



**HAL**  
open science

# Convergence of knowledge in a stochastic cultural evolution model with population structure, social learning and credibility biases

Sylvain Billiard, Maxime Derex, Ludovic Maisonneuve, Thomas Rey

► **To cite this version:**

Sylvain Billiard, Maxime Derex, Ludovic Maisonneuve, Thomas Rey. Convergence of knowledge in a stochastic cultural evolution model with population structure, social learning and credibility biases. *Mathematical Models and Methods in Applied Sciences*, World Scientific Publishing, 2020, 30 (14), pp.2691-2723. 10.1142/S0218202520500529 . hal-02357188v2

**HAL Id: hal-02357188**

**<https://hal.archives-ouvertes.fr/hal-02357188v2>**

Submitted on 20 Jul 2020

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

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

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

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

ABSTRACT. Understanding how knowledge emerges and propagates within groups is crucial to explain the evolution of human populations. In this work, we introduce a mathematically oriented model that draws on 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 by which innovations are gradually added to an existing pool of knowledge, most often over multiple generations [4, 28]. The ability to learn from others (social learning) is pivotal to this process because it allows innovations to be passed from individual to individual and from generation to generation.

Since the beginning of the twentieth century, much energy has been devoted to understand the mechanics of learning and cultural innovation. Among the oldest seminal papers, [14, 23] used the so-called “stimulus sampling” approach, which formalizes the effects of repetition on the learning process following a ‘trial and error’ procedure. Stochastic models were subsequently used to take the effect of randomness into account in the learning process. Despite their simplicity, these models correctly describe simple repetitive experiments [6, 7]. These two approaches were then extended and generalized as learning theory in the early 2000s in the seminal papers by Cucker and Smale [8, 10]. These papers introduced and extended the modern theory of mathematical learning process by splitting it into two fundamental mechanisms: cognitive learning where an individual increases its knowledge by its own experience, and social learning by which an individual learns from the interaction with other individuals. After these works, many new models have been published, either stochastic or deterministic, using various approaches ranging

---

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 Skłodowska-Curie grant agreement number 748310 and acknowledges funding from ANR under grant ANR-17-EURE-0010 (Investissements dAvenir program). SB has been supported by the chair Modélisation Mathématique et Biodiversité of Veolia Environnement-Ecole Polytechnique-Museum National dHistoire Naturelle-Fondation X.

from microscopic agent based models [27, 9, 24, 19], to mesoscopic kinetic theory of active particles [3, 15, 26, 11] to macroscopic fluid models [18, 17] (see [5] for a detailed review).

In this work, we develop a new mathematical model that aims to describe the dynamics of knowledge creation and propagation among interacting individuals. Our approach generalizes the Cucker-Smale theory of learning by introducing different possible biases and asymmetries in knowledge transmission due to population structure, individuals' status, or individuals' credibility. As a result, the influence of individuals over each other varies through time and various hypotheses from [10] are relaxed.

**1.2. Outline of the paper.** The model is properly introduced and simple applications are given in Section 2. In Section 3, we 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 [10] 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. Conceptualizations are defined 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 their interactions with others (social learning) and random exploration (individual learning). Our model is an extension of the model of Cucker, Smale and Zhou describing the evolution of language [10], and can be seen as an hybrid between a learning algorithm [8] and an individual based model [2, 1].

We suppose that individuals influence each other according to a social learning matrix  $\Lambda \in \mathcal{M}_N(\mathbb{R})$ . This matrix depends on individuals' perception of each other's credibility through a credibility matrix  $C \in \mathcal{M}_N(\mathbb{R})$ . As a consequence, individuals' credibility and how knowledge is transmitted between individuals change at each time step. The social learning matrix  $\Lambda$  also depends on the structure of the population described by a structure matrix  $\Gamma \in \mathcal{M}_N(\mathbb{R})$  (*e.g.* a professor has a strong impact on her students, while students have a lesser impact on their professor or individuals learn preferentially from spatially closer individuals). 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  $\mathbf{e}$ :

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

- 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 linking the experience set  $\mathcal{E}$  to the concept set  $\mathcal{C}$ .

Each knowledge-like function 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 with colors. Let  $\mathcal{E}$  be  $[0, 1000] \subset \mathbb{R}$  representing the set of wavelengths in nanometers and  $\mathcal{C} = \{0, \text{red}, \text{yellow}, \text{green}, \text{blue}, \text{purple}\}$  contains color names and 0. In this case  $\mathcal{E}$  is a continuous space and  $\mathcal{C}$  is a discrete space. A knowledge-like function assigns 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.

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

We recall that  $[380, 750]$  is the range of the visible spectrum. Let's suppose an individual who never experienced wavelengths in  $[0, 380] \cup [750, 1000]$ . His knowledge-like function associates all experiences in  $[0, 380] \cup [750, 1000]$  with 0. An individual with a knowledge-like function  $f$  assigns each element of  $\mathcal{E}$  to a color as shown in Figure 1.

