CONVERGENCE OF KNOWLEDGE IN A CULTURAL EVOLUTION MODEL WITH POPULATION STRUCTURE, RANDOM SOCIAL LEARNING AND CREDIBILITY BIASES

SYLVAIN BILLIARD, MAXIME DEREX, LUDOVIC MAISONNEUVE, AND THOMAS REY

Abstract. Understanding how knowledge is created and propagates within groups is crucial to explain how human populations have evolved through time. Anthropologists have relied on different theoretical models to address this question. In this work, we introduce a mathematically oriented model that shares properties with individual based approaches, inhomogeneous Markov chains and learning algorithms, such as those introduced in [F. Cucker, S. Smale, *Bull. Amer. Math. Soc*, 39 (1), 2002] and [F. Cucker, S. Smale and D. X Zhou, *Found. Comput. Math.*, 2004]. After deriving the model, we study some of its mathematical properties, and establish theoretical and quantitative results in a simplified case. Finally, we run numerical simulations to illustrate some properties of the model.

Keywords: Individual based model, inhomogeneous Markov chains, convergence to equilibrium, numerical simulations, concentration inequalities, cultural evolution, language evolution, cumulative culture.

2010 Mathematics Subject Classification: 92D25, 68T05, 92H10.

1. Introduction

1.1. **On social learning.** Computers, spaceships and scientific theories have not been invented by single, isolated individuals. Instead, they result from a collective process in which innovations are gradually added to an existing pool of knowledge, most often over multiple generations [3, 15]. The ability to learn from others (social learning) is pivotal to that process because it allows innovations to be shared and be built upon by other individuals.

This process, termed cumulative culture, has been extensively studied by evolutionary anthropologists, both theoretically and experimentally [8, 14, 6, 12]. Most existing theoretical models, however, rely on strong assumptions and omit important aspects of social dynamics. For instance, previous models typically assume that individuals learn from the most skilled member of their social group. Yet, in real life, many reasons can prevent this strategy to come about: individuals might fail to evaluate each other's skills and hierarchical or spatial structures might preclude individuals from accessing to the most useful sources of social information, among others.

The aim of this work is to develop a more general mathematical model of knowledge evolution by relaxing hypotheses and incorporating more realistic forms of social interaction dynamics, such as those taking place in hierarchically or spatially structured populations.

1.2. **Outline of the paper.** In this work, we develop a new mathematical model that aims to describe the dynamics of knowledge creation and propagation among interacting individuals. The model is properly introduced and simple applications are given in Section 2. In Section 3, we

TR was partially funded by Labex CEMPI (ANR-11-LABX-0007-01) and ANR Project MoHyCon (ANR-17-CE40-0027-01). MD has received funding from the European Union's Horizon 2020 research and innovation programme under Marie Sklodowska-Curie grant agreement number 748310. Support from the ANR-Labex Institute for Advanced Study in Toulouse is acknowledged.

study some of the mathematical properties of the model, and establish theoretical, quantitative results in a simplified case describing the evolution of knowledge among interacting individuals. Finally, we develop a numerical method to simulate our model in Section 4. This method allows us to run numerical analyses of the model in cases where we do not have analytical results and to present numerical illustrations of the classical model of [5] on the evolution on language, which is contained in our model.

Acknowledgments. TR would like to thanks Mylène Maida for useful discussions on the inhomogeneous Markov chain structure of the model. LM would like to thanks Dorian Ni for his feedback on the model.

2. Presentation of the mathematical model

In this section, we shall present the model describing the evolution of knowledge within a finite population. Many different definitions of knowledge have been proposed. Here, we consider that knowledge results from conceptualizations that appropriately reflect the structure of the world and model conceptualization as functions linking a set of possible experiences to a set of possible concepts. We call these functions knowledge-like functions.

Time is supposed discrete. At each time step the knowledge-like function of individuals changes according to a learning dynamic that depends on both social and individual learning. Our model is an extension of the model of Cucker, Smale and Zhou describing the evolution of language [5], and can be seen as an hybrid between a learning algorithm [4] and an individual based model [2, 1].

We suppose that each individual influences each other through a social learning matrix $\Lambda \in \mathcal{M}_N(\mathbb{R})$. This matrix depends on both the structure of the population (*e.g.* a professor has a strong impact on its students, while students have less impact on their professor) and the credibility that each individual grants to each other. These influences are described by a structure matrix $\Gamma \in \mathcal{M}_N(\mathbb{R})$ and a credibility matrix $C \in \mathcal{M}_N(\mathbb{R})$, respectively. Knowledge-like functions also evolve by individual learning which is described as a stochastic process that we will detail in the following. The learning algorithm then takes into account both social and individual learning.

Let us first start with some useful notations that we shall use in the following:

- The space of square matrices of size $N > 0$ with coefficients in $\mathbb{K}$ will be denoted by $\mathcal{M}_N(\mathbb{K})$.
- The vector of $\mathbb{R}^N$ composed of 1s will be denoted by $\mathbf{e}$:

$$\mathbf{e} = (1, \dots, 1)^T \in \mathbb{R}^N. \tag{2.1}$$

- The distance from a function f to a set $\mathcal{X}$ is defined by

$$d(f, \mathcal{X}) = \inf_{g \in \mathcal{X}} d(f, g).$$

2.1. Modeling Knowledge.

Definition 1. *A knowledge setting $\mathcal{K}$ is a triple $(\mathcal{E}, \mathcal{C}, \mathcal{F})$ where :*

1. *$\mathcal{E}$ is a closed and bounded subset of $\mathbb{R}^n$.*
2. *$\mathcal{C} \subset \mathbb{E}^l$ with $l \in \mathbb{N}^*$, $\mathbb{E}$ an euclidean space, and $0 \in \mathcal{C}$.*
3. *$\mathcal{F}$ is a subset of the set of the functions from $\mathcal{E}$ to $\mathcal{C}$.*

The set $\mathcal{E}$ represents all the possible experiences, and $\mathcal{C}$ represents all the concepts (an illustration is presented in Fig 1).

Definition 2. *A knowledge-like function $f \in \mathcal{F}$ is a function from the experience set $\mathcal{E}$ to the concept set $\mathcal{C}$.*

Each knowledge-like function represents represents the knowledge of one individual. Let e be in $\mathcal{E}$, when there is a $c \in \mathcal{C}$ such as $f(e) = c$ and $c \neq 0$, we say that the knowledge-like function conceptualizes e. We assume that individuals conceptualize all experiences they go through. Elements that are not conceptualized (i.e. not experienced) by individuals are sent to the zero of the set $\mathcal{C}$ by their knowledge-like function.

Example. The knowledge-like function associated to colors. Let $\mathcal{E}$ be $[0, 1000] \subset \mathbb{R}$ representing the set of wavelengths in nanometers. We remind that $[380, 750]$ is the set of the visible spectrum. An individual associates each element of $\mathcal{E}$ to a color as shown in Figure 1. The set of concepts $\mathcal{C}$ contains the name of the color and 0. In this case $\mathcal{E}$ is a continuous space and $\mathcal{C}$ is a discrete space.

A knowledge-like function associates a color to each wavelength, or 0 if the individual has not conceptualized this color. For example f defined below is a knowledge-like function.

$$f'(e) = \begin{cases} \text{purple} & \text{if } e \in [380, 430], \\ \text{blue} & \text{if } e \in [430, 520], \\ \text{green} & \text{if } e \in [520, 565], \\ \text{yellow} & \text{if } e \in [565, 610], \\ \text{red} & \text{if } e \in [610, 750], \\ 0 & \text{otherwise.} \end{cases} \tag{2.2}$$

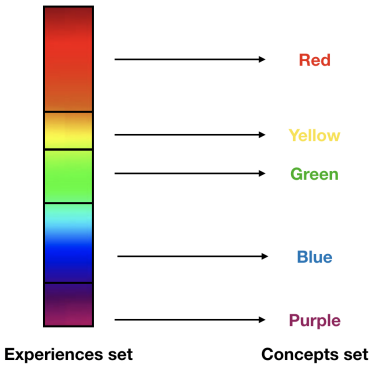


FIGURE 1. Illustration of a knowledge-like function of the visible spectrum. The arrows point to the color associated with each wavelength and illustrate the knowledge-like function f (2.3).

Remark 1. *In Japanese the kanji 青 (ao) names both colors green and blue. As in other languages, the Japanese language did not differentiate between these two colors at the beginning of its history. This language could be modeled by the knowledge-like function f':*

$$f(e) = \begin{cases} \textit{purple} & \textit{if } e \in [380, 430], \\ \textit{green} & \textit{if } e \in [430, 565], \\ \textit{yellow} & \textit{if } e \in [565, 610], \\ \textit{red} & \textit{if } e \in [610, 750], \\ 0 & \textit{otherwise.} \end{cases} \tag{2.3}$$

This knowledge-like function f' is different from f given in (2.2)*. Within the same population, individuals can have different knowledge-like functions.*

2.2. **Individual Based Model.** Let N be the number of individuals in the population. Each individual i is associated with a knowledge-like function $k^i \in \mathcal{F}$.

Definition 3. *A structure matrix $\Gamma = (\gamma_{ij})_{1\leq i,j\leq N}$ is a square matrix of size N describing the influence that individuals have on each other. More precisely for each $(i,j) \in \{1,...,N\}^2, \gamma_{ij} \in \mathbb{R}$ describes the strength of the influence of j on i.*

Remark 2. *For all $i \in \{1,...,N\}$ the greater the γ_{ii}, the less the individual i will be influenced by others. So γ_{ii} can be interpreted as the inertia of the individual i.*

Examples of structure matrices.

- We consider a population of N individuals structured in age, sorted such that individual 1 is the youngest and N the oldest. It has been shown that older individuals tend to have a higher inertia [7], which can be modelled by the condition $\gamma_{11} < ... < \gamma_{NN}$.
- Let us consider the relationship between a parent and her offspring. The offspring learns a lot from her parent but the situation is not symmetric. Let $s \in (0,1)$ describes the influence of a parent on her offspring. We have
$$\Gamma = \begin{pmatrix} 1 & 0.1 \\ s & 1-s \end{pmatrix}.$$
- We consider now the relationship between two students and their professor. Because of her status, the professor has a high influence on her students, while being very little influenced by them. Assuming that the relationship between the students is symmetric, we have
$$\Gamma = \begin{pmatrix} 1 & 0.1 & 0.1 \\ 1 & 0.2 & 0.2 \\ 1 & 0.2 & 0.2 \end{pmatrix}.$$

2.3. **Likelihood landscape.** In our model, some conceptualizations (i.e. knowledge-like functions) appropriately reflect the structure of the world, while other do not. For instance, in an environment in which blue berries are safe to eat while green berries are unsafe, color categorizations that discriminate between blue and green are superior because they appropriately capture the structure of the environment. Individuals don't know a priori how to categorize their environment. An individual who, by chance, only ever ate blue/safe berries might consider that discriminating between blue and green makes no sense. Yet, an individual who got sick after eating green/unsafe berries is likely to refine her color conceptualization to avoid being sick again. Sometimes, alternative and irreconcilable conceptualizations are equally likely. As an illustration let us consider the shape illustrated in figure 2. One might consider that it represents (1) two faces (in black) or (2) one cup (in white). Additional observations will not allow individuals to decide whether one conceptualization is more likely than the other.

