents prefer their children to learn from better models.

- Parents cannot serve as models of types which are different from their own and the best they can do when there is a better type out there is to choose bi = 0.

- This seemingly permissive parenting style is different from the one of unbiased parents who choose zero effort because all variants are viewed as similar and parents are indifferent between their type and the expected model.

- The easiest comparative static results are those with unbiased parents since in this case all parents exert zero effort and changes in the Luce values only affect children’s copying probabilities.

- In this case, the effect of an increase in a variant's Luce value will always increase that variant's long-run share and can also lead to the increase of some other variants' shares, as stated in the following corollary.

COROLLARY 1. *Assume that proposition 2 holds and that all parents are unbiased. Then,*

$$x_i^* = x_i(\mathcal{E}) = \frac{v_i \sum_{k \in M^i} v_k}{\sum_{j=1}^{n} v_j \sum_{k \in M^j} v_k} \qquad i = 1, 2, ..., n, \tag{23}$$

and

$$\frac{\partial x_i(\mathcal{E})}{\partial v_j} > 0, \tag{24}$$

when $i = j$, *or if* $i \neq j$ *when* $i \in M^j$ *and*

$$\frac{\sum_{s \in M^j} v_s}{\sum_{j=1}^{n} v_j \sum_{k \in M^j} v_k} \sum_{s \in M^i} v_s < \frac{1}{2}. \tag{25}$$

Condition (51) is more likely to be satisfied by those variants in M^j which have sets of role models with low aggregate Luce value. Note, that under mutual observation,

$$\sum_{s \in M^i} v_s = \sum_{s=1}^{n} I(s, M^i) v_s = \sum_{s=1}^{n} I(i, M^s) v_s, \tag{26}$$

and

$$\sum_{s=1}^{n} I(i, M^s) = \sum_{s=1}^{n} I(s, M^i) = \left| M^i \right|, \tag{27}$$

where $|M^i|$ is the cardinality of M^i. Therefore, the number of elements in M^i is also the number of sets in $\mathcal{M}$ that contain i, and $|M^i|$ is a proxy of i's visibility. The sum $\sum_{s \in M^i} v_s$ will be smaller the less visible variant i is and the less attractive the variants in M^i are. Changes in the attractiveness of a variant (i.e, changes in its Luce value) can result in an increase not only of the equilibrium share of that variant but also of the less visible variant in its set of role models. This is illustrated by the following simple example.

Example 2. *Assume* $\mathcal{M} = \{\{1,2\},\{1,2,3\},\{2,3,4,5\},\{3,4,5\},\{3,4,5\}\}$, *unbiased parents and* $v = \{1, v_2, 1, 1, 1\}$. *It follows from (21) that,*

$$x_1^* = \frac{1 + v_2}{10 + v_2(4 + v_2)}, \tag{28}$$

which is increasing in v_2 *when* $v_2 < \sqrt{7} - 1 \approx 1.65$,

$$x_2^* = \frac{v_2(2 + v_2)}{10 + v_2(4 + v_2)}, \tag{29}$$

which is increasing in v_2 *for all* $v_2 > 0$, *and*

$$x_3^* = \frac{3 + v_2}{10 + v_2(4 + v_2)}, \tag{30}$$

which is decreasing in v_2 *for all* $v_2 > 0$.

When all families have the same set of role models and variants differ only in the Luce values, the long-run share (21) simplifies to,

$$x_i(\mathcal{E}) = \frac{v_i}{\sum_{j=1}^{n} v_j} \qquad i = 1, 2, .., n. \tag{31}$$

Once the differences in information about the available opportunities disappear, the only determinant of the long-run distribution are the Luce values.

When all variants have the same Luce value ($v_i = \bar{v}$ for all i) and all parents are unbiased, the long-run share (21) can be written as,

$$x_i(\mathcal{E}) = \frac{|M^i|}{\sum_{j=1}^{n} |M^j|} \qquad i = 1, 2, ..., n, \tag{32}$$

and types with larger sets of models will have larger long-run shares.

- Compare, for instance, a star structure and a circular city with a center.
- In the star, the center (variant n) can see all variants and the $(n-1)$ variants in the periphery can only see the center.