**Remark 1.** In Japanese the color word (*ao*) includes what Western people would call green and blue. As with other languages, the Japanese language did not initially differentiate between these two colors. This language could be modeled by the knowledge-like function  $f'$ :

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

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

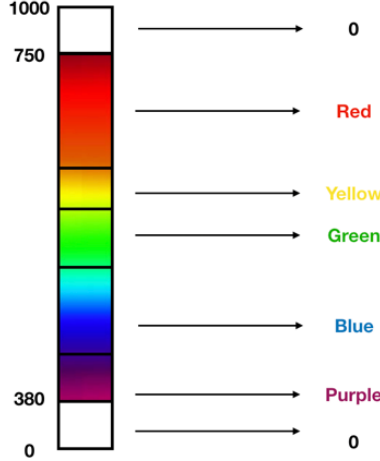


FIGURE 1. Illustration of a knowledge-like function of the visible spectrum. The arrows point to the color associated with each wavelength and illustrate the knowledge-like function  $f$  (2.3).

**2.2. Learning algorithm.** Let  $N$  be the number of individuals in the population. Each individual  $i$  is associated with a knowledge-like function  $k_i \in \mathcal{F}$ . We consider a dynamical, discrete time model: knowledge-like functions evolve over time because of social and individual learning. Let us denote by  $k^t := (k_1^t, \dots, k_N^t) \in \mathcal{F}^N$  the state of the population at time  $t > 0$ . As time evolves, individuals modify their conceptualization through the learning algorithm presented in [8]:

**Definition 3.** *The learning algorithm computes the knowledge-like function at the next time step using a least-square procedure*

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

where  $S_i^t = \{(e_1^{i,t}, c_1^{i,t}), \dots, (e_m^{i,t}, c_m^{i,t})\}$  with  $(e_j^{i,t}, c_j^{i,t}) \in \mathcal{E} \times \mathcal{C}$  for  $j \in \{1, \dots, m\}$ .

For each individual  $i$ , the elements of  $S_i^t$  represent what will shape the knowledge of  $i$  at the next step. The sampling  $S_i^t$  is done randomly using the probability measure  $\rho^{i,t}$  defined in (2.4). The elements of  $S_i^t$  result from social or individual learning. The relative importance of individual vs. social learning is controlled by a parameter  $\tau$  such that:

**Definition 4.** *Let us denote by  $\tau \geq 0$  the proportion of individual learning, fixed and independent of  $i$ . The sampling measure is defined by*

$$(2.4) \quad \rho^{i,t} = (1 - \tau)\rho_\Lambda^{i,t} + \tau\rho_{\mathcal{I}}^{i,t},$$

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.

We assume that the extent to which individuals are influenced by each other is determined by the social learning matrix  $\Lambda = (\lambda_{i,j}^t)_{1 \leq i, j \leq N} \in \mathcal{M}_N(\mathbb{R})$  (see 2.7 below for the calculation).

**Definition 5.** The probability measure  $\rho_\Lambda^{i,t}$  representing social learning is defined by:

$$(2.5) \quad \rho_\Lambda^{i,t}(e, c) \propto \sum_{j=1}^N \lambda_{ij}^t \mathbb{1}_{\{k_j(e)=c\}},$$

where  $\lambda_{ij}^t$  describes the influence of individual  $j$  on individual  $i$  (see 2.7 below for the calculation of the social learning matrix  $\Lambda^t$ ) at time  $t$ .

According to (2.5), 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 6.** The probability measure  $\rho_{\mathcal{I}}^{i,t}$  representing individual learning is given by

$$(2.6) \quad \begin{cases} \int_{c \in \mathcal{C}} \rho_{\mathcal{I}}^{i,t}(e, c) dc \propto \int_{e' \in \mathcal{E}} \mathbb{1}_{\{k_i^t(e') \neq 0\}} \mathbb{1}_{\{e \in E\}} \exp - \frac{\|e - e'\|^2}{2\sigma_E^2} de', \\ \rho_{\mathcal{I}}^{i,t}(c|e) \propto \mathbb{1}_{\{c \in \mathcal{C}\}} \exp - \frac{\|c - f_i^t(e)\|^2}{2\sigma_C^2}, \end{cases}$$

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  $\phi$  by

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

In this 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  $\pi : \mathcal{E} \times \mathcal{C} \rightarrow \mathcal{E}$  is the projection on  $\mathcal{E}$ .

Thus, the individual learning phase is 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. The concept  $c$  is drawn following a probability law centered and concentrated on  $k_i^t(e)$ .

**2.3. Social learning.** Social learning depends on the influence of each individual over each other described by the social learning matrix  $\Lambda \in \mathcal{M}_N(\mathbb{R})$ . This matrix captures the effect of the structure of the population and reflects, for instance, spatial structure or differences in social statuses. Social learning also depends on the credibility that individuals assign to each other based on their own knowledge-like function (as described in Section 2.4). Population structure and individuals credibility are described by the matrices  $\Gamma = (\gamma_{i,j}^t)_{1 \leq i,j \leq N}$  and  $C = (c_{i,j}^t)_{1 \leq i,j \leq N}$  respectively.

