World Scientific
www.worldscientific.com

EXPLORING THE DIMENSIONS OF CONVENTION EMERGENCE IN MULTIAGENT SYSTEMS

DANIEL VILLATORO*,‡, SANDIP SEN†,§
and JORDI SABATER-MIR*,¶

*Artificial Intelligence Research Institute (IIIA),
Spanish National Research Council (CSIC),
Bellatera, Barcelona, Spain

†Department of Mathematical and Computer Science,
University of Tulsa, Tulsa, Oklahoma, USA
‡dvillatoro@iiia.csic.es
§sandip-sen@utulsa.edu
¶jsabater@iiia.csic.es

Social conventions are useful self-sustaining protocols for groups to coordinate behavior without a centralized entity enforcing coordination. The emergence of such conventions in different multi agent network topologies has been investigated by several researchers, although exploring only specific cases of the convention emergence process. In this work we will provide multi-dimensional analysis of several factors that we believe determines the process of convention emergence, such as: the size of agents memory, the population size and structure, the learning approach taken by agents, the amount of players in the interactions, or the convention search space dimension. Although we will perform an exhaustive study of different network structures, we are concerned that different topologies will affect the emergence in different ways. Therefore, the main research question in this work is comparing and studying effects of different topologies on the emergence of social conventions. While others have investigated memory for learning algorithms, the effects of memory on the reward have not been investigated thoroughly. We propose a reward metric that is derived directly from the history of the interacting agents. Another research question to be answered is what effect does the history based reward function and the learning approach have on convergence time in different topologies. Experimental results show that all the factors analyzed affect differently the convention emergence process, being such information very useful for policy-makers when designing self-regulated systems.

Keywords: Conventions; multiagent systems; social learning; topology.

1. Introduction

As stated in [8], "A normative multiagent system is a multiagent system organized by means of mechanisms to represent, communicate, distribute, detect, create, modify, and enforce norms, and mechanisms to deliberate about norms and detect norm

violation and fulfillment." This definition, even though is very general, points out every aspect that still needs to be treated in the research of normative systems. In the same work, authors differentiate between two types of normative points of view:

(1) The *legalistic view* of normative multiagent systems is a top-down view which considers the normative system as a regulatory instrument to regulate emerging behavior of open systems without enforcing the desired behavior. Agents are often motivated by sanctions to stick to norms, rather than by their sharing of the norms. Even if agents are allowed some freedom to create norms, this freedom is mostly restricted to the possibility for agents to create contracts to regulate the interaction among them. The *legalistic view* of a multiagent system has merited in depth-study in the literature. We can relate this legalistic point of view with systems where norms are predefined by an authority on the system prior to runtime [7, 14, 25].

(2) The *interactionist view* on normative multiagent systems represents a bottom-up view. In this autonomous, individual-centric view, norms can be seen as regularities of behavior which emerge without any enforcement system as agents conform to them either because their goals happen to be aligned, or because they feel themselves as part of the group or because they share the values of other agents. Sanctions, or formal measures towards norm violating agents carried out by agents whose task is to sanction norm violations, are not always necessary because social blame and spontaneous exclusion of non-conforming agents are often sufficient to incentivize conformity to norms.

For our understanding, the research on the interactionist view of norms is divided into two different stages: (a) how norms appear in the mind of one or several individuals and (b) how these new norms are spread over the society until they become accepted social norms. Both processes are complex and have to be carefully analyzed. We are interested in studying the second stage: the spreading and acceptance of social norms, defined as *norm support* [5]. Our understanding of norm support deals with the problem of which norm is established as the dominant when more than one norm exists for the same situation. Specifically, this is the case of study in the convention emergence literature [10, 16, 19, 24, 26–29]. Conventions are a special type of norms, used as a mechanism for sustaining social order, increasing the predictability of behavior in the society and specify the details of those unwritten laws. Following Coleman's theory [11], conventions emerge to solve coordination problems, where there exist no conflict between the individual and the collective interests, as what is desired is that everyone behaves in the same way, without any major difference on which action agents are coordinated. Therefore, the selection of the focal action in such norms is arbitrary. One clear example of these kind of norms is the selection of which side of the road to drive on, both options are equally good as long as all drivers agree. Examples of conventions pertinent to

MAS would be the selection of a coordination protocol, communication language, or (in a multitask scenario) the selection of the problem to be solved.