In our model, we assume that individuals evaluate the likelihood of their conceptualization according to their own experience. To do so we define a likelihood landscape as following:

Definition 4. *$\forall c \in \mathcal{C}$, we define a function*
$$\begin{aligned} L(\cdot, c) : \mathcal{E} &\to [0,1], \\ e &\mapsto L(e,c). \end{aligned}$$
with
$$\forall e \in \mathcal{E},\ L(e,0) = \frac{1}{2}.$$
The map L is called the likelihood landscape.

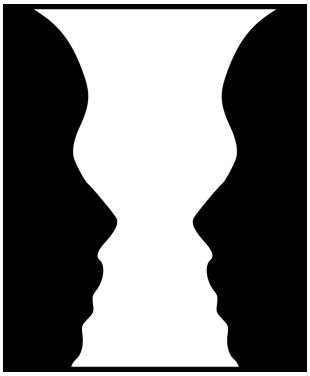


FIGURE 2. On this figure one may see two faces (in black) or a cup (in white). Image courtesy of Bryan Derksen, under license CC BY-SA 3.0

For all e, c in $\mathcal{E} \times \mathcal{C}$, $L(e, c)$ represents how well the concept c explain the experience (or observation) x.

Examples of likelihood landscapes.

- Let us consider the evolution of two different concepts in a population: flat earth (F), and round earth (R). Individuals can have experiences where the Earth seems flat (f) and others where the Earth seems round (r) (seeing a picture of the Earth, a boat vanishing behind the horizon, etc). In this case $\mathcal{E} = \{f, r\}$ and $\mathcal{C} = \{0, F, C\}$. We define the likelihood landscape as :

$$L(e, F) = \begin{cases} 1 & \text{if } e = f, \\ 0 & \text{if } e = r, \end{cases}$$

and

$$L(e, R) = \begin{cases} 1 & \text{if } e = f, \\ 1 & \text{if } e = r. \end{cases}$$

  When the earth seems flat (f) the earth could be flat or round (because round surfaces can appear flat when observed up close), so both concepts (F and R) are likely. However, when the earth seems round only the concept that the earth is round is likely.
- Let us consider again the example of color developed above (Fig. 1), where E is the set of all wavelengths of the visible spectrum and $\mathcal{C}$ is the set of colors. Moreover, let us consider that it does not make sense to discriminate between colors. In that case, we would define the likelihood landscape as $L(e, c) = 1$ for all $(e, c) \in \mathcal{E} \times \mathcal{C} \setminus \{0\}$.

2.4. **Credibility.** In addition to the population structure Γ from Definition 3, the influence of individuals on each other also depends on their credibility, through the credibility matrix C. The level of credibility attributed to an individual by another depends on both the knowledge-like functions, and the likelihood landscape.

More precisely, c_{ij} describes the credibility individual i attributes to individual j. If the credibility given to individual j by individual i is high relatively to that one attributed to other individuals (including herself), that means that individual i is more prone to adopt individual j's conceptualization. We will describe in Section 2.6 how this adoption changes one's knowledge-like function. We also consider that the credibility c_{ii} that an individual i gives to its own categorization can be affected by her own new experiences. In other words, individuals are able of self-criticism. The lower the self-credibility, the more likely an individual is to be influenced by other individuals (self-credibility directly affects an individual's inertia).

Definition 5. *A credibility matrix $C = (c_{ij})_{1\leq i,j\leq N}$ is a square matrix of size N defined by:*

$$c_{ij} = \max(\tilde{c}_{ij}, c_{\min}), \tag{2.4}$$

where $c_{\min} \geq 0$ is a fixed parameter , and

$$\tilde{c}_{ij} = \frac{1}{1 + \mathbb{1}_{\{i\neq j\}} \int_{c\in k_j(X)} \mathrm{d}c} \exp\left(\int_{e\in E} \ln\left(L(e, k_j(e)) \mathbb{1}_{\{k_i(e)\neq 0\}} \right) \mathrm{d}e \right). \tag{2.5}$$

Remark 3. *The term $\left(1 + \mathbb{1}_{\{i\neq j\}} \int_{c\in k_j(X)} \mathrm{d}c\right)^{-1}$ penalizes individuals who use on a wider range of concepts, which means that conceptualization that rely on smaller number of concepts are more likely to spread. The second term corresponds to the evaluation of the likelihood of the knowledge-like function of individual j on the experiences experienced by individual i. Note that for all e, $L(e, 0) = \frac{1}{2}$ (corresponding to cases where individual j has not experienced e), which decreases individual's credibility. This means that an individual i considers an individual j less credible if j has not experienced an experience individual i has gone through.*

Finally, the constant $c_{\min}$ corresponds to the minimal credibility: if $c_{\min} > 0$, individuals with low credibility can still influence other individuals.

Remark 4. *If the set $\mathcal{E}$ contains a finite number of elements the credibility formula reduces to*

$$\tilde{c}_{ij} = \frac{1}{1 + \mathbb{1}_{\{i\neq j\}} \int_{c\in k_j(X)} \mathrm{d}c} \prod_{e\in E} \mathbb{1}_{k_i(e)\neq 0} L(e, k_j(e)),$$

namely, the second part of the formula is similar to a measure of likelihood in probability theory [10]. *In this formula, associating few experiences with unlikely concepts penalizes credibility a lot.*

Application to the round vs. flat earth example. Let $\mathcal{E} = \{f_1, f_2, f_3, r_1, r_2\}$ and $\mathcal{C} = \{0, F, R\}$, together with $c_{\min} := 0$. For any $i \in \{1, 2, 3\}$, f_i are experiences where the Earth as likely to be flat as it is round (e.g. a human watching the horizon), and r_1, r_2 are experiences where the Earth is unlikely to be flat and likely to be round. We consider a population of 4 individuals with different knowledge-like functions k_1, k_2, k_3 and k_4 such as:

$$k_1(e) = \begin{cases} F & \text{if } e = f_1, \\ 0 & \text{otherwise,} \end{cases} \qquad k_2(e) = \begin{cases} F & \text{if } e = f_1 \text{ or } e = r_1, \\ 0 & \text{otherwise ,} \end{cases}$$

$$k_3(e) = \begin{cases} R & \text{if } e = f_1 \text{ or } e = r_1, \\ 0 & \text{otherwise,} \end{cases} \qquad k_4(e) = \begin{cases} F & \text{if } e = f_1, \\ R & \text{if } e = r_1, \\ 0 & \text{otherwise.} \end{cases}$$

We now compute the credibility matrix,

$$C = \begin{pmatrix} 1 & 1 & 1 & 1 \\ \frac{1}{2} & 0 & 1 & 1 \\ \frac{1}{2} & 0 & 1 & \frac{1}{2} \\ \frac{1}{2} & 0 & 1 & 1 \end{pmatrix}.$$

Let us normalize C such that its lines sum up to 1, in order to easily read the influences on an individual j in the row i:

$$C = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{5} & 0 & \frac{2}{5} & \frac{2}{5} \\ \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{5} & 0 & \frac{2}{5} & \frac{2}{5} \end{pmatrix}.$$

Individual #1 has only experienced f_1, she judges the other individuals (and herself) regarding that sole experience. All individuals associate f_1 with an appropriate conceptualization. So individual #1 gives the same credibility to all individuals. Individuals #2, 3 and 4 all have experienced f_1 and r_1. They evaluate individual #1 as less credible than themselves because #1 has not experienced r_1. Individual #3 and 4 evaluate #2 as not credible at all because she associates r_1 with a concept that is not probable anymore (i.e. the earth is flat while their experience shows it is round). #2 judges herself not credible because she associates r_1 with an unlikely concept. Individual #3 evaluates #4 as less credible than herself because #4 uses several concepts (in our model, less parsimonious conceptualizations are penalized).

2.5. **Social learning.** The social learning matrix $\Lambda \in \mathcal{M}_N(\mathbb{R})$ represents the influence of individuals on each other. This matrix captures the effect of the structure of the population described in Γ and the effect due to credibility C.

The influence of individual j on i depends on the structural influence γ_{ij} of j on i, and on the credibility c_{ij} that i gives to j. We shall assume in this work that these phenomena are multiplicative.

Definition 6. *The social learning matrix $\Lambda = (\lambda_{ij})_{1\leq i,j\leq N}$ is a square matrix of size N defined by:*

$$\lambda_{ij} = \begin{cases} \frac{\gamma_{ij}c_{ij}}{\sum_{l=1}^{N}\gamma_{il}C_{il}} & \text{if } \sum_{l=1}^{N}\gamma_{il}C_{il} \neq 0, \\ 1/N & \text{otherwise.} \end{cases} . \tag{2.6}$$

2.6. **Dynamics.** Finally, we consider a dynamical, discrete time model: knowledge-like functions evolve over time, altogether with associated quantities such as individuals' credibility. Let us denote by $k^t := (k_1^t, ..., k_N^t) \in \mathcal{F}^N$ the state of the population at time $t > 0$. As time evolves, individuals modify their conceptualization by the learning algorithm presented in [4]:

$$S_i^t \mapsto k_i^{t+1}, \tag{2.7}$$

which computes knowledge-like function from a sample

$$S_i^t = \{(e_1^{i,t}, c_1^{i,t}), ..., (e_m^{i,t}, c_m^{i,t})\}. \tag{2.8}$$

The sampling S_i^t is done using the probability measure $\rho^{i,t}$ defined in (2.9). For each individual i, the components of S_i^t represent the influences that will shape the knowledge of i at the next step. The elements of S_i^t can come from social or individual learning.

Definition 7. *Let us denote by $\tau \geq 0$ the proportion of individual learning, fixed and independent on i. The sampling measure is defined by*

$$\rho^{i,t} = (1-\tau)\rho_\Lambda^{i,t} + \tau\rho_\mathcal{I}^{i,t}, \tag{2.9}$$

where $\rho_\Lambda^{i,t}$ and $\rho_\mathcal{I}^{i,t}$ are two probability measures representing the effects of social and individual learning respectively, and defined below.

Definition 8. *The probability measure $\rho_\Lambda^{i,t}$ representing social learning is defined by:*

$$\rho_\Lambda^{i,t}(e,c) \propto \sum_{j=1}^{N} \lambda_{ij}^t \mathbb{1}_{\{k_j(e)=c\}}, \tag{2.10}$$

where λ_{ij}^t describes the influence of the individual j on the individual i through 2.6 at time t.

According to (2.10), drawing an element $(e, k_i^t(e))$ using the probability measure $\rho_\Lambda^{i,t}$ is equivalent to randomly drawing an individual i weighted by the coefficient $(\lambda_{ij}^t)_{1\leq j\leq N}$, and randomly choosing an experience e in $\mathcal{E}$.

