Language Evolution, Coalescent Processes, and the Consensus Problem on a Social Network

Hariharan Narayanan* and Partha Niyogi

May 11, 2013

Abstract

In recent times, there has been an increased interest in theories of language evolution that have an applicability to the study of dialect formation, linguistic change, creolization, the origin of language, and animal and robot communication systems in general. One particular question that has attracted some interest has the following general form: *how might a group of linguistic agents arrive at a shared communication system purely through local patterns of interaction and without any global agency enforcing uniformity?* In this paper, we consider a natural model of language (or more precisely, word) evolution on a social network, prove several theoretical properties, and establish connections to related phenomena in biology, social sciences, and physics.

keywords: Cognitive Science, Neuroscience, Learning Theory

1 Introduction

In recent times, there has been an increased interest in theories of language evolution that have an applicability to the study of dialect formation, linguistic change, creolization, the origin of language, and animal and robot communication systems in general (see [11, 14, 7] and references therein). One particular question that has attracted some interest has the following general form: *how might a group of linguistic agents arrive at a shared communication system purely through local patterns of interaction and without any global agency enforcing uniformity?* The linguistic agents in question might be humans, animals, or machines in a multi-agent society. For an example of interesting simulations that suggest how a shared vocabulary might emerge in a population , see Liberman (2005) (other simulations are also provided by [18, 5, 1, 2, 19] among others). In this paper, we consider a generalization of Liberman's model, prove several theoretical properties, and establish connections to related phenomena in biology, social sciences, and physics.

*Departments of Statistics and Mathematics, University of Washington, harin@uw.edu

Our model is as follows. For simplicity, we consider how a common word for a particular concept might emerge through local interactions even though the agents had different initial beliefs about the word for this concept. For example agents might use the phonological forms “dog”,“kukur”, “farama” etc. to describe the concept of a canine animal. Thus we imagine a situation where every time an event in the world occurs that requires the agents to use a word to describe this event, they may start out by using different words based on their initial belief about the word for this event or object. By observing the linguistic behavior of their neighbors agents might update their beliefs. The question is - will they eventually arrive at a common word and if so how fast.

1.1 Model

1. Let $\mathbf{W}$ be a set of words (phonological forms, codes, signals, etc.) that may be used to denote a certain concept (meaning or message).

2. Let each agent hold a belief that is a probability measure on $\mathbf{W}$. At time t, we denote the belief of agent i to be $\mathbf{b}_i^{(t)}$.

3. Agents are on a communication network which we model as a weighted directed graph where vertices correspond to agents. We further assume that the weight of each directed edge is positive and that there exists a directed path from any node to any other. An agent (say i) can only observe the linguistic actions of its out-neighbors, i. e.nodes to which a directed edge points from i. We denote weight of the edge from i to j by A_{ij}.

4. The update protocol for the $\mathbf{b}_i^{(t)}$ as a function of time is as follows:

   (a) At each time t, each agent i chooses a word $w = w_i^{(t)} \in \mathbf{W}$ (randomly from to its current belief $\mathbf{b}_i^{(t)}$) and produces it. Let $X_i^{(t)}$, denote the probability measure concentrated at $w_i^{(t)}$. Since $w_i^{(t)}$ is a random word $X_i^{(t)}$ is correspondingly a random measure.

   (b) At every point in time, each agent can observe the words that their neighbors produce but they have no access to the private beliefs of these same neighbors.

   (c) Let P be the matrix whose ij^{th} entry satisfies

   $$P_{ij} = \frac{A_{ij}}{\sum_{k=1}^{n} A_{ik}}.$$

   At every time step, every agent updates its belief by a weighted combination of its current belief and the words it has just heard, i.e.,

   $$\mathbf{b}_i^{(t+1)} = (1-\alpha)\mathbf{b}_i^{(t)} + \alpha \sum_{j=1}^{n} P_{ij} X_j^{(t)},$$

where α is a fixed real number in the interval $(0, 1)$. We assume $A_{ii} = 0$ for each i.

