RESEARCH ARTICLE

Emergence of linguistic conventions in multi-agent reinforcement learning

Dorota Lipowska[1], Adam Lipowski [ID][2] *

1 Faculty of Modern Languages and Literature, Adam Mickiewicz University, Poznań, Poland, **2** Faculty of Physics, Adam Mickiewicz University, Poznań, Poland

* lipowski@amu.edu.pl

Abstract

Recently, emergence of signaling conventions, among which language is a prime example, draws a considerable interdisciplinary interest ranging from game theory, to robotics to evolutionary linguistics. Such a wide spectrum of research is based on much different assumptions and methodologies, but complexity of the problem precludes formulation of a unifying and commonly accepted explanation. We examine formation of signaling conventions in a framework of a multi-agent reinforcement learning model. When the network of interactions between agents is a complete graph or a sufficiently dense random graph, a global consensus is typically reached with the emerging language being a nearly unique object-word mapping or containing some synonyms and homonyms. On finite-dimensional lattices, the model gets trapped in disordered configurations with a local consensus only. Such a trapping can be avoided by introducing a population renewal, which in the presence of superlinear reinforcement restores an ordinary surface-tension driven coarsening and considerably enhances formation of efficient signaling.

Introduction

Functioning of societies is to a large extent regulated by various norms and conventions shared by its members [1]. In some cases, these rules are centrally imposed or coordinated, e.g., a dress code in a company or the side of the road that one should drive on. But some conventions, such as the color of cloth that we wear in grief or greeting our friends with a handshake, appeared more spontaneously.

Perhaps the most important convention of this kind, which emerged in the absence of any explicit, centralized coordination, is language. Human language provides a highly efficient communication system acquired by individuals in cultural interactions. Some researchers try to explain how such a system could have appeared and evolved using evolutionary game theory [2, 3], evolutionary linguistics [4] or cognitive science [5]. A promising approach considers language as a signaling system, which emerged via a reinforcement learning process. Such a framework originates from Lewis signaling game [1]. In the simplest version, there are two players and a fixed number of signals. The speaker sends a signal (which is to correspond to the state of the world) and the hearer interprets the signal (i.e., takes some action). If he does it

correctly, both players receive some payoff, which might influence their further actions. The model can actually be considered as a certain urn model [6, 7]. Some mathematical subtleties concerning, e.g., the convergence of the above scheme, were analyzed by Skyrms [8] and Beggs [9], while an adaptation focusing on language evolution was proposed by Lenaerts *et al.* [10]. Attempts to compare several related approaches where learned signaling might emerge were also made [11].

In all these studies, the examined number of agents was rather small ($\lesssim$ 30), however, one should note that the reinforcement learning leads to nontrivial results already in two-agent models [1, 8, 12]. Nonetheless, having in mind the evolution of language, much larger populations of agents should certainly be examined. In such a case, for a population of agents, one has to specify the network of their interactions. While for a small group, a complete graph, where each agent interacts with each of the others, seems the most natural topology, for larger groups of agents, some other structures such as planar or heterogeneous graphs can also be relevant (e.g., when studying the emergence of a linguistic coherence in large-scale communities such as a city population or a nation). The emergence of language is often modeled as a process of reaching an agreement (consensus) about linguistic forms used in a population. Opinion formation or ferromagnetism are also manifestations of such an agreement dynamics. For such processes, the structure of the network usually plays an important role, determining whether the consensus will be reached at all, and affecting the way it could be reached [13–15]. Networks examined in the present paper (Cartesian lattices, complete graphs, random graphs) are only mathematically and computationally appealing idealizations of real networks. Certainly, placing our models on more realistic networks, which take into account a node-distribution heterogeneity, directionality, small-worlds, modular structure or assortativity [16], would be desirable.

A model that is often examined in the context of language emergence is the Naming Game [17]. Due to its computational simplicity, the Naming Game allows for analytical as well as numerical approaches, and global aspects of its dynamics are now relatively well understood [18]. In particular, it is found that typically in the Naming Game, a consensus emerges and reaching such a state resembles the coarsening in the Ising model. The similarity is not accidental because due to the presence of a surface tension [19], both models operate with the so-called curvature-driven dynamics [20]. Let us notice that the coarsening dynamics of the Naming Game, which gradually eliminates certain languages and eventually leads to a global consensus, can be found very appealing in some linguistic contexts. There are even some indications that the curvature-driven dynamics may underlie such linguistic processes as, e.g., an evolution of dialects [21]. The simplicity of the Naming Game implies, however, simplicity of an emerging language, and in many of its versions agents negotiate the name of just a single object. On the other hand, for models that have a potential to generate more complex languages, global aspects of their dynamics are rather poorly understood. Such models could incorporate agents, which, using the reinforcement learning, would try to establish a language reflecting their multi-object and multi-agent world. An objective of the present paper is to specify whether and how an efficient communication might emerge in such a system.

Methods

Reinforcement learning via urn model

The basic building block of our model is a Pólya urn model. In the simplest version of this model, a ball is drawn randomly from an urn with black and white balls [6, 7]. Then the ball is put back into the urn along with an extra ball of the same color (reinforcement), and the process is repeated *ad infinitum*. In this scheme, the probability to select a ball of a given color is

proportional to the number of such balls in the urn. We can also consider a generalized version of this model with the selection probability proportional to the number of balls raised to a certain power α [22]. In this case, the behavior of the model strongly depends on α. For $\alpha < 1$, the model converges toward an equal number of balls of each color, but for $\alpha > 1$, a monopolistic solution appears with the urn dominated by one color. The monopolistic solution is in fact a simple manifestation of a spontaneous symmetry breaking, the phenomenon of much interest in statistical mechanics or particle physics. The basic Pólya urn model is equivalent to the $\alpha = 1$ case, thus determining the transition between these two different regimes.

Our intention is to study a multi-agent model of a signaling game with communicating agents as interacting urns. In the simplest (single-object) version, agents engage in pairwise interactions to negotiate the word to be associated with an object. After a weighted selection of a word, the speaker and the hearer increase its weights (reinforcement learning), which affects subsequent selections. It seems plausible that in the $\alpha > 1$ regime, a monopolistic solution would emerge with agents almost always selecting the same word. There is, however, a number of questions, which one can ask concerning such a linguistic consensus. For example, is it a global consensus, where the entire population of agents uses the same word, or rather a local one corresponding to a certain multi-word solution. Most likely the answer will depend on the topology of interactions between agents, e.g., networks of long-range connectivity should favour the global consensus. Furthermore, agents may be involved in more complicated interactions, e.g., negotiating simultaneously the names for several objects (multi-object version). In that case they need some recognition mechanism, and the resulting language is likely to be more complex.

It is difficult to advocate that in the linguistic contexts, $\alpha > 1$ should be used. In economy, the emergence of a monopoly is sometimes associated with a certain positive superlinear feedback known as Metcalfe's Law [23]. For example, in social networks, the greater the number of users with a certain service, the more valuable the service becomes to the community, and hence its total value is likely to increase quadratically ($\alpha = 2$) with the number of its users. One might expect that a similar superlinear feedback appears during language formation processes. Most of the results presented in our paper are for $\alpha = 2$; some of our results demonstrate that the behaviour of the model is qualitatively similar as long as $\alpha > 1$. For $\alpha = 1$ the convergence toward consensus is typically much slower and in some cases the model does not evolve toward consensus at all.

Single-object version

In the simplest version of our model, we have a population of N agents, which try to establish a name for a given object. Each agent A has an inventory of the same N_w words W_i with their corresponding weights $w_i(A)$ ($i = 1, 2, \ldots, N_w$; initially all $w_i(A) = 1$). In an elementary step, a randomly selected agent (the speaker) interacts with one of its randomly selected neighbors (the hearer) communicating a word. The probability that the speaker A will select the i-th word depends on its weight and is given as

$$s_i(A) = w_i^\alpha(A) \Big/ \sum_{k=1}^{N_w} w_k^\alpha(A). \tag{1}$$

After the interaction, both the speaker and the hearer increase their weights of the communicated word by 1. Such an elementary step of our model is illustrated in Fig 1. In our simulations, a unit of time (t = 1) comprises N elementary steps (i.e., in a unit of time, each agent is on average selected once as a speaker).