**Definition 7.** A structure matrix  $\Gamma = (\gamma_{ij})_{1 \leq i,j \leq N}$  is a square matrix of size  $N$  describing the structural influence that individuals have on each other. More precisely for each  $(i, j) \in \{1, \dots, N\}^2$ ,  $\gamma_{ij} \in \mathbb{R}$  describes the strength of the influence of  $j$  on  $i$ .

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

**Definition 8.** A credibility matrix  $C = (c_{ij})_{1 \leq i,j \leq N}$  is a square matrix of size  $N$  describing how credible the individuals are to each other, hence weighting their mutual influence.

*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 [13], which can be modelled by the condition  $\gamma_{11} < \dots < \gamma_{NN}$ .
- 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 teacher. Because of her status, the teacher has a high influence on her students but is influenced by them to a lesser extent. Assuming that the relationship between the students is symmetric, we have

$$\Gamma = \begin{pmatrix} 1 & 0.1 & 0.1 \\ 1 & 0.2 & 0.2 \\ 1 & 0.2 & 0.2 \end{pmatrix}.$$

Thus the influence of individual  $j$  on  $i$  depends on the structural influence  $\gamma_{ij}$  of  $j$  on  $i$ , and on the credibility  $c_{ij}$  that  $i$  attributes to  $j$ . We shall assume in this work that these phenomena are multiplicative.

**Definition 9.** *The social learning matrix  $\Lambda = (\lambda_{ij})_{1 \leq i, j \leq N}$  is a square matrix of size  $N$  defined by:*

$$(2.7) \quad \lambda_{ij} = \begin{cases} \frac{\gamma_{ij}c_{ij}}{\sum_{l=1}^N \gamma_{il}c_{il}} & \text{if } \sum_{l=1}^N \gamma_{il}c_{il} \neq 0, \\ 1/N & \text{otherwise.} \end{cases}.$$

**2.4. Likelihood landscapes and how they affect credibility.** In our model, some conceptualizations (i.e. knowledge-like functions) appropriately reflect the structure of the world, while others 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) a pair of 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 10.**  $\forall c \in \mathcal{C}$ , we define a function

$$L(\cdot, c) : \mathcal{E} \rightarrow [0, 1], \\ e \mapsto L(e, c).$$

with

$$\forall e \in \mathcal{E}, L(e, 0) = \frac{1}{2}.$$

The map  $L$  is called the likelihood landscape.

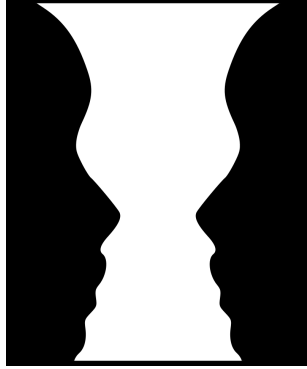


FIGURE 2. On this figure one may see a pair of faces (in black) or a cup (in white). Image courtesy of Bryan Derksen, under license CC BY-SA 3.0

For all  $e, c$  in  $\mathcal{E} \times \mathcal{C}$ ,  $L(e, c)$  represents how well the concept  $c$  explains 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  $\mathcal{C} = \{0, F, R\}$ . 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 is not useful for individuals to discriminate between colors. In that case, we would define the likelihood landscape as  $L(e, c) = 1$  for all  $(e, c) \in \mathcal{E} \times \mathcal{C} \setminus \{0\}$ .

As introduced earlier (Definition 8), the influence individuals have on each other 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 compared to those attributed to other individuals (including herself), that means that individual  $i$  is more prone to adopt individual  $j$ 's conceptualization. We described in Section 2.2 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 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 11.** A credibility matrix  $C = (c_{ij})_{1 \leq i, j \leq N}$  is a square matrix of size  $N$  defined by:

$$(2.8) \quad c_{ij} = \max(\tilde{c}_{ij}, c_{\min}),$$

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

$$(2.9) \quad \tilde{c}_{ij} = \frac{1}{1 + \mathbb{1}_{\{i \neq j\}} \int_{c \in k_j(X)} dc} \exp \left( \int_{e \in E} \ln(L(e, k_j(e)) \mathbb{1}_{\{k_i(e) \neq 0\}}) de \right).$$

**Remark 3.** The term  $\left(1 + \mathbb{1}_{\{i \neq j\}} \int_{c \in k_j(X)} dc\right)^{-1}$  penalizes individuals who use 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  $j$  based on individual  $i$ 's experiences. 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  $i$  already had.

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

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

namely, the second part of the formula is similar to a measure of likelihood in probability theory [22]. 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 is as likely to be flat as 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 four individuals with different knowledge-like functions  $k_1, k_2, k_3$  and  $k_4$  such that:

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

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

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 rows sum up to 1, in order to easily grasp the influences on an individual  $j$  in row  $i$ :

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

Individual #1 has only experienced  $f_1$ , she judges the other individuals (and herself) based on her 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 unlikely to them (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).

### 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 towards 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 distinct processes that we shall analyze separately.

