Emergence of Multi-Step Conventions

Marina Katoh[1], Feyza M. Hafızoğlu[2][0000−0002−9598−1061], Jacob Brue[1][0009−0001−6656−3960], and Sandip Sen[1][0000−0001−6107−4095]

[1] The University of Tulsa, Tulsa OK 74104, USA
[2] İstanbul Commerce University, 34480 İstanbul, Türkiye

Abstract. Emergence of conventions and social norms has been an active area of research in multiagent systems to facilitate coordination in agent societies. Various learning approaches, interaction frameworks, topological connections, and information availability assumptions have been investigated to facilitate the emergence of conventions. Most of these scenarios involve repeated bilateral interactions between learning agents choosing actions simultaneously and often modeled as stage games. Many real-life conventions, however, involve sequential decision making by two or more parties. In this paper, we investigate convention emergence in agent populations repeatedly playing bilateral sequential games. We investigate the development of conventions for exchange of greetings. We show what assumptions and biases can consistently produce stable and beneficial conventions to emerge in sequential interaction scenarios. Our experimental results and concomitant analysis sheds light on the dynamics of the emergence of multi-step conventions with sequential interactions.

Keywords: coordination, convention emergence, multi-step conventions

1 Introduction

Whenever we speak a language or greet someone, we follow social conventions. In other words, conventions determine the way we speak or greet. Conventions and social norms can be considered as the grammar of social interactions [3]. Similar to a grammar, conventions and norms help us to differentiate what is acceptable and what is not in a society. Without these shared rules, it becomes either impossible or very costly to achieve goals due to social conflicts.

Emergence of conventions and social norms have been an active area of research in multi-agent systems to facilitate coordination in agent societies [1, 8, 13, 16]. Various learning approaches [2, 18], interaction frameworks [12, 19], topological connections [14, 17], and information availability assumptions [10] have been investigated to facilitate the emergence of conventions.

Most of these scenarios involve repeated bilateral interactions between learning agents choosing actions simultaneously and often modeled as stage games. Many real-life conventions, however, involve sequential decision making by two or more parties. For instance, greeting each other, a sequence of coordinated actions between a parent and a child [5], dialogues in a group [12]. Throughout this

paper, we adopt the term either "multi-step" or "sequential" conventions which can be defined as a sequence of coordinated actions of players.

It is critical to understand the dynamics of multi-step conventions to obtain a better insight on the emergence of social conventions. Despite the prevalence of such sequential decision making scenarios and criticality of the subject, research on multi-step conventions did not exist and the research on the convention emergence has been restricted to simultaneous decision making scenarios during interactions.

In this paper, we study the emergence of multi-step conventions, a novel interaction model, in agent populations repeatedly playing bilateral sequential games. We investigate the development of conventions for scenarios like exchange of greetings (shaking hands, kissing, hugging, bowing, or a simple "hi!"). To do so, we consider the sequential coordination game where the players choose their actions sequentially. Hence, the second player can observe the action chosen by the first player and then chooses her action accordingly. The players obtain a positive reward in case of no conflict. To make action decisions, agents learn from a combination of their past interactions and observations of their neighbors. We carefully conducted an extensive set of experiments to examine the influence of key factors such as decision models, different topologies and neighborhood models, number of actions available, and number of agents, on the emergence of multi-step conventions. Our experimental results and concomitant analysis shed light on the dynamics of the emergence of multi-step conventions with sequential interactions.

The remainder of the paper is structured as follows. We first present related work. Following, the society model that is considered in this research is outlined, and our empirical methodology is explained. The results of our experiments are presented in the subsequent section, where we also discuss the findings. The paper concludes with a summary and directions for future research.

2 Related Work

Over the past two decades, a considerable amount of literature has been published on the emergence of norms and conventions in multi-agent systems. What we know about the emergence of social norms is largely based on experimental studies that investigate the conditions under which norms are followed by the majority of society. These studies investigated the mechanisms that lead to the emergence of norms and conventions and the influence of individual and environmental factors.

First, it is critical to highlight the difference between social norms and conventions, which is often blurred. According to Lewis [11], conventions are the equilibria of coordination games. In these games, there are multiple equilibria and only one of them will be the conventions as a consequence of interactions of the individuals. An individual's interest on a specific action is conditional upon the action choices of other individuals in the society, i.e., an action is chosen only if most people follow it [3]. In contrast to conventions, it may not be an

individual's immediate interest to conform with the social norm. For example, a player may be tempted to defect even if the social norm is to cooperate. In this case, the individual's interest conflict with the collective interests in contrast to conventions [10].

To date, various mechanisms have been suggested towards achieving convention emergence in agent societies. Reinforcement learning is the most prevalent learning technique towards forming social conventions [1, 10, 16] in multi-agent systems literature. Airiau *et al.* [2] showed that emergence of conventions can be achieved through social learning, i.e., learning from interaction experiences. Yu *et al.* [18] proposes a novel spiking neural learning model correlating microscopic neural activities with global social norms.