Definition 9. *The probability measure $\rho_{\mathcal{I}}^{i,t}$ representing individual learning is given by*

$$\begin{cases} \displaystyle\int_{c\in C} \rho_{\mathcal{I}}^{i,t}(e,c)\,\mathrm{d}c \propto \int_{e'\in E} \mathbb{1}_{\{k_i^t(e')\neq 0\}}\mathbb{1}_{\{e\in E\}} \exp -\frac{\|e-e'\|^2}{2\sigma_E^2}\,\mathrm{d}e', \\ \rho_{\mathcal{I}}^{i,t}(c|e) \propto \mathbb{1}_{\{c\in C\}} \exp -\dfrac{\|c-f_i^t(e)\|^2}{2\sigma_C^2}, \end{cases} \tag{2.11}$$

where $\rho(c|e)$ denotes the conditional probability measure on $\mathcal{C}$, defined for every $(e,c)\in\mathcal{E}\times\mathcal{C}$, and every integrable function ϕ by

$$\int_{\mathcal{E}\times\mathcal{C}} \phi(e,c)\,\mathrm{d}\rho = \int_{\mathcal{E}} \left(\int_{\mathcal{C}} \phi(e,c)\,\mathrm{d}\rho(c|e)\right)\,\mathrm{d}\rho_{\mathcal{E}}.$$

In this last expression, $\rho_{\mathcal{E}}$ denotes the marginal probability measure on $\mathcal{E}$, namely

$$\rho_{\mathcal{E}}(e) := \rho(\pi^{-1}(e)), \quad \forall e\in\mathcal{E},$$

where $\pi : \mathcal{E}\times\mathcal{C}\to\mathcal{E}$ is the projection on $\mathcal{E}$.

The individual learning phase is then equivalent for each individual i to draw an element e' experienced by i and to draw an experience e following a normal law centered and concentrated on e'. Because of the shape of this probability law, individuals tend to explore the set $\mathcal{E}$ close to the elements they already explored (namely, innovating). The concept c is drawn following a probability law centered and concentrated on $k_i^t(e)$.

Definition 10. *The learning algorithm finally computes the knowledge-like function at the next time step using a least-square procedure* [4]

$$k_i^{t+1} \in \arg\min_k \sum_{(e,c)\in S_i^t} (k(e)-c)^2.$$

3. The case of globally shared knowledge: convergence without individual learning

In this section, we are interested in the convergence of the learning dynamics with high probability to a common shared conceptualization among individuals, i.e. when everybody carries the same knowledge-like function k. This result is obtained assuming no individual learning : in all this section, we shall assume that the rate of individual learning $\tau = 0$. The more realistic case where individuals also learn individually is explored below using numerical simulations.

One step idealistic processes. In order to establish the main result, let us decompose the stochastic process k^t into two other processes that we shall analyze separately.

Definition 11. *Let us define the application $\mathcal{T} : \mathcal{F}\times\mathbb{N}\mapsto\mathcal{F}$ by*

$$\mathcal{T}(f,t) = \Lambda^t f. \tag{3.1}$$

We can then define the one step deterministic idealistic process as

$$K_{\mathcal{T}}^t := \mathcal{T}(k^t,t), \quad K_{\mathcal{T}}^0 = k^0 \in \mathcal{F}^N. \tag{3.2}$$

Using these idealistic processes, the time evolution of the knowledge-like function is given by

$$k^t = \Delta k^t + K_{\mathcal{T}}^t, \tag{3.3}$$

where $\Delta k^t = k^t - K_{\mathcal{T}}^t$.

Definition 12. *Let $\mathcal{M}_{\mathcal{F}} = \{(k,...,k), k\in\mathcal{F}\}$ be the space of all the common shared conceptualizations.*

We first prove the contraction of the idealized process $K^t_{\mathcal{T}}$ from (3.1) in the space $\mathcal{M}_{\mathcal{F}}$ using some algebraic properties of primitive matrices, as well as results about inhomogeneous Markov chains. Secondly, we shall prove the convergence of the process Δk^t with high probability using learning theory. Then, under certain hypothesis (such as $\tau = 0$), we prove the convergence of k^t towards the set $\mathcal{M}_{\mathcal{F}}$ with high probability.

3.1. **Primitive matrices and their applications.** The behavior of processes is mainly driven by the influence matrix Λ. In this Section, we study the relationship between the properties of the influence matrix and the interactions taking place within the population.

Definition 13. *A matrix $A \in \mathcal{M}_N(\mathbb{R})$ is said to be primitive if $A \geq 0$ and if $\exists k \in \mathbb{N}^*$ such as $A^k > 0$.*

Definition 14. *Let $A \in \mathcal{M}_N(\mathbb{R})$. Let i,j be in $\{1, ..., N\}$,*

- *We say that i communicates with j (denoted $i \xrightarrow{A} j$) if there exists $n \geq 0$ and $i_1, ..., i_n \in \{1, ..., N\}$ such that*
$$a_{ii_1} \prod_{l=1}^{n-1} (a_{i_l i_{l+1}}) a_{i_n j} > 0.$$
If i does not communicate with j we write $i \nrightarrow j$.
- *We say that i communicates with j with k intermediates if there exists $i_1, ..., i_k \in \{1, ..., N\}$ such that*
$$a_{ii_1} \prod_{l=1}^{k} (a_{i_l i_{l+1}}) a_{i_k j} > 0,$$
and if there exists not $i_1, ..., i_{k-1} \in \{1, ..., N\}$ such that
$$a_{ii_1} \prod_{l=1}^{k-1} (a_{i_l i_{l+1}}) a_{i_{k-1} j} > 0,$$
- *Let $I, J \in \mathscr{P}(\{1, ..., N\})$. We say that $I \xrightarrow{A} J$ if*
$$\exists i, j \in I \times J \text{ such that } i \xrightarrow{A} j.$$

If A is a primitive matrix of size N, for each i, j in $\{1, ..., N\}, i \to j$. Moreover if there is k such that $A^k > 0$, then for each i, j in $\{1, ..., N\}$ i communicates with j with at most k intermediates.

Examples. Let us consider a matrix A defined by

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

As we can see on the graph of the matrix A (Fig. 3), every individual communicates with each other with at most 3 intermediates. Then, A is a primitive matrix, because

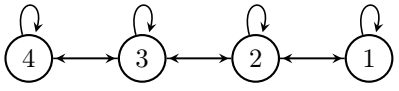


FIGURE 3. Graph representing the matrix A.

$$A^3 = \begin{pmatrix} 4 & 5 & 3 & 1 \\ 5 & 7 & 6 & 3 \\ 3 & 6 & 5 & 7 \\ 1 & 3 & 5 & 4 \end{pmatrix} > 0.$$

However the converse is not true. Indeed if one considers the matrix

$$B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

then $1 \to 2$ and $2 \to 1$ (see fig. 4). Nevertheless, B is not a primitive matrix because for all

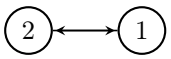


FIGURE 4. Graph representing the matrix B.

$k \in \mathbb{N}$,

$$B = \begin{cases} B & \text{if } k \text{ is odd,} \\ I_2 & \text{if } k \text{ is even.} \end{cases}$$

According to Perron-Frobenius theorem, one has

Proposition 1. *Let $A \in \mathcal{M}_N(\mathbb{R})$ be a primitive stochastic matrix. 1 is an eigenvalues of A, and all the other eigenvalues are less than 1 in modulus.*

We can now establish results relating graphs and eigenvalues of matrices:

Proposition 2. *Let A be a stochastic matrix of size N, if*

$$\forall i, j \in \{1, ..., N\}, i \xrightarrow{A} j \text{ or } j \xrightarrow{A} i,$$

holds then 1 is an eigenvalue of A, and its multiplicity is 1.

Proof. We suppose that 1 is not an eigenvalue of multiplicity 1 of A. Since A is a stochastic matrix one has $A\mathbf{e} = \mathbf{e}$ with $\mathbf{e}$ given by (2.1), so 1 is an eigenvalue of A. In particular, its order of multiplicity is bigger than 1, and there exists $X \in \mathbb{R}^N \setminus \mathbb{R}\mathbf{e}$ such that $AX = X$.

Let P be a permutation matrix such that the coordinates of the vector PX are ranked from the lowest to the highest. Let $A' = P^T AP$ and $X' = P^T X$. Let n_l and n_h be respectively the number of coordinates equals to the lowest and to the highest coordinates values. We have $n_l \geq 1$, $n_h \geq 1$ and $n_l + n_h \leq N$.

First case : $n_l + n_h = N$

$$\begin{aligned} AX = X &\iff A'X' = X'. \\ &\implies \forall i \in \{1, ..., N\}, \sum_{j=1}^{N} a'_{ij} X'_j = X'_i. \\ &\implies \begin{cases} \forall i \in \{1, ..., n_l\}, \forall j \in \{n_l + 1, ..., N\}, a'_{ij} = 0, \\ \forall i \in \{n_l + 1, ..., N\}, \forall j \in \{1, ..., n_l\}, a'_{ij} = 0. \end{cases} \\ &\implies \exists (B, C) \in \mathcal{M}_{n_l}(\mathbb{R}) \times \mathcal{M}_{n_h}(\mathbb{R}), A = \left(\begin{array}{c|c} B & 0 \\ \hline 0 & C \end{array}\right). \end{aligned}$$

So $1 \nrightarrow N$ et $N \nrightarrow 1$ with the matrix A'. Let $i = \sigma^{-1}(1)$ and $j = \sigma^{-1}(N)$, we have $i \nrightarrow j$ and $j \nrightarrow i$ with the matrix A.

Second case : $n_l + n_h < N$ Let $G_1 = \{1, ..., n_l\}$, $G_2 = \{n_l + 1, ..., N - n_l\}$, and $G_3 = \{N - n_h + 1, ..., N\}$.

$$\begin{aligned} AX = X &\iff A'X' = X'. \\ &\implies \forall i \in \{1, ..., N\}, \sum_{j=1}^{N} a'_{ij} X'_j = X'_i. \\ &\implies \begin{cases} \forall i \in \{1, ..., n_l\}, \forall j \in \{n_l + 1, ..., N\}, a'_{ij} = 0, \\ \forall i \in \{N - n_h + 1, ..., N\}, \forall j \in \{1, ..., n_l\}, a'_{ij} = 0. \end{cases} \\ &\implies \exists (B, C, D, E, F), A = \left(\begin{array}{c|c|c} B & 0 & 0 \\ \hline C & D & E \\ \hline 0 & 0 & F \end{array} \right). \end{aligned}$$

As shown on the figure 5, $G_1 \nrightarrow G_3$ and $G_3 \nrightarrow G_1$.

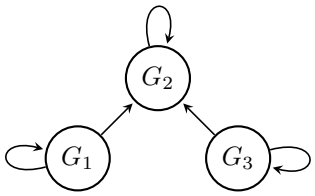


FIGURE 5. Illustration of the influence relationship between the three clusters G_1, G_2, and G_3

We conclude as before.