**Definition 12.** Let us define the application  $\mathcal{T} : \mathcal{F} \times \mathbb{N} \mapsto \mathcal{F}$  by

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

We can then define the one step deterministic idealistic process as

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

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

$$(3.3) \quad k^t = \Delta k^t + K_{\mathcal{T}}^t,$$

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

**Definition 13.** Let  $\mathcal{M}_{\mathcal{F}} = \{(k, \dots, k), k \in \mathcal{F}\}$  be the space of all the common shared conceptualizations.

We first prove the contraction of the idealized process  $K_{\mathcal{T}}^t$  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  $\Delta k^t$  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 the idealistic processes  $K_{\mathcal{T}}^t$  is mainly driven by the social learning matrix  $\Lambda$ . In this Section, we study the relationship between the properties of the influence matrix and the interactions taking place within the population.

**Definition 14.** 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 15.** Let  $A \in \mathcal{M}_N(\mathbb{R})$ . Let  $i, j$  be in  $\{1, \dots, N\}$ ,

- We say that  $i$  communicates with  $j$  (denoted  $i \xrightarrow{A} j$ ) if there exists  $n \geq 0$  and  $i_1, \dots, i_n \in \{1, \dots, N\}$  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  $i \not\xrightarrow{A} j$ .

- We say that  $i$  communicates with  $j$  with  $k$  intermediates if there exists  $i_1, \dots, i_k \in \{1, \dots, N\}$  such that

$$a_{ii_1} \prod_{l=1}^k (a_{i_l i_{l+1}}) a_{i_{k+1} j} > 0,$$

and if there exists no  $i_1, \dots, i_{k-1} \in \{1, \dots, N\}$  such that

$$a_{ii_1} \prod_{l=1}^{k-1} (a_{i_l i_{l+1}}) a_{i_k j} > 0,$$

- Let  $I, J \in \mathcal{P}(\{1, \dots, N\})$ . 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  $\{1, \dots, N\}$ ,  $i \rightarrow j$ . Moreover if there is  $k$  such that  $A^k > 0$ , then for each  $i, j$  in  $\{1, \dots, N\}$   $i$  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

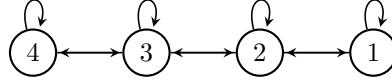


FIGURE 3. Graph representing the matrix  $A$ .

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

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

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

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

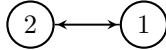


FIGURE 4. Graph representing the matrix  $B$ .

$k \in \mathbb{N}$ ,

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

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 eigenvalue 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 **the property***

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

*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  $A' = P^T A P$  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 :  $n_l + n_h = N$*

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

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

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

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

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

We conclude as before.

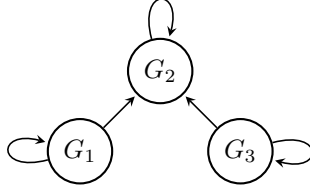


FIGURE 5. Illustration of the influence relationship between the three clusters  $G_1$ ,  $G_2$ , and  $G_3$

□

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

Let  $\Gamma$  and  $C$  be respectively the structure and credibility matrices defined in Def. 7 and equation (2.9). By construction,  $C$  is a stochastic matrix. The social learning matrix  $\Lambda$  is defined according to (2.7) by

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

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

*Proof.*  $\forall i, j \in \{1, \dots, N\}, \gamma_{ij} > 0 \implies \lambda_{ij} > 0$  using (2.7). Thus  $\forall i, j \in \{1, \dots, N\}, i \xrightarrow{\Gamma} j$  or  $j \xrightarrow{\Gamma} i \implies i \xrightarrow{\Lambda} j$  or  $j \xrightarrow{\Lambda} i$ .

□

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

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

$$\begin{aligned} \forall i, j \in \{1, \dots, N\}, \gamma_{ij}^k > 0 &\iff \forall i, j \in \{1, \dots, N\}, \\ &\sum_{0 \leq l_1, \dots, l_{k-1} \leq N} \gamma_{il_1} \gamma_{l_1 l_2} \dots \gamma_{l_{k-1} j} > 0, \\ &\iff \forall i, j \in \{1, \dots, N\}, \exists l_1, \dots, l_{k-1} \in \{1, \dots, N\}, \\ &\gamma_{il_1} \gamma_{l_1 l_2} \dots \gamma_{l_{k-1} j} > 0. \end{aligned}$$

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

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

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

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

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  $\Lambda$  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  $(\alpha_i)_{i=1\dots N}$  of its eigenvalues:*

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

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

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

*Proof.* Let  $\mathcal{A} = \{A \in \mathcal{M}^N(\mathbb{R}), \forall i \in \{1, \dots, N\}, \sum_j a_{ij} = 1, \forall i, j \in \{1, \dots, N\}, a_{ij} \geq \underline{m}_\lambda\}$ . Being composed of stochastic matrices, the set  $\mathcal{A}$  is bounded (by 1) for the norms induced by both the 1 and  $\infty$  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

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

the application that returns the eigenvalues of a matrix, sorted in a nonincreasing (in modulus) order. According to the Theorem II.5.1 of [20],  $\mathcal{L}$  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 the 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 with the eigenvalue 1 and  $\mathcal{W}$  the eigenspace associated with 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  $(x_{\mathcal{M}}, x_{\mathcal{W}}) \in \mathcal{M} \oplus \mathcal{W}$ ,*

