Faster Convention Emergence by Avoiding Local Conventions in Reinforcement Social Learning

Muzi Liu[1], Ho-fung Leung[1], and Jianye Hao[2]

[1] The Chinese University of Hong Kong, Hong Kong, China
{liumuzi@link, lhf@}cuhk.edu.hk
[2] Tianjin University, Tianjin, China
jianye.hao@tju.edu.cn

Abstract. In this paper, we propose a refinement of multiple-R [1], which is a reinforcement-learning based mechanism to create a social convention from a significantly large convention space for multi-agent systems. We focus on the language coordination problem, where agents develop a lexicon convention from scratch. As a lexicon is a set of mappings of concepts and words, the convention space is exponential to the number of concepts and words. We find that multiple-R suffers from local conventions, and refine it to the independent-R mechanism, which excludes neighbors' rewards from the value update function, and thus avoids local conventions. We also explore how local conventions influence the dynamics of convention emergence. Extensive simulations verify that independent-R outperforms the state-of-the-art approaches, in the sense that a more widely adopted convention emerges in less time.

Keywords: Artificial intelligence in modeling and simulation · Multi-agent systems

1 Introduction

Coordination among autonomous agents is essential for cooperative goal achievement in open multi-agent systems (MAS), as incompatible actions usually incur resource cost to the participating agents [2]. In the process of cooperative goal achievement, the conformity to a convention helps to simplify agents' decision-making process and hence improve the efficiency of agent societies [3]. Since a centralized entity that directly enforces a convention requires the imposition of global rules, a convention that emerges in a decentralized manner is more feasible for coordination in MASs [3, 4].

The study of social convention emergence explores how agents involved in repeated coordination games can reach consensus through local interactions [4]. Some researchers focus on characterizing the dynamics of convention emergence process. Airiau and Sen [5] explore the emergence of convention through social learning [6], and study the effect of network characteristics on the emergence speed. They observe and explain the formation of stable local conventions, which hinders global convention emergence. Similar phenomena are also

investigated in [7], where stable local conventions are stated to benefit coordination. Other works concentrate more on developing efficient mechanisms of convention emergence. Most of the approaches are spreading-based. They combine local optimization and imitation to achieve efficient convention emergence [8, 9]. Some researchers also branch out to reinforcement learning-based (RL-based) approaches. RL-based approaches usually utilize reinforcement social learning, where an agent learns its policy over repeated interactions with multiple agents [5, 6]. Some hierarchical learning frameworks are also proposed to further improve the efficiency of the emergence of social conventions [10, 11].

The above mentioned approaches are designed for relatively small number of alternative conventions. Challenges appear along with the increase of convention space, including issues related to convention quality and emergence efficiency [1, 12]. One particular research problem that captures the challenge of convention space explosion is the *language coordination problem* [13], which is an imitation of how human develop languages from scratch. For this problem, Hasan, Raja and Bazzan propose a topology-aware mechanism (TA) with a dynamic MAS [12], where agents update lexicons only based on the information of current episode, causing frequent oscillation of lexicon qualities. Recently, RL-based approaches for lexicon coordination problem have been proposed by Wang et al. [1]. Inspired by the classic Q-learning algorithm and its variants [14], they propose two efficient mechanisms, multiple-Q and multiple-R (MR), that ensure high-quality final conventions.

However, in the RL-based mechanism proposed by Wang et al., when an agent updates its policy, it averages the rewards of its neighbors to adjust its policy, which we find leads to the emergence of local conventions among this neighborhood. The emergence of local conventions significantly hinders global convention emergence. We find that excluding neighbors' rewards from an agent's value update function reduces the emergence of local conventions.

Similar to previous works [1, 5, 12, 15, 16], we focus on developing better solutions for the challenging language coordination problem, in which the convention space is exponential to the number of concepts and words. In this work, we modify MR through understanding and characterizing the dynamics of the convention emergence process. Our refined RL-based strategy, independent-R (IR), overcomes the limitations caused by local conventions. We improve the MR mechanism by valuing each agent's independent and newly obtained information rather than the information given by neighbors. Simulations show that IR outperforms the state-of-the-art approaches, in the sense that a more widely adopted convention emerges in a shorter time. We further compare our approach with MR, and show that MR suffers from the emergence of local conventions while our approach does not.

2 Related Work

The studies about social conventions in multi-agent systems have long been attracting researchers. Since Shoham and Tennenholts addresses the question

about the efficiency of convention evolution led by local decisions of agents in multi-agent systems [17], many researchers have studied and proposed possible strategies to shrink the time of convention stabilization [1, 4, 6, 12, 13, 15–17].