Such norms are conflict resolution strategies that develop from the population interactions instead of a centralized entity dictating agent protocol. Moreover, history of interaction is then instrumental for norm evolution. Learning algorithms incorporate the history of interaction into their decisions, but reward metrics are typically static and independent of the agent histories. Norm evolution is dependent upon the exertion of social pressure by the group on aberrant individuals. It is through learning via repeated interactions that social pressure is applied to individuals in the group. However, a reward metric based on the current interaction does not necessarily model the full context or capture the persistent nature of social pressure in human societies. In particular, society often uses past history to judge individuals and hence actions have future consequences in addition to immediate effects. Accordingly, we propose a reward structure based upon the agent's interaction history as a more appropriate alternative to the single interaction reward metric normally used in the literature. In our model agents are rewarded based upon the conformity of action between two agents, such that the agent who has the larger of the majority action choice in the stored interaction history of the agents receives higher reward. Hence, both interaction agents' history of actions are used to calculate each individuals' payoff from an interaction. We investigate how this history, and in particular, its size (memory size) affects different types of society structure.

We are also keenly interested in understanding how agent relationships and social connections affect the success and rate of adoption of social norms. As confirmed in the literature [16], the social topology that restricts agent interactions plays a crucial role on any emergent phenomena resulting from the interactions amongst agents. In this paper, we will experiment on different types of topologies in order to observe, compare and analyze their effects and dynamics of reaching social conventions, proposing a new type of topological structure named *Fully Connected Stars Network*. Concerning the topological effects of the network of agents on the process of convention emergence, this work highlights the formation and stability of metastable subconventions. Our approach to studying norm emergence in agent societies assumes the *social learning* [19, 24] framework, where agents learn their action preferences from repeated chance interactions, based on the topology of the agent network, with different members in the population. We observe that agent groups in such societies can develop subconventions depending on their position in the interaction topology. Subconventions are conventions adopted by a subset of agents in a social network who have converged to a different convention than the majority of the population. Metastable subconventions may adversely affect the speed of the emergence of more general conventions. We identify factors that promote the emergence and maintenance of subconventions.

Additionally, we explore scale-up properties of the convention emergence process for the following features:

- Action space: We also explore the effects of scale up with different numbers of actions in the conventions search space,
- Multiplayer interactions: We analyze the effect of the number of players involved in an interaction.

The structure of this article is as follows: we review the previous and related work in the area of emergence of social conventions in multiagent systems in Sec. 2; in Sec. 3 we introduce the agent interaction and reward model that we have used; experimental results are presented in Sec. 4; conclusions from the analysis of the results are presented in Sec. 6, and finally we present the future work plans in Sec. 7.

2. Previous Work

One of the most influencing authors in the research area of social norms would be Axelrod. In his seminal work [5] Axelrod analyzes the factor that influence the emergence and spread of norms within societies. Other than an important theoretical and philosophical contribution, the author present an experimental section where it analyses with an evolutionary model the effects of the metanorm mechanism in the establishment of norms. However, neither topological effects or online social learning was considered in its simulation framework.

The precursors on the research of convention emergence in MAS were Shoham and Tennenholtz. In their seminal work [26], they introduce the computational definition of social convention in a game-theoretical framework (not far from the one given by Coleman), accompanied with a formal specification of it. As an application of their theoretical framework they approach the Coordination Game, and the emergence of the social convention of agreement by using the Highest Cumulative Reward (HCR) rule. Agents using this reward rule will choose the action that has provided them with the highest reward in the last t timesteps. They perform experiments to check the right value of t to ensure efficient emergence. Despite their important contribution on the settlement of this research area on multiagent systems, neither these authors (like Axelrod) leave unexplored the relationship of the topological conditions on the social learning of conventions.

Up to our knowledge, Kittock's research [16] is the first one considering the restrictive topology of interaction on the diffusion of social conventions. This research is continued by Delgado *et al.* [13]. Interesting results where obtained by these authors when using strategy update rules like *Highest Cumulative Reward*.

Social network analysis has been an important topic of research in the problems related with social learning and information diffusion. Even though the problem we study in this work (emergence of social conventions) might seem simpler than other applications [1, 15, 17], it models the dynamics of a number of pertinent problems in MAS.

Sen and Airiau [20, 24] explored norm emergence where interaction rewards were not dependent on previous interactions. That work is focused on the problem of coordination of two cars arriving at an intersection. Each agent can choose to "go" or "yield" to the other agent. The reward metric is designed so that if each agent chooses the same action, they receive small payoff but if agents choose opposite actions, they receive a large payoff. So if the row and the column agents both "go" they both receive a poor payoff, but only one player choosing to go will yield a relatively high payoff for both. Depending on which player chooses to yield, two possible effective social norms can be established. The history of interaction does not directly affect the reward agents receive. Reward is only affected by the agents' action choice in the current interaction. However, learning takes place via social pressure from repeated interaction, thus the history of interaction indirectly influences agent's action choice.