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

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

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

*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 (namely Theorem 3.1 in [29]) 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 [10], and has similar conclusions. Nevertheless, one has to bear in mind that with its set of hypotheses, the original result from [10] 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,  $\mathcal{M} := \mathbb{R}e$  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  $f_{\mathcal{W}} \in \mathcal{W}$ ,*

$$\begin{aligned} d(f_{\mathcal{W}}, \mathcal{M}) &= \sqrt{\int_{\mathcal{E}} \sum_{i=1}^l \|f_{\mathcal{W},i}(e)\|^2 de}, \\ d(f_{\mathcal{M}} + f_{\mathcal{W}}, \mathcal{M}) &= d(f_{\mathcal{W}}, \mathcal{M}), \\ d(A(f_{\mathcal{M}} + f_{\mathcal{W}}), \mathcal{M}) &\leq (1 - N\underline{m}_A) d(f_{\mathcal{M}} + f_{\mathcal{W}}, \mathcal{M}). \end{aligned}$$

*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

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

□

**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  $d_{\Lambda^t}$  and  $\underline{m} > 0$  independent from  $t$  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,*

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

**3.4. Learning theory.** The results presented in this part are inspired by [8]. This article deals with the inference of functions **to fit** 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  $\Delta k^t$  with high probability.

**3.4.1. Sample error.** We study the learning process from random samples governed by the probability measure  $\rho$  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 including zero.

**Definition 16.** *We define the least square error of  $f$  as*

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

for  $f : \mathcal{E} \rightarrow \mathcal{C}$ , 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  $f : \mathcal{E} \rightarrow \mathcal{C}$ ,*

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

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

*Proof.* Adding and subtracting  $f_{\rho}$  yields

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

We have

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

For the second term we have

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

□



As a consequence of the proposition 4, the regression function  $f_\rho$  minimizes the mean square error  $\varepsilon$ .

**Definition 17.** Let  $f_{\mathcal{F}}$  be the target function that minimizes  $\varepsilon$ :

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

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

**Definition 18.** We define the empirical error  $\varepsilon_S$  of  $f$  on the sample  $S$  by

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

and  $f_S$  the empirical target function, namely a minimizer of  $\varepsilon_S$ :

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

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 [25]:

**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  $f \in \mathcal{F}$ ,  $\|f(e) - c\|_c \leq M$  almost everywhere;
- (3)  $\rho$  is a probability measure on  $\mathcal{Z}$ .

Then for all  $\eta > 0$ ,

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

where  $\mathcal{N}(\mathcal{F}, s)$  is the so-called covering number, namely the minimal  $l \in \mathbb{N}$  such that there exists  $l$  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 the sample for the learning of the individual  $i$  at time  $t$  to be drawn, 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 :

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

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  $\varepsilon$  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  $\mathcal{M}_{\mathcal{F}} = \{(k, \dots, k), k \in \mathcal{F}\}$ .

**Theorem 2.** *We suppose that  $\tau = 0$ ,  $\Gamma > 0$ ,  $c_{\min} > 0$ , and  $\mathcal{F}$  is compact and convex. There exists 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  $m \geq m_t = m_t(\mathcal{M}_{\mathcal{F}}, k^0, M, \alpha_*, \delta)$ , 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

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

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

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

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

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

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

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

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

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

In particular,

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

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

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

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:

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

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:

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

with

$$\|x\|_B^2 = \sum_{i=1}^l \|((x_1(e))_i, \dots, (x_N(e))_i)\|^2, \quad \forall x \in (\mathbb{R}^l)^N.$$

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,

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

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

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

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

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

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.  $\square$

**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 [10], 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  $\delta$  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

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

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

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

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

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

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

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

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

$\square$

#### 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 new experiences and observations, and original conceptualizations which make possible the evolution of knowledge for both individuals and populations.

**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  $\mathcal{E} = \{1, \dots, 5\}$  and  $\mathcal{C} = [-10, 10]$ . We consider a relationship between two individuals (labeled 1 and 2). The structure matrix  $\Gamma$  is given by

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

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

Parameter  $\alpha$  can be interpreted as cognitive inertia (see Remark 2). The higher the  $\alpha$ , the less change in individuals' knowledge-like functions along the dynamics. As likelihood landscape (10), we take  $L(e, c) = 1$  for all  $e \in \mathcal{E}$  and  $c \in \mathcal{C} \setminus \{0\}$ , 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  $\alpha$  varies in  $(0, 1)$ , all 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 the population converges to a common shared knowledge, which is consistent with our mathematical results.

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

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

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  $\alpha$ . We ran 100 simulations. The average dynamics is presented in Figure 6.