□

3.2. **Eigenvalues of the matrix of influence Λ.** In this Section we study the the quantitative properties of the eigenvalues of the influence matrix, in order to understand the dynamics of the idealized process $K^t_{\mathcal{T}}$.

Let Γ and C be respectively the structure and credibility matrices defined in Def. 3 and equation (2.5). By construction, C is a stochastic matrix. The structure matrix Λ is defined according to (2.6) by

$$\lambda_{ij} = \begin{cases} \frac{\gamma_{ij} c_{ij}}{\sum_{l=1}^{N} \gamma_{il} C_{il}} & \text{if } \sum_{l=1}^{N} \gamma_{il} C_{il} \neq 0, \\ 1/N & \text{otherwise.} \end{cases}$$

Proposition 3. *If for all $i, j \in \{1, ..., N\}$, one has $i \xrightarrow{\Gamma} j$ or $j \xrightarrow{\Gamma} i$, and the credibility matrix $C > 0$, then for all $i, j \in \{1, ..., N\}$, $i \xrightarrow{\Lambda} j$ or $j \xrightarrow{\Lambda} i$.*

Proof. $\forall i, j \in \{1, ..., N\}, \gamma_{ij} > 0 \implies \lambda_{ij} > 0$ using (2.6). Thus $\forall i, j \in \{1, ..., N\}, i \xrightarrow{\Gamma} j$ or $j \xrightarrow{\Gamma} > 0 \implies i \xrightarrow{\Lambda} j$ or $j \xrightarrow{\Lambda} i$.

□

Lemma 1. *If Γ is primitive and $c_{\min} > 0$, then Λ is a primitive matrix.*

Proof. Γ is primitive so there exists $k \in \mathbb{N}$ such that $\Gamma^k > 0$. In particular,

$$\begin{aligned}\forall i,j \in \{1,...,N\}, \gamma_{ij}^k > 0 &\iff \forall i,j \in \{1,...,N\}, \\ &\quad \sum_{0 \leq l_1,...,l_{k-1} \leq N} \gamma_{il_1}\gamma_{l_1l_2}...\gamma_{l_{k-1}j} > 0, \\ &\iff \forall i,j \in \{1,...,N\}, \exists l_1,...,l_{k-1} \in \{1,...,N\}, \\ &\quad \gamma_{il_1}\gamma_{l_1l_2}...\gamma_{l_{k-1}j} > 0.\end{aligned}$$

Moreover $C > 0$ implies that $\forall i,j \in \{1,...,N\}, \gamma_{ij} > 0$. One can then conclude that $\lambda_{ij} > 0$. □

Lemma 2. *Let us assume that $\Gamma > 0$. If $c_{\min} > 0$, the social learning matrix Λ is bounded by below: there exists $\underline{m}_\lambda > 0$ such that $\forall i,j \in \{1,...,N\}$, $\lambda_{ij} \geq \underline{m}_\lambda$.*

Proof. Let us set $\underline{m}_\gamma = \min_{i,j} \gamma_{ij}$. Since $\gamma_{ij} > \underline{m}_\gamma$ for all $i,j \in \{1,...,N\}$, one has that

$$\lambda_{ij} > \mathbb{1}_{\sum_{l=0}^N \gamma_{il}C_{il}=0}\frac{1}{N} + \mathbb{1}_{\sum_{l=0}^N \gamma_{il}C_{il}\neq 0}\frac{\underline{m}_\gamma c_{\min}}{N(1-N\underline{m}_\gamma)}.$$

It is then enough to choose

$$\underline{m}_\lambda = \min\left(\frac{1}{N}, \frac{\underline{m}_\gamma c_{\min}}{N(1-N\underline{m}_\gamma)}\right).$$

□

Lemma 3. *Let Λ be a stochastic matrix of size N that is bounded by below by $\underline{m}_\lambda$ as in Lemma 2. Let us consider the sorted collection $(\alpha_i)_{i=1...N}$ of its eigenvalues:*

$$\alpha_1 = 1 > |\alpha_2| \geq ... \geq |\alpha_N|.$$

Then there exists a universal constant $0 < \overline{M}_\lambda < 1$ such that

$$\forall i \in \{2,...,N\}, |\alpha_i| \leq 1 - \overline{M}_\lambda.$$

Proof. Let $\mathcal{A} = \{A \in \mathcal{M}^N(\mathbb{R}), \forall i \in \{1,...,N\}, \sum a_{ij} = 1, \forall i,j \in \{1,...,N\}, a_{ij} \geq \underline{m}_\lambda\}$. Being composed of stochastic matrices, the set $\mathcal{A}$ is bounded (by 1) for the norms induced by both the 1 and ∞ vector norms on $\mathbb{R}^N$. Moreover, it is closed by construction. In particular, $\mathcal{A}$ is a compact subset of $\mathcal{M}^N(\mathbb{R})$.

Let

$$\begin{aligned}\mathcal{L}\colon \mathcal{M}^N(\mathbb{R}) &\to \mathbb{C}^N \\ A &\mapsto Sp(A),\end{aligned}$$

the application that returns the eigenvalues of a matrix, sorted in a nonincreasing (in modulus) order. According to the Theorem II.5.1 of [9], $\mathcal{L}$ is a continuous function on the set of stochastic matrices. In particular, $\mathcal{A}$ being compact, the numerical range of $\mathcal{L}$,

$$R(\mathcal{L}) := \{\mathcal{L}(A), \forall A \in \mathcal{A}\} \subset \mathbb{C}^N,$$

is also compact. Continuous function reach their bounds on compact sets, so that one can take

$$\begin{aligned}\overline{M}_\lambda &= 1 - \sup_{\mathcal{L}\in\mathcal{A}}\{|\mu_2|, \mu \in R(\mathcal{L})\} \\ &= 1 - \max\{|\mu_2|, \mu \in R(\mathcal{L})\}.\end{aligned}$$

□

3.3. **Contraction of a stochastic primitive matrix.** We shall now study in this section properties of stochastic primitive matrices, in order to understand the behavior of the process $K_{\mathcal{T}}^t$.

Lemma 4. *Let $A \in \mathcal{M}_N(\mathbb{R})$ be a stochastic matrix that is bounded by below by $\underline{m}_A$ as in 2. Let $\mathcal{M} := \mathbb{R}\mathbf{e}$ be the eigenspace associated to the eigenvalue 1 and $\mathcal{W}$ the eigenspace associated to the remaining eigenvalues. One has*

1. *$\mathbb{R}^N = \mathcal{M} \oplus \mathcal{W}$, both spaces being stable by A;*
2. *There is a norm $\|\cdot\|$ on $\mathcal{W}$ and a distance d on $\mathbb{R}^N$ such that for all $(x_{\mathcal{M}}, x_{\mathcal{W}}) \in \mathcal{M} \oplus \mathcal{W}$,*

$$d(x_{\mathcal{W}}, \mathcal{M}) = \|x_{\mathcal{W}}\|, \tag{3.4}$$

$$d(x_{\mathcal{M}} + x_{\mathcal{W}}) = d(x_{\mathcal{W}}, \mathcal{M}), \tag{3.5}$$

$$d(A(x_{\mathcal{M}} + x_{\mathcal{W}}), \mathcal{M}) \leq (1 - N\underline{m}_A) d(x_{\mathcal{M}} + x_{\mathcal{W}}, \mathcal{M}). \tag{3.6}$$

Proof. (1) Classical decomposition result.

(2) We take $\|x_{\mathcal{W}}\| = \max_{i,j\in\{1,\dots,N\}}\{|x_{\mathcal{W},i} - x_{\mathcal{W},j}|\}$ for all $x_{\mathcal{W}} \in \mathcal{W}$. The matrix A being stochastic, one can use results on inhomogeneous Markov chains together (namely Theorem 3.1 in [16]) together with the upper bound (3) on the eigenvalues of A to have that

$$\|Ax_{\mathcal{W}}\| \geq (1 - N\underline{m}_A) \|x_{\mathcal{W}}\|.$$

Let $x = x_{\mathcal{M}} + x_{\mathcal{W}}$ and $y = y_{\mathcal{M}} + y_{\mathcal{W}}$ be in $\mathcal{M} \oplus \mathcal{W}$, we define d as

$$d(x, y) = \|x_{\mathcal{M}} - y_{\mathcal{M}}\|_2 + \|x_{\mathcal{W}} - y_{\mathcal{W}}\|,$$

with $\|\cdot\|_2$ being the euclidean norm on $\mathbb{R}^N$.

□

Remark 5. *This result is inspired from the Lemma 1 from* [5]*, and has similar conclusions. Nevertheless, one has to bear in mind that with its set of hypotheses, the original result from* [5] *is wrong. Indeed, being only stochastic and weakly irreducible is not enough to have the existence of a norm with the desired, precise contraction property (ii)-(c). For example,*

$$\Lambda := \begin{pmatrix} 0 & 1/2 & 1/2 \\ 3/4 & 0 & 1/4 \\ 1/8 & 7/8 & 0 \end{pmatrix}$$

is a stochastic, weakly irreducible matrix which has 2 complex eigenvalues, preventing the validity of (ii)-(c). Our set of hypotheses, as well as our new proof, prevent this.

Corollary 1. *If $A \in \mathcal{M}_N(\mathbb{R})$ is as in Lemma 4, $\mathcal{M} := \mathbb{R}\mathbf{e}$ be the eigenspace associated to the eigenvalue 1 and $\mathcal{W}$ the eigenspace associated to the remaining eigenvalues:*

1. *Then $\mathcal{F}^N = \mathcal{M} \oplus \mathcal{W}$, both spaces being stable by A;*
2. *There is a norm $\|\cdot\|$ on $\mathcal{W}$ and a distance d on $\mathbb{R}^N$ such that, for all $f_{\mathcal{M}} \in \mathcal{M}$ and for all $f_{\mathcal{W}} \in \mathcal{W}$,*

$$d(f_{\mathcal{W}}, \mathcal{M}) = \sqrt{\int_{\mathcal{E}} \sum_{i=1}^{l} \|f_{\mathcal{W},i}(e)\|^2 \, \mathrm{d}e},$$

$$d(f_{\mathcal{M}} + f_{\mathcal{W}}, \mathcal{M}) = d(f_{\mathcal{W}}, \mathcal{M}),$$

$$d(A(f_{\mathcal{M}} + f_{\mathcal{W}}), \mathcal{M}) \leq (1 - N\underline{m}_A)\, d(f_{\mathcal{M}} + g_{\mathcal{W}}, \mathcal{M}).$$

Proof. Let $f = f_{\mathcal{M}} + f_{\mathcal{W}}$ and $g = g_{\mathcal{M}} + g_{\mathcal{W}}$ in $\mathcal{M} \oplus \mathcal{W}$, we define as

$$d(g, \mathcal{M}) = \sqrt{\int_{\mathcal{E}} \sum_{i=1}^{l} (\|f_{\mathcal{M},i}(e) - g_{\mathcal{M}(e),i}\|_2^2 + \|f_{W,i}(e) - g_{W,i}(e)\|^2) \, \mathrm{d}e},$$

□

Corollary 2. *Let $A \in \mathcal{M}_N(\mathbb{R})$ as in Lemma 4, and $f \in \mathcal{F}^N$ such that $f = f_{\mathcal{M}} + f_{\mathcal{W}} \in \mathcal{M} \oplus \mathcal{W}$. There exists a distance d on $\mathcal{F}$ such that*

$$d(Af, f_{\mathcal{M}}) \leq (1 - N\underline{m}_A) \, d(f, f_{\mathcal{M}}).$$

Proof. Consequence of the previous corollary. □

We recall that the application $\mathcal{T} : \mathcal{F} \times \mathbb{N} \mapsto \mathcal{F}$ and the process $K_{\mathcal{T}}^t$ are defined by $\mathcal{T}(f, t) = \Lambda^t f$ and $K_{\mathcal{T}}^t = \mathcal{T}(k^t, t)$. This may be interpreted as an idealistic step of learning. Moreover, we assumed that $\tau = 0$, namely no individual learning occurs in the model.