- In the circular city each variant in the periphery can see the two direct neighbors as well as the center. In the first case,

$$x_n^* = \frac{n}{3n-2}, \tag{33}$$

while in the second,

$$x_n^* = \frac{n}{5n-4}. \tag{34}$$

- In both cases the majority is in the periphery, but the center is more populated in the star.
- As n becomes larger (33) tends to 1=3 while (34) tends to 1=5.
- Under MO if one element, let us say j, is added to Mi, then i has to be added to M_j.
- The following proposition analyses the effect of such a change on the long-run distribution (21).

COROLLARY 2. *Assume that proposition 2 holds for* $\mathcal{E} = \langle \mathcal{M}, \mathbf{V}, v \rangle$ *and that all parents are unbiased. Let* $\hat{\mathcal{E}} = \langle \hat{\mathcal{M}}, \hat{\mathbf{V}}, \hat{v} \rangle$ *where* $\hat{M}^i = M^i \cup \{j\}$, $\hat{M}^j = M^j \cup \{i\}$ *and* $\hat{M}^s = M^s$ *for all* $s \neq i, j$, $\hat{\mathbf{V}} = \mathbf{V}$ *and* $\hat{v} = v$. *Then,*

1) $x_s(\hat{\mathcal{E}}) < x_s(\mathcal{E})$ *for all* $s \neq i, j$

2) $x_s(\hat{\mathcal{E}}) \leq x_s(\mathcal{E})$ *for* $s = i, j$ *whenever* $x_s(\mathcal{E}) \geq 1/2$, *otherwise,* $x_s(\hat{\mathcal{E}}) > x_s(\mathcal{E})$.

Proof. See Appendix.

- The proposition shows that either i, j or both will have larger shares and that all other variants are negatively affected by the change.
- Moreover, if only one type increases its share, it will be the one with the smallest initial long run share.
- The following simple example illustrates this result.

Example 3. *Assume* $T = \{1, 2, 3, 4, 5\}$, $v = \{1, 1, v_3, 1, 1\}$, *unbiased parents and that initially* $\mathcal{M} = \{\{1, 2\}, \{1, 2, 3\}, \{2, 3, 4, 5\}, \{3, 4, 5\}, \{3, 4, 5\}\}$. *Making 1 and 3 mutually known will lead to a fall in the shares of all variants but variant 1 whenever* $v_3 \geq 2\sqrt{2}$. *Otherwise both the share of 1 and that of 3 will increase while all other shares will decrease.*

- Corollary 2 implies that opening a contact with a variant which is, from the children's point of view less attractive (with a lower Luce value), more isolated (with a smaller set of role models) and/or in contact with less attractive variants (have low Luce values) can lead to a decline in the equilibrium value of all variants but the less attractive and isolated one.

- This predicts that “unattractive” variants will increase their shares when entering in contact with more attractive ones and that integration policies trying to exploit the exemplary effect of good variants may end up being counterproductive.

- This can explain why variants like drug use or membership in some sects do not disappear but increase their shares when entering in contact with other variants, the same way that a bad apple spoils the bunch.

- Corollary 1 shows that increases in the Luce value of a variant will always increase its share if parents are unbiased.

- This is not necessarily the case when parents are biased.

- In this case the share of a variant i can be decreasing in v_i, as we see in the following simple example.

Example 4. *Assume* $\mathcal{M} = \{\{1,2,3\},\{1,2,3\},\{1,2,3\}\}$, $v = \{1, v_2, 1\}$, *all parents are negatively biased against variant 2* ($V_i = (V,1,V)$ *for all* i, *and* $V > 1$) *and* $c(b) = c^2(b)/2$. *Families 1 and 3 will exert the same positive effort (equal to* $v_2(V-1)/(2+v_2)$*). Type 2 families will exert no effort. Variant's 2 equilibrium share*

$$x_2^* = \frac{v_2}{2(1 - v_2(V-1)(2+v_2)^{-1})^{-1} + v_2}, \tag{35}$$

is increasing (decreasing) in v_2 *for* $v_2 < \bar{v}(V)$ ($v_2 > \bar{v}(V)$), *where* $\bar{v}(V)$ *is the unique* v_2 *which solves,*

$$V = \frac{2(2 + 4v_2 + v_2^2)}{v(4+v_2)}. \tag{36}$$

For instance $\bar{v}(14/5) = 1$ *and when* $V = 14/5$ *(35) reaches a maximum at* $v_2 = 1$. *Increases in* v_2 *beyond 1 lead to a lower equilibrium share of variant 2.*

- The driving force behind this result is the parents' effort. As children are more likely to copy the variant disliked by all parents, type 1 and 3 increase their effort to transmit their own variant while type 2 parents do not exert any effort because the family variant is the least preferred one.