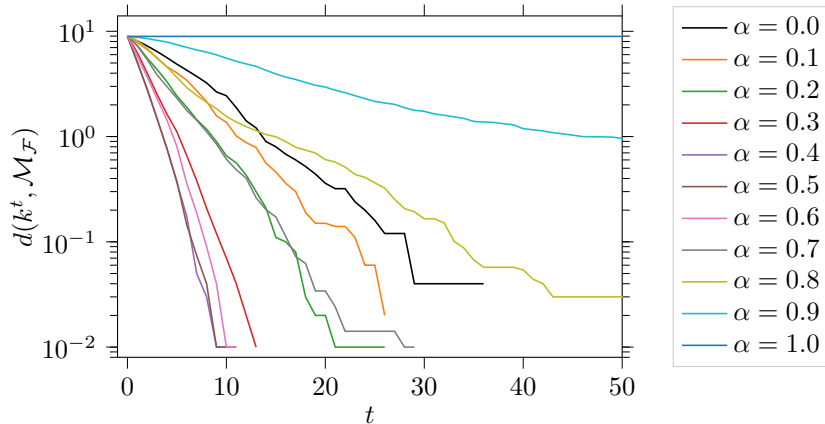


FIGURE 6. **Test 1.** Evolution of the distance between  $k^t$  and the set  $\mathcal{F}$  through time.

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, which is 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 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.

*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 pure 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  $(e, c) \in \mathcal{E} \times \mathcal{C} \setminus \{0\}$

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

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

We call  $k_{eq}$  the common shared knowledge at the equilibrium. We define  $\Delta_i$  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).

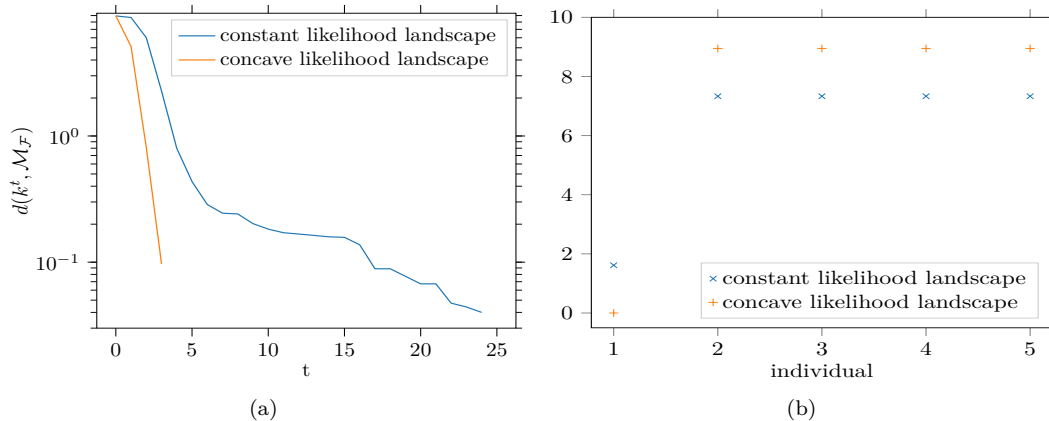


FIGURE 7. **Test 2.**(a) Evolution of the distance between  $k^t$  and the set  $\mathcal{F}$  through time. (b)  $\Delta_i$  for each individual  $i$  in the population. The blue crosses represent the case where the likelihood is constant, and the orange pluses show the case where the likelihood is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

Figure 7(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 towards a common shared knowledge. When the likelihood landscape is concave, the professor has a strong influence on her students and the convergence towards a common shared knowledge is faster.

Figure 7(b) shows the values of  $\Delta$ . In both cases, the common shared knowledge is farther away from the students' initial knowledge than from the professor's.

**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  $\mathcal{E} = \{1, \dots, 25\}$  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  $1_{\mathcal{F}}$  defined as:

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

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

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

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

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

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

When each 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 likelier 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 our simulations, the individual learning rate was fixed at  $\tau = 0.02$ . 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 evolution model.** Cucker, Smale and Zhou developed a model to describe the evolution of language [10]. In their model a language-like function is a function that links a space of objects to a space of signals. Our work is strongly inspired by their model and extends it by introducing individual learning and a credibility matrix that allows influences between individuals to vary through time. Although the interpretation of our variables differs, our results are consistent with those of [10] that proved the convergence of the languages of different individuals to a common shared language (although under different hypotheses, see Remark 5).

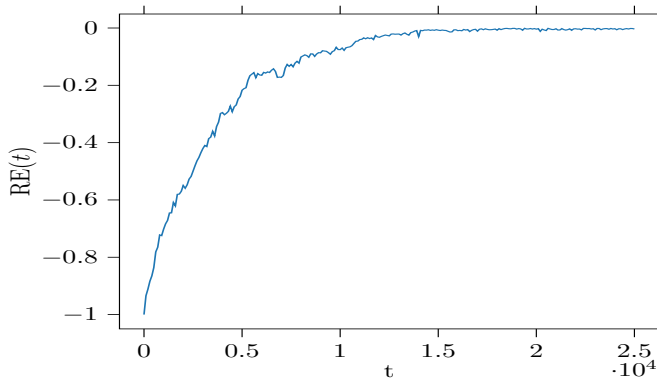


FIGURE 8. **Test 3.** Evolution of the relative entropy through time.

*Test 4. On the evolution of language.* We modified our numerical method in order to simulate the model of language evolution developed in [10]. We consider two different linguistic communities of two individuals with few interactions. We take  $\mathcal{E} = \{1, \dots, 5\}$  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

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

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.