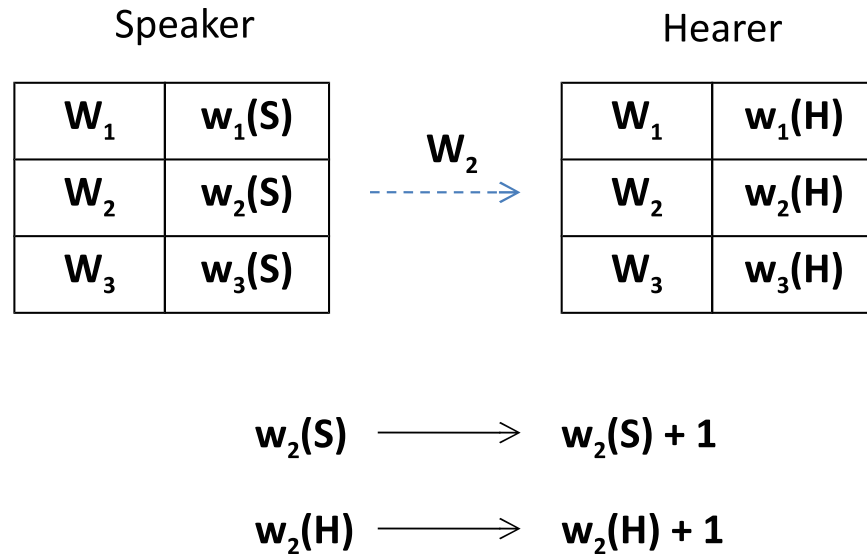| Speaker			Hearer	

Fig 1. An elementary step of a single-object version of the model ($N_w = 3$). Using the probabilities defined in Eq (1), the speaker selects one of its words (here: W_2). Next both the speaker and the hearer increase their weights of the selected word by 1.

Multi-object version

We also examine a more general version of our model, in which agents try to establish names for a set of N_o objects. Their inventories are more complex now as they contain the same set of N_w words W_i (coupled with their respective weights) for each object. In other words, each inventory consists now of N_o copies of inventories from the single-object version and thus each agent A has $N_w N_o$ weights $w_{i,j}(A)$, where $i = 1, \ldots, N_w$ and $j = 1, \ldots, N_o$. First, a randomly selected speaker chooses an object with a uniform probability $1/N_o$. Then the speaker selects the word to be communicated taking into account the weights associated with the words for the chosen object. By analogy with Eq (1), the probability that agent A will select the i-th word for the j-th object equals

$$s_{i,j}(A) = w_{i,j}^{\alpha}(A) / \sum_{k=1}^{N_w} w_{k,j}^{\alpha}(A).$$

(2)

Next, the role of the hearer (H) is to assign an object to the communicated word. This word, say W_i, appears in the hearer's inventory N_o times with weights $w_{i,j}(H)$, where j denotes the object. The hearer uses these weights to guess which object the speaker is talking about. Hence, the hearer recognizes the j-th object as that communicated by the i-th word with probability

$$r_{i,j}(H) = w_{i,j}^{\alpha}(H) / \sum_{k=1}^{N_o} w_{i,k}^{\alpha}(H).$$

(3)

Provided that the object recognized by the hearer is the same as that chosen by the speaker, both agents increase the corresponding weights by 1. An elementary step of this version of the model is illustrated in Fig 2.

The above specified rules are consequences of a number of simplifying assumptions and certainly more realistic versions might be considered. For example, one might assume that the words in agents' inventories are not necessarily identical and agents could learn new words
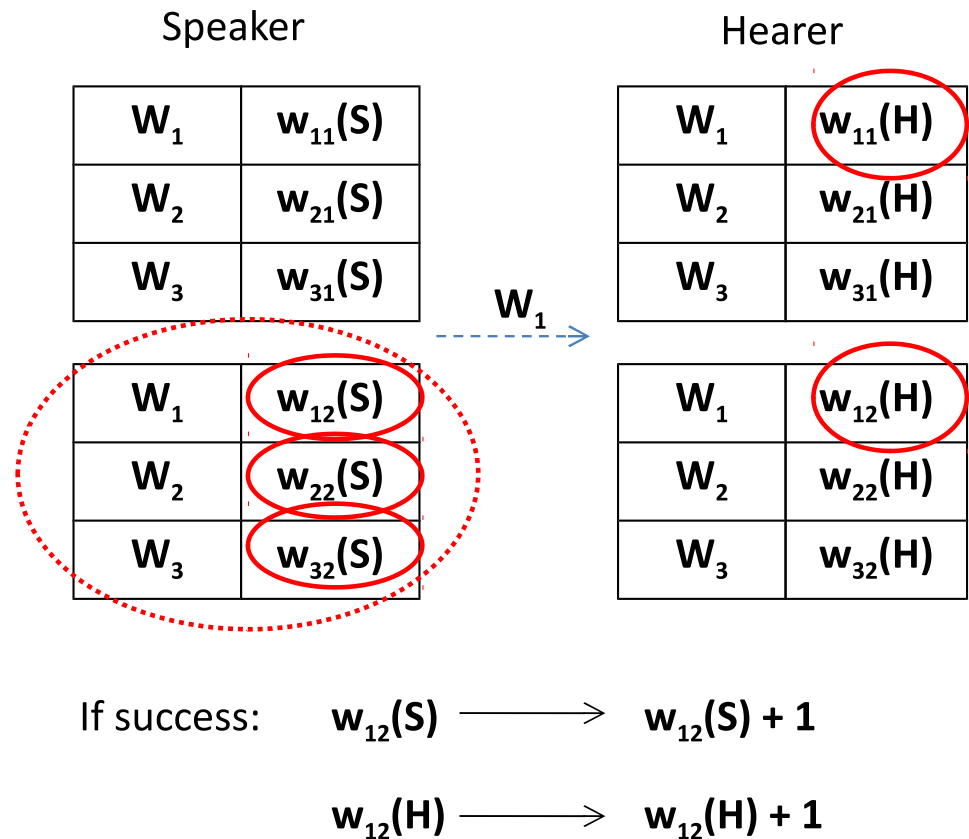
Fig 2. An elementary step of a multi-object version of the model ($N_w = 3$, $N_o = 2$). With a uniform probability $1/N_o$, the speaker chooses an object (the corresponding section of the inventory is encircled by the dotted line). Using the relevant weights (in solid circles), the speaker calculates the probabilities defined in Eq (2) and selects one of its words (here: W_1). Next the hearer tries to guess the object the speaker is talking about by calculating the probabilites (3) based on its weights of the communicated word (in circles). When the hearer's guess is correct, both agents increase their corresponding weights by 1.

from each other. Most likely such a change would require a more sophisticated recognition mechanism and perhaps a notion of a distance between words would have to be used. Further analysis of such a version, although it seems more realistic and potentially interesting, is left for the future.

Population renewal

We also introduce a simple modification of our model (both in its single and multi- object versions), which takes into account a population renewal. The modification seems to be plausible, especially for modeling a formation of a communication system in a population of humans. In such population, when considered at a timescale of, say, hundreds of years, we should take into account a generational turnover (and possibly migrations [24]). A child learns the language of its parents but it might also acquire a (possibly different) language of its neighbors. Certainly, for a young person this is more likely to happen than for an adult. Let us notice that in urn models, due to the accumulation of weights after a large number of iterations, it is almost impossible to shift their balance (i.e., change the language). To allow for such a shift, we introduce a population renewal: With (usually small) probability p, the agent selected to be a

speaker is replaced with a new agent (with all weights equal to 1), while with probability $1 − p$, the speaker acts as previously defined.