Some works [1, 12, 15, 16] propose mechanisms designed for a particular challenging problem in large and open MASs, the *language coordination problem*. In language coordination problem, a population of agents tries to develop a lexicon convention through repeated local interactions. The problem was originated from Luc Steels' model [13], where a group of distributed agents develops a vocabulary to name themselves and to identify each other using spatial relations. In Luc Steels' model, agents develop their own vocabularies privately and then try to align its vocabulary with others. A survey of existing approaches for convention formation in normative multi-agent system is provided in [18]. SRA [15] and FGJ [16] are two spreading-based mechanisms that help agents agree on the best convention from multiple alternatives. In SRA, a partial-transfer strategy is used to guarantee that a convention emerges. SRA makes use of a sophisticated agent architecture design to create high-quality conventions. On the other hand, FGJ introduces a set of influencer agents equipped with high-quality lexicons, so that the agent population is guided towards the adoption of high-quality conventions. While the above two methods are both set in static networks, Hasan et al. extend SRA by leveraging a topology-aware utility computation mechanism and equipping agents with the ability of reorganizing their neighborhood [12]. These approaches require a large amount of time to converge to a dominant convention for large convention space.

In addition to spreading, some RL-based approaches have been proposed recently [1], and remarkably accelerate convention emergence in large convention space. In these models, agents learn to align their lexicons with neighbors through repeated interactions [1]. However, as the learning mechanisms take advantage of neighbors' learned information, they suffer from the emergence of local conventions, which slows down the emergence of a global convention.

3 Language Coordination Game

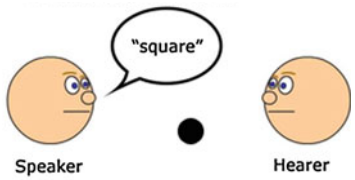
We consider a situation where a language emerges among a population of agents in a bottom-up manner. A language is simplified to a lexicon, which is a set of mappings from words to concepts. Every agent has a lexicon. Interactions occur between a pair of agents. Through the interactions, agents accumulate their knowledge of others' lexicons and adjust their own lexicons accordingly. They try to align with others by learning from experience. Therefore, after repeated interactions, lexicons in the whole population tend to converge. Ideally, every agent will eventually have the same lexicon, which we consider as a convention.

To formally define language coordination game, we introduce three components as follows: (a) language coordination problem, (b) interaction model, and (c) convention emergence problem.

3.1 Language Coordination Problem

To describe the situation where agents develop a language in a decentralized fashion, we first model a language as a *lexicon*. A lexicon contains a set of *words* (W) and a set of *concepts* (C), where each word is mapped to a concept. In a lexicon, it is possible that one concept is mapped to multiple words (*synonymy*), or multiple concepts are mapped to a single word (*homonymy*). We consider a lexicon with one-to-one mappings has the highest quality. Initially, each agent has a lexicon, where each word is arbitrarily mapped to a concept. Through repeated interactions, agents learn to adjust their lexicons so that they can succeed in more interactions.

Interactions, which can be either successful or failed, are modeled as two-player $|W|$-action coordination games. For each interaction, there are two agents that participate in it. The two participating agents are called *speaker* and *hearer*, whose lexicons are $L_{speaker}$ and L_{hearer} respectively. The speaker and the hearer interact under a certain concept, which is selected before the interaction. The concept selection may follow certain frequency distribution, which represents the appearing frequency of each concept. Suppose the selected concept is c. During an interaction, the speaker sends a word $w_{speaker}$ to the hearer, where concept-word mapping $(c, w_{speaker}) \in L_{speaker}$. Upon receiving $w_{speaker}$, the hearer also finds a word w_{hearer}, where $(c, w_{hearer}) \in L_{hearer}$. If $w_{speaker} = w_{hearer}$, the interaction succeeds; otherwise the interaction fails. Following [1] and [12], we model an agent interaction as a two-player $|W|$-action coordination game. Each agent selects a word as its action, thus the number of possible actions is the number of words $|W|$. In a successful interaction, the hearer receives a positive reward; in a failed interaction, the hearer receives a negative reward. Speaker receives no reward. For instance, one simple 2-action coordination game on the concept of "round shape" is shown in Table 1, where the possible words to express the shape are "circle" and "square". The two desirable outcomes ("circle", "circle") and ("square", "square") are both Nash equilibria, indicating the two agents have the same word mapped to the concept of round shape.