Theorem 1. *If $\Gamma > 0$, and the minimum credibility $c_{\min} > 0$ is fixed, then for all times t, there exists a distance d_{Λ^t} and $\underline{m} > 0$ independent on t such that*

$$d(K_{\mathcal{T}}^{t+1}, k_{\mathcal{M}}^t) \leq (1 - \underline{m}) \, d(k^t, k_{\mathcal{M}}^t),$$

where $k_{\mathcal{M}}^t$ is the projection of k^t on $\mathcal{M}$.

Proof. Consequence of lemmas 1 and 2, and corollary 2. □

We define an idealistic deterministic process K^t by $K^{t+1} = \Lambda_t K^t$ and $K^0 = k^0$. Theorem 1 implies that the idealistic, deterministic process converges to a common shared knowledge:

Corollary 3. *Under the hypotheses of Theorem 1, there exist $(K_{\mathcal{M}}^0, K_{\mathcal{W}}^0)$ in $\mathcal{M} \otimes \mathcal{W}$ such that $K^0 = K_{\mathcal{M}}^0 + K_{\mathcal{W}}^0$. Then,*

$$\lim_{t \to +\infty} (K^t, K_{\mathcal{M}}^0) = 0.$$

3.4. **Learning theory.** The results presented in this part are inspired by [4]. This article deals with the inference of functions to match with random samples. In our case, the functions are knowledge-like functions and samples come from social learning and individual learning. Nevertheless, our theoretical results shall only deal with the case where individuals learn from social sources only (under the hypothesis $\tau = 0$). The results of this theory implies the convergence of the process Δk^t with high probability.

3.4.1. *Sample error.* We study the learning process from random samples governed by the probability measure ρ on $\mathcal{Z} = \mathcal{E} \times \mathcal{C}$. We recall that $\mathcal{E}$ is a compact subset of $\mathbb{R}^n$, and that $\mathcal{C}$ is a subset of an euclidean space containing zero.

Definition 15. *We define the least square error of f as*

$$\varepsilon(f) = \int_{\mathcal{Z}} \|f(e) - c\|_{\mathcal{C}}^2 \, \mathrm{d}\rho,$$

for $f : \mathcal{E} \to \mathcal{C}$, where $\|\cdot\|_{\mathcal{C}}$ is a norm on $\mathcal{C}$ associated with the inner product $\langle \cdot, \cdot \rangle_{\mathcal{C}}$ of the ambient euclidean space $\mathbb{E}^l$.

Proposition 4. *For every $f : \mathcal{E} \to \mathcal{C}$,*

$$\varepsilon(f) = \int_{\mathcal{E}} \|f(e) - f_\rho(e)\|_{\mathcal{C}}^2 \, \mathrm{d}\rho_{\mathcal{E}} + \varepsilon(f_\rho),$$

where $f_\rho(e) := \int_{\mathcal{C}} c \, \mathrm{d}\rho(c|e)$, for any $e \in \mathcal{E}$.

Proof. Adding and subtracting f_ρ yields

$$\begin{aligned}\varepsilon(f) &= \int_{\mathcal{Z}} \|f(e) - f_\rho(e)\|^2 \,\mathrm{d}\rho + \int_{\mathcal{Z}} \|f_\rho(e) - c\|^2 \,\mathrm{d}\rho + 2\int_{\mathcal{Z}} \langle f(e) - f_\rho(e), f_\rho(e) - c\rangle_{\mathcal{C}} \,\mathrm{d}\rho \\ &= A + \varepsilon(f_\rho) + 2B.\end{aligned}$$

We have

$$\begin{aligned}A &= \int_{\mathcal{E}} \left(\int_{\mathcal{C}} \|f(e) - f_\rho(e)\|^2 \,\mathrm{d}\rho(c|e) \right) \mathrm{d}\rho_{\mathcal{E}} \\ &= \int_{\mathcal{E}} \left(\|f(e) - f_\rho(e)\|^2 \int_{\mathcal{C}} \mathrm{d}\rho(c|e) \right) \mathrm{d}\rho_{\mathcal{E}} \\ &= \int_{\mathcal{E}} \|f(e) - f_\rho(e)\|^2 \,\mathrm{d}\rho_{\mathcal{E}}.\end{aligned}$$

For the second term we have

$$\begin{aligned}B &= \int_{\mathcal{E}} \left(\int_{\mathcal{C}} \langle f(e) - f_\rho(e), f_\rho(e) - c\rangle_{\mathcal{C}} \,\mathrm{d}\rho(c|e) \right) \mathrm{d}\rho_{\mathcal{E}} \\ &= \int_{\mathcal{E}} \langle f(e) - f_\rho(e), f_\rho(e) - \int_{\mathcal{C}} c \,\mathrm{d}\rho(c|e) \rangle_{\mathcal{C}} \,\mathrm{d}\rho_{\mathcal{E}} \\ &= \int_{\mathcal{E}} \langle f(e) - f_\rho(e), f_\rho(e) - f_\rho(e) \rangle \,\mathrm{d}\rho_{\mathcal{E}} = 0.\end{aligned}$$

□

As a consequence of the proposition 4, the regression function f_ρ minimizes the mean square error ε.

Definition 16. *Let $f_{\mathcal{F}}$ be the target function that minimizes ε:*

$$f_{\mathcal{F}} \in \arg\min_{f \in \mathcal{F}} \varepsilon(f).$$

During the learning phase, the probability measure ρ is not assumed to be known. The learning process is a minimisation procedure on a sample $S = ((e_1, c_1), ..., (e_m, c_m))$, $m \in \mathbb{N}^*$.

Definition 17. *We define the empirical error ε_S of f on the sample S by*

$$\varepsilon_S(f) = \frac{1}{m} \sum_{i=1}^{m} \|f(e_i) - c_i\|^2 \,\mathrm{d}\rho,$$

and f_S the empirical target function, namely a minimizer of ε_S:

$$f_S \in \arg\min_{f \in \mathcal{F}} \varepsilon_S(f).$$

This minimizer is of course not unique. Nevertheless, when the size m of the sample is large enough, the empirical target function will approximate the target function. More precisely, one has the following classical concentration inequality from [13]:

Proposition 5. *We assume that:*

1. *$\mathcal{F}$ is a compact and convex set;*
2. *there exists $M \in \mathbb{R}_+^*$, such that for all $f \in \mathcal{F}$, $\|f(e) - c\|_{\mathcal{C}} \leq M$ almost everywhere;*
3. *ρ is a probability measure on $\mathcal{Z}$.*

Then for all $\eta > 0$,

$$\mathrm{Prob}\left\{\int_{\mathcal{E}} \|f_S(e) - f_{\mathcal{F}}(e)\|_{\mathcal{C}}^2 \,\mathrm{d}\rho_{\mathcal{E}} \leq \eta\right\} \geq 1 - \mathcal{N}(\mathcal{F}, \frac{\eta}{24M}) e^{\frac{-m\eta}{288M^2}}$$

where $\mathcal{N}(\mathcal{F}, s)$ is the so-called covering number, namely the minimal $l \in \mathbb{N}$ such that there exist l disks in $\mathcal{F}$ with radius s covering $\mathcal{F}$. Since $\mathcal{F}$ is compact, this number is finite.

We can now get back to our model. We recall that the probability measure $\rho^{i,t}$ allows to draw the sample for the learning of the individual i at time t, and that $\tau = 0$. The probability measure $\rho^{i,t}$ then depends only on social learning. During the learning phase of our model we have :

$$\begin{aligned} f_{\rho^{i,t}}(e) &= \int_{\mathcal{C}} c \,\mathrm{d}\rho(c|e) \\ &= \sum_{j=1}^{N} \Lambda_{ij}^t \int_{\mathcal{C}} c \mathbb{1}_{k_j(e)=c} \,\mathrm{d}c \\ &= \sum_{j=1}^{N} \Lambda_{ij}^t k_j(e) \,\mathrm{d}c. \end{aligned}$$

Since $\mathcal{F}$ is convex, $f_{\rho^{i,t}} \in \mathcal{F}$. If $\mathcal{E}$ is finite, or in the other case, if $\mathcal{F}$ is a set a continuous functions, we have $f_{\mathcal{F}}^{i,t} = f_{\rho^{i,t}}$ with $f_{\mathcal{F}}^{i,t}$ being the minimiser of the error ε with $\rho = \rho^{i,t}$.

3.5. **Main result.** Combining our results on the one-step idealistic process, together with the ones on learning theory, we are able to study the convergence of the full process k^t with high probability.

We recall that $\mathcal{M}_{\mathcal{F}} = \{(k, ..., k), k \in \mathcal{F}\}$.

Theorem 2. *We suppose that $\tau = 0$, $\Gamma > 0$, $c_{\min} > 0$, and $\mathcal{F}$ is compact and convex. There exist some constants $\alpha_* < 1$, $M > 0$, and $A \geq 0$ such that for each $0 < \delta < 1$, and $t \geq 0$, if the sample size $m \geq m_t = m_t\left(\mathcal{M}_{\mathcal{F}}, k^0, M, \alpha_*, \delta\right)$, then*

$$d(k^t, \mathcal{M}_{\mathcal{F}}) \leq A\, \alpha_*^t\, d(k^0, \mathcal{M}_{\mathcal{F}}),$$

with confidence at least $1 - \delta$.

Proof. Let d be the distance defined in corollary 1.