Topology is a significant construct in the life cycle of conventions. Hasan *et al.* [7] demonstrate that convention emerge efficiently despite the large convention space when the agents use a neighborhood reorganization mechanism. Similary, Centola and Baronchelli [4] performed an interesting series of experiments with human subjects showing that simple changes in the network structure lead to global conventions. Franks *et al.* [6] proposed recruiting a number of influencer agents with certain conventions, to facilitate the emergence of high-quality conventions efficiently. Hu and Leung [8] demonstrated that agents can achieve coordination via establishing diverse stable local conventions which is still a solution to coordination issue as an alternative to the global conventions.

There are various aspects of agent societies effecting the emergence of conventions. Though majority of the studies consider one-to-one interactions as pairs among the agent populations [1, 8, 13, 16], one-to-many interactions may occur in real life. Wang *et al.* [17] studied the emergence of conventions when higher-order interaction occurs in a group with two or more agents.

Overall, an extensive research on convention emergence in multi-agent systems have been conducted. While a large body of work has focused on modeling interactions as stage games where the players choose actions simultaneously, we instead considered repeated interactions where the players take actions sequentially with complete information.

3 Preliminaries

3.1 The Sequential Game of Coordination

Figure 1 presents the coordination game where players can gain positive (zero) reward as a result of coordination (anti-coordination) in their actions. Specifically, the positive reward is chosen to be one, as shown in the leaves of the game tree in Figure 1, for the sake of simplicity. In this game, the first player makes a decision. After observing the first player's action, i.e., complete information, the second player decides what to play. Then the corresponding payoffs are distributed among the players.

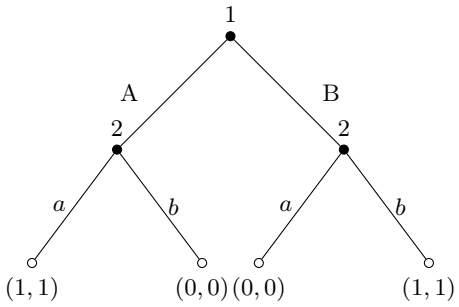


Fig. 1. The sequential game of coordination: If Agent 1 and 2 select the same options (A or B) both receive a reward of 1. Otherwise, no reward is provided.

3.2 The Interaction Model

Our model considers a population of agents, N, where each agent is connected to a subset of the agents according to a static network topology, T. The agents in the population repeatedly play the coordination game with their neighbors for a number of episodes, E.

At the beginning of each episode, each agent is paired with one of its neighbors randomly to play the coordination game as follows. Each agent will have the opportunity to be the first player in each episode. The first player chooses an action $a_1 \in \mathcal{A}$, where $\mathcal{A}$ is the action space.

After observing the action of the first player, the second player chooses its action $a_2 \in \mathcal{A}$. It should also be noted that in each episode, some agents may have more than one opportunity to be the second player due to the fact that the first player is randomly paired with one of its neighbors. The action sequence $\langle a_1, a_2 \rangle$ determines the outcome from this interaction. Outcomes that conform to conventions (i.e. $a_1 = a_2$) are rewarded higher than outcomes that fail to conform.

Network Topologies: We investigate the emergence of conventions in the presence of three well-known representative topologies: (a) *Toroidal grid*, (b) *Small-world*, (c) *Scale-free*, and (d) *Complete* graph (an agent can interact with any other agent in the population).

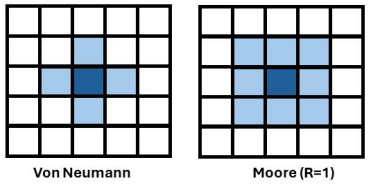


Fig. 2. von Neumann and Moore neighborhood

Neighborhood: In case of grid topology, we considered two different neighborhood models: i. *von Neumann* and ii. *Moore* (see Figure 2). According to the von Neumann neighborhood, all the agents that are adjacent to the central agent are considered as neighbors [15]. On the other hand, in a Moore neighborhood, all eight agents surrounding the central agent are considered neighbors.

Rewards: If the two players choose the same action, they will receive the same, high reward. The focus is on whether or not the agents are coordinated; the actual action they coordinate on does not affect the reward. It should also be noted that as long as the coordination occurs, it does not matter whether the chosen action is the global convention or not.

Algorithm 1 describes the simulation of sequential game of coordination.

Algorithm 1: Sequential Coordination Game

```
   Input  : N, T, E, k_obs, ε_init, ε_end
1  Initialize Agents with N, T, E, k_obs, ε_init, ε_end
2  for episode = 1 to E do
3      foreach agent i in N do
4          j ⟵ selectANeighborRandomly(i, T)
5          a_1 ⟵ Agents[i].getAction(0)
6          a_2 ⟵ Agents[j].getAction(1, a_1)
7          r_1, r_2 ⟵ playGame(a_1, a_2)
8          Agents[i].updateReward(r_1)
9          Agents[j].updateReward(r_2)
```

3.3 Agent's Decision Model

In our framework, the decision making process followed by an agent differs based on whether it is the first or the second player in the coordination game. Comparing the two decisions, the first player's decision process plays a more critical role in determining the convention. On the other hand, coordination, choosing the same action with the first player, is the second player's best interest after observing the first player's move. Thus, the second player employs a simpler decision process while the first player employs a comprehensive one which considers the three criteria explained below.