3.2 Interaction Model

We consider a population of N agents, where all the agents are connected following a static network topology. In each interaction, an agent randomly selects

Table 1. A coordination game on the concept of "round shape"

Agent 1's action	Agent 2's action	reward
circle	circle	$+r$
circle	square	$-r$
square	circle	$-r$
square	square	$+r$

a neighbor to interact with. The interaction model of the multi-agent system is represented by an undirected graph, $G = (V, E)$, where G is the network structure, V is the set of nodes ($|V| = N$), E is the set of edges. If $v_i, v_j \in V$ and $(v_i, v_j) \in E$, then v_i, v_j are *neighbors*. $N(i)$ is the *neighborhood* of node i, where $N(i) = \{v_j | (v_i, v_j) \in E\}$. A *path* between v_i, v_j is a sequence of edges connecting a sequence of nodes which begins with v_i and ends with v_j. The *distance* between two nodes is the shortest path between them. Three representative networks are considered: random, small-world [19], and scale-free network [20].

3.3 Convention Emergence Problem

We model the language coordination problem as a convention emergence problem. A desirable convention is a one-to-one lexicon adopted by all agents in the network. As every lexicon can become the convention, the convention space contains all possible lexicons, the size of which is $|W|^{|C|}$. The convention space is large even for a moderate number of words and concepts. Thus, following [1], we decouple a lexicon into concept-word mappings. Hence each lexicon can be defined as a Markov strategy, and the dynamics of interactions in each episode can be modeled as a two-player Markov Game $\langle S, \{A_i\}_{i \in V}, \{R_i\}_{i \in V}, T \rangle$, where

- S is the set of states, which corresponds to the concepts.
- V is the set of all agents.
- $\{A_i\}_{i \in V}$ is the collection of action sets. $\forall i \in V$, A_i contains all the words.
- $\{R_i\}_{i \in V}$ is the set of payoff functions. $R_i : S \times A_i \times A_j \to R$, where agent i and j are the interacting agents. It is positive if the two agents select the same action and negative otherwise. R_i satisfies the Gaussian distribution.
- T is the state transit function: $S \times A_i \times A_j \to Prob(S)$, where agent i and j are the interacting agents. T models concept usage frequencies.

4 A Convention Emergence Framework

4.1 Overall Algorithm

Algorithm 1 describes the overall framework of repeated language coordination games. Initially, each agent has an initial lexicon, in which words are randomly mapped to each concept. To develop a convention, agents discard old lexicons

Algorithm 1 Convention emergence framework

```
1: for each episode do
2:   for each agent do
3:     lexiconInitialization()
4:     while transition < λ do
5:       conceptTransition()
6:       neighborSelection()
7:       actionSelection()
8:       policyUpdate()
9:       lexiconUpdate()
10:    end while
11:  end for
12: end for
```

(line 3) and construct new lexicons in each episode. The constructed lexicons are in fact composed by the concept-word mappings that are selected in interactions of current episode. To construct a lexicon, in each episode, each agent initiates interactions for λ times as a hearer (line 4). Each interaction focuses on one concept, which is determined by a concept transition function (line 5).

During an interaction (line 6 to line 9), the hearer randomly selects a neighbor as the speaker (line 6). Based on its policy, the hearer selects a word as its action for the focused concept(line 7); the speaker also selects a word as the speaker's action. If both of the hearer and the speaker select the same word, the interaction is successful. In the case of successful interaction, the hearer receives a positive reward; otherwise, the hearer receives a negative reward. According to the reward it receives, the hearer updates its policy (line 8) to get higher rewards in future interactions. The word selected by the hearer will be mapped to the concept of this interaction, and the mapping becomes a part of the hearer's lexicon (line 9). The details of policy update will be introduced in Section 4.3.

The mappings from the selected words to the focused concepts in all λ interactions of the current episode compose the hearer's new lexicon. A lexicon convention emerges when all agents in the network has the same lexicon.

4.2 Convention Establishment

To establish a convention in an agent society, there are mainly two types of approaches. One is spreading-based approach, where agents spread conventions through information transfer. A typical transfer strategy is copy-transfer: an agent simply replicates its neighbors' lexicons. Another class of techniques is reinforcement learning, where agents learn conventions by repeatedly interacting with other agents. Up to now, the most efficient method to establish conventions is a RL-based approach, which is called multiple-R (MR) [1]. However, instead of learning from interactions, the MR mechanism mainly uses neighbors' learning results to guide an agent's own policy. The interference from neighbors largely affects agents' policies, and thus they always align with neighbors instead of the whole network. In this situation, local conventions are likely to emerge and hinder the emergence of a global convention.

Grounded on the above analysis, we propose independent-R learning strategy, which is a RL-based strategy basing on the up-to-date MR strategy [1]. The main

improvement is that we avoid the interference from neighbors, and thus prevents the emergence of local conventions.

4.3 Independent-R Learning Strategy

Value function update The following strategy corresponds to the policy update in the independent-R mechanism (line 8 in Algorithm 1).