We recall that the application $\mathcal{T} : \mathcal{F} \times \mathbb{N} \mapsto \mathcal{F}$ is defined by $\mathcal{T}(f, t) = \Lambda^t f$. Let $K_{\mathcal{T}}^t = \mathcal{T}(k^t, t)$. Notice in particular that the process $K_{\mathcal{T}}^t$ is different from k^t. By the triangle inequality we have, using (3.3) and (3.2), that

$$d(k^t, \mathcal{M}_{\mathcal{F}}) \leq d(k^t, K_{\mathcal{T}}^t) + d(K_{\mathcal{T}}^t, \mathcal{M}_{\mathcal{F}}).$$

The contractivity of the second term is yielded by Theorem 1: there exists $\alpha_t < 1$ such that

$$d(K_{\mathcal{T}}^t, \mathcal{M}_{\mathcal{F}}) \leq \alpha_t d(k_{\mathcal{T}}^{t-1}, \mathcal{M}_{\mathcal{F}}).$$

Now, we need to estimate the other term. We recall that $\mathcal{E}$ is compact in $\mathbb{R}^N$. By the compactness of $\mathcal{E}$ and $\mathcal{F}$ we have that

$$\sup_{f \in \mathcal{F}, e \in \mathcal{E}} \|f(e)\|_{\mathcal{C}} < \infty.$$

In particular, there exists $M > 0$ such that

$$\max_{(e,c) \in \mathcal{E} \times \mathcal{C}, f \in \mathcal{F}} \|f(e) - c\|_{\mathcal{C}} \leq M$$

Using Proposition 5, for each $\eta > 0$ and $i \in \{1, ..., N\}$,

$$\text{Prob}\left\{\int_{\mathcal{E}} \|k_i^t(e) - K_{\mathcal{T},i}^t(e)\|_{\mathcal{C}}^2 \, \mathrm{d}\rho_i \leq \eta\right\} \geq 1 - \mathcal{N}(\mathcal{F}, \frac{\eta}{24M}) e^{\frac{-m\eta}{288M^2}}. \tag{3.7}$$

Let us now define the norm $\|\cdot\|_{\mathcal{F}_\rho^N}$ on $\mathcal{F}^N$ by:

$$\|F\|_{\mathcal{F}_\rho^N} = \sqrt{\sum_{i=1}^N \int_{\mathcal{E}} \|F_i(e)\|_{\mathcal{C}}^2 \, \mathrm{d}\rho_i}, \qquad \text{for all } F \in \mathcal{F}^N.$$

For all $1 \leq i \leq N$, one has:

$$\int_{\mathcal{E}} \|k_i^t(e) - K_{\mathcal{T},i}^t(e)\|_{\mathcal{C}}^2 \, \mathrm{d}\rho_i \leq \eta \implies \sum_{i=1}^N \int_{\mathcal{E}} \|k_i^t(e) - K_{\mathcal{T},i}^t(e)\|_{\mathcal{C}}^2 \, \mathrm{d}\rho_i \leq N\eta. \tag{3.8}$$

In particular,

$$\cup_{i=1}^N \left\{\int_{\mathcal{E}} \|k_i^t(e) - K_{\mathcal{T},i}^t(e)\|_{\mathcal{C}}^2 \, \mathrm{d}\rho_i \leq \eta\right\} \subset \{\|k^t - K_{\mathcal{T}}^t\|_{\mathcal{F}_\rho^N} \leq N\eta\}.$$

Thus, gathering (3.7) and (3.8), and using the convexity of the exponential,

$$\begin{aligned}
\text{Prob}\left\{\|k^t - K_{\mathcal{T}}^t\|_{\mathcal{F}_\rho^N} \leq N\eta\}\right\} &\geq \text{Prob}\left\{\cup_{i=1}^N \{\int_{\mathcal{E}} \|k_i^t(e) - K_{\mathcal{T},i}^t(e)\|_{\mathcal{C}}^2 \, \mathrm{d}\rho_i \leq \eta\}\right\} \\
&\geq (1 - \mathcal{N}(\mathcal{F}, \frac{\eta}{24M}) e^{\frac{-m\eta}{288M^2}})^N \\
&\geq 1 - N\mathcal{N}(\mathcal{F}, \frac{\eta}{24M}) e^{\frac{-m\eta}{288M^2}}. \qquad (3.9)
\end{aligned}$$

Let $d_{\mathcal{F}_\rho^N}$ be the distance on $\mathcal{F}^N$ defined by the norm $\|\cdot\|_{\mathcal{F}_\rho^N}$. For all f and g in $\mathcal{F}^N$, we have:

$$d_{\mathcal{F}_\rho^N}(f, g)^2 = \sum_{i=1}^N \int_{\mathcal{E}} \|f_i(e) - g_i(e)\|_{\mathcal{C}}^2 \, \mathrm{d}\rho_i = \int_{\mathcal{E}} \|f(e) - g(e)\|_A^2 \, \mathrm{d}\rho_i,$$

with

$$\|x\|_A^2 = \sum_{i=1}^N \|x_i\|_{\mathcal{C}}^2 \quad \forall x \in (\mathbb{R}^l)^N.$$

For all f and g in $\mathcal{F}^N$ we also have:

$$\begin{aligned}
d(f, g)^2 &= \int_{\mathcal{E}} \sum_{i=1}^l \|((f_1(e) - g_1(e))_l, \ldots, (f_N(e) - g_N(e))_l)\|_{\mathcal{C}}^2 \, \mathrm{d}\rho_i \\
&= \int_{\mathcal{E}} \|f(e) - g(e)\|_B^2 \, \mathrm{d}\rho_i,
\end{aligned}$$

with

$$\|x\|_B^2 = \sum_{i=1}^l \|((x_1(e))_l, \ldots, (f_N(e) - g_N(e))_l)\|^2, \quad \forall x \in (\mathbb{R}^l)^N.$$

All the norm being equivalent on $(\mathbb{R}^l)^N$, there exist C'_A and C_A such that

$$C'_A \|x\|_B \leq \|x\|_A \leq C_A \|x\|_B, \quad \forall x \in (\mathbb{R}^l)^N.$$

Hence,

$$C'_A d(f, g) \leq d_{\mathcal{F}^N(f,g)} \leq C_A d(f, g), \quad \forall l f, g \in \mathcal{F}.$$

Using (3.9) with confidence at least $1 - N\mathcal{N}(\mathcal{F}, \frac{\eta}{24M})e^{\frac{-m\eta}{288M^2}}$, we finally have

$$d(k^t, \mathcal{M}_\mathcal{F}) \leq C_A\sqrt{N\eta} + \alpha_t d(k^{t-1}, \mathcal{M}_\mathcal{F}).$$

Iterating on the discrete times, one has with confidence at least $1 - tN\mathcal{N}(\mathcal{F}, \frac{\eta}{24M})e^{\frac{-m\eta}{288M^2}}$ that

$$d(k^t, \mathcal{M}_\mathcal{F}) \leq C_A\sqrt{N\eta}\left(\sum_{i=0}^{t-1}\prod_{j=1}^{i}\alpha_j\right) + \prod_{i=1}^{t}\alpha_i d(k^0, \mathcal{M}_\mathcal{F}).$$

Let $\alpha_* = \max_{i=0,\dots,N}\max_t \alpha_i(t)$. According to Lemma 3, one has $\alpha_* < 1$, yielding that

$$\begin{aligned} d(k^t, \mathcal{M}_\mathcal{F}) &\leq C_A\sqrt{N\eta}\left(1 + \alpha_+ \ldots + \alpha_*^{t-1}\right) + \alpha_*^t d(k^0, \mathcal{M}_\mathcal{F}). \\ &\leq \frac{C_A}{1-\alpha_*}\sqrt{N\eta} + \alpha_*^t d(k^0, \mathcal{M}_\mathcal{F}). \end{aligned}$$

Thus, for any $0 < \delta < 1$, choosing the parameter m such that

$$m \geq \frac{288M^2}{\eta}\left(\ln\left(\frac{tN\mathcal{N}(\mathcal{F}, \frac{\eta}{24M})}{\delta}\right)\right), \tag{3.10}$$

yields with confidence at least $1 - \delta$ that

$$d(k^t, \mathcal{M}_\mathcal{F}) \leq \frac{C_A}{1-\alpha_*}\sqrt{N\eta} + \alpha_*^t d(k^0, \mathcal{M}_\mathcal{F}).$$

Taking $\eta = \alpha_*^{2t} d(k^0, \mathcal{M}_\mathcal{F})^2/N$ finishes the proof. □

Remark 6. *When time t goes to infinity, so does the number of sample m_t needed for the convergence in Theorem 2 to occur.*

Indeed, using Section 7.1 of [5]*, there exists $C_\mathcal{F} > 0$ and $a > 0$ such that*

$$\ln \mathcal{N}(\mathcal{F}, \epsilon) \leq C_\mathcal{F}\left(\frac{1}{\epsilon}\right)^a.$$

Plugging this into (3.10) *yields that*

$$\begin{aligned} m_t &\leq \frac{288NM^2}{(1-\alpha_*)^2\alpha_*^{2t}d(k^0, \mathcal{M}_\mathcal{F})^2}\ln\left(tNC_\mathcal{F}\left(\frac{24NM}{(1-\alpha_*)^2\alpha_*^{2t}d(k^0, \mathcal{M}_\mathcal{F})^2}\right)^a + \ln\left(\frac{1}{\delta}\right)\right) \\ &\leq \frac{288NM^2}{(1-\alpha_*)^2\alpha_*^{2t}d(k^0, \mathcal{M}_\mathcal{F})^2}\left(\ln(tN) + C_\mathcal{F}\left(\frac{24NM}{(1-\alpha_*)^2\alpha_*^{2t}d(k^0, \mathcal{M}_\mathcal{F})^2}\right)^a + \ln\left(\frac{1}{\delta}\right)\right) \end{aligned}$$

Choosing appropriately δ as a function of t, one can show that f^t tends to $\mathcal{M}_\mathcal{F}$ almost surely. One can then define the minimal sampling size $m(t)$ by

$$m(t) = \frac{288NM^2}{(1-\alpha_*)^2\alpha_*^{2t}d(k^0, \mathcal{M}_\mathcal{F})^2}\left(\ln(t^2N) + C_\mathcal{F}\left(\frac{24NM}{(1-\alpha_*)^2\alpha_*^{2t}d(k^0, \mathcal{M}_\mathcal{F})^2}\right)^a\right),$$

which tends to $+\infty$ when $t \to +\infty$.

Corollary 4. *Let f_m^t be the process at the t when the size of the sample in the dynamics is m. One has that*

$$\sup_{\epsilon>0}\lim_{t\to\infty}\text{Prob}\left\{d(f_{m(t)}^t, \mathcal{M}_\mathcal{F}) \leq \epsilon\right\} = 1$$

Proof. Let $\epsilon > 0$. For all t big enough, one has

$$A\alpha_*^t d(f^0_{m(t)}, \mathcal{M}_{\mathcal{F}}) < \epsilon.$$

Taking $\delta = \frac{1}{t}$ yields

$$\text{Prob}\left\{ d(f^t_{m(t)}, \mathcal{M}_{\mathcal{F}}) \leq \epsilon \right\} \geq 1 - \frac{1}{t}.$$

□

4. Numerical simulations

Let us now both illustrate the mathematical results of the paper, such as Theorem 2, and show that some generalizations also hold when individual learning is possible ($\tau > 0$). Individual learning allows innovations, new experiences and observations, and original conceptualizations which make possible the evolution of knowledge for both the individuals and the population. By analogy, individual learning plays the same role for the evolution of knowledge than genetic mutations for the biological evolution of species [11].

4.1. **Illustration of the main theorem.**

Our model aims to be used by theoretical anthropologists. To show its usefulness, we illustrate the results of our main theorem in specific cases.