Criterion 1: Q-learning Agents employ the Q-learning mechanism which has the following three states:

1. The agent is the first player (no previous action).
2. The agent is the second player and the first player chose the first action.
3. The agent is the second player and the first player chose the second action.

In each state, two actions defined in the coordination game are available, for a total of six q-values per agent.

When updating the Q-value of a given state&action pair, the learning rate α_t is $1/(s_t+1)$, where s_t is the number of times the Q-value has previously been updated. This forms a running average when utilized in the Q-update equation. r_t is the reward episode t.

$$Q_{(s,a)} = (1 - \alpha_t) \cdot Q_{(s',a')} + \alpha_t \cdot r_t.$$

The exploration likelihood ϵ is linearly decreased from a high initial value, ϵ_{init}, until a lower threshold, ϵ_{end} is reached.

$$\epsilon_t = (\epsilon_{\text{init}} - \epsilon_{\text{end}}) \cdot \text{decay}(t) + \epsilon_{\text{end}}.$$

ϵ_t is the exploration probability in episode t. $decay(t)$ controls the interpolation from ϵ_{init} to ϵ_{end} of the exploration likelihood over time. In our framework, the following decay function is used.

$$decay(t) = max(1 - (t/t_{me}), 0)$$

t_{me} is the episode at which the exploration probability reaches its minimum ϵ_{end}. ϵ_t is the exploration probability of choosing a random action, while $1 - \epsilon_t$ is the exploitation probability of choosing the action with maximum Q-value.

$$P_q(a) = \begin{cases} 1 & \text{if } \text{argmax}_{a'} Q_{(s,a')} = a \\ 0 & \text{else} \end{cases}$$

Criterion 2: Experiences as The Second Player Each agent has a record of all the actions it has observed as the second player from its first player partner in its binary interactions. When deciding which action to choose, it samples from the frequency distribution of the observed actions taken by its partner over the past K interactions as the second player (represented by $\{a_1, a_2, ..., a_k\}$).

$$P_p(a) = \frac{|\{a = a_i | 1 \leq i \leq k\}|}{k}$$

Criterion 3: Observations of Neighbors For this criterion, the agent samples from the frequency distribution of the most recent action taken by all of its neighbors as the first player. The probability of selecting action a for an agent with neighborhood n is:

$$P_n(a) = \frac{|\{a = a_j | j \in n\}|}{|n|}$$

3.4 Epsilon Greedy Exploration

Agents also employ $\epsilon-$Greedy exploration. With probability ϵ, agents will choose a random action from a uniform distribution $P_U(a) = \frac{1}{\text{count}(A)}$.

Varying Weights of Decision Traits: Each criterion is assigned a specific weight to reflect the degree of influence it has on the action choice made by the agent. The weight for the criterion 1 (Q-value) is w_q, the weight for the criterion 2 (past partners) is w_p, and the weight for the criterion 3 (neighbors) is w_n. The full probability distribution is:

$$P(a) = \epsilon_t P_U(a) + (1 - \epsilon_t)(w_q P_q(a) + w_p P_p(a) + w_n P_n(a))$$

The second player chooses actions based only on the Q-value estimates of the actions in $\mathcal{A}$ at its state $\mathcal{S}$ defined by the action of the first player, subject to epsilon greedy exploration (just like the first player with $w_q = 1$).

4 Experimental Setting

We run simulations with a population of $N = 100$ homogeneous agents that are situated according to one of the topologies.

The ϵ_{init} and ϵ_{end} are set to 0.9 and 0.001, respectively. When evaluating actions observed as second player, the agents take into account the past ten actions, i.e. $k_{obs} = 10$, observed. The number of convergence will be denoted as $\mathbb{N}_{conv}$, while the number of experiments, $\mathbb{N}_{exp}$, and the number of episodes, E per experiment will vary depending on the type of experiment being conducted. Figure 3 indicates the default values of the parameters for experimental setting.

Convergence is considered to have taken place if 90% of the agents have matching selections in any of the episodes being run in an experiment [9].

ϵ_{init}: 0.9	$N = 100$
ϵ_{end}: 0.001	$k_{obs} = 10$

Fig. 3. Default values of the parameters in the experiments

5 Results

In this section, we will evaluate the results of our investigation on the emergence of convention among agents in a sequential decision-making scenario.

5.1 Effect of Varying Weights on Convergence

We begin by assessing the effect that each weight within the criteria of the decision allocation framework has on the convergence of convention among agents.

Analysis on w_n: Table 1 indicates that the neighborhood criteria w_n produced the greatest number of convergence within the given number of episodes being run per experiment. In other words, the agents' observations of the actions undertaken by all neighboring entities within the environment had the most significant influence on the decision-making process of the agent, thereby playing a pivotal role in shaping the selections it ultimately made.