In an interaction focused on concept s, assume the hearer agent i takes word a as action, and receives a reward $r(s,a)$. $r(s,a)$ is positive if the interaction succeeds, or negative otherwise. The hearer takes average of the rewards it has received for (s,a) in recent n episodes and updates $R_i(s,a)$ accordingly. Let $n_i(s,a)$ be the number of interactions in recent n episodes, where the interaction is under concept s and agent i selects word a. Initially, $R_i(s,a)$ is 0. Formally:

$$R_i(s,a) \leftarrow R_i(s,a) + \frac{r(s,a) - R_i(s,a)}{n_i(s,a)}. \tag{1}$$

Then, agent i updates its Q-table:

$$Q_i(s,a) \leftarrow Q_i(s,a) + \alpha(R_i(s,a) - Q_i(s,a)), \tag{2}$$

where α is the learning rate. $Q_i(s,a)$ may fluctuate a lot if α is too large.

Action selection In the agent interaction, the pair of participating agents will select actions following certain action selection strategy (line 7 in Algorithm 1). We adopt the action selection design in [1] as follows.

For an interaction of concept s, let RA_i be the set of words that has been mapped to the concepts in set $S \setminus \{s\}$ in current episode. To determine the word a_i for this interaction, agent i uses *ε-greedy* strategy to choose the word with the highest utility from set $A_i \setminus RA_i$, which is the set of words that are not mapped to any concept yet in current episode:

$$a_i \leftarrow \begin{cases} argmax_{a \notin RA_i} Q_i(s,a) & \text{with probability } 1-\varepsilon \\ \text{a random word} \in A_i \setminus RA_i & \text{with probability } \varepsilon \end{cases} \tag{3}$$

If there are multiple actions with maximum utilities, agent will randomly select one from these actions.

4.4 Comparison with multiple-R

Independent-R (IR) only uses an agent's own rewards to update its strategy, while MR uses a weighted average of neighbors' rewards. In MR, agent i updates its recent average rewards $R_i(s,a)$ as same as (1). However, in addition to $R_i(s,a)$, agent i also computes the weighted average of neighbors' rewards:

$$\overline{r} = \sum_{j \in N(i) \cup \{i\}} f(i,j) R_j(s,a), \tag{4}$$

where $f(i,j) = \frac{degree(j)}{\sum_{k \in N(i) \cup \{i\}} degree(k)}$. Then, agent i updates its Q-value $Q'_i(s,a)$:

$$Q'_i(s,a) \leftarrow Q'_i(s,a) + \alpha(\overline{r} - Q'_i(s,a)). \tag{5}$$

We note that the information of neighbors' degree takes effect implicitly when agents interact, hence it is not used in the value update function (2) of IR. For each interaction, the speaker is randomly selected by the hearer from its neighborhood. Thus, agents who have larger degree will be selected by more hearers and participate in more interactions. Through repeated interactions, lexicons adopted by high degree agents naturally have more influence.

We also note that taking others' rewards into account reduces the effects of the agents' own rewards, especially the reward received from the most recent interactions. When an agent i is updating $Q'_i(s,a)$, the reward $r(s,a)$ it just received will first be divided by $n_i(s,a)$ to compute $R_i(s,a)$ as (1), and then the average reward $R_i(s,a)$ is further multiplied by a fraction $f(i,j)$ in (4). Thus neighbors' rewards and the agent i's own past rewards together play a big part in Q-values, suppressing the influence of immediate reward $r(s,a)$.

5 Experiments and Results Analysis

We conduct experiments on several typical networks, including random, scale-free, and small-world networks. The performance of IR is compared with the state-of-the-art approaches including multiple-R with teacher-student mechanism (MR+TS) [1], TA [12], FGJ [16], and SRA [15].

We define dominant lexicon convention as the one shared by the most agents. Based on previous works, the metrics used for comparison are as follows:

- **Efficiency**: Efficiency measures how fast a network converges into a dominant lexicon convention. The criterion used for measurement are followings:
  - **Average Communicative Efficacy (ACE)**: The proportion of successful interactions in total. It measures the level of coordination.
  - **Proportion of Agent Compliant with Convention (ACC)**: The proportion of agents adopting dominant lexicon. It measures the level of convention emergence.
- **Effectiveness**: A mechanism is effective if it is able to converge into a widely adopted lexicon convention within a reasonable amount of time. It is reflected by reaching high ACC in a reasonable number of episodes.