- The overall effect of an increase in v_2 is a decrease in x_2^* when v_2 is large enough, i.e., $v_2 > \bar{v}(V)$.

- Similar results can be obtained with family biased parents. In this case increases in v_2 are accompanied by increases in b_1^* and b_3^* (since now it is less likely that the child copies the family type) and a fall in b_2^* (since now it is more likely that the child copies the family type from the models).

- It is then possible to construct examples in which and increases (a fall) in the Luce value of a variant is accompanied by a fall (an increase) in its long run share.

- As example 4 shows public policies that agree with the parents may end-up back-firing since parents "relax" and this can lead to an increase of the undesirable variable.

- Think of variant 2 as some bad trait (drug consumption) and a public education campaign which reduces its appeal (reduction in v_2).

- If initially v_2 is larger than v(V), the campaign will end up with more drug addicts.

5. Conformism

- So far, it has been assumed that the probabilistic choice function describing children's learning is independent of the actual distribution of types in the population and depends only on the “attractiveness” of the different types as captured by the Luce values.
- This assumption implies that the dynamic system is a Markov chain.
- This, together with full teachability, guarantees cultural diversity i.e; the coexistence of different cultural variants in steady state.
- This result does survive the introduction of conformism, provided that the latter is not too strong.

For sake of tractability assume that all children have the same sets of role models ($M^i = T = \{1, .., n\}$ for all i), that parents are unbiased ($b_i^* = 0$ for all i) and that the the probability of copying variant $i \in T$, $\rho(i, T)$, is now equal to

$$\rho(i, x) = \frac{v_i x_i^\alpha}{\sum_{j=1}^n v_j x_j^\alpha}, \tag{37}$$

where $x = (x_1, .., x_n)$ is the vector of population shares, $\alpha > 0$ is a conformism parameter and $v_i > 0$ is variant i's intrinsic attractiveness. When popularity does not matter, i.e; $\alpha = 0$, (37) corresponds to Luce rule (20) and the long distribution is given by (31).[12]

- The choice function (37) incorporates two ideas.

- The first one is that cultural variants differ in their attractiveness and that these differences, which are captured by the v-values, make that ceteris paribus, some variants are more likely to be copied than others.

- The second ideas is that individuals often use the frequency of the different variants in the population to evaluate their worth.

- This introduces a frequency-dependent bias (Boyd and Richerson, 1985) which tends to favor the more frequent variants.

- This conformity bias is captured by the second term in the numerator.

- The transmission process described by (37) for $\alpha > 0$ is conformist because when all variants are equally attractive ($v_i = v$ for all i), more frequent variants are copied with higher probability.

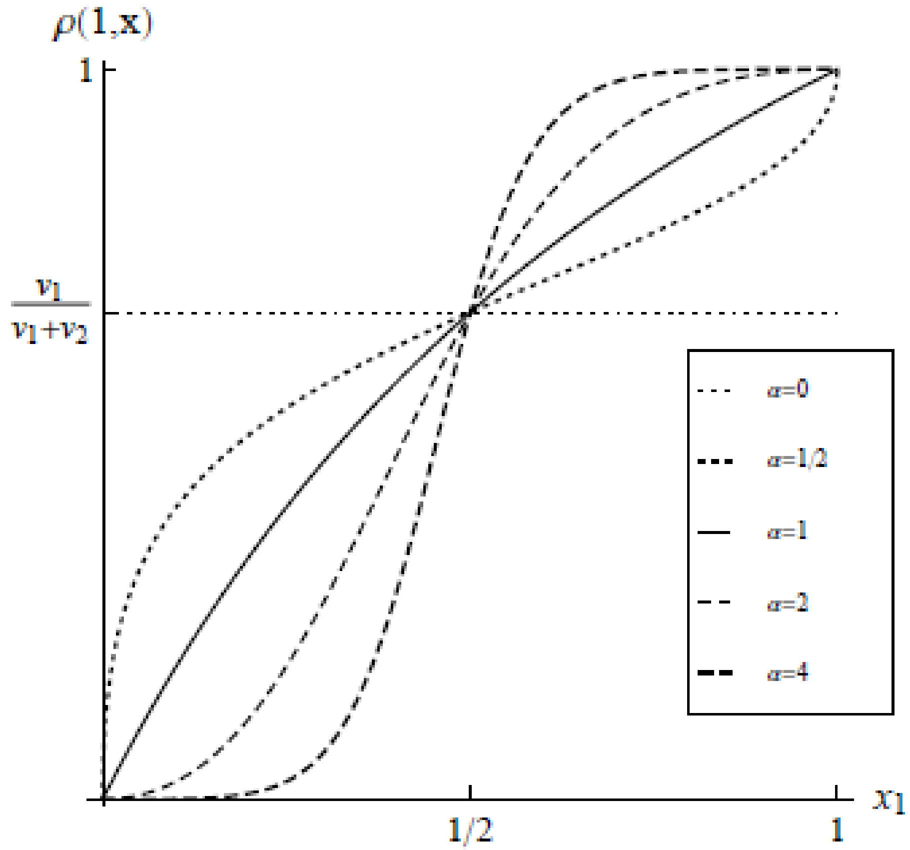
- The weight given in (37) to popularity will be larger the larger : We shall refer to $\alpha \geq 1$ as "strong" conformism and to $0 \leq \alpha < 1$ as "weak" conformism.