Test 1. Impact of self-inertia. As a first numerical test, we aim to illustrate Theorem 2. Let $\mathcal{E} = \{1, ..., 5\}$ and $\mathcal{C} = [-10, 10]$. We consider a relationship between two individuals (labeled 1 and 2). The structure matrix 3 is given by

$$\Gamma = \begin{pmatrix} \alpha & 1-\alpha \\ 1-\alpha & \alpha \end{pmatrix},$$

where $\alpha \in [0, 1]$.

The parameter α can be interpreted as cognitive (see Remark 2) or self-inertia. The higher the α, the less individuals' knowledge-like functions change along the dynamics. As likelihood landscape (4), we take $L(e, c) = 1$ for all $e \in \mathcal{E}$ and $c \in \mathcal{C}\backslash\{0\}$, and take $c_{\min} = 0.1$. We define $\mathcal{F}$ as the set of continuous functions from $\mathcal{E}$ to $\mathcal{C}$ so $\mathcal{F}$ is convex. As $\mathcal{E}$ contains a finite number of elements and $\mathcal{C}$ is compact, then $\mathcal{F}$ is compact. We set $\tau = 0$ so the dynamics is only driven by social learning.

When α varies in $(0, 1)$, all the hypotheses of Theorem 2 are met (even though, strictly speaking, we do not illustrate exactly the theorem because we cannot compute m_t). Our numerical simulations show that according to this result the population converges to a common shared knowledge.

At the initial state, the knowledge-like functions of individuals 1 and 2 are k_1^0 and k_2^0, respectively, given by

$$k_1^0(e) = 2 \quad \forall e \in \mathcal{E}, \qquad k_2^0(e) = 6 \quad \forall e \in \mathcal{E}.$$

Let d be the distance defined in Corollary 1. By using numerical simulations we follow the evolution of $d(k^t, \mathcal{M}_{\mathcal{F}})$ through time for different values of the parameter α. We ran 100 simulation replicates. The mean dynamics is presented in Figure 6.

When $\alpha \neq 1$ the population rapidly converges to a common shared knowledge (Fig. 6) as predicted by Theorem 2. We notice that this convergence is exponential, as expected given that the process is driven by an inhomogeneous Markov chain. We notice that the convergence is faster when $\alpha = 0.5$.

When $\alpha = 1$ the matrix does not respect the hypothesis of Theorem 2 since $\gamma_{12} = \gamma_{21} = 0$. It corresponds to the case where the individuals do not communicate with each other. Thus individuals knowledge-like functions do not vary through time, and the process does not converge towards a common shared knowledge.

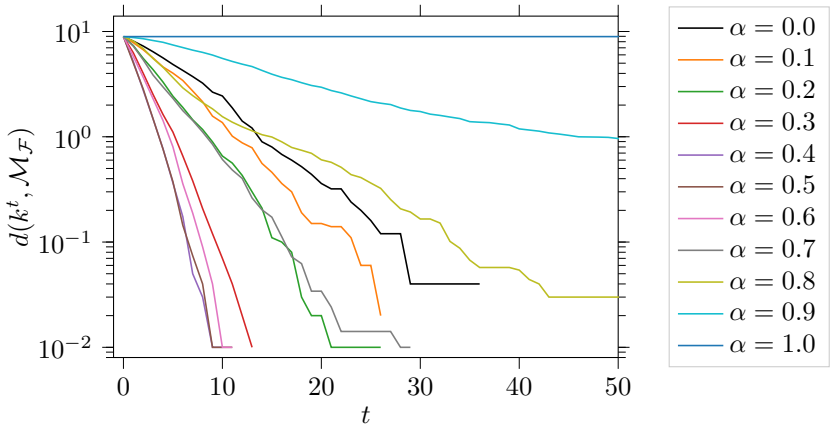


FIGURE 6. **Test 1.** Evolution of the distance between k^t and the set $\mathcal{F}$ through time.

Test 2. A professor and its audience.

Now let us consider a population of 5 individuals: 1 professor (1) and 4 students (2, 3, 4, 5). We keep the same setting as previously, namely no individual learning ($\tau = 0$), since it is a purely teaching situation.

Let the structure matrix be

$$\Gamma = \begin{pmatrix} 1 & 0.01 & 0.01 & 0.01 & 0.01 \\ 1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 1 & 0.1 & 0.1 & 0.1 & 0.1 \end{pmatrix}.$$

At the initial state, knowledge-like functions are defined by:

$$k_1^0(e) = 5 \quad \forall e \in \mathcal{E},$$

and

$$k_i^0(e) = 1 \quad \forall e \in \mathcal{E}, \forall i \in \{2, \dots, 5\},$$

so at the initial time, all students have the same knowledge.

We consider two cases. First, the likelihood landscape is assumed constant. Second, it is considered concave assuming the concept c is fixed: we set for all $(e, c) \in \mathcal{E} \times \mathcal{C} \backslash \{0\}$

$$L(e, c) = e^{\frac{(e-6)^2}{10}},$$

so the professor has a knowledge-like function that is more likely than that one of her students.

We call k_{eq} the common shared knowledge at the equilibrium. We define Δ_i as the distance between the initial knowledge of individual i and the common shared knowledge. We have

$$\Delta_i = d_\mathcal{C}(k_i^0, k_{eq}),$$

with $d_\mathcal{C}$ the distance induced by the inner product on $\mathcal{C}$.

We ran 100 numerical simulations as previously. Results are shown in Figures 7(a) and 7(b).

Figure 7(a) shows the evolution of the distance to space $\mathcal{F}$ with time. In both cases, whether the likelihood landscape is fixed or concave, the population rapidly converges to a common shared knowledge. When the likelihood landscape is concave, the professor has a strong influence on her students and the convergence to a common shared knowledge is faster.

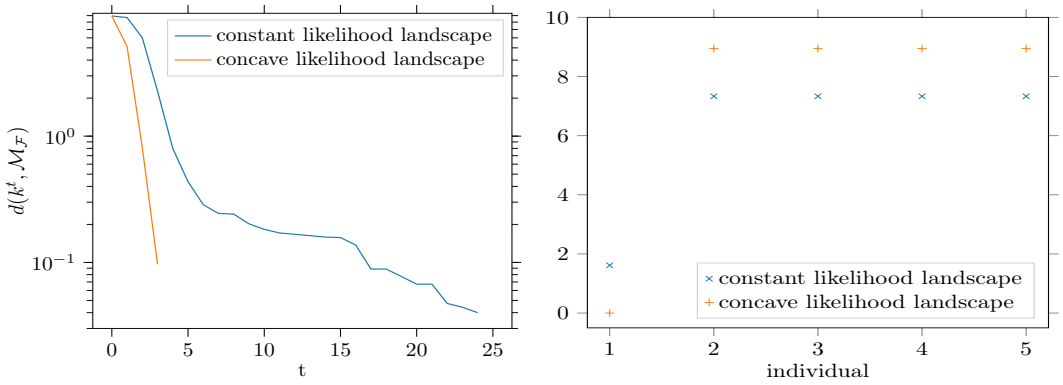


FIGURE 7. **Test 2.**(a) Evolution of the distance between k^t and the set $\mathcal{F}$ through time. (b) Δ_i for each individual i in the population. The blue crosses represent the case where the likelihood is constant, and the orange pluses show the case where the likelihood is concave.

Figure 7(b) shows the values of $\Delta_.$. In both cases, the common shared knowledge is close to the professor's initial one. This common shared knowledge is farther from the students' initial knowledge than from the professor's. When the likelihood landscape favors the professor influence, the common shared knowledge is closer to the initial professor knowledge.

4.2. **Creation of knowledge.** We now consider the case where individual learning is present, namely $\tau > 0$. Although we couldn't prove a convergence result for this case, we can still use numerical approaches when the parameters of the model do not allow analytical resolution.

Test 3. Creation of knowledge among interacting individuals. In this part we set $\mathcal{E} = \{1, ..., 25\}$ and $\mathcal{C} = \mathbb{R}$. We define the likelihood landscape as

$$L(e, c) = \begin{cases} \frac{1}{2} & \text{if } c = 0, \\ \exp -(x-1)^2 & \text{otherwise,} \end{cases}$$

such that the function $1_\mathcal{F}$ is defined as:

$$\forall e \in E, 1_\mathcal{F}(e) = 1,$$

is the most likely function. We consider a population of ten individuals and we set a initial state where $K^0 = (0_\mathcal{F}, ..., 0_\mathcal{F})$. Let Γ be the square matrix of size N full of 1. Thus at the initial state, individuals are "newborn", that is, they do have not conceptualized any experiences. We investigate convergence of knowledge towards the function $1_\mathcal{F}$ and its dynamics by simulation runs.

We define the relative entropy (RE) of the population as

$$\text{RE(t)} = -\frac{1}{N} \sum_{i=1}^{N} d_\mathcal{F}(k_i^t, 1_\mathcal{F}), \tag{4.1}$$

with for all f, g in $\mathcal{F}$,

$$d_\mathcal{F}(f, g) = \sqrt{\sum_{e=1}^{N_E} \frac{(f(e) - g(e))^2}{N_E}}. \tag{4.2}$$

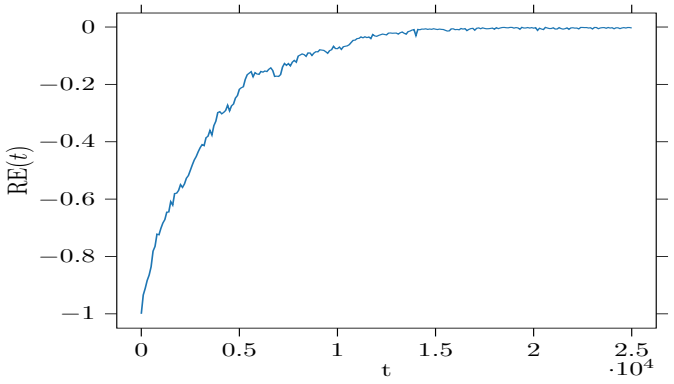


FIGURE 8. **Test 3.** Evolution of the relative entropy through time.

When every individual in the population has $1_{\mathcal{F}}$ as knowledge-like function, the relative entropy is maximal and equals 0. We use the relative entropy as a measure of knowledge in the population *i.e.* the higher the relative entropy, the more likely the individuals' knowledge. This allows us to quantify the effect of parameters on the evolution of knowledge. Figure 8 shows that the relative entropy increases with time.

In simulations, the individual learning rate $\tau = 0.02$. Individual learning results in new experiences and observations, while social learning promotes the spread of adequate conceptualizations. The combined effect of individual and social learning allows the population to evolve towards better solutions (Figure 9).

4.3. **Comparison with a language model.** Cucker, Smale and Zhou developed a model to describe the evolution of language [5]. Our work is stongly inspired by their work. For our purpose, we needed to substantially modify this model by introducing the credibility matrix and individual learning. However, interpretation of the variables of the model is different: in their model a language-like function is a function from a space of objects to a space of signals. As in our case, they proved the convergence of the languages of different individuals to a common shared language, although under different hypotheses (see also Remark 5).