5.1 Simulation Setup

We conduct experiments on three network topologies: random, scale-free, and small-world, where scale-free and small-world networks are generated by Barabasi-Albert model [20] and Watts and Strogatz small-world model [19] respectively. The random networks are generated by randomly connecting two nodes continuously from a uniform spanning tree of the complete graph. Each network

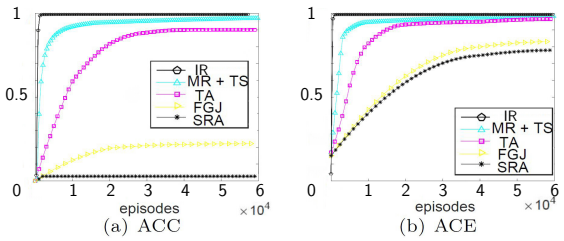


Fig. 1. ACC and ACE figure for small-world network

consists of $1,000$ agents and $20,000$ edges, so that the average degree is 20. Following [1, 12], both the number of concepts and words are 10, thus the size of convention space is 10^{10}. For each agent, the concept transition number λ is 20. The learning rate α is 0.02. We adopt the random frequency distribution as the concept usage frequency. Specifically, the 20 concept transitions are divided into two sections, and each concept is selected exactly once in each section. Thus, the sequence of concept transitions in a section is a random permutation of the concept set. In this way, each agent initiates equal number of interactions in each episode for each concept. Each realization executes for $60,000$ episodes and all the presented results are the averages over 50 realizations of each network.

5.2 Simulation Results

The results of simulation for small-world networks are shown in Fig. 1. Similar results also apply to random and scle-free networks. The figure presents how the proportion of agent compliant with convention (ACC) and average communication efficacy (ACE) evolves over time. We add the performances of other approaches in small-world networks in Fig. 1 for comparisons. The results of other approaches are provided in [1]. The results support that IR is the most effective and efficient approach compared with state-of-the-art approaches.

For effectiveness, the figure shows that when convention stabilizes, the ACC of IR is the highest, which is over 0.99. It suggests that the dominant lexicon is adopted by almost every agent in IR. As the dominant lexicon convention is widely accepted, IR is the most effective approach over the existing approaches.

For efficiency, from Fig. 1a, we can observe that IR is the most efficient method, since it needs much less number of episodes to reach the same ACC as other approaches. For example, TA approach requires almost $30,000$ episodes, and MR+TS requires almost $10,000$ episodes for ACC to reach 0.9, while IR only needs hundreds of episodes. The performances of FGJ and SRA are even worse, as their ACC cannot reach 0.9 in given $60,000$ episodes.

The ACE figure also shows a similar pattern. For example, when ACE is 0.9, the fastest approach MR+TS takes $12,968$ time-steps for random network, while IR requires only 524 time-steps. As expected, ACE grows rapidly along with ACC, since the higher rate of successful interactions is a result of higher proportion of agents adopting a same lexicon.

In conclusion, the simulation results show that IR is more efficient and more effective than state-of-the-art approaches, in the way that it requires less time steps to have more agents adopting the dominant lexicon convention.

6 The Emergence of Local Conventions

From the experimental results, we observe that IR outperforms other approaches significantly, as it is more efficient and more effective. However, there is only a slight difference between MR and IR, namely the Q-value update function.

The reason why IR outperforms MR is that local conventions may appear in MR, and hinder the emergence of global convention. It is mentioned in [1] that the ACE of MR is high even when the ACC is low, which means a lot of interactions succeed even though there exist no dominant lexicons. It supports our finding that local conventions emerge in the network, and thus most of the interactions happen between agents who are in the same group, leading to high proportion of successful interactions even when there is no dominant lexicon.

6.1 Local Conventions Emergence

To verify the emergence of local conventions in MR, we present simulation results to confirm the existence of local conventions in MR mechanism.

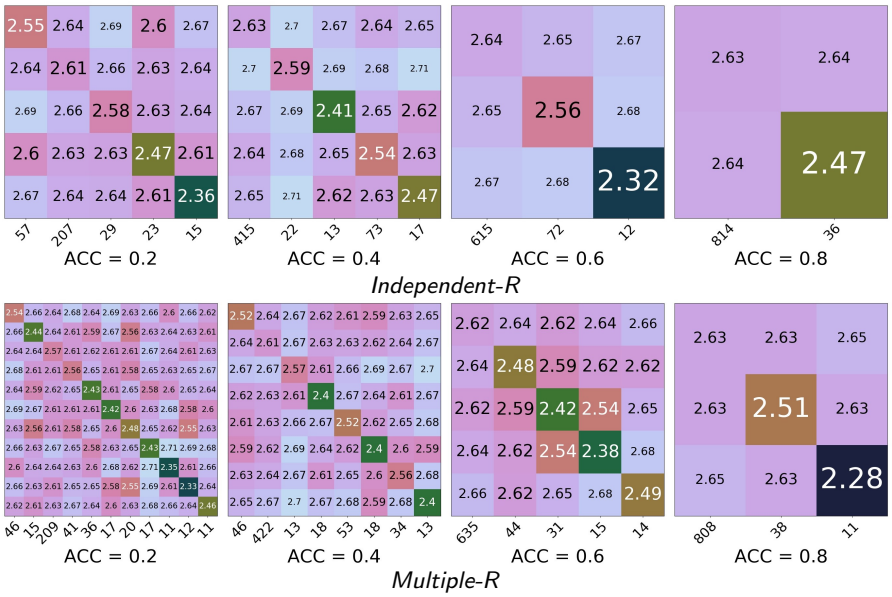
To describe local conventions, we use *group* to refer to the set of agents that adopt a same lexicon, and *group distance* to measure the distance between pairs of groups. Specifically, for a lexicon p, let $G_p = (V_p, E_p)$ be a subgraph of $G = (V, E)$, where V_p is the set of agents that adopt lexicon p, and E_p is the set of edges $(v_i, v_j) \in E$, where $v_i, v_j \in V_p$. Let $d(v_i, v_j)$ be the distance between nodes v_i and v_j. Then the group distance between G_p and G_q is defined by $D(G_p, G_q) = \frac{\sum_{v_i \in V_p, v_j \in V_q} d(v_i, v_j)}{|V_i| \times |V_j|}$.