At a time t, let the beliefs of the agents be represented by a vector

$$\mathbf{b}^{(t)} := (\mathbf{b}_1^{(t)}, \ldots, \mathbf{b}_n^{(t)})^T.$$

Similarly, let the point measures on words $X_i^{(t)}$ be organized into a vector

$$X^{(t)} := (X_1^{(t)}, \ldots, X_n^{(t)})^T.$$

Then the reassignment of beliefs can be expressed succinctly in matrix form where the entries in the vectors involved are measures rather than numbers as

$$\mathbf{b}^{(t+1)} = (1 - \alpha)\mathbf{b}^{(t)} + \alpha P X^{(t)}. \tag{1}$$

1.2 Remarks:

1. If beliefs were directly observable and agents updated based on a weighted combination of their beliefs and that of their neighbors,

$$\mathbf{b}^{(t+1)} = (1 - \alpha)\mathbf{b}^{(t)} + \alpha P\mathbf{b}^{(t)}, \tag{2}$$

   the system has a simple linear dynamics, where all beliefs converge to a weighted average of the initial beliefs. Thus eventually, everyone has the same belief (see [3] for pioneering work and [6] for a recent elaboration in an economic context.)

2. Our focus in this paper is on the situation where the beliefs are *not observable* but only the linguistic actions $X_i^{(t)}$ are (and only to the immediate neighbors). Therefore, the corresponding dynamics follows a Markov chain. The state space of this chain (defined by Equation 1) is the set of all n-tuples of belief vectors. Since this is continuous, the standard mixing results with finite state spaces do not apply directly.

1.3 Results:

Our main results are summarized below.

1. With probability 1 (w.p.1), as time tends to infinity, the belief of each agent converges in total variation distance to one supported on a single word, common to all agents.

2. w.p.1, there is a finite time T such that for all times $t > T$, all agents produce the same fixed word.

3. The rate at which beliefs converge depends upon the mixing properties of the Markov chain whose transition matrix is P.

4. The rate of convergence is *independent* of the size of $\mathbf{W}$. One might think that a population where every agent has one of two words for the concept would arrive at a shared word faster than one in which every agent had a different word for the concept. This intuition turns out to be incorrect.

5. The proof of these results exposes a natural connection with coalescent processes and has a parallel in population genetics.

6. Our analysis brings out two different interpretations of the behavior of a linguistic agent. In the most direct interpretation, the agent's linguistic knowledge of the word is internally encoded in terms of a belief vector. This belief vector is updated with experience. In a second interpretation an agent's representation of its linguistic knowledge is in terms of a memory stack in which it literally stores every single word it has heard weighted by how long ago it heard it and the importance of the person it heard it from. Such an interpretation is consistent with exemplar theory An external observer looking at this agent's linguistic actions will not be able to distinguish between these two different internal representations that the agent may have.

2 Convergence to a Shared Belief: Quantitative results

We will define an auxiliary markov Chain to model the exemplar based view of the evolution of the memory stack. We require the original n states S corresponding to agents and an additional n states $\hat{S}$ to model whether a word was uttered at time t, or was embedded in the memory of some agent at that time.

Let $\tilde{P}$ be the transition matrix on the state space $\tilde{S} = S \cup \hat{S}$, where for $i, j \in S := \{1, \ldots, n\}$ and $\hat{S} = \{\hat{1}, \ldots, \hat{n}\}$.

$$\tilde{P}(i \to j) = \tilde{P}(\hat{i} \to j) = \alpha P_{ij},$$

$$\tilde{P}(i \to \hat{i}) = \tilde{P}(\hat{i} \to \hat{i}) = 1 - \alpha.$$

Definition 1. *Let $T_{mix}(\epsilon)$ denote the mixing time of $\tilde{P}$, defined as the smallest t for which, for each specific choice of $v, w \in \tilde{S}$,*

$$\sum_{u \in \tilde{S}} |\tilde{P}^{(t)}(v \to u) - \tilde{P}^{(t)}(w \to u)| < \epsilon.$$

Here $\tilde{P}^{(t)}(b \to c)$ denotes the probability that a Markov Chain governed by $\tilde{P}$ starting in b lands in c at the t^{th} time step.