Results

Single-object version

We analysed the behavior of our model for several interaction networks, namely Cartesian lattices, complete graphs and random graphs. The results obtained indicate that on Cartesian lattices the model gets stuck in a disordered structure, where consensus is only local (Fig 3). Most of the agents reach a monopolistic regime and communicate with their neighbors with the same words. There is only a small fraction of interfacial agents, which persist in a more symmetric state. For $N_w = 2$, it is tempting to confront our results with some other statistical-mechanics models. In particular, the snapshot configurations in Fig 3 suggest that initally our model coarsens, similarly to the Naming Game and low-temperature Ising models. However, contrary to these models, the evolution of our model gets trapped in a disordered state much before reaching the uniform (mono-word) state.

To examine in more detail the behavior of the model, we calculated for $N_w = 2$ the quantities m_G and m_L defined as

$$m_G = |\frac{1}{N}\sum_A (s_1(A) − s_2(A))|,$$

$$m_L = \frac{1}{N}\sum_A |s_1(A) − s_2(A)|$$

(4)

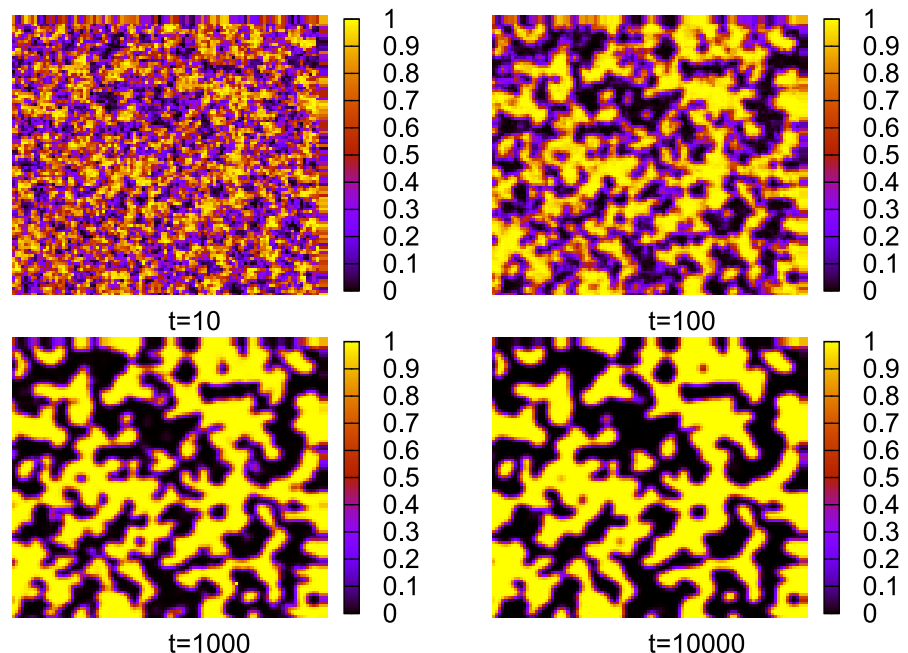where summation is over all agents A in our model.

Fig 3. Spatial distribution of s_1, the probability that an agent will select word W_1 (Eq (1)). Results for a single-object model on a square lattice with $N = 10^2 \cdot 10^2 = 10^4$, $\alpha = 2$, $N_w = 2$. The dynamics traps the model in a disordered state (the configurations for, e.g., $t = 10^3$ and $t = 10^4$ differ only slightly). Since s_1 is generally close to unity or to zero, it means that almost every agent developed a strong preference toward one of the words.

The quantities m_G and m_L allow us to examine the global (m_G) and local (m_L) symmetries of the model. We do not present the results for $\alpha < 1$, in which case the model remains symmetric (for each agent, $s_1 = s_2 = 1/2$), and consequently, $m_G = m_L = 0$. For $\alpha > 1$, however, the asymmetry in an agent's inventory implies that $s_1 \neq s_2$ and thus $m_L > 0$. When the system is disordered, as in Fig 3, then there is no global preference toward any of the two words and $m_G = 0$.

Numerical results for $\alpha = 2$ support such an analysis. In two dimensions, relatively large m_L (Fig 4) indicates that most of the agents operate in a monopolistic regime. Moreover, m_G remains close to zero, which confirms the disordered nature of the regime (Fig 5). Results for the one- and three-dimensional Cartesian lattices show a similar behavior. Much different behavior is seen, however, for a complete graph, where each agent interacts with every other one. In this case, m_G quickly reaches unity, which indicates that basically all agents communicate using the same word. It means that on the complete graph not only the local symmetry is broken ($m_L > 0$) but also the global one ($m_G > 0$). For $\alpha = 1$, simulations for both a square lattice and a complete graph show that m_L is small (and decreases in time) and thus even the local symmetry is preserved (Fig 4).

Having in mind modeling the emergence of communicative consensus in a population of agents, it is desirable to examine the behavior of our model also on heterogeneous networks. The simplest ones are perhaps random graphs. We examined our model on Erdös-Rényi random graphs [25, 26] of an average node degree z. For large z ($z = 10$), the model behaves similarly as on a complete graph and quickly reaches a global consensus about the communicated word (Fig 6). For smaller z ($z = 2, 3$), the model remains trapped in a disordered phase, where consensus is reached only locally (similarly as on a square lattice).
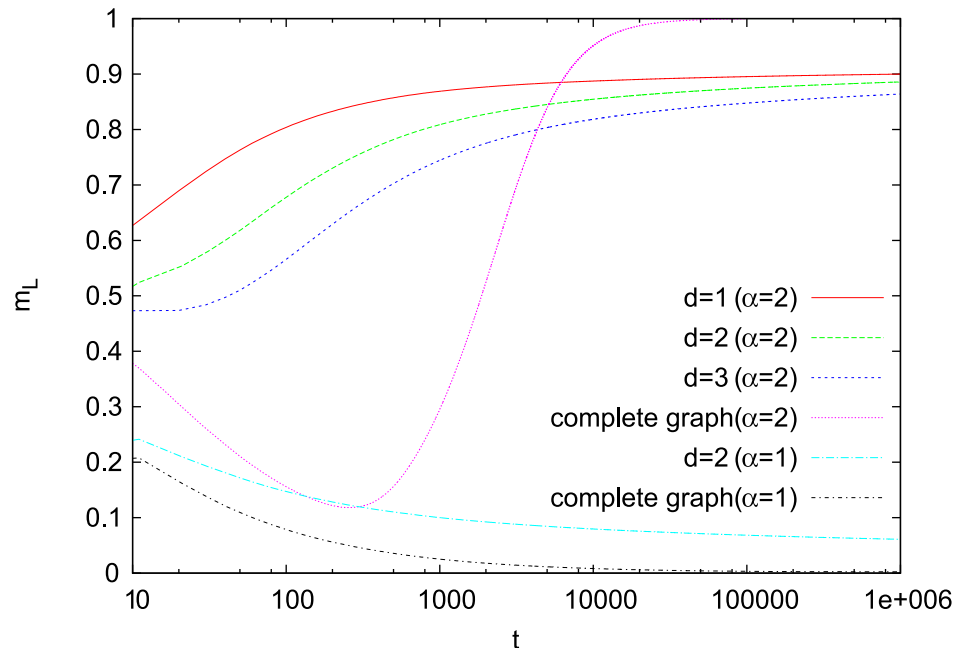