Most of these research [12, 13, 16, 26] have studied rather simplified versions of the emergence of conventions, e.g. using a convergence threshold of 90%, restricted case of only two possible conventions. For our understanding, these two factors need to be relaxed and observe the different effects of reaching total convergence (100%) and with larger sets of possible conventions than 2 (as done in [23]), as well as a more exhaustive study on the different topologies.

Moreover, other authors in MAS have studied different aspects of normative systems and architectures for self-organization ([3, 9, 16, 18, 22]), however, for the scope of this work we will not need such elegant but complex agent architectures.

3. Model

The social learning situation for norm emergence that we are interested in is that of learning to reach a social convention. We adopt the following definition of a social convention from the definition of a social law [26]: *A social law is a restriction on the set of actions available to agents.* A social law that restricts agents' behavior to one particular action is called a social convention.

For the sake of generalization, our framework is built with the most accepted convention emergence model (used by [13, 16, 19, 24, 26, 29]). We represent the interaction between two agents as an n-person m-action game. At each time step, each agent is paired with another agent from the population (society) and chooses from one of several alternatives (the choice can either be that of an action to execute, which we use in this paper, or a particular state to be in). In our case, as in the case in [13], a social convention will be reached if all the n agents are in the same state or choose the same action, i.e. the actual state or action chosen as a convention is not important. For our purpose, an agent choosing a particular action is equivalent to it being in a corresponding state.

We model agent environments by networks, where each agent is represented by a node and the links in the network represent the possibility of interaction between nodes (or agents). We consider the following three different agent network
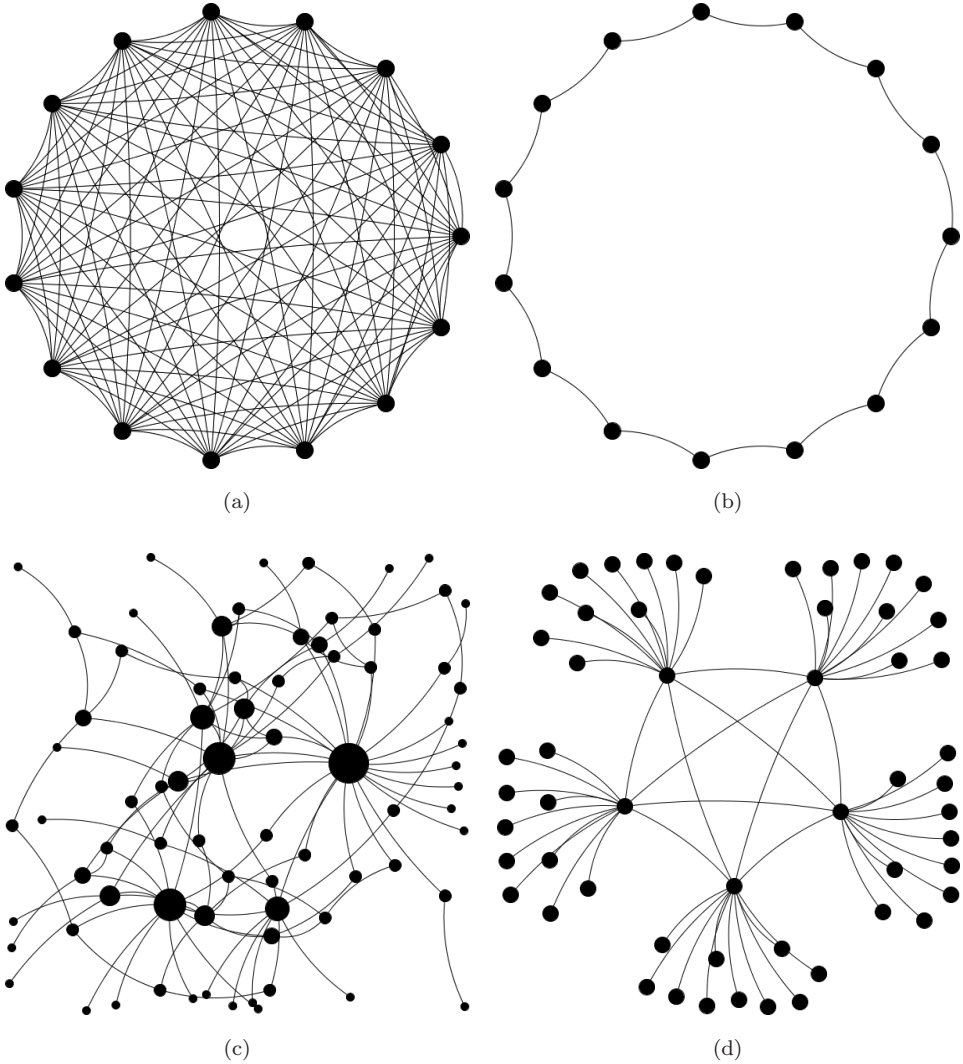
(a)