The following is the main result of this paper.

Theorem 1. *1. The probability that all agents produce the same word at times $T, T+1, \ldots$ tends to 1 as T tends to ∞. More precisely, if*

$$\begin{aligned} \tau &= (4n/\alpha^2)T_{mix}(\frac{\alpha}{4})\ln(4n/\alpha^2) \\ M &= e, \end{aligned}$$

then

$$\mathbb{P}[\forall_{\substack{t \geq T \\ u \in S}} \; X_u^t = X_1^T] > 1 - \frac{MnTe^{-\frac{T}{\tau}}}{1 - e^{-\frac{T}{\tau}}}. \tag{3}$$

2. As time $t \to \infty$ all produced words converge (almost surely) to a word whose probability distribution is

$$\sum_{i=1}^{n} \pi_i \mathbf{b}_i^{(0)},$$

where $(\pi_1, \ldots, \pi_n)$ is the stationary distribution of the Markov chain whose transition matrix is P.

2.1 A Model of Memory

Let $B^{(t)}$ denote the vector of belief measures corresponding to agents 1 to n at time t. The evolution of the $B^{(t)}$ is a Markov chain. It can be seen that its only absorbing states are of the form $(\mathbf{b}_1^{(t)}, \ldots, \mathbf{b}_n^{(t)})^T$, where $\forall i, \mathbf{b}_i^{(t)} = \delta_w$, and δ_w is the point measure concentrated on some word $w \in X$. Formally, δ_w is the measure on $\mathbf{W}$, which assigns to a measurable set A the measure $\delta_w(A)$ according to the following rule.

$$\begin{aligned} \delta_w(A) &= 1 \quad \text{If } w \in A \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

Therefore, if the Markov Chain were finite, a simple argument would suffice. In our case however, we have a Markov Chain whose state space is (possibly) uncountably infinite. Thus in principle, its dynamics could be hard to analyze. Our proof is based on coalescent processes, which have also been extensively used to study biological evolution [8, 10]. In analyzing the evolution of beliefs, we trace the origin of words backwards in time and find that all surviving words, are copies of a single word produced at some point in time sufficiently far in the past. Observe that if the process had begun at time 0, the beliefs at time $t+1$ would be given by the formula below, which is obtained iteratively.

Observation 1.

$$B^{(t+1)} = \sum_{i=0}^{t} \alpha(1-\alpha)^i P X^{(t-i)} + (1-\alpha)^{t+1} B^{(0)}. \tag{4}$$

$X^{(t)} = (X_1^{(t)}, \dots, X_n^{(t)})^T$ is a random vector whose entries are point measures, where $X_i^{(t)} = \delta(w_i^{(t)})$ and $w_i^{(t)}$ is chosen from the measure $\mathbf{b}_i^{(t)}$ on X, independent of the choice of other coordinates of the vector $X^{(t)}$. This observation, motivates a model of memory that we define in the following paragraph.

Let each agent's memory be modeled as a stack. At the top level of the stack of agent i are all the words heard at time t. Below this are all words heard at time $t-1$ and so on tracing backwards in time until the first words heard at an initial time 1. At the lowest level, corresponding to time 0, is the initial belief $\mathbf{b}_i^{(0)}$ which is a probability distribution on the set of words. We may imagine this to be a form of vestigial memory.