Fig 4. Time dependence of m_L. Results for a single-object model with $N_w = 2$ on complete graphs ($N = 10^5$) and Cartesian lattices with $d = 1$ ($N = 10^5$), $d = 2$ ($N = 300^2 = 9 \cdot 10^4$), and $d = 3$ ($N = 50^3 = 125 \cdot 10^3$). In the case of the ordinary reinforcement ($\alpha = 1$), none of the words is even locally preferred on a complete graph (since $m_L \to 0$), and only a small asymmetry is seen for a square lattice. The results presented (also in the following figures) are averages over 20 independent runs. Statistical errors are typically smaller than plotting symbols and are omitted.
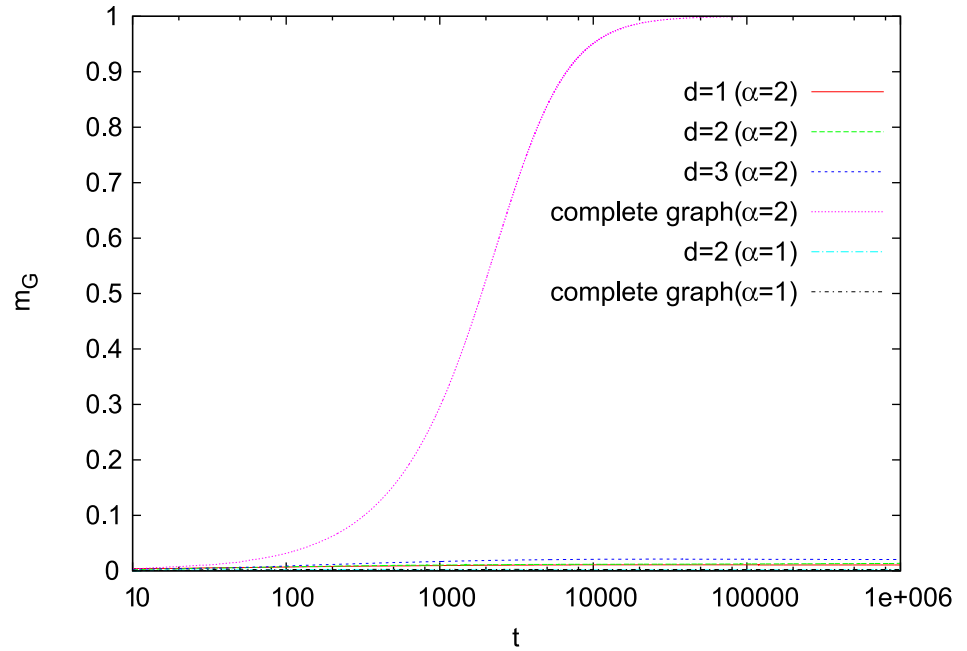
Fig 5. Time dependence of m_G. Results for a complete graph and Cartesian lattices with the same simulation parameters as in Fig 4. Only for a complete graph and $\alpha = 2$, a global symmetry gets broken and one word dominates in the entire population of agents.
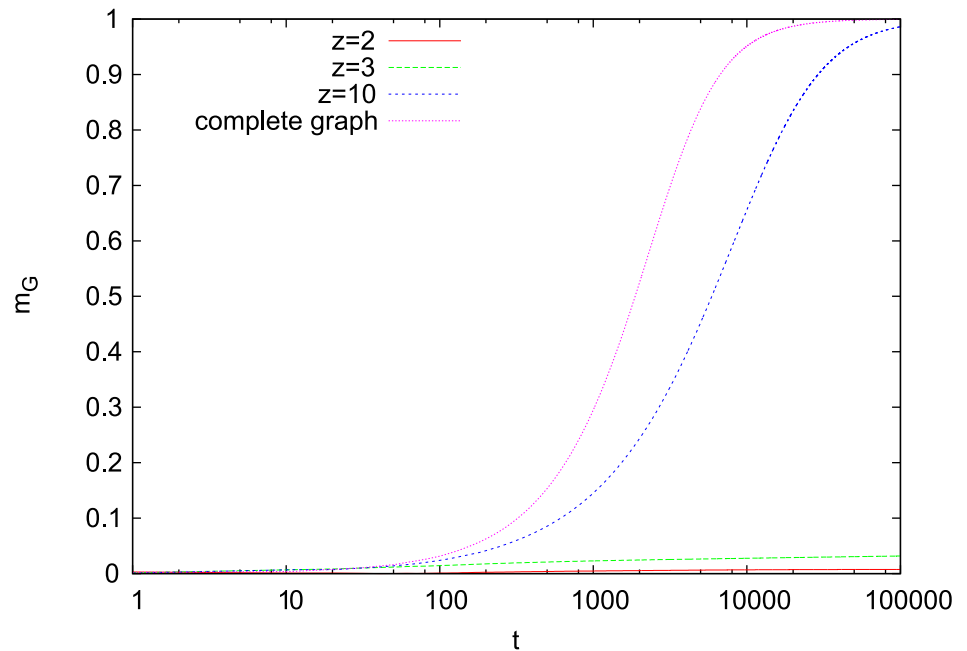
Fig 6. Time dependence of m_G (random graphs). Results for random graphs of an average node degree z and a complete graph ($\alpha = 2$, $N_w = 2$, $N = 10^5$). Only for sufficiently large z, the behavior on the random graphs is similar to that on the complete graph. Averaging over 20 runs includes generation of independent graphs.

Let us recall that for random graphs, $z = 1$ marks the percolation transition [26]. To study the formation of a global consensus, one needs to consider only $z > 1$, since for $z < 1$ the graph decomposes into separate components. Random graphs for finite z are tree-like hence they are effectively infinite-dimensional. One might thus expect that a statistical-mechanics model placed on such graphs behaves similarly for any $z > 1$. Such an expectation is supported with the exact solution of the Ising model on random graphs [27], which shows that for any $z > 1$, the model has a finite-temperature critical point belonging to the mean-field universality class. Also the Naming Game exhibits a similar behavior and numerical simulations show that for any $z > 1$, it reaches a global consensus [14]. However, for directed random graphs, the consensus dynamics does depend on z and a global consensus appears but only for $z > 1.96$ in the Naming Game and for $z > 1.85$ in the Ising model [14]. The present model seems to have a similar behavior with the average node degree z playing an important role. A global consensus characterized by nonzero m_G appears for $z = 10$ while for $z = 2$ and 3, which is still above the percolation threshold, the model gets trapped in a disordered configuration. A precise location of the transition between these two regimes remains, however, beyond the scope of the present paper.

There is a number of models with dynamics driving the system toward consensus, such as, for example, the Voter, Ising, or Naming Game models. All of them evolve toward consensus but they differ in the details of the evolution. One of the important quantities characterizing their dynamics is a surface tension [14], which keeps the interface (i.e., the boundary between different phases) bounded and is responsible for shrinking droplet excitations. The dynamics of the Ising model or the Naming Game exhibit a number of similarities, such as, for example, a power-law coarsening, which is to a large extent a consequence of the surface tension present in these models [20]. The absence of a surface tension results in a quite different dynamics. Indeed, the Voter model, known to have a tensionless dynamics, exhibits, for example, in the two-dimensional case, a logarithmically slow coarsening and in the three-dimensional version, it does not coarsen at all. Let us notice that in certain disordered systems (spin glasses), the dynamics might also be tensionless [28, 29]. The dynamics of our model on regular networks, which (as shown in Fig 3) remains disordered, evolves very slowly and has a well developed interface, may also be tensionless. Possible relations with some other disordered (and maybe glassy) systems is interesting but beyond the scope of the present work. As we show in section on population renewal, one can modify the dynamics of our model so that it does not get trapped in a disordered state and most likely evolves as, e.g., an Ising model. It may indicate that in such a way we restore the surface tension into the dynamics.

We do not present here the corresponding numerical results, but we analysed our model also for $N_w > 2$ and observed a qualitatively similar behavior. The model gets stuck in a disordered structure for finite-dimensional Cartesian lattices but rather quickly reaches a mono-word phase on a complete graph or sufficiently dense random graphs.