Table 1. Effects of varying weights for Criteria 1-3 from the decision allocation framework. A_1 and A_2 represent the counts of convergence to the two actions, while NC represents the number of non-converging experiments ($\mathbb{N}_{exp} = 50, E = 2000$)

Weights			von Neumann			Moore			All			Small-World			Scale-Free		
w_q	w_p	w_n	A_1	A_2	NC	A_1	A_2	NC	A_1	A_2	NC	A_1	A_2	NC	A_1	A_2	NC
0.33	0.33	0.33	0	0	50	1	0	49	0	0	50	0	0	50	0	0	50
1	0	0	0	0	50	0	0	50	0	0	50	0	0	50	0	0	50
0	1	0	17	19	14	22	10	18	22	13	15	9	16	25	13	7	30
0	**0**	**1**	**25**	**25**	**0**	**19**	**31**	**0**	**25**	**25**	**0**	**23**	**27**	**0**	**20**	**30**	**0**
0.1	0.45	0.45	20	22	8	27	12	11	25	22	3	18	7	25	24	14	12
0.45	0.1	0.45	0	0	50	0	0	50	0	0	50	0	0	50	0	0	50
0.45	0.45	0.1	0	0	50	0	0	50	0	0	50	0	0	50	0	0	50

Analysis on w_p: Agents that relied on their partners' previous observations as second players (w_p) also had a reasonable chance for convergence. As results from Table 1 indicate, some convergence occurred, but the number of it occurring was less than when criteria three was solely evaluated (i.e. $w_n = 1$). To gain insight to the underlying cause, we examined the effect that increasing the number of episodes being run per experiment would have on the overall number of convergence taking place.

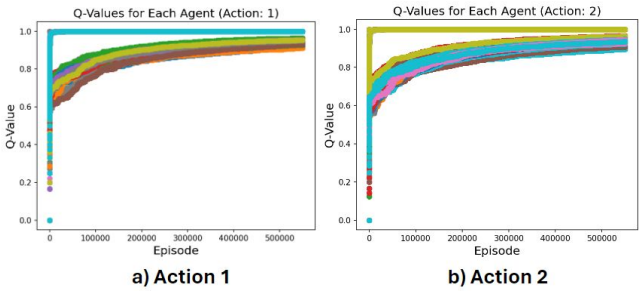
Table 2 reveals that after running 50 experiments of 10,000 episodes, an increase in the number of convergence is experienced in almost all of the neighborhood types. This indicated that the time needed to reach convergence is longer when $w_p = 1$ than when $w_n = 1$.

Table 2. Effect of increasing E to 10,000 per experiment when considering only w_p during action selection ($\mathbb{N}_{exp} = 50$)

Weights			von Neumann			Moore			All			Small-World			Scale-Free		
w_q	w_p	w_n	A1	A2	NC	A1	A2	NC	A1	A2	NC	A1	A2	NC	A1	A2	NC
0	1	0	28	22	0	26	24	0	25	25	0	19	31	0	25	25	0

Analysis on w_q: While w_p and w_n had a pivotal role in driving the convergence process, w_q, as reflected in Table 1, had no discernible effect on the system's dynamics, resulting in no meaningful convergence behavior. With the rewards of action 1 and 2 being equal as long as both agents agree upon the same action, the Q-Value estimates for both actions will be substantially identical over time, thus resulting in the absence of convergence. Figure 4 illustrates by showing how all agents have identical Q-values for both actions, indicating an absence of preference of either actions.

Fig. 4. Each Color Represents the an Agent's Q-Value as the first player per episode for a) Action 1 and b) Action 2 (von Neumann neighborhood: $w_q = 1$, $E = 550000$)

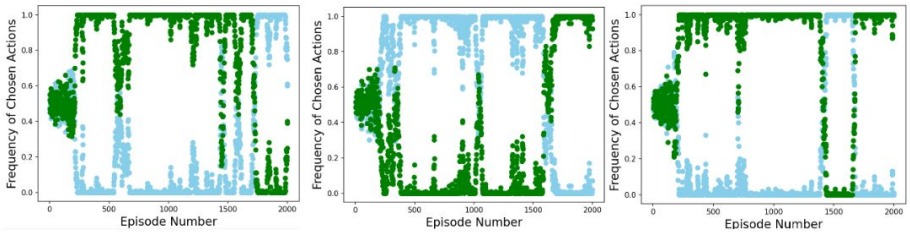


a) Action 1 **b) Action 2**

5.2 Switches in Convergence Patterns

As indicated in the previous portion of Section 5, a convergence was claimed to have taken place when at least 90% of the agents demonstrated alignment in their selections during any given time of the experiment. Upon closer examination of the convergence pattern when $w_n = 1$, an unusual pattern was observed when first player considered actions of its neighbors as first players, which is illustrated in Figure 5. When the first player exclusively took its neighboring first player actions into account, there was an observance of fluctuations in convergence patterns, with the transitions of convergence patterns frequently taking place after 1000 or more episodes.