Let agent j be adjacent to agent i. We shall describe the process by which agent j produces word $X_j(t)$. Let S_j be the stack held by agent j, and $S_j^{(t)}, \dots, S_j^{(0)}$ be the levels in its stack from top to bottom. After j produces $X_j(t)$, i places $X_j(t)$, and all other $X_{j'}(t)$ produced by neighbors of i at time step t on the top of its stack. In order to describe the mechanism by which $X_j(t)$ is generated, let us introduce a binomial random variable Y where

$$\mathbb{P}[Y = i] = \alpha(1-\alpha)^i.$$

If $Y \leq t-1$, $X_j(t)$ is chosen to be the word produced by j' at time $t-1-Y$ (which is stored in S_{t-1-Y}) with probability $P_{jj'}$. If $Y \geq t$, $X_j(t)$ is chosen from the distribution in $\mathbf{b}_j^{(0)}$. This process has been illustrated in Figure 2.1. Note that in this model words are formal objects. While any two words present in the stack positions $S_j^{(t)}$ for $t = 1, 2, \dots$ are considered distinct, there is a natural "parent-child" structure existing on the set of words. Under this scheme, let the probability distribution of $X_i^{(t)}$ be denoted $\tilde{\mathbf{b}}_i^{(t)}$. Denoting by $\tilde{B}^{(t)}$ the vector $(\tilde{\mathbf{b}}_1^{(t)}, \tilde{\mathbf{b}}_2^{(t)}, \dots, \tilde{\mathbf{b}}_n^{(t)})$.

Observation 2. *A direct computation shows that in the model just described*

$$\tilde{B}^{(t+1)} = \sum_{i=0}^{t} \alpha(1-\alpha)^i P X^{(t-i)} + (1-\alpha)^{t+1} \tilde{B}^{(0)}. \tag{5}$$

This along with the fact that the randomness used in the generation of $X_j^{(t)}$ is independent of the randomness in the generation of all other words, tells us that the model of memory just described results in a system with the same dynamics as that introduced earlier. This particular model of memory may be viewed as an implementation of the ideas implicit in exemplar based accounts of linguistic behavior.

3 Proofs

By observations 1 and 2, in order to obtain an upper bound on $\mathbb{P}[X_i^{(t_1)} \neq X_j^{(t_2)}]$, it is sufficient to trace the ancestry of both words backwards in time and show that the probability that they do not have a common ancestor is small. Our

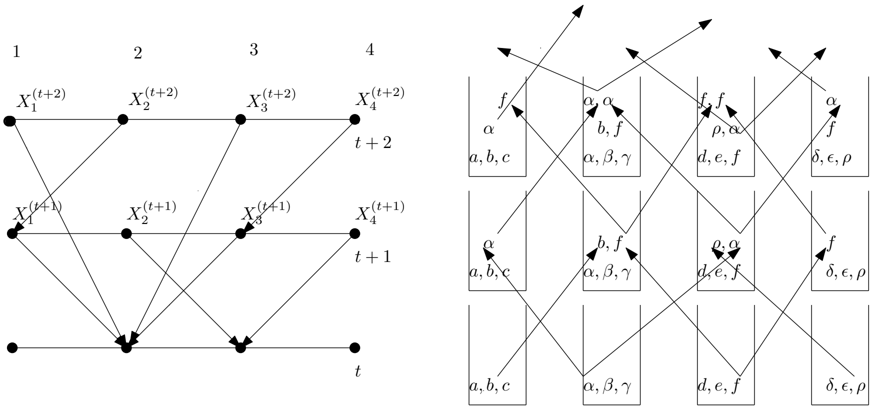

Figure 1: A coalescent process obtained by tracing the origin of words backwards in time, and the associated memory stacks of agents 1 to 4 for time steps t to $t+2$. Each agent produces α at time $t+2$ due to coalescence to a single word α produced by agent 2 at time t.

results are best stated in terms of the coalescence time of a set of random walks. In Figure 2, we illustrate how the path tracing the origin of a word backwards in time can be encoded as a Markov chain on a state space $S \cup \hat{S} = \{1, \ldots, n, \hat{1}, \ldots, \hat{n}\}$. We use the states $\hat{1}, \ldots, \hat{n}$ as additional "memory" states.