Multi-object version

In the previous section, we analysed a model, in which agents try to establish a name for a single object. Here we examine its multi-object generalization. On a square lattice, the multi-object version behaves similarly to the single-object one, namely it gets trapped in a disordered configuration, where only some local consensus appears. Indeed, simulations for $N_o = 2$ show that only small groups of agents communicate with the same word (Fig 7). We do not present here our numerical results, but a group of agents may reach a fairly good consensus while talking on a certain object, while much worse agreement with respect to another one (in other words, the panels in Fig 7, which present the dominant words used by agents for the first
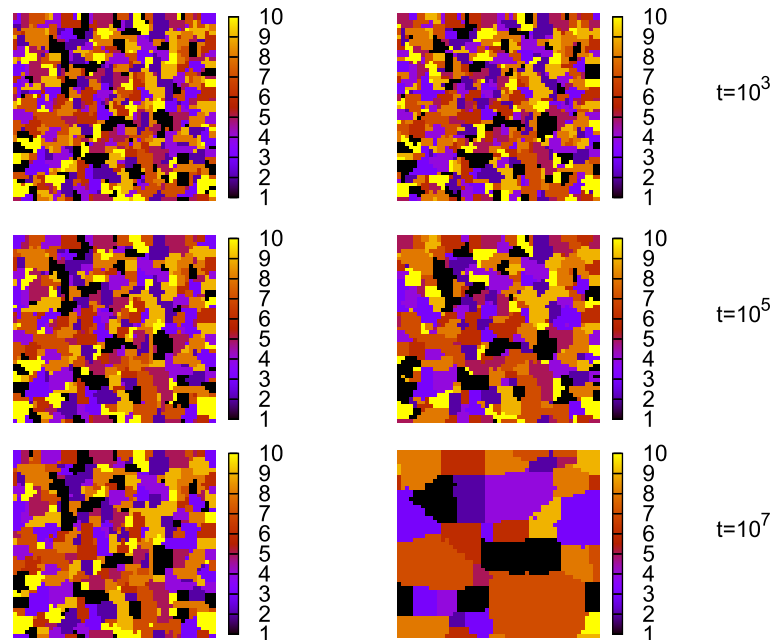
Fig 7. Distribution of the dominant words that agents use to talk about the first object. *Left*: simulations on a square lattice with $N_o = 2$, $N_w = 10$, $N = 50 \cdot 50 = 2.5 \cdot 10^3$, $\alpha = 2$. *Right*: the same simulations but with a population renewal (with probability $p = 10^{-5}$).

object, would be uncorrelated with those for the second object). As in the single-object version, after some initial transient, the evolution of the model nearly stops (in Fig 7 the configurations for $t = 10^5$ and $t = 10^7$ are almost identical). On a complete graph, a much better consensus is reached. During simulations, we measured a success rate defined as a fraction of successful communication attempts in a unit of time. Numerical results for $N_o = 10$ show that when N_w, i.e., the number of words in agents' inventories, is large enough ($N_w = 50$ and 70), the model reaches rather fast a regime, where the success rate is nearly 1 (Fig 8). For smaller N_w (20, 30, and 40), the success rate is much lower, even after long simulations. It suggests that large- and small-N_w regimes may be qualitatively different. Let us also notice that for $\alpha = 1$, the regime with the success rate close to unity is reached in time approximately a decade longer than for $\alpha = 2$. Previous simulations in a similar numerical setup but for a much smaller number of agents and shorter time scale suggested that the ordinary reinforcement ($\alpha = 1$) does not lead to an optimal communication system [11].

Fig 9 provides yet another indication that the dynamics for large and small N_w considerably differ. In this figure, we present a total weight associated with a given word, defined as follows:

$$w_i^{total} = \sum_{A,j} w_{i,j}(A), \quad i = 1, 2, \ldots, N_w \quad (5)$$

where summation is over all agents (A) and objects (j). In Fig 9, what is actually plotted is a normalized total weight given as $w_i^{total} / \sum_k w_k^{total}$. For $N_o = 10$ and $N_w = 50$, we can notice 10 peaks corresponding to the 10 words that are mainly in use. Taking into account a very large success rate (Fig 8), it means that the agents established a single word for each object, which became dominant in their inventories related to this object. As a result only this word is selected by speakers when they decide to talk about the object and just this word leads then to a correct recognition of the object by hearers. The resulting language provides a nearly perfect
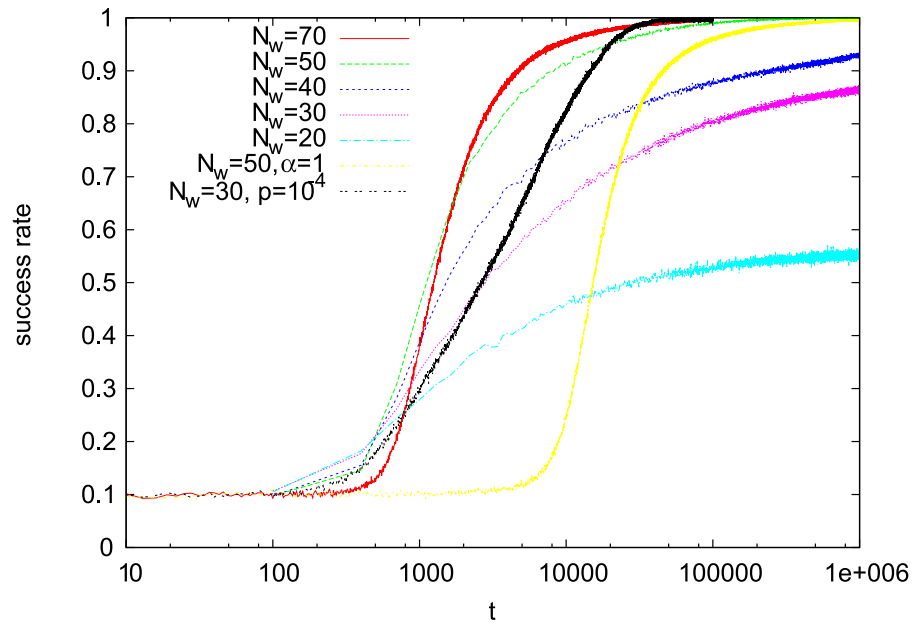
Fig 8. Time dependence of the success rate. Results for the model on the complete graph of size $N = 10^4$ with $N_o = 10$, several values of N_w, and $\alpha = 2$. For $N_w = 50$ and $\alpha = 1$ (yellow line), we can see a much slower convergence to a consensus than for $N_w = 50$ and $\alpha = 2$. The black line shows the success rate for the version with a population renewal (with probability $p = 10^{-4}$).

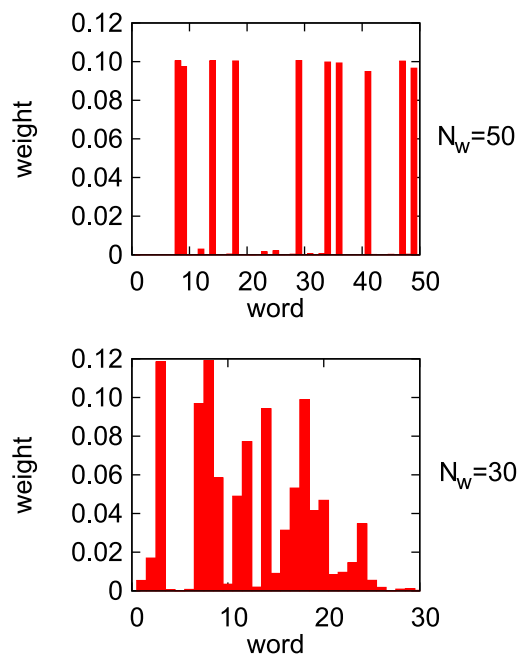https://doi.org/10.1371/journal.pone.0208095.g008

Fig 9. Distribution of the total (normalized) weights associated with particular words. Results for simulations with $N_o = 10$ on the complete graph of size $N = 10^4$ and simulation time $t = 10^6$ ($\alpha = 2$). Simulations for $t = 10^5$ lead to nearly identical distributions.

https://doi.org/10.1371/journal.pone.0208095.g009

one-to-one mapping between objects and words. Let us emphasize that such a global language emerges spontaneously, as a result of two-agent interactions only.

Much different behavior can be seen for smaller N_w. In Fig 9, some peaks can be also distinguished for $N_w = 30$, but in addition there is an entire spectrum of less important but clearly nonnegligible words, which are being used by agents. The one-to-one correspondence between words and objects is missing in this case and the resulting language has a much smaller success rate (Fig 8). Since agents increase the weights of the communicated word only when the object is recognized correctly, it means that also nondominant words lead sometimes to a correct recognition—otherwise their weights (in relation to dominant words) would diminish to zero. It is an analogue of synonymy, a common feature of natural languages. Synonymy, however, does not reduce the success rate, while homonymy (or polysemy [30]) does. Homonymy appears when a certain word has significant weights associated with more than one object. Communicating such a word may result in an incorrect recognition of the object and the success rate smaller than 1 indicates that homonyms are also present in the emerging language of our model. The structure of our model is quite complex and some intermediate scenarios are also possible. Namely, with respect to some objects, agents may develop a one-to-one relation between objects and words (like for $N_w = 50$), while with respect to some others, a more complex language containing synonyms and homonyms may be used.