The group distance between a group and itself reflects how closely the agents inside the group are connected. If $D(G_p, G_p)$ is significantly smaller than $D(G_p, G_{p\prime})$ for all other $p\prime$, it indicates that p is a local convention among G_p. A small group distance between G_p and itself indicates that most of the agents inside G_p are neighbors with one another. Thus agents inside G_p will mostly interact with agents inside G_p. As they all adopt lexicon p, these interactions always succeed. By repeated interactions, their lexicons are confirmed by each other. In this way, the local convention is reinforced inside such a group.

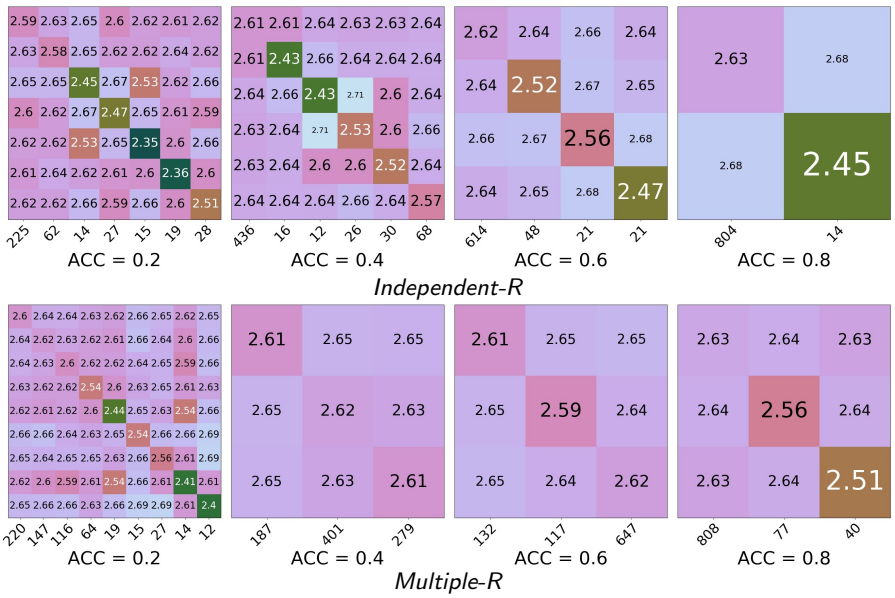
From Fig. 1 we can observe that ACC increases monotonically during convention emergence, and converges to 1 with the number of episodes increasing. We record the group distances when ACC reaches 0.2, 0.4, 0.6, 0.8 in MR and IR. Only the groups having more than 10 agents are recorded.

Fig. 2 shows how the group distances evolve in two typical simulations. Fig. 2a shows the situation where there are a lot of small groups; and Fig. 2b shows the situation where there are few groups with significant sizes.

There are two rows of graphs shown in Fig. 2a and Fig. 2b. The upper rows show the data collected using IR approach, while the lower rows show the data



(a) Large amount of groups with small sizes



(b) Small amount of groups with large sizes

Fig. 2. Group distance figure. Each block represents a group distance. The two coordinates are the two measured groups. Group sizes are labeled on x-axis. The upper row is the group distances in IR and the lower row is in MR when ACC = 0.2, 0.4, 0.6, 0.8 in the same network. Only the groups having more than 10 agents are recorded.

collected using MR in the same network. In each row, there are four graphs. From the left to right, the four graphs show the data corresponding to the episode when ACC $= 0.2, 0.4, 0.6, 0.8$ respectively. In each graph, there are small blocks with numbers labeling on. The number and color of a block at coordinate (g_1, g_2) show the group distance $D(g_1, g_2)$. The exact size of each group is labeled on the x coordinate. For a group g, if the color of the block (g, g) is apparently darker from the blocks in the same row, then $D(g, g)$ is significantly smaller than $D(g, g')$ for all g', and the lexicon of group g is more likely a local convention.