- Figure 3 illustrates the effect of the biases on the probability of transmission for the two-variant case and $v_1 > v_2$.

- On the horizontal axes I have plotted the proportion of agents with variant 1 and on the vertical the probability that 1 is copied as a function of x_1 for different values of α.

- The dotted lines correspond to $\alpha < 1$, the horizontal being for $\alpha = 0$ (Luce rule). The dashed lines correspond to $\alpha > 1$ and the solid one to $\alpha = 1$ (the exact values of are given in the figure's legend).

Figure 3: Strong and weak conformism



With the random choice function (37), the dynamics (6) are no longer a Markov chain and (16) can not be used to obtain the long-run distribution.

Substituting (37) in (6), subtracting $x_i(t)$ from both sides, making $b_i^* = 0$, and eliminating the time indexes I obtain the following system of difference equations,

$$\Delta x_i = \frac{v_i x_i^{\alpha}}{\sum_{j=1}^{n} v_j x_j^{\alpha}} \sum_{j=1}^{n} x_j - x_i, \; i = 1, 2, ...n. \tag{38}$$

Which can be written, after cross multiplying and rearranging, as

$$\Delta x_i = \frac{x_i^{\alpha}}{\sum_{j=1}^{n} v_j x_j^{\alpha}} \sum_{j=1}^{n} x_j v_j \left(\frac{v_i}{v_j} - \frac{x_i^{1-\alpha}}{x_j^{1-\alpha}} \right), \; i = 1, 2, ...n. \tag{39}$$

Note that in any rest point x^* of (39),

$$\frac{v_i}{v_j} = \left(\frac{x_i^*}{x_j^*}\right)^{1-\alpha} \tag{40}$$

whenever $x_i^*, x_j^* > 0$ and $\alpha \neq 1$. This observation proves to the following lemma.

LEMMA 2. *Assume that* $\alpha \neq 1$. *The interior state* $x^* = (x_1^*, ..., x_n^*)$ *where*

$$x_i^* = \frac{v_i^{\frac{1}{1-\alpha}}}{\sum_{j=1}^n v_j^{\frac{1}{1-\alpha}}}, \; i = 1, 2, ...n., \tag{41}$$

is the unique rest point of (38) with full support.

When $\alpha = 1$, (38) can be written as

$$\Delta x_i = \frac{v_i}{\sum_{j=1}^n v_j x_j} - 1, \; i = 1, 2, ...n. \tag{42}$$

- It is easy to see from (42) that in any rest point with more than one variant, all the existing variants must have the same v-value.

- The following proposition characterizes the long-run behavior of (38) distinguishing between culturally homogenous societies where all individuals are of the same type (x_i^* for all but one type) and plural societies in which different types coexist in equilibrium ($x_i^* > 0$ for more than one i).

PROPOSITION 3. *Assume that* $M^i = T$, *unbiased parents and that children copy cultural traits according to (37),* $x(0) \neq x^*$ *where* x^* *is given by (41) and* $x_i(0) > 0$ *for all* i. *Then,*

a) *Strong conformism* ($\alpha \geq 1$) *always leads to cultural homogeneity. The variant which is actually selected depends on the initial conditions when* $\alpha > 1$ *and it is the one with the highest* v *when* $\alpha = 1$.

b) *Weak conformism* ($\alpha \in (0, 1)$) *leads always to plural societies, the share of each trait in the long-run is given by (41).*

Proof. See Appendix.