Population renewal

As we have already seen, both the single- and multi-object versions of our model on a two-dimensional lattice get trapped in a disordered regime, with only a local consensus reached. In such a regime, the coarsening dynamics, which could lead to formation of larger clusters of agents that reached a consensus, becomes very slow. Some other models with a consensus dynamics, such as the Ising model or the Naming Game, are known to have much faster coarsening dynamics, which could be attributed to the surface tension generated in these dynamics. In this section, we examine our model modified in such a way that the dynamics does not get trapped in a disordered state and induces perhaps some kind of a surface tension. Namely, we introduced a simple mechanism of a population renewal, which means that in each step, the selected agent either (with some probability p) is replaced with a new one (with all weights reset to 1) or else the agent acts as a speaker.

The time evolution of a single-object model with a population renewal on a square lattice is shown in Fig 10. Certainly, the evolution in this case is different than in the absence of population renewal (Fig 3). It seems that there is a tendency to reduce the length of interfaces in this model, just as in some models with a surface tension.

Additional arguments that the dynamics generates some kind of an effective surface tension come from the analysis of time dependence of $1 - m_L$ (Fig 11). Let us notice that significant contributions to this quantity come mainly from interfacial agents. Provided that the characteristic cluster size is l, we easily find a scaling relation $1 - m_L \sim l^{-1}$ [31]. From the time dependence of $1 - m_L$ (Fig 11), we conclude that $l \sim t^{0.41}$ and such a value seems to be independent of the renewal probability $p > 0$. Only for $p = 0$, we obtain a much slower increase of l, perhaps logarithmic (in time), which reflects the trapping of the model in a disordered configuration as, for example, in Fig 3. A small deviation from the Ising model increase $l \sim t^{1/2}$ [20] may be attributed perhaps to a diffusive structure of an interface in our model. Let us notice that a similar increase $l \sim t^{0.45}$ was observed also in certain opinion-formation models that are expected to have a dynamics with an effective surface-tension [32].

We also analysed how the coarsening dynamics depends on α. Our results for the renewal probability $p = 10^{-3}$ are shown in Fig 12. It seems that the asymptotic decay of $1 - m_L$ is
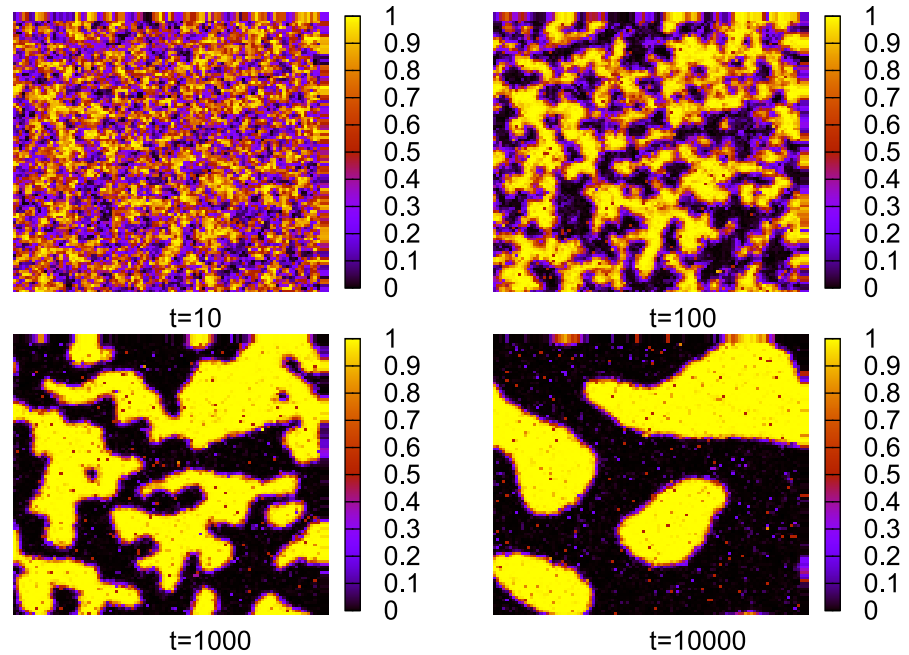
Fig 10. Spatial distribution of s_1, the probability that an agent will select word W_1 (Eq (1)). Results for a single-object model on a square lattice with $N = 10^2 \cdot 10^2 = 10^4$, $\alpha = 2$, $N_w = 2$, and with the renewal probability $p = 10^{-4}$. In this case, contrary to Fig 3, clusters of agents with the same dominant word grow steadily.
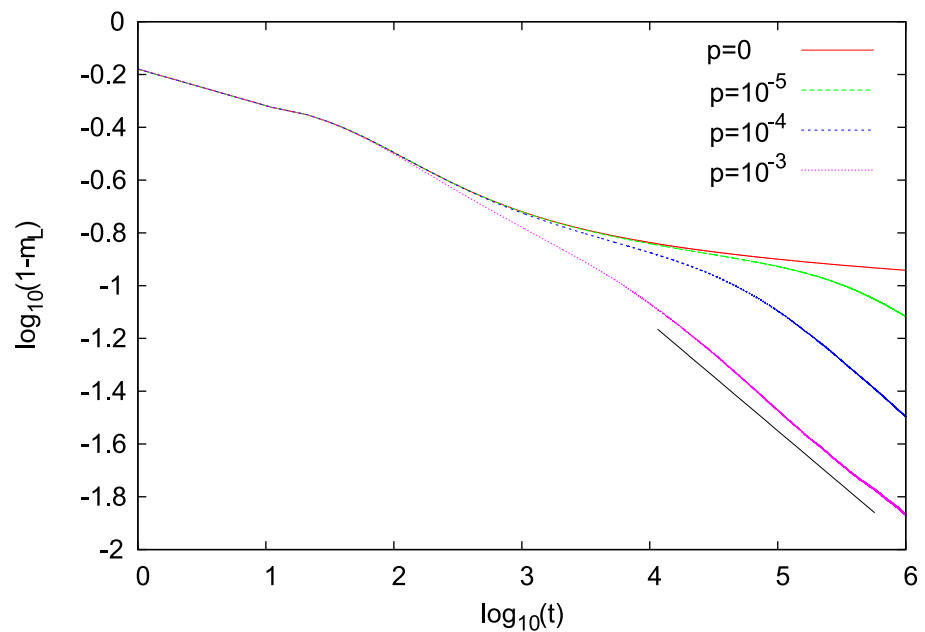
Fig 11. Time dependence of $1 - m_L$ for several values of the renewal probability p. Results for a single-object, square-lattice version of our model. Simulations were made for $\alpha = 2$, $N_w = 2$, and $N = 200 \cdot 200 = 4 \cdot 10^4$, and the results are averages of 20 independent runs. The line segment has a slope corresponding to $t^{-0.41}$.
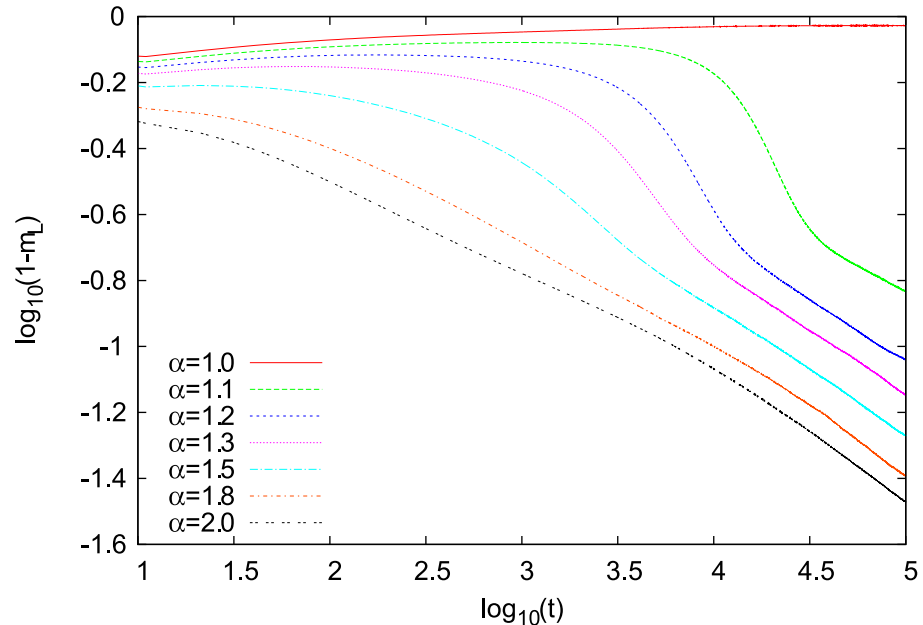
Fig 12. Time dependence of $1 - m_L$ for several values of α. Results for a single-object square-lattice version of our model for the renewal probability $p = 10^{-3}$. Simulations were made for $N_w = 2$ and $N = 200 \cdot 200 = 4 \cdot 10^4$, and the results are averages of 20 independent runs.