Fig. 5. Switches in convergence pattern when $w_n = 1$ and $\epsilon_{end} = 0.001$, (x-axis: episode number and y-axis: fraction of agents selecting each action; $E = 2000$)

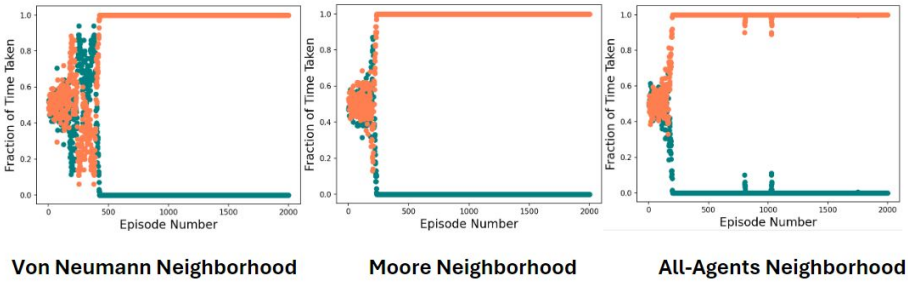


Von Neumann Neighborhood **Moore Neighborhood** **All-Agents Neighborhood**

The results of Figure 5 indicate that despite a low ϵ_{end} value of 0.001, it can have an influence on the convergence pattern, particularly over the course of thousands of episodes. Therefore, a smaller ϵ_{end} reduces the likelihood that convergence-switching occurs. Figure 6 illustrates how a decrease in ϵ_{end} value from 0.001 to 0.00001 significantly decreases the amount of convergence-switching observed within the 2000 episodes. However, it should be noted that it does

not completely eliminate convergence-switching. Rather, it decreases the rate at which the switching occurs. The only way to truly eliminate convergence switching is to set ϵ_{end} to zero.

Fig. 6. Convergence pattern when $w_n = 1$ and $\epsilon_{end} = 0.00001$ ($E = 2000$)



When considering observations made only by the second player (i.e. w_p=1), the convergence-switching pattern can still be observed, but the episodes required to change to the new norm afterwards occurs at a significantly slower rate. Due to the slow switching rate, there are periodic episodes at which convergence among agents is not observed.

5.3 Closer Look Into Convergence Pattern When Only Observation as Second Player is Considered for Decision Making

While the results from Table 2 indicates a slight increase in the number of convergence when the number of episodes run per experiment increases. It does not provide a comprehensive explanation as to why there is a disparity in the convergence rate between the cases when $w_p = 1$ and $w_n = 1$.

When the convergence pattern was compared graphically between the two cases ($w_p = 1$ and $w_n = 1$), the pattern in $w_p = 1$ resembled an elongated version of the pattern in $w_n = 1$. 20,000 episodes were run for the case in which $w_p = 1$ in order to be able to observe a similar pattern that can be observed when $w_n = 1$ with only 2000 episodes. As shown in Figure 7, the complete switching of convergence when $w_p = 1$ takes longer duration, making it highly more likely that no convergence is observed for a period of time throughout the experiment.

To better understand the cause of the slower convergence rate at $w_p = 1$, an assessment was made to investigate whether the number of past actions, k_{obs}, an agent observed of its neighbor had a significant influence on the convergence rate when $w_p = 1$.

Figure 8 illustrates that as the value of k_{obs} increases, the likelihood of observing convergence within a given number of episode decreases. However, at $k_{obs} = 1$, the convergence pattern resembles that of $w_n = 1$ convergence patterns in Figure 5, including the convergence switch patterns.

Fig. 7. Convergence pattern comparison between the cases when $w_p = 1$ (in 20,000 episodes) and $w_n = 1$ (in 2,000 episodes)

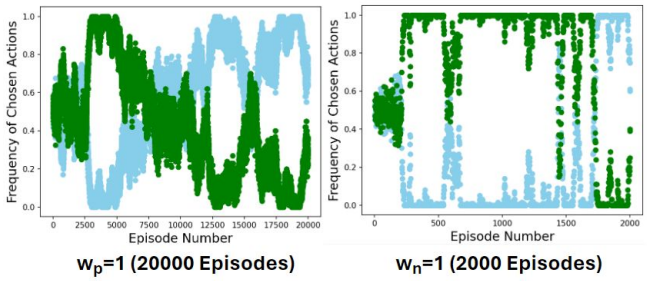
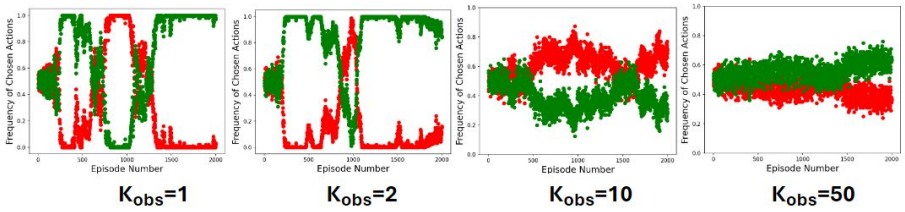


Fig. 8. Effect of varying k_{obs} values in von Neumann neighborhood (x-axis: episode number in increments of 500, and y-axis: fraction of agents selecting each action; $E = 2000$)



Due to the fact that convergence pattern can slightly vary under the same k_{obs} value, more experiments were conducted to confirm that the value of k_{obs} was indeed having a significant impact on the convergence pattern when $w_p = 1$. Figure 9 and Table 3 depict the results of 50 experiments with 2000 episodes being run for each k_{obs} value. As the k_{obs} value increased, the amount of episodes it took for the first convergence to take place likewise increased. This result aligns with the patterns shown in both Figure 7 and Figure 8, where increase in k_{obs} leads to slower convergence movement in $w_p = 1$ compared to that of when $w_n = 1$.