Observation 3. *Since the random variable Y introduced in section 2.1 can be interpreted as the length of a run of heads in a biased coin (whose probability of coming heads is $1-\alpha$), we can account Y using additional memory states.*

We define a variant of the meeting time between two Markov Chains as follows. Let $u, v \in S \cup \hat{S}$.

Definition 2. *For $t \geq 0$, let Y_t and Z_t be two independent random walks on $S \cup \hat{S}$ each of which has $\tilde{P}$ as its transition matrix and have initial states $Y_0 = u, Z_0 = v$. For $\Delta > 0$, let $M_{uv}(\Delta)$ be the smallest time $t > 0$ for which $Y_{t+\Delta} = Z_t \in S$.*

Theorem 1. *1. The probability that all agents produce the same word at times $T, T+1, \ldots$ tends to 1 as T tends to ∞. More precisely, if*

$$\begin{aligned} \tau &= (4n/\alpha^2) T_{mix}(\frac{\alpha}{4}) \ln(4n/\alpha^2) \\ M &= e, \end{aligned}$$

then

$$\mathbb{P}[\forall_{\substack{t \geq T \\ u \in S}} \; X_u^t = X_1^T] > 1 - \frac{MnTe^{-\frac{T}{\tau}}}{1 - e^{-\frac{T}{\tau}}}. \tag{6}$$

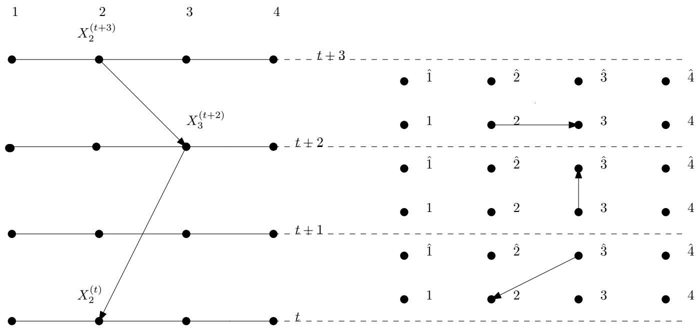

Figure 2: The ancestry of $X_2^{(t+3)}$ has been traced backwards in time to $X_2^{(t)}$. On the right,is an encoding of this path in terms of the transitions in a Markov Chain with "auxiliary states" $\hat{1}, \ldots, \hat{n}$. $\hat{3}$ is occupied at time step $t+1$ because the agent 3 produced a word at a time $t+2$ from past memory.

2. *As time $t \to \infty$, all produced words converge (almost surely) to a random word chosen from the probability distribution*

$$\sum_{i=1}^{n} \pi_i \mathbf{b}_i^{(0)},$$

where $(\pi_1, \ldots, \pi_n)$ is the stationary distribution of the Markov chain whose transition matrix is P.

Proof. To prove the first part, we observe that

$$\mathbb{P}\left[\neg\left(\forall_{\substack{t\geq T\\ u\in S}}\ X_u^t = X_1^T\right)\right]$$

$$\leq \sum_{j=1}^{\infty}\left(\mathbb{P}[X_1^{jT} \neq X_1^{(j+1)T}] + \sum_{k=0}^{T-1}\sum_{u=1}^{n}\mathbb{P}[X_u^{jT+k} \neq X_1^{jT}]\right)$$

by the union bound. The following application of Lemmas 1 and 2 completes the proof.

$$\mathbb{P}\left[\neg\left(\forall_{\substack{t\geq T\\ u\in S}} X_u^t = X_1^T\right)\right]$$

$$\leq \sum_{j=1}^{\infty}\left(\mathbb{P}[X_1^{jT} \neq X_1^{(j+1)T}] + \sum_{k=0}^{T-1}\sum_{u=1}^{n}\mathbb{P}[X_u^{jT+k} \neq X_1^{jT}]\right)$$

$$\leq \sum_{j=1}^{\infty} \left(\mathbb{P}[M_{11}(T) \geq jT] + \sum_{k=0}^{T-1} \sum_{u=1}^{n} \mathbb{P}[M_{u1}(k) \geq jT] \right)$$

$$\leq \frac{MnTe^{-\frac{T}{\tau}}}{1 - e^{-\frac{T}{\tau}}},$$

where M and τ are the constants that appear in Lemma 2.