This result shows that the cultural pluralism of the previous sections survives conformism, provided that this is not strong. Under weak conformism initial conditions do not matter but the long-run shares depend not only on the Luce values v, but also on the conformist parameter α.

- Figure 4 illustrates the results of proposition 3 in a two-variant world. I assume that initially the majority of agents in the population are of type 1 and that $v = (1,3/2)$.

- The figure shows $x_2(t)$ for different initial conditions and conformism parameter α.

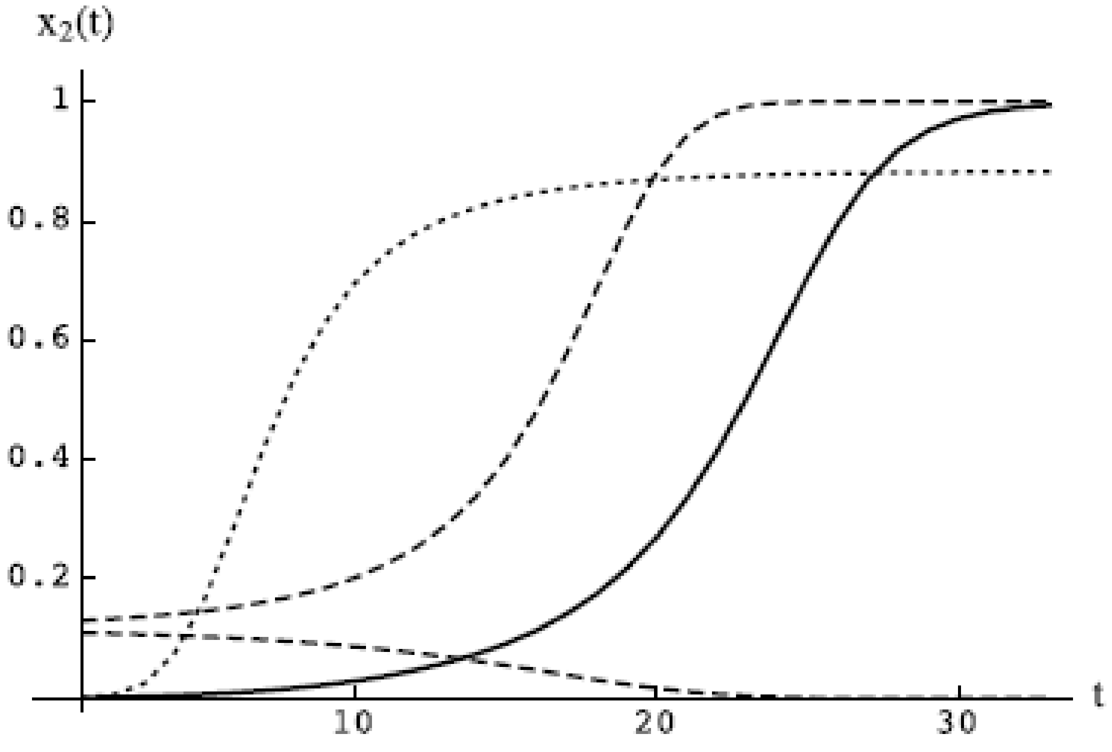
- The dashed lines corresponds to a population of strong conformists ($\alpha = 1.2$) for two different initial conditions ($x_2(0) = 0.11$ *and* $x_2(0) = 0.13$).

- As proposition 3 shows, when conformism is strong initial conditions matter and determine whether variant 1 or 2 takes over the whole population. Small differences can lead to very different long-run outcomes.

- The dotted line corresponds to a population of weak conformists (α= 0.8) which converges to a mixed population given by (41).

- Finally, the solid line corresponds to a heterogeneous population with 25% of weak conformists (α= 0.8) and 75% of strong conformists (α = 1.2) and initial condition $x_2(0)$= 1/1000.

Figure 4: Strong and weak conformism



- Conformism does not necessarily lead, as it is often argued, to cultural homogeneity nor does it prevent the spread of new types.

- Under weak conformism, new superior variants (with a higher v value) spread in the population the way the computer has done at expenses of the typewriter without the need to rely on unreasonably high proportion of initial adopters, as required by strong conformism.