Table 4 demonstrates that with $k_{obs} = 1$, the results have strong resemblance to the number of convergence that occurred when $w_n = 1$ as shown in Table 1. This indicates that considering past actions from earlier nodes lead to increased noise, thereby overfitting to past experiences and delaying convergence. Based on the effect the value of k_{obs} had on the convergence pattern when $w_p = 1$, the noise had a more significant effect on the number of convergence observed when $w_p = 1$ than when $w_n = 1$.

5.4 Effect of Varying the Number of Actions Available

In this section of the study, we will elaborate on the effects of increasing the number of actions available for agents to select on convergence patterns. We

Fig. 9. First episode at which convergence occurs for different k_{obs} values ($\mathbb{N}_{exp} = 50$, $E = 2000$)

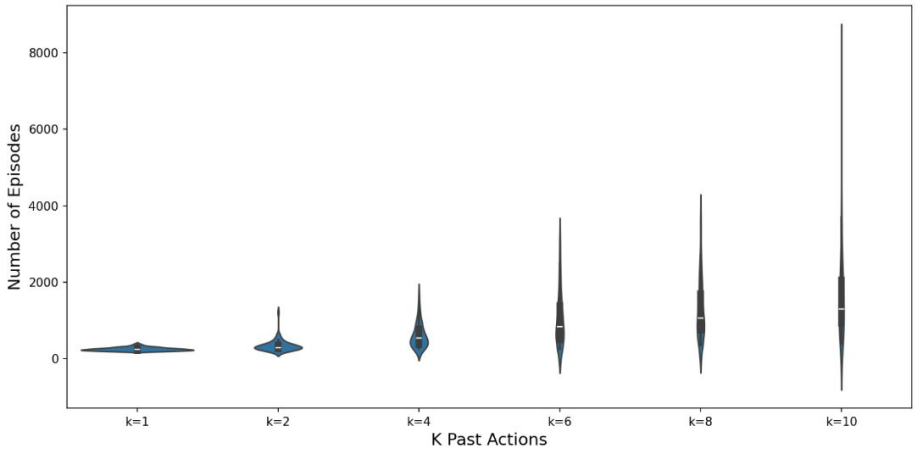


Table 3. Mean episode of first convergence with standard deviation (E = 2000, $\mathbb{N}_{exp} = 50$)

k_{obs}	Mean	Std
1	249.28	44.515
2	342.50	156.999
4	608.78	320.686
8	1284.36	778.200
10	1716.7	1324.780
20	6644.12	4544.923
30	12517.18	10368.045

varied the number of actions up to 100. Based on the depictions in Figure 10, where $w_p = 1$, convergence was observed for all scenarios that were tested. As the number of actions increased, the episode at which the first convergence occurred was delayed to a later point in the experiment.

Similar to when $w_p = 1$, convergence is observed for all scenarios when $w_n = 1$. However, as the number of actions increased, the number of switches in convergence per experiment oftentimes slightly increased. However, as shown in Figure 11, there can be periodic moments in which the convergence switching rate is slower than when there are less actions (e.g. episodes 1000-8500). Nevertheless, after the around episode 8500, the convergence switching rate drastically increases.

It should be noted that whenever finding the max Q-value at the initial state, where all Q-values are zero, the Q-value *arg max* should be chosen randomly rather than selecting the first Q-value it sees. As Figure 12 depicts, if this precaution is not followed in scenarios such as $w_n = 1$, it will result in a

Table 4. Mean episode of first convergence with standard deviation (E=2000, $\mathbf{N}_{exp} = 50$)

Weights			von Neumann			Moore			All			Small-World			Scale-Free		
w_q	w_p	w_n	A1	A2	NC	A1	A2	NC	A1	A2	NC	A1	A2	NC	A1	A2	NC
0	1	0	22	28	0	28	22	0	23	27	0	25	25	0	27	23	0

Fig. 10. Effects of increasing the action space in von Neumann neighborhood (x-axis: Episode number, y-axis: Fraction of agents who have selected each action, $w_p = 1$, $E = 10000$)

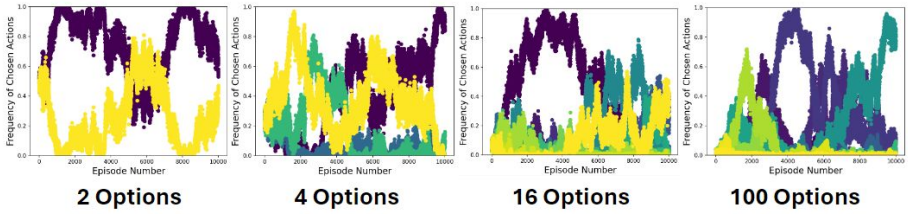
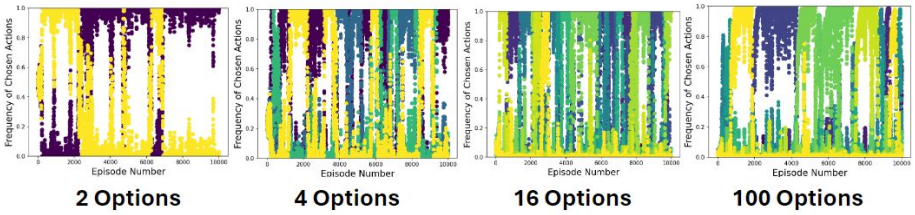


Fig. 11. Effects of increasing the action space in von Neumann neighborhood (x-axis: Episode number, y-axis: Fraction of agents who have selected each action, $w_n = 1$, $E = 5000$)



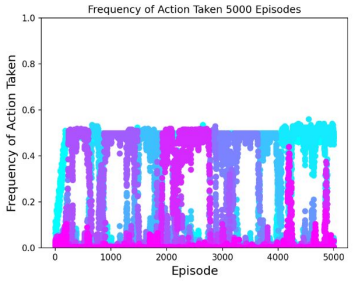
situation where half of the population converges to an action, while the other half converges to other actions. In the scenario for Figure 12, there are 100 actions, where each action number is assigned to the designated agent at the initial episode.