To prove the second part, we use the linearity of expectation to show that the expected value of the beliefs follows a simple rule. Namely

$$\begin{aligned}
\mathbb{E}\mathbf{b}^{(t+1)} &= (1-\alpha)\mathbb{E}\mathbf{b}^{(t)} + \alpha P \mathbb{E} X^{(t)} \\
&= ((1-\alpha)I + \alpha P)\mathbb{E}\mathbf{b}^{(t)} \\
&= \dots \\
&= ((1-\alpha)I + \alpha P)^{t+1}\mathbb{E}\mathbf{b}^{(0)}.
\end{aligned}$$

By well known results on Markov chains,

$$\lim_{t \to \infty} ((1-\alpha)I + \alpha P)^t = (1, \dots, 1)^T (\pi_1, \dots, \pi_n),$$

where π_i is the stationary probability of the state i under the chain P. Therefore, for each j,

$$\lim_{t \to \infty} \mathbb{E}\mathbf{b}_j^{(t)} = \sum_{i=1}^{n} \pi_i \mathbf{b}_i^{(0)},$$

By the first part of this theorem, as $t \to \infty$, $\mathbf{b}^{(t)}$ converges almost surely to a measure that is concentrated on a single common word w. Given a signed measure μ, let

$$|\mu| = \sup_{\|f\|_\infty \leq 1} \int f d\mu.$$

Then,

$$\begin{aligned}
\left|\mathbb{E}[\delta_w] - \mathbb{E}[X_i^T]\right| &\leq \mathbb{P}\left[\neg \left(\forall_{\substack{t \geq T \\ u \in S}} X_u^t = X_1^T \right)\right] \\
&\leq \frac{MnTe^{-\frac{T}{\tau}}}{1 - e^{-\frac{T}{\tau}}},
\end{aligned}$$

It follows that this common word w must have the distribution $\sum_{i=1}^{n} \pi_i \mathbf{b}_i^{(0)}$.

□

Lemma 1. *The probability that the word produced by agent u at time step t_1 is different from that produced by agent v at time step t_2 greater than t_1 can be bounded from above as follows.*

$$\mathbb{P}[X_u^{(t_1)} \neq X_v^{(t_2)}] \leq \mathbb{P}[M_{uv}(t_2 - t_1) \geq t_1].$$

Proof. In the model of memory introduced in section 2.1 we described a parent-child relationship between words, where a child word is identical to a parent word. The evolution of the Markov chain defined in this section corresponds to the genealogy of a word. The event that the words $X_u^{(t_1)}$ and $X_v^{(t_2)}$ have a common ancestor produced at some time ≥ 0 is the event that $M_{uv}(t_2 - t_1) \leq t_1$. The lemma follows from the fact that two words that have a common ancestor are the same. $\square$

Lemma 2. *The random variable $M_{uv}(\Delta)$ has an exponential tail bound uniform over u, v and Δ. More precisely, there exist constants $M, \tau > 0$ independent of u, v and Δ such that*

$$\mathbb{P}[M_{uv}(\Delta) \geq T] < Me^{-\frac{T}{\tau}}.$$

(In fact, this is satisfied for $\tau = \frac{4n}{\alpha^2} T_{mix}(\frac{\alpha}{4}) \ln\left(\frac{4n}{\alpha^2}\right)$ and $M = e$.)

Proof. The stationary measure $\tilde{\mu}$ satisfies for each i, the identity $\alpha\tilde{\mu}(\hat{i}) = (1 - \alpha)\tilde{\mu}(i)$.