## 5. CONCLUSION

The aim of this work was to develop a more general mathematical model of knowledge evolution than the existing ones e.g. [12, 21, 16]. 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 interactions.

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.



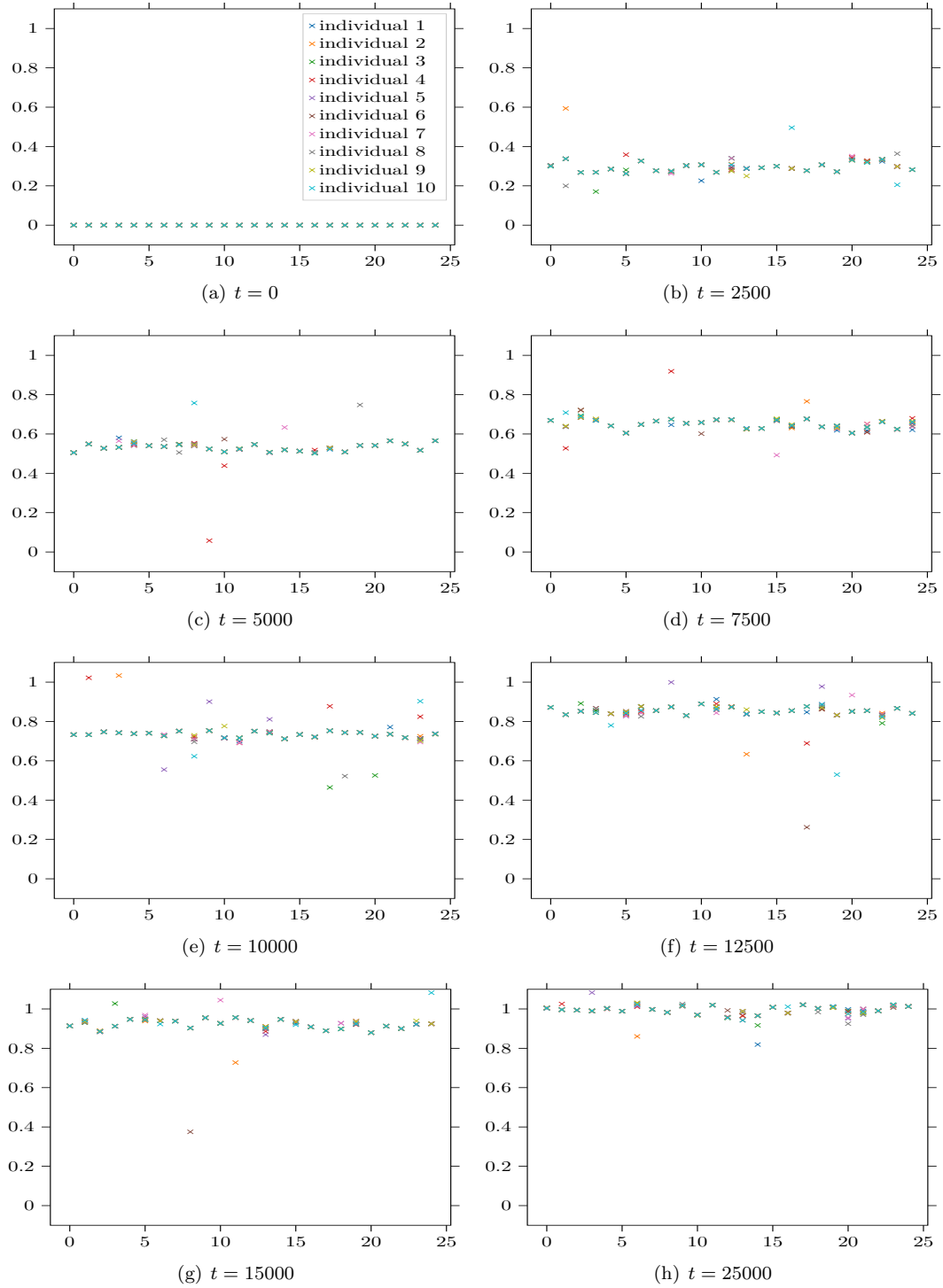


FIGURE 9. **Test 3.** Knowledge-like functions for every individuals in the population at different given times.

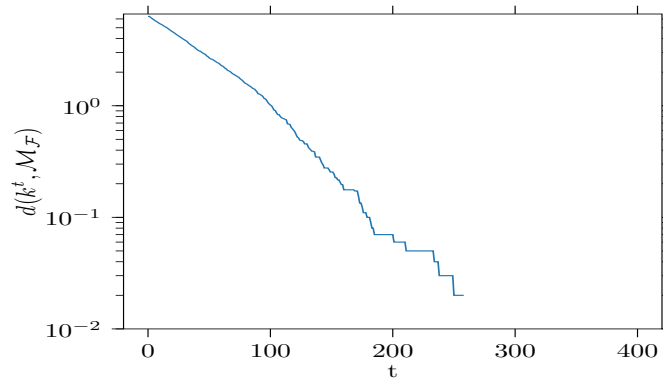


FIGURE 10. **Test 4.** Evolution of the distance between  $k^t$  and the set  $\mathcal{F}$  through time.