In Fig. 2a, we can observe that compared to graphs of IR, there are more blocks in the graphs of MR, which means there are more groups in MR. Also, the diagonal lines on graphs of MR are clearer than those of IR, indicating that groups in MR are more closely connected than in IR. These closely connected groups in MR shows that local conventions appear in these groups as we expect.

Fig. 2b shows the situation where there are several groups with significant size. For example, when ACC $= 0.4$, there are only one group of size 400 and several small groups smaller than 70 in IR; while in MR, there are two groups of size around 200 and one group of size 400. These large groups persist and compete with each other during the whole process of convention emergence.

From simulations and the above analysis, we conclude that compared to MR, agents equipped with IR produce less local convention groups with smaller sizes, and thus speed up the emergence of global convention.

6.2 Analysis of Multiple-R Approach

As we mentioned before, the only difference between MR and IR is that MR collects rewards from neighbors to update Q-values while IR does not. Nonetheless, local conventions appear and hinder the emergence of global convention in MR more frequently than in IR. It is reasonable to infer that collecting neighbors' rewards increases the emergence of local conventions in MR.

In MR, for an agent i, as the value update function (5) includes the average rewards $\overline{r}$ of i's neighbors, and the rewards of high degree neighbors are largely weighted by $f(i, j)$ in (4), agents always quickly align with high degree neighbors. Also, it is highly likely that other agents in i's neighborhood $N(i)$ are connected to the same high degree node and adopt the same lexicon as well. Thus agent i and its neighbors always receive positive rewards from their interactions, and the lexicon they adopt is reinforced. Therefore, it is difficult for lexicons outside of i's neighborhood to affect agent i after the local convention is formed. The quick adoption to lexicons of the high degree agents and the ineffectiveness of distant lexicons lead to a large amount of local conventions in the network.

In addition, it is harder and slower for a global convention to supersede local conventions in MR. To supersede a local convention group, the boundary agents of the group need to adopt the global convention first, so that they can influence their neighbors inside the group by interactions. However, as the immediate rewards from agent interactions only takes a small part in $\overline{r}$, the influence of agent interactions is overwhelmed by high degree neighbors in (4). The boundary agents will hardly switch to a new lexicon, since they are affected more by their

high degree neighbors inside the group. If the boundary agents are affected more by their neighbors inside the group, they will hardly switch to a new lexicon that is not adopted by their high degree neighbors. Thus the local conventions are difficult to assimilate, which will significantly slow down the emergence of a global convention.

6.3 Effect of Local Conventions

To describe the situation where local convention hinders the emergence of global convention, we consider two kinds of possible cases depending on the amount of non-dominant local conventions, which are elaborated below.

Large amount of small groups Consider the situation where there is one dominant lexicon convention and many non-dominant local conventions. The dominant lexicon has the largest chance to become the global convention.

To achieve global convention, the dominant lexicon needs to supersede other small local conventions in the network. However, for the agents inside a closely connected group, their lexicons are reinforced by each other through repeated interactions. The group can only be superseded by changing the lexicons of the agents that are located in boundary areas of a group first. To make the boundary agents adopt the dominant lexicon, the ideal situation is that those boundary agents continuously interact with neighbors outside the group, so that the influences from outside neighbors can be reinforced and take effect. However, as the neighbor to interact with is randomly selected for each interaction, it may take a lot of time steps before that happens. After changing one boundary agent i, it's neighborhood $N(i)$ then becomes a set of boundary agents, who can be affected by outside neighbors. By repeating the process of superseding peripheral agents, a local convention group may then be superseded by the dominant convention.

As it is hard to supersede the local convention groups, the local conventions are the obstruction of the emergence of a global convention in the situation where there are a large amount of small groups.

Small amount of large groups Another possible situation is that there are few local conventions with sizes close to the dominant lexicon.

In this case, local conventions may supersede the original dominant lexicon. If a local convention keeps absorbing agents from the dominant lexicon, it then becomes dominant instead. However, the new dominant lexicon is not stable either, since it does not have an overwhelmingly large number of conforming agents. Similar competitions then continue between the new dominant lexicon and other local conventions.

As the lexicons compete for domination, they in fact prevent each other from spreading. None of the competitors has a chance to steadily increase its size. Thus the ACC will fluctuate for a while, as the dominant lexicon is changing back and forth. The competition continues until a local convention suddenly gets an overwhelming advantage by chance and obtains consistent domination.

Therefore, under this situation, it is clear that the local conventions prevent the convention from growing. When there are few local conventions with large sizes, they are also obstructions of the convention emergence.