Let $\tau_1 = T_{mix}(\frac{\alpha}{4}) \ln(\frac{4n}{\alpha^2})$. Let us denote by $q_u(i)$ the probability $\mathbb{P}[Z_\tau = i \big| Z_0 = u]$. Then,

$$\begin{aligned}
&\sup_{u,v} \mathbb{P}[\neg(Y_{\tau+\Delta} = Z_\tau \in S) \big| Y_\Delta = u, Z_0 = v] \\
&= 1 - \inf_{u,v} \sum_{i \in S} q_u(i) q_v(i) \\
&\leq 1 - \inf_{u,v} \sum_{i \in S} \min(q_u(i), q_v(i))^2 \\
&\leq 1 - \inf_{u,v} \frac{(\sum_{i \in S} \min(q_u(i), q_v(i)))^2}{n} \\
&\leq 1 - \frac{\alpha^2}{4n}.
\end{aligned}$$

Now, using the Markov property and conditioning repeatedly, we see that

$$\begin{aligned}
\mathbb{P}[M_{uv}(\Delta) \geq T] \;&\leq\; \mathbb{P}[\neg(Y_\Delta = Z_0 \in S)] \times \\
&\prod_{i=1}^{\lfloor \frac{T}{\tau_1} \rfloor} \sup_{u,v} \mathbb{P}[\neg(Y_{\Delta + i\tau_1} = Z_{i\tau_1} \in S) \big| \\
&(Y_{\Delta + (i-1)\tau_1}, Z_{(i-1)\tau_1}) = (u, v)] \\
&\leq \mathbb{P}[\neg(Y_\Delta = Z_0 \in S)] \prod_{i=1}^{\lfloor \frac{T}{\tau_1} \rfloor} (1 - \frac{\alpha^2}{4n}) \\
&\leq \left(1 - \frac{\alpha^2}{4n}\right)^{\frac{T}{\tau_1} - 1} \leq e^{1 - \frac{T}{\tau}}.
\end{aligned}$$

where

$$\tau = \frac{4n}{\alpha^2} T_{mix}(\frac{\alpha}{4}) \ln \left(\frac{4n}{\alpha^2} \right),$$

which proves the Lemma. □

3.1 Concluding Remarks

The general theme of predicting the macroscopic behavior of a system from the local behavior of its microscopic components arises in many different areas of physics, biology, and the social sciences. It is also a fundamental issue in the analysis of distributed systems in computer science.

In Spin systems, which originated as models for Ferromagnets, atoms are pictured to be in a 2-Dimensional square array, each possessing a spin "up" or "down." The effect that an atom has on the spin of a neighbor is a function of temperature. Typically, coherence is observed at low temperatures, while at high temperatures atoms tend not to align, which is in agreement with the demagnetization that ferromagnets undergo at high temperatures. The model we consider, involving the convergence in beliefs has many high level similarities though we do not address the question of what might be the analog of temperature in our model, how to take the thermodynamic limit, and if and how phase transitions may arise.

Another closely related model is the voter model studied in probability theory with its origins in the social sciences. Each agent lives on the vertex of the graph, has a belief which is a discrete variable, and is observable to its neighbors. Each agent changes its belief with a certain probability based on the observed beliefs of its neighbors. Another kind of belief propagation model is that described by Jackson (2007). In both cases, the beliefs are observable in contrast to our setting. Our communication graphs model the pattern of local interaction among agents and may arise through modes of social network formation studied in the field of social network theory [12].

Linear update rules are often used in distributed systems, to achieve coherence among different agents or to share knowledge gathered individually. In a model that has been intensively studied, a number of sensors form a network, each of which measures a quantity such as temperature [3]. Neighbors communicate during each time step and make linear updates in a synchronous or asynchronous manner. The rate at which consensus is attained is studied. There is also a related body of work on Coordination and Distributed Control. A model of flocking has been considered in [4], where a group of birds, have a certain initial velocity, and the evolution of their velocities is governed by a differential equation wherein each bird modifies its velocity to bring it closer to that of its neighbors. The update rule involves a graph Laplacian. Some results are derived concerning the initial conditions that result in flocking behavior.