characterized by the same exponent for any $\alpha > 1$. A qualitatively different behavior is seen only for $\alpha = 1$, where the model does not coarsen. These results suggest that the behavior of our model, also with respect to other properties, should not depend on a particular choice of α as long as $\alpha > 1$ (superlinear reinforcement).

The population renewal affects also a multi-object version of our model. In the absence of the renewal, the model on the complete graph (for $N_o = 10$ and $N_w = 30$) develops a language with a reduced success rate (Fig 8) and with no clear object-word mapping (Fig 9). However, even a tiny renewal probability ($p = 10^{-4}$) considerably increases the success rate of communication (Fig 8). We do not present our numerical data here but in this case, a clear object-word mapping does emerge, similar to that in the upper panel of Fig 9. Also a multi-object square-lattice version of the model behaves differently for $p > 0$. Indeed, the snapshot configurations (Fig 7) show that in this case the model does not get trapped (as for $p = 0$) but coarsens similarly to the single-object version with population renewal (Fig 10).

The overall behavior of our models is summarized in Table 1.

Discussion and conclusions

The objective of the present study was to examine the emergence of linguistic conventions in a multi-agent model with reinforcement learning. Models of this kind have a potential to generate a complex language, which reflects their multi-object and multi-agent structure, but global aspects of their dynamics are rather purely understood. This is much in contrast to some simpler models with agreement dynamics, like Ising model or Naming Game, which due to the surface tension have curvature-driven dynamics [20], or voter model, which lacks the surface tension and has a much different dynamics [33].

In the single-object version, agents do not need to recognize the object and each communication attempt is in a sense successful. It turns out that it is the structure of the network of interactions between agents that plays a decisive role and determines the asymptotic state of

Table 1. Behavior of our models as a function of dynamics and network structure.

Dynamics	Network	Behavior
Single-Object		
$\alpha > 1$, no renewal	Cartesian	local consensus
	Complete graph	global consensus
	Random Graph	global/local consensus for large/small z
$\alpha > 1$, renewal	Cartesian	global consensus, power-law coarsening
$\alpha = 1$, no renewal	Cartesian	no local consensus
	Complete graph	no local consensus
$\alpha = 1$, renewal	Cartesian	no local consensus
Multi-Object		
$\alpha > 1$, no renewal	Cartesian	partial local consensus
$\alpha > 1$, no renewal, large N_w	Complete graph	unique object-word mapping
$\alpha > 1$, no renewal, small N_w	Complete graph	language with homonyms and synonyms
$\alpha > 1$, renewal, small N_w	Complete graph	unique object-word mapping
$\alpha > 1$, renewal	Cartesian	unique object-word mapping, power-law coarsening

Global consensus means that all agents communicate using the same word. Local consensus means that only neighbouring agents use the same word. Partial local consensus means that neighbouring agents use the same words to communicate only with respect to some objects while with respect to others, there may be no efficient communication.

the model. While for a complete graph or sufficiently dense random graphs, a global consensus is reached and all agents use the same word for communication, on finite-dimensional lattices or sparse random graphs, only a local consensus is reached and the model gets trapped in a disordered configuration.

In the Naming Game, which is an alternative model describing formation of linguistic consensus, agents also try to establish a name for a given object. However, in this case a global consensus is much easier to reach except on networks with a strong community structure [34, 35]. Such a strong tendency to reach a consensus could be explained using the notion of an effective surface tension, which is generated in the Naming Game. It turns out that when a population renewal is introduced, the surface tension emerges also in our model and its evolution toward consensus is much enhanced. However, such a curvature-driven evolution appears only for $\alpha > 1$, which indicates the importance of a superlinear reinforcement. Our study thus shows that a physical intuition developed for some statistical-mechanics models may be also used to understand to some extent reinforcement learning systems.

Having in mind linguistic contexts, the multi-object version of our model is more interesting. The results show that in this case the structure of the network plays an important role as well. On complete graphs, an efficient global communication may be established, such that all agents unambiguously match each object with the same corresponding word. On finite-dimensional lattices, such a mapping is again only local (and partial) and the model gets trapped in a disordered configuration. In the multi-object version, in addition to the network structure, other parameters are also important, namely the number of objects N_o and the number of words that agents have at their disposal N_w. Our simulations suggest that a unique object-word matching may emerge only when N_w is considerably greater than N_o. If this is not the case, the resulting communication is less efficient and the emerging language contains some homonyms and synonyms. Of course, such behavior should be by no means considered as undesirable or unrealistic, since all natural languages contain such forms. Further studies

concerning, for example, the frequency and durability of homonyms and synonyms in our model would be desirable, but are left for the future.

In the multi-object version, the population renewal also enhances formation of an efficient communication. The population renewal basically resets the weights of an agent, and thus it plays a role similar to forgetting, which is a factor already known to improve the performance of reinforcement learning systems [36]. Our snapshot configurations show that also in this case, the population renewal most likely induces a certain surface tension, similarly to the single-object version. Hence, one of the merits of our work is the demonstration that reinforcement learning systems with the population renewal and superlinear reinforcement ($\alpha > 1$) reveal certain similarities to some other models with the agreement dynamics (such as the Naming Game) and exhibit a power-law coarsening. However, without (or perhaps with a sufficiently small) population renewal or for (sub-)linear reinforcement ($\alpha \leq 1$), the dynamics of these two systems considerably differ.

Let us notice that the surface tension might be of some importance also in linguistic processes and, for example, some recent works show that the boundaries of dialect regions are controlled by a length-minimizing effect analogous to the surface tension [21]. Moreover, the fast extinction of natural languages, especially those of a small number of users, indicates that some coarsening does take place. Hopefully, some simple models can be propounded, which might provide some insight into such a linguistic dynamics. Of course, the processes of emergence, diversification or extinction of languages are very complex and affected by a large number of factors such as, for example, politics, geography, economy, or technological development. Thus computational modeling may provide their very crude, qualitative description at most.

Population renewal supplies a new generation of language users. Similarly as children, they quickly learn a language of the neighbors they interact with. Let us notice that some linguists strongly advocate the view that profound language changes occur in the process of language learning and children perhaps play an important role in this process, for example, making mistakes [37, 38]. However, such a view can be questioned because the modifications children generate seldom survive till their adulthood [39, 40] due to, for example, usually lower social and economic status of youngsters [41]. In our opinion, it would be certainly interesting, as well as feasible, to consider an aged-structured version of our model and analyse the role of the young generation. As we have already noticed, young users are needed to generate a surface tension and coarsening (which in the context of human language evolution is probably more realistic than a population trapped in a multilanguage regime). In the aged-structured population, a peer communication is an expected feature, but with such preference being too strong, the population could split up into separate linguistic communities. Analysing the emergence of a young generation dialect and its possible influence on the language of adults is, however, left as a future problem.

Author Contributions

Conceptualization: Dorota Lipowska, Adam Lipowski.

Formal analysis: Dorota Lipowska, Adam Lipowski.

Investigation: Dorota Lipowska, Adam Lipowski.