7 Conclusion and Future Work

In this work, we propose a new RL-based approach, independent-R, basing on the previous multiple-R mechanism, which avoids local conventions and thus accelerates global convention emergence. We find that using neighbors' average rewards to adjust an agent's own rewards causes the emergence of local conventions, which largely hinder the global convention emergence. We value the information obtained from interactions by an agent itself, instead of the information given by neighbors, so that local conventions are less likely to appear. Extensive simulations indicate that IR outperforms the state-of-the-art approaches, in the sense that a more widely adopted convention emerges in a shorter time. We also analyze local conventions in detail to discuss how local conventions may hinder the efficiency of convention emergence.

As future work, one of the worthwhile directions is to capture the influential factors for the formation of local conventions in large convention space, or to explore possible mechanisms to utilize local conventions.

References

1. Y. Wang, W. Lu, J. Hao, J. Wei, and H.-F. Leung, "Efficient convention emergence through decoupled reinforcement social learning with teacher-student mechanism," in *Proc. 17th Int. Conf. on Autonomous Agents and MultiAgent Systems*, AAMAS '18, 2018.
2. J. M. Marchant, N. Griffiths, and M. Leeke, "Manipulating conventions in a particle-based topology," in *Coordination, Organizations, Institutions and Norms in Agent Systems Workshop : A workshop of the 12th International Conference on Autonomous Agents and Multiagent Systems : AAMAS2015*, AAMAS, 2015.
3. O. Sen and S. Sen, "Effects of social network topology and options on norm emergence," in *Coordination, Organizations, Institutions and Norms in Agent Systems V* (J. Padget, A. Artikis, W. Vasconcelos, K. Stathis, V. T. da Silva, E. Matson, and A. Polleres, eds.), 2010.
4. M. Mihaylov, K. Tuyls, and A. Nowé, "A decentralized approach for convention emergence in multi-agent systems," *Autonomous Agents and Multi-Agent Systems*, 2014.
5. S. Airiau, S. Sen, and D. Villatoro, "Emergence of conventions through social learning," *Autonomous Agents and Multi-Agent Systems*, 2014.
6. S. Sen and S. Airiau, "Emergence of norms through social learning," in *Proc. 20th Int. Joint Conference on Artifical Intelligence*, 2007.
7. S. Hu and H. fung Leung, "Achieving coordination in multi-agent systems by stable local conventions under community networks," in *Proc. 26th Int. Joint Conference on Artificial Intelligence*, 2017.

8. J. M. Pujol, J. Delgado, R. Sangüesa, and A. Flache, "The role of clustering on the emergence of efficient social conventions," in *Proc. 19th Int. Joint Conference on Artificial Intelligence*, IJCAI'05, 2005.
9. J. Delgado, "Emergence of social conventions in complex networks," *Artificial Intelligence*, vol. 141, pp. 171–185, 10 2002.
10. C. Yu, H. Lv, F. Ren, H. Bao, and J. Hao, "Hierarchical learning for emergence of social norms in networked multiagent systems," in *AI 2015: Advances in Artificial Intelligence*, 2015.
11. T. Yang, Z. Meng, J. Hao, S. Sen, and C. Yu, "Accelerating norm emergence through hierarchical heuristic learning," in *ECAI*, 2016.
12. M. R. Hasan, A. Raja, and A. Bazzan, "Fast convention formation in dynamic networks using topological knowledge," in *Proc. 29th AAAI Conf. on Artificial Intelligence*, AAAI'15, 2015.
13. L. Steels, "A self-organizing spatial vocabulary," *Artif. Life*, 1995.
14. H. V. Hasselt, "Double q-learning," in *Advances in Neural Information Processing Systems 23* (J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta, eds.), pp. 2613–2621, 2010.
15. N. Salazar, J. A Rodrigues-Aguilar, and J. L Arcos, "Robust coordination in large convention spaces," *AI Communications*, pp. 357–372, 2010.
16. H. Franks, N. Griffiths, and A. Jhumka, "Manipulating convention emergence using influencer agents," *Autonomous Agents and Multi-Agent Systems*, 2013.
17. Y. Shoham and M. Tennenholts, "Emergent conventions in multi-agent systems: initial experimental results and observations (preliminary report)," in *Proc. of Knowledge Representation and Reasoning*, 1992.
18. B. T. R. Savarimuthu and S. Cranefield, "Norm creation, spreading and emergence: A survey of simulation models of norms in multi-agent systems," *Multiagent and Grid Systems*, 2011.
19. D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
20. A.-L. Barabasi and R. Albert, "Albert, r.: Emergence of scaling in random networks. science 286, 509-512," *Science (New York, N.Y.)*, vol. 286, pp. 509–12, 11 1999.