In their model, influences between individuals do not vary through time. In reality, we expect influences between individuals to be dynamic, and that is why we introduced the credibility matrix which changes at each time step and modifies the interactions within the population.

Test 4. On the evolution of language. We modified our numerical method in order to simulate the model of language evolution developed in [5]. We consider two different linguistic communities of two individuals with few interactions. We take $\mathcal{E} = \{1, ..., 5\}$ and $\mathcal{C} = [-10, 10]$. The individuals of the first and the second communities have the language-like function k_1 and k_2, respectively. Where k_1 and k_2 correspond to two different languages. This language-like function is defined as

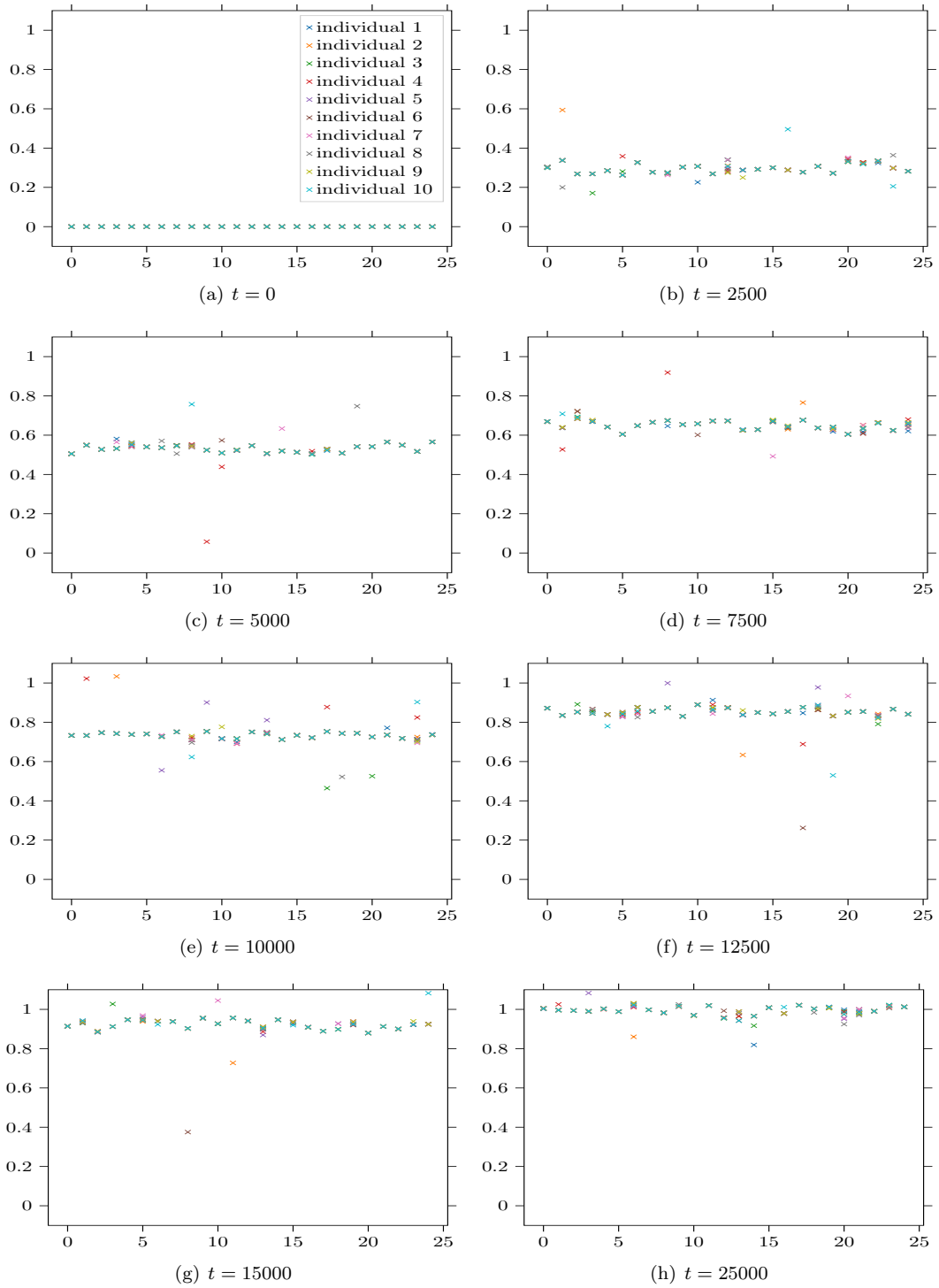
$$\forall e \in \mathcal{E}, k_1(e) = 5,$$

and

$$\forall e \in \mathcal{E}, k_2(e) = 7.$$

We take

$$\Gamma = \begin{pmatrix} 1 & 1 & 0.01 & 0.01 \\ 1 & 1 & 0.01 & 0.01 \\ 0.01 & 0.01 & 1 & 1 \\ 0.01 & 0.01 & 1 & 1 \end{pmatrix},$$



(a) $t = 0$ (b) $t = 2500$

(c) $t = 5000$ (d) $t = 7500$

(e) $t = 10000$ (f) $t = 12500$

(g) $t = 15000$ (h) $t = 25000$

FIGURE 9. **Test 3.** Knowledge-like functions for every individuals in the population at different given times.

so the two linguistic communities hardly interact. Numerical simulations show that the two communities converge to a common shared language: Figure 10 shows that the distance between the process k^t and the set $\mathcal{F}$ tends to 0.

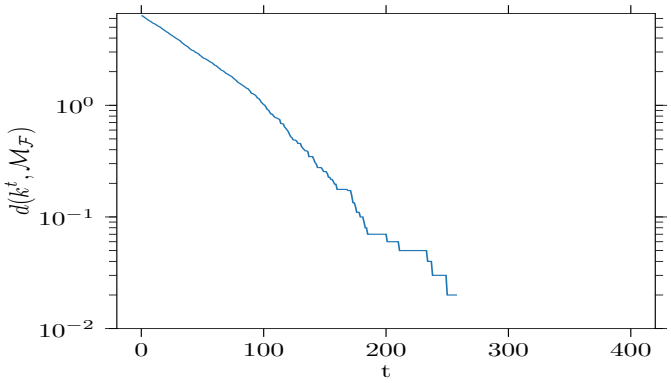


FIGURE 10. **Test 4.** Evolution of the distance between k^t and the set $\mathcal{F}$ through time.

5. Conclusion

The aim of this work was to develop a more general mathematical model of knowledge evolution than the existing ones e.g. [8, 14]. Existing models have been widely used to investigate the impact of population size on the evolution of knowledge. However, they rely on strong assumptions and omit important aspects of social dynamics. Here, we developed a hybrid model, between an individual based stochastic model and a learning algorithm, that relaxes hypotheses and incorporates various forms of social interaction dynamics.

Analytical results show that interacting individuals converge with high probability towards a common shared knowledge, when no innovation occurs (i.e. no individual learning). Numerical simulations show that these results hold when individuals combine individual and social learning and that conceptualizations that appropriately reflect the structure of the world emerge across time. This model can be used to investigate knowledge evolution in hierarchically or spatially structured populations of variable sizes.

References

[1] Albi, G., Bellomo, N., Fermo, L., Ha, S. H., Kim, J., Pareschi, L., Poyato, D., and Soler, J. Vehicular traffic, crowds, and swarms. from kinetic theory and multiscale methods to applications and research perspectives. Math. Mod. Meth. Appl. Sci. 29, 10 (2019), 1901–2005.

[2] Ambrosio, L., Fornasier, M., Morandotti, M., and Savaré, G. Spatially inhomogeneous evolutionary games. preprint arXiv 1805.04027.

[3] Boyd, R., Richerson, P. J., and Henrich, J. The cultural niche: Why social learning is essential for human adaptation. Proceedings of the National Academy of Sciences 108, Supplement 2 (2011), 10918–10925.

[4] Cucker, F., and Smale, S. On the mathematical foundations of learning. Bulletin of the American Mathematical Society 39, 1 (2002), 1–49.

[5] Cucker, F., Smale, S., and Zhou, D.-X. Modeling Language Evolution. Foundations of Computational Mathematics 4, 3 (2004), 315–343.

[6] Derex, M., Beugin, M.-P., Godelle, B., and Raymond, M. Experimental evidence for the influence of group size on cultural complexity. Nature 503 (Nov. 2013), 389–391.

[7] Gopnik, A., O'Grady, S., Lucas, C. G., Griffiths, T. L., Wente, A., Bridgers, S., Aboody, R., Fung, H., and Dahl, R. E. Changes in cognitive flexibility and hypothesis search across human life history from childhood to adolescence to adulthood. Proceedings of the National Academy of Sciences 114, 30 (2017), 7892–7899.

[8] Henrich, J. Demography and Cultural Evolution: How Adaptive Cultural Processes can Produce Maladaptive Losses: The Tasmanian Case. American Antiquity 69, 2 (2004), 197–214.
[9] Kato, T. Perturbation theory for linear operators, second edition, vol. 132. Springer Science & Business Media, 1995.
[10] Le Gall, J.-F. Intégration, Probabilités et Processus Aléatoires. Lecture notes, downloaded from `https://www.math.u-psud.fr/~jflegall/IPPA2.pdf` in Oct. 2019.
[11] Mesoudi, A. Cultural Evolution, How Darwinian theory can explain human culture and synthesize the social sciences. University of Chicago Press, 2011.
[12] Muthukrishna, M., Shulman, B. W., Vasilescu, V., and Henrich, J. Sociality influences cultural complexity. Proceedings of the Royal Society B: Biological Sciences 281, 1774 (2014), 20132511.
[13] Pollard, D. Convergence of stochastic processes. Springer Series in Statistics. Springer-Verlag, New York, 1984.
[14] Powell, A., Shennan, S., and Thomas, M. Late pleistocene demography and the appearance of modern human behavior. Science (New York, N.Y.) 324 (07 2009), 1298–301.
[15] Richerson, P. J., and Boyd, R. Not by genes alone. University of Chicago Press, 2010.
[16] Seneta, E. Non-negative Matrices and Markov Chains. 0172-7397. Springer, New York, NY, 1981.

Sylvain Billiard
Univ. Lille, CNRS, UMR 8198 - Evo-Eco-Paleo, F-59000 Lille, France
Email: sylvain.billiard@univ-lille.fr

Maxime Derex
Institute for Advanced Study in Toulouse, CNRS, UMR 5314
F-31015 Toulouse, France
Email: maxime.derex@iast.fr

Ludovic Maisonneuve
National Museum of Natural History, UMR 7205, Institut de Systématique, Evolution et Biodiversité
F-75005 Paris, France
Email: ludovic.maisonneuve@mnhn.fr

Thomas Rey
Univ. Lille, CNRS, UMR 8524, Inria – Laboratoire Paul Painlevé
F-59000 Lille, France
Email: thomas.rey@univ-lille.fr