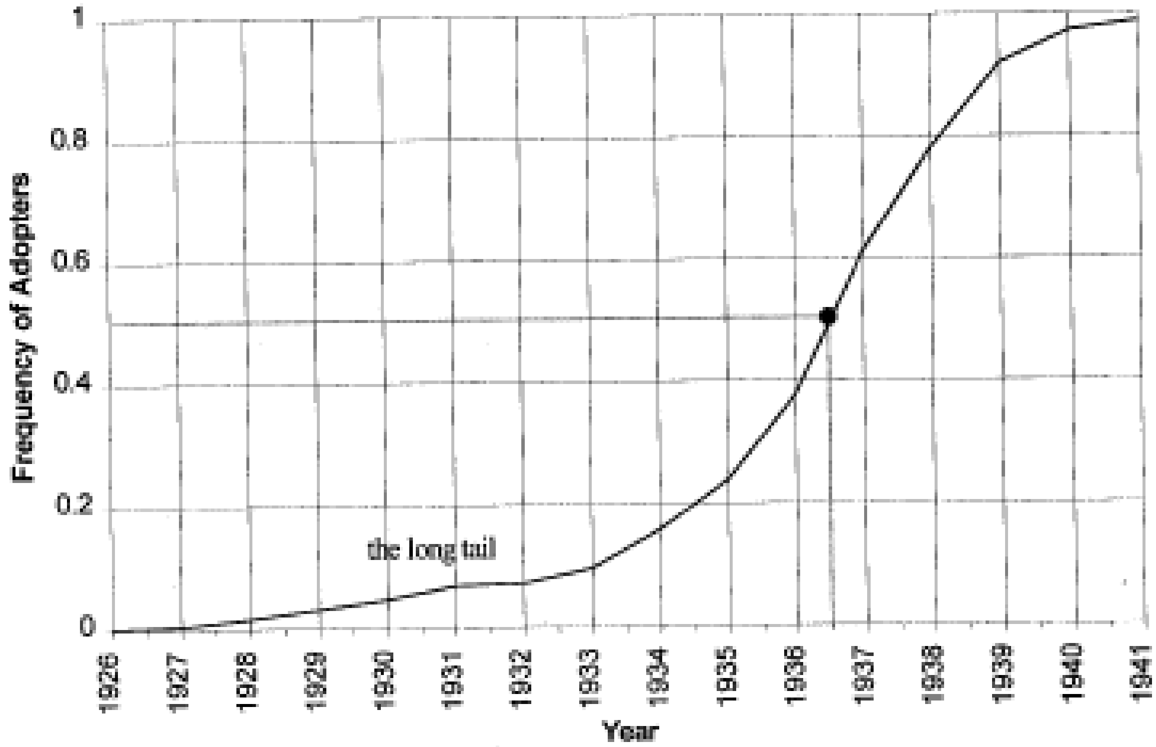
- As Figure 4 illustrates if individuals differ in their degree of conformity, new ideas will initially spread through weak-conformists who respond more to the objective advantages (as captured by the v-values) than to popularity.

- Once a large enough share has adopted the new variant the strong conformists, who follow the crowd, will follow suit and the new variant will become the predominant one.

- This generates the typical S-shaped diffusion curves with long tails, so that adoption is slow at the beginning, picks up in the middle and slow down as the innovation becomes dominant.

- Figure 5 shows the "true" diffusion curve for the hybrid corn seed in Iowa reported in Heinrich (2006).

Figure 5: Diffusion of hybrid corn seed in Iowa



6. Conclusion

- Acquiring a different culture is like moving to a new place with a different view of society.

- This process of moving from one place to another is what drives the inter generational transmission of culture which I have studied in this paper.

- The paper explores the implications of observational learning and stochastic choice for the long-run distribution of cultural variants, under the assumption that learning takes place primarily in the family environment, that this environment differs across families and that children can only copy what they see.

- I have given conditions that guarantee the existence of a unique stationary distribution with cultural diversity and analyzed how this distribution is affected by changes in the different parameters of the model and in the actual sets of role models.

- Those changes can be due to policy interventions like public schooling, integration policies and information campaigns.

- The model can also incorporate conformism. I show that cultural heterogeneity, namely the co-existence of several cultural variants, is only possible if agents do not put too much weight on popularity when deciding what types to copy.

- Although this paper focuses on cultural traits, the theory can be applied to other contexts such as the diffusion of new ideas and technologies.

- New superior technologies can spread in a world which is not too conformist, and this is true irrespective of the initial conditions.

- When agents are strong conformists, initial conditions matter, and new technologies may need a large initial share of adopters in order to grow in the population.

- On this ground, conformism is often seen as a hindering factor in the creation and diffusion of novel ideas. This paper shows that this is not necessarily the case.