#### 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] BELLOMO, N., AND BELLOUQUID, A. On multiscale models of pedestrian crowds from mesoscopic to macroscopic. *Communications in Mathematical Sciences* 13, 7 (2015), 1649–1664.
- [4] 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.
- [5] BURINI, D., DE LILLO, S., AND GIBELLI, L. Collective learning modeling based on the kinetic theory of active particles. *Physics of life reviews* 16 (2016), 123–139.
- [6] BUSH, R. R., AND ESTES, W. K. *Studies in Mathematical Learning Theory*. Stanford University Press, 1959.
- [7] BUSH, R. R., AND MOSTELLER, F. *Stochastic models for learning*. John Wiley & Sons, Inc., 1955.
- [8] CUCKER, F., AND SMALE, S. On the mathematical foundations of learning. *Bulletin of the American Mathematical Society* 39, 1 (2002), 1–49.
- [9] CUCKER, F., AND SMALE, S. Emergent behavior in flocks. *IEEE Transactions on automatic control* 52, 5 (2007), 852–862.
- [10] CUCKER, F., SMALE, S., AND ZHOU, D.-X. Modeling Language Evolution. *Foundations of Computational Mathematics* 4, 3 (2004), 315–343.
- [11] DIMARCO, G., AND TOSCANI, G. Kinetic modeling of alcohol consumption. *Journal of Statistical Physics* 177, 5 (2019), 1022–1042.
- [12] GHIRLANDA, S., AND ENQUIST, M. Cumulative culture and explosive demographic transitions. *Quality & Quantity* 41, 4 (2007), 591–600.
- [13] 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.
- [14] GULLIKSEN, H. A rational equation of the learning curve based on thorndike’s law of effect. *The Journal of General Psychology* 11, 2 (1934), 395–434.
- [15] HA, S.-Y., AND TADMOR, E. From particle to kinetic and hydrodynamic descriptions of flocking. *Kinetic & Related Models* 1, 3 (2008), 415.
- [16] HENRICH, J. Demography and Cultural Evolution: How Adaptive Cultural Processes can Produce Maladaptive Losses: The Tasmanian Case. *American Antiquity* 69, 2 (2004), 197–214.
- [17] IRELAND, T., AND GARNIER, S. Architecture, space and information in constructions built by humans and social insects: a conceptual review. *Philosophical Transactions of the Royal Society B: Biological Sciences* 373, 1753 (2018), 20170244.
- [18] JACOB, E. B., BECKER, I., SHAPIRA, Y., AND LEVINE, H. Bacterial linguistic communication and social intelligence. *TRENDS in Microbiology* 12, 8 (2004), 366–372.
- [19] JIA, Y., AND VICSEK, T. Modelling hierarchical flocking. *New Journal of Physics* 21 (2019), 093048.

- [20] KATO, T. Perturbation theory for linear operators, second edition, vol. 132. Springer Science & Business Media, 1995.
- [21] KOBAYASHI, Y., AND AOKI, K. Innovativeness, population size and cumulative cultural evolution. Theoretical Population Biology 82, 1 (2012), 38 – 47.
- [22] 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.
- [23] MILLER, G. A., AND FRICK, F. C. Statistical behavioristics and sequences of responses. Psychological Review 56, 6 (1949), 311.
- [24] MOTSCH, S., AND TADMOR, E. Heterophilous dynamics enhances consensus. SIAM review 56, 4 (2014), 577–621.
- [25] POLLARD, D. Convergence of stochastic processes. Springer Series in Statistics. Springer-Verlag, New York, 1984.
- [26] REY, T., AND TAN, C. A Rescaling Velocity Method for Collisional Kinetic Equations – Application to Flocking. SIAM J. Numer. Anal. 54, 2 (2016), 641–664.
- [27] REYNOLDS, C. W. Flocks, herds and schools: A distributed behavioral model. In Proceedings of the 14th annual conference on Computer graphics and interactive techniques (1987), pp. 25–34.
- [28] RICHERSON, P. J., AND BOYD, R. Not by genes alone. University of Chicago Press, 2010.
- [29] 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 DEREK

INSTITUTE FOR ADVANCED STUDY IN TOULOUSE, CNRS, UMR 5314  
F-31015 TOULOUSE, FRANCE  
EMAIL: MAXIME.DEREK@IAST.FR

LUDOVIC MAISONNEUVE

NATIONAL MUSEUM OF NATURAL HISTORY, UMR 7205, INSTITUT DE SYSTÉMATIQUE, EVOLUTION ET  
BIODIVERSITÉ  
F-75005 PARIS, FRANCE  
EMAIL: LUDOVIC.MAISONNEUVE.2015@POLYTECHNIQUE.ORG

THOMAS REY

UNIV. LILLE, CNRS, UMR 8524, INRIA LABORATOIRE PAUL PAINLEVÉ  
F-59000 LILLE, FRANCE  
EMAIL: THOMAS.REY@UNIV-LILLE.FR