Methodology: Dorota Lipowska, Adam Lipowski.

Software: Dorota Lipowska.

Visualization: Adam Lipowski.

Writing – original draft: Dorota Lipowska, Adam Lipowski.

Writing – review & editing: Dorota Lipowska.

References

1. Lewis D. Convention: A philosophical study. Oxford, UK: Blackwell; 2002.

2. Nowak M, Krakauer D, Kingdom U. The evolution of language. Proceedings of the National Academy of Sciences. 1999; 96(July): 8028–8033. https://doi.org/10.1073/pnas.96.14.8028

3. Nowak M, Komarova N. Towards an evolutionary theory of language. Trends in Cognitive Sciences. 2001; 5(7): 288–295. https://doi.org/10.1016/S1364-6613(00)01683-1 PMID: 11425617

4. Oliphant M. The dilemma of Saussurean communication. BioSystems. 1996; 37(1-2): 31–38. https://doi.org/10.1016/0303-2647(95)01543-4 PMID: 8924637

5. Barr DJ. Establishing conventional communication systems: Is common knowledge necessary? Cognitive Science. 2004; 28(6): 937–962. https://doi.org/10.1207/s15516709cog2806_3

6. Eggenberger F, Pólya G. Über die Statistik vorketter vorgänge. Zeit. Angew. Math. Mech. 1923; 3: 279–289. https://doi.org/10.1002/zamm.19230030407

7. Pemantle R. A survey of random processes with reinforcement. Probab. Surveys. 2007; 4: 1–79. https://doi.org/10.1214/07-PS094

8. Skyrms B. Signals: Evolution, learning, and information. Oxford, UK: Oxford University Press; 2010. https://doi.org/10.1093/acprof:oso/9780199580828.001.0001

9. Beggs AW. On the convergence of reinforcement learning. Journal of Economic Theory. 2005; 122: 1–36. https://doi.org/10.1016/j.jet.2004.03.008

10. Lenaerts T, Jansen B, Tuyls K, De Vylder B. The evolutionary language game: An orthogonal approach. Journal of Theoretical Biology. 2005; 235: 566–582. https://doi.org/10.1016/j.jtbi.2005.02.009 PMID: 15935174

11. Spike M, Stadler K, Kirby S, Smith K. Minimal requirements for the emergence of learned signaling. Cognitive Science. 2017; 41(3): 623–658. https://doi.org/10.1111/cogs.12351 PMID: 26988073

12. Barrett JA. Numerical Simulations of the Lewis Signaling Game: Learning Strategies, Pooling Equilibria, and the Evolution of Grammar. UC Irvine: Institute for Mathematical Behavioral Sciences. 2006. Available from: https://escholarship.org/uc/item/5xr0b0vp

13. Baronchelli A. The emergence of consensus: A primer. R. Soc. Open Sci. 2018; 5(2): 172189. https://doi.org/10.1098/rsos.172189 PMID: 29515905

14. Lipowski A, Lipowska D, Ferreira AL. Agreement dynamics on directed random graphs. J. Stat. Mech. 2017; 063408. https://doi.org/10.1088/1742-5468/aa727a

15. Castellano C, Fortunato S, Loreto V. Statistical physics of social dynamics. Rev. Mod. Phys. 2009; 81: 591. https://doi.org/10.1103/RevModPhys.81.591

16. Toivonen R, Onnela JP, Saramäki J, Hyvönen J, Kaski K. A model for social networks. Physica A: Statistical Mechanics and its Applications. 2006; 371(2):851–60. https://doi.org/10.1016/j.physa.2006.03.050

17. Steels L. A self-organizing spatial vocabulary. Artificial Life. 1995; 2(3): 319–332. https://doi.org/10.1162/artl.1995.2.3.319 PMID: 8925502

18. Baronchelli A, Felici M, Loreto V, Caglioti E, Steels L. Sharp transition towards shared vocabularies in multi-agent systems. J. Stat. Mech: Theory Exp. 2006; 2006(06): P06014. https://doi.org/10.1088/1742-5468/2006/06/P06014

19. Dall'Asta L, Castellano C. Effective surface-tension in the noise-reduced voter model. EPL (Europhysics Letters). 2007; 77(6): 60005. https://doi.org/10.1209/0295-5075/77/60005

20. Bray AJ. Theory of phase-ordering kinetics. Advances in Physics. 2002; 51(2): 481–587. https://doi.org/10.1080/00018730110117433

21. Burridge J. Spatial Evolution of Human Dialects. Phys. Rev. X. 2017; 7: 031008.

22. Drinea E, Frieze A, Mitzenmacher M. Balls and bins models with feedback. In: Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete algorithms (SODA'02). Society for Industrial and Applied Mathematics, Philadelphia, PA, USA; 2002. pp. 308-315.

23. Shapiro C, Varian HR. Information Rules. Boston, MA: Harvard Business School Press; 1999.

24. Lipowska D, Lipowski A. Language competition in a population of migrating agents. Phys. Rev. E. 2017; 95(5): 052308. https://doi.org/10.1103/PhysRevE.95.052308 PMID: 28618596

25. Erdös P, Rényi A. On Random Graphs I. Publicationes Mathematicae. 1959; 6: 290–297.

26. Newman MEJ. Networks: An Introduction. Oxford: Oxford University Press; 2010.

27. Dorogovtsev SN, Goltsev AV, Mendes JFF. Ising model on networks with an arbitrary distribution of connections. Phys. Rev. E. 2002; 66: 016104. https://doi.org/10.1103/PhysRevE.66.016104

28. Houdayer J, Martin OC. A geometrical picture for finite-dimensional spin glasses. EPL (Europhysics Letters). 2000; 49(6): 794. https://doi.org/10.1209/epl/i2000-00221-7

29. Lipowski A, Johnston D. Tensionless structure of a glassy phase. Phys. Rev. E. 2001; 65(1): 017103. https://doi.org/10.1103/PhysRevE.65.017103

30. Panman O. Homonymy and polysemy. Lingue. 1982; 58: 105–136.

31. A similar reasoning relates an excess energy and a characteristic length during coarsening of Ising-type models. See, e.g., Shore JD, Holzer M, Sethna JP. Logarithmically slow domain growth in nonrandomly frustrated systems: Ising models with competing interactions. Phys. Rev. B. 1992; 46(18): 11376. https://doi.org/10.1103/PhysRevB.46.11376

32. Dall'Asta L, Galla T. Algebraic coarsening in voter models with intermediate states. J. Phys. A. 2008; 41: 435003. https://doi.org/10.1088/1751-8113/41/43/435003

33. Dornic I, Chaté H, Chave J, Hinrichsen H, Critical coarsening without surface tension: The universality class of the voter model, Phys. Rev. Lett. 2001; 87: 045701. https://doi.org/10.1103/PhysRevLett.87.045701 PMID: 11461631

34. Kozma B, Barrat A. Consensus formation on adaptive networks. Phys. Rev. E. 2008; 77: 016102. https://doi.org/10.1103/PhysRevE.77.016102

35. Lipowska D, Lipowski A. Naming game on adaptive weighted networks. Artificial Life. 2012; 18: 311–323. https://doi.org/10.1162/artl_a_00067 PMID: 22662912

36. Barrett J, Zollman KJS. The role of forgetting in the evolution and learning of language. Journal of Experimental and Theoretical Artificial Intelligence. 2009; 21(4): 293–309. https://doi.org/10.1080/09528130902823656

37. Lightfoot D. The development of language: Acquisition, change, and evolution. Malden, MA: Blackwell; 1999.

38. Lightfoot D. How new languages emerge. Cambridge: Cambridge University Press; 2006.

39. Kerswill P. Children, adolescents, and language change. Language Variation and Change. 1996; 8: 177–202. https://doi.org/10.1017/S0954394500001137

40. Diessel H. Language change and language acquisition. In: Bergs A, Brinton L. editors. Historical linguistics of English: An international handbook, vol. 2. Berlin: Mouton de Gruyter; 2012. pp. 1599–1613.

41. Labov W. Principles of language change, vol. II: Social factors. Malden, MA: Blackwell; 2001.