There are two connections to evolutionary theory that are worth mentioning. First, our proof of convergence exposes a natural coalescent process over words. Coalescent processes are, of course, widely used in modeling and making inferences about genetic evolution [8, 10]. Second, researchers have considered game-theoretic models of evolution [9] and more recent research in this tradition has addressed evolutionary games on graphs [16, 13, 17]. The question of how agents may learn an appropriate strategy for a coordination game on a graph has many high level similarities to the problem studied in this paper.

Finally, there have been a large number of models on achieving coherence in a linguistic population. Many of these rely on simulations. Among mathematical studies, two strands are worth noting. The model of language evolution proposed in has many similarities with languages of agents evolving on a graph. But it is worth noting that in that model, if at each time step, the number of linguistic examples (observations) collected by each agent is bounded from above by a constant (independent of time), the community fails to achieve a consensus language. A second strand is the collection of results obtained in [15, 11]. While there are many synergies with that body of work, there is nothing that is directly comparable.

4 Acknowledgements

We are grateful to Mark Liberman for a stimulating discussion.

References

[1] A. Baronchelli, M. Felici, E. Caglioti, V. Loreto, L. Steels, Sharp transition towards shared vocabularies in multi-agent systems, J. Stat. Mech. (P06014).

[2] A. Baronchelli, V. Loreto, L. Steels, In-depth analysis of the naming game dynamics: The homogeneous mixing case, International Journal of Modern Physics C 19 (5) (2008) 785–812.

[3] D. P. Bertsekas, J. N. Tsitsiklis, Parallel and Distributed Computation: Numerical Methods, Athena Scientific, 1997.

[4] F. Cucker, S. Smale, Emergent Behavior in Flocks, IEEE Transactions on Automatic Control 52 (5) (2007) 852–862.

[5] B. de Boer, Evolution of speech and its acquisition, Adaptive Behavior 13 (4) (2005) 281–292.

[6] M. O. Jackson, Social and Economic Networks, Princeton University Press, 2008.

[7] M. D. Hauser, The Evolution of Communication, MIT Press/BradfordBooks, 1997.

[8] R. R. Hudson, Gene genealogies and the coalescent process, Oxford Surveys in Evolutionary Biology (1991) 1–44.

[9] J. M. Smith, Evolution and the Theory of Games, Cambridge University Press, Cambridge, UK, 1982.

[10] J. F. C. Kingman, On the Genealogy of Large Populations, Journal of Applied Probability 19 (1982) 27–43.

[11] S. Kirby, Function, Selection, and Innateness: The Emergence of Language Universals, Oxford Univ Press, 1999.

[12] J. M. Kleinberg, Navigation in a small world, Nature 406 (6798) (2000) 845.

[13] E. Lieberman, C. Hauert, M. A. Nowak, Evolutionary dynamics on graphs, Nature 433 (7023).

[14] P. Niyogi, The Computational Nature of Language Learning and Evolution, MIT Press, Cambridge, MA, 2006.

[15] M. A. Nowak, N. L. Komarova, Towards an evolutionary theory of language, Trends in Cognitive Sciences 5 (7) (2001) 288–295.

[16] M. Nowak, K. Sigmund, Evolutionary dynamics of biological games, Science 303 (5659) (2004) 793 – 799.

[17] H. Ohtsuki, C. Hauert, E. Lieberman, M. A. Nowak, A simple rule for the evolution of cooperation on graphs and social networks, Nature 441 (7092) (2006) 502–505.

[18] L. Steels, The puzzle of language evolution, Kognitionswissenschaft 8 (1999) 2000.

[19] L. Steels, A. Mcintyre, Spatially Distributed Naming Games, Advances in complex systems 1 (4) (1999) 301–323.