5.5 Effect of Varying the Number of Agents

Additionally, we conducted experiments to explore the impact of changing the number of agents, N, on the overall convergence pattern. Twenty experiments, 2000 episodes each, were conducted for each selected value of N to analyze the mean episode at which the first convergence takes place, μ_{fc}.

As demonstrated in Table 5, as the number of agents increased, μ_{fc} also rose. Along with the linear regression model of $y = 0.622783x + 140.91105$, a strong correlation (r=0.99) between the number of agents and μ_{fc} can be observed.

Fig. 12. Effects of not randomly choosing max Q-value when there are more than one max value (100 actions, $w_n = 1$, E=5000, von Neumann neighborhood)



N	200		300		400		500		600		700	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
	278.86	68.85	306.08	95.35	395.32	168.99	464.50	224.94	494.00	269.605	588.22	298.72

Table 5. Effect of increasing the number of agents on the convergence count in 20 experiments in von Neumann neighborhood ($w_n = 1$, $E = 2000$)

6 Discussions

6.1 Q-Learning Criterion

The lack of norm convergence based on solely Q-learners is explained by the lack of pressure to select a pure single action strategy. Each agent quickly learns the best action as the second player to each first player action, so there is not enough time for a population of agents with suboptimal second actions to influence the strategy of the first move. In addition, while the exploration rate allows agents to learn about both actions for sufficient episodes, after the epsilon value decreases agents will arbitrarily prefer one action to the other. Since this does not depend on the population of other agents, each agent independently follows one action as first player and sticks to it. Continued exploration alongside a lack of pressure to form a norm results in a highly mixed, stable population with no norm.

6.2 Neighborhood Criterion

The norm switching behavior demonstrated in Figure 5 can be traced to the action choice function that the agents use.

At any individual episode of the simulation, each agent A_i takes action a_i and observes a set of other agents which form a neighborhood of adjacent agents $A_j \in n(A_i)$ with count $|n(A_i)| = n_i$. Then each agent in turn chooses to adopt a new action a_i', with proportion $\frac{|\{a_i'=a_j | A_j \in n(A_i)\}|}{n_i}$ This means that the expected number of agents that copy the strategy of agent a_i is $\Sigma_{A_j \in n(A_i)} \frac{1}{n_j}$. When $n_i = n_j$, as in the fully connected case, this expected value is 1.

In addition, in the case of any arbitrary set of connections, the average expected number of agents that will copy the action of an agent is also 1, independent of the number of the actions of other agents.

$$\frac{1}{N}\Sigma_i(\Sigma_{A_j \in n(A_i)}\frac{1}{n_j}) = \frac{1}{N}\Sigma_i n_i \frac{1}{n_i} = \frac{1}{N}N = 1$$

Therefore, if an agent is selected uniformly at random, the expected number of agents influenced by the strategy of that agent at the next step is 1. The actual distribution leads to most agents' influence disappearing after one or two rounds, with a small number of agents influencing the whole population. This is a direct result of using $\frac{|\{a_i'=a_j | A_j \in n(A_i)\}|}{n_i}$ as the probability for selecting an action from a neighborhood.

The approximate rate of convergence to a majority norm can be found by calculating the expected number of rounds until the number of agents whose initial strategy affects the current pool of strategies decreases to 1. For a fully connected pool of 100 agents, 50% of the time this occurs before round 171.

The iterative algorithm solves issues with parity that could otherwise cause a connected agent network without odd cycles to behave like two distinct subnetworks with opposite parities, like a checkerboard.

6.3 Partner History Criterion

Agents using the partner history criterion exhibit a wide range of behaviors depending on the hyper-parameters k_{obs} and ϵ. With very low k_{obs} like 1 and 2, they behave similar to agents using the w_n neighborhood criteria. However, a large k_{obs} has a similar effect to having a higher agent count, resulting in slower convergence. When population behavior change is slow, the norm switching behavior that can be triggered by epsilon-greedy exploration behavior and general stochastic nature of the agent choices can cause extended periods of simulation with no convergence.

6.4 Neighborhoods

While neighborhood shape plays a role in convergence time, ultimately connected agent graphs were influenced more strongly by other hyper-parameters like action criteria and exploration probability.

6.5 Takeaways

In scenarios where selective pressure does not bias one solution over another, learning is not sufficient to form a norm. We show that in some conditions an imitation scheme is sufficient to facilitate norm emergence. The findings from this study should help inform future work in norm emergence for cooperative sequential games, which are a good analogue for many more complex interactions, such as conversation. In these scenarios, understanding how norms emerge

and change can lead to improvements in cooperative agent design. In these scenarios, predicting which convention will emerge is a harder problem for future exploration. Some real-life scenarios are best modeled by agents that follow the norm with a high probability, while other real-life scenarios are best modeled by agents that break from the norm with a high probability. Finally, there are real-life scenarios where alternative norms occasionally take over a population. We show that sampling actions uniformly from other agents within a local or global region results in norm switching behavior when mutations appear within the population.

7 Conclusions

The purpose of this study was to investigate the emergence of the conventions on repeated the sequential coordination games where the first and second players take their actions in order, rather than simultaneously. The agents make their decision of action based on their game experiences, observations, and the first player's action (if the agent is the second player).

We observed that both observation of neighbors and past experiences have a significant effect on the emergence of conventions. Comparing the two, convention emerges faster when the agents consider their observations of neighbors. As expected, q-learning could not achieve emergence at all as the second player follows the first player's choice regardless of the action chosen. The second major finding was that the switches in emerged convention occurs due to the exploration policy. Furthermore, the frequency of the switches is influenced by the criterion considered either observations or past experiences, the window length of observations. Increasing the number of actions and agents had a negative effect on convergence.

Given the fact that occurrence of non-simultaneous actions in an interaction are prevalent in real life scenarios, this research offers valuable insights for understanding the emergence patterns in situations where the actions are taken sequentially. This study has been one of the first attempts to thoroughly examine the stage games in the context of convention emergence. This sequential interaction model must be studied further in other scenarios such as solving social dilemmas.

References

1. Abeywickrama, D.B., Griffiths, N., Xu, Z., Mouzakitis, A.: Emergence of norms in interactions with complex rewards. Autonomous Agents and Multi-Agent Systems **37**(1), 2 (2023)
2. Airiau, S., Sen, S., Villatoro, D.: Emergence of conventions through social learning: Heterogeneous learners in complex networks. Autonomous Agents and Multi-Agent Systems **28**, 779–804 (2014)
3. Bicchieri, C., Muldoon, R., Sontuoso, A.: Social norms (2011)

4. Centola, D., Baronchelli, A.: The spontaneous emergence of conventions: An experimental study of cultural evolution. Proceedings of the National Academy of Sciences **112**(7), 1989–1994 (2015)
5. Duncan Jr., S., Farley, A.M.: Achieving parent-child coordination through convention: Fixed- and variable-sequence conventions. Child Development **61**(3), 742–753 (1990)
6. Franks, H., Griffiths, N., Jhumka, A.: Manipulating convention emergence using influencer agents. Autonomous Agents and Multi-Agent Systems **26**, 315–353 (2013)
7. Hasan, M., Raja, A., Bazzan, A.: Fast convention formation in dynamic networks using topological knowledge. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 29 (2015)
8. Hu, S., Leung, H.f.: Achieving coordination in multi-agent systems by stable local conventions under community networks. In: IJCAI. pp. 4731–4737 (2017)
9. Kittock, J.E.: Emergent conventions and the structure of multi-agent systems. In: Proceedings of the 1993 Santa Fe Institute Complex Systems Summer School. vol. 6, pp. 1–14. Citeseer (1993)
10. Leung, C.w., Turrini, P.: Learning partner selection rules that sustain cooperation in social dilemmas with the option of opting out. In: Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024). AAMAS (2024)
11. Lewis, D.: Convention: A philosophical study. John Wiley & Sons (2008)
12. Mills, G.: The emergence of procedural conventions in dialogue. In: Proceedings of the Annual Meeting of the Cognitive Science Society. vol. 33 (2011)
13. Morris-Martin, A., De Vos, M., Padget, J.: Norm emergence in multiagent systems: a viewpoint paper. Autonomous Agents and Multi-Agent Systems **33**, 706–749 (2019)
14. Sen, O., Sen, S.: Effects of social network topology and options on norm emergence. In: International Workshop on Coordination, Organizations, Institutions, and Norms in Agent Systems. pp. 211–222. Springer (2009)
15. Toffoli, T., Margolus, N.: Cellular automata machines: a new environment for modeling. MIT press (1987)
16. Wang, Y., Lu, W., Hao, J., Wei, J., Leung, H.f.: Efficient convention emergence through decoupled reinforcement social learning with teacher-student mechanism. In: Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems. pp. 795–803 (2018)
17. Wang, Z., Li, R., Jin, X., Ding, H.: Emergence of social norms in metanorms game with high-order interaction topology. IEEE Transactions on Computational Social Systems **10**(3), 1057–1072 (2023). https://doi.org/10.1109/TCSS.2022.3144978
18. Yu, C., Chen, Y., Lv, H., Ren, J., Ge, H., Sun, L.: Neural learning for the emergence of social norms in multiagent systems. In: 2017 IEEE International Conference on Agents (ICA). pp. 40–45 (2017). https://doi.org/10.1109/AGENTS.2017.8015298
19. Yuan, Y., Guo, T., Zhao, P., Jiang, H.: Adherence improves cooperation in sequential social dilemmas. Applied Sciences **12**(16), 8004 (2022)