(b)

(c)

(d)

Fig. 1.   Underlying topologies. (a) Fully connected network. (b) Ring network or one-dimensional lattice with neighborhood size 2. (c) Scale-free network. (d) Fully-connected stars network.

topologies or environment types: (i) a *one-dimensional lattice* with different neighborhood sizes, where an agent is connected to all neighbors (examples can be seen in Figs. 1(a) and (b)); (ii) a *scale-free network*, whose node degree distribution asymptotically follows a power law (an example can be seen in Fig. 1(c)), representing the topology of social networks [6, 21]; (iii) to further our understanding of the norm emergence process, and to capture some typical real-world scenarios, e.g. a community of closely knit researchers and their students, we use a rather novel network topology, namely the *fully connected stars network*: such a network has a

relatively small number of hubs or core nodes which are fully connected forming a clique, and each of these core nodes is also connected with a number of leaf nodes (an example can be seen in Fig. 1(d)).

The *one-dimensional lattice* provides a structure in which agents are connected with their n nearest neighbors. Different values of the neighborhood size (n) produce different network structures. For example, when $n = 2$ the network will have a ring structure (as in Fig. 1(b)) and agents will only be connected with their immediate neighbors on either side, corresponding to a ring topology. On the other hand, when $n = PopulationSize$, the connections result in a fully connected network (as in Fig. 1(a)) where each agent is connected with all other agents. The motivation for the usage of this type of networks is to simulate the ordered structure that many computational systems (memory clusters, sensor networks, etc.) are organized.

On the other hand, in the *scale-free network* there are many vertices with small degrees and only few vertices with large degrees. This makes the network diameter (average minimum distance between pairs of nodes) significantly smaller than *one-dimensional lattices* with small neighborhood sizes. Scale-free networks do represent the topology of real networks [2], and we find valuable to have results at least in synthetic scale-free networks.

As in [16], we use agents with a memory M_k of size M (same memory size for all the agents). For agent k, the memory M_k will record information on the history of its decisions: The value of the position i of the memory M_k will be a tuple $\langle a_k^i, t^i \rangle$ where t^i is the time the ith memory event took place, and a_k^i is the decision taken by agent k at time t^i ($1 \leq i \leq M$). Thus, the memory of each agent will work as a record of the history for the last *memory size* actions taken by the agent. Note, that the memory is quite rudimentary and does not even record the identity of the other agent(s) in an interaction. This is a necessary ingredient for social learning, where agents learn policies to play against arbitrary member(s) of the population, and not targeted towards particular individual(s).

Agents cannot observe the other agent's memory, current decision, or immediate reward, and hence cannot calculate the payoff for any action before actually interacting with the opponent. When two agents interact, the instantaneous reward that an agent receives is calculated based on the action it selected and the action history of both agents as shown in Algorithm 1 (the algorithm calculates reward for agent 1 and assumes only two actions available per agent, but can be readily extended to an arbitrary number of actions). Here A_x and B_x are the number of A and B actions in memory that agent x has taken, $Action_x$ is the last action taken by agent x, and for which it is rewarded, $MajorityAction$ is selected to be whichever action is selected more frequently by the two players combined, $MajorityActions_x$ is the number of actions equal to the majority action that agent x has previously taken, and $TotalMajorityActions$ is the number of times the majority action was chosen by both players in their finite histories.

Agents use a learning algorithm to estimate the worth of each action. Agents will choose their action in each interaction in a semi-deterministic fashion. A certain

percentage of the decisions will be chosen randomly, representing the exploration of the agent, and for the rest of the decisions, the agents deterministically choose the action estimated to be of higher utility. In all the experiments presented in this article, the exploration rate has been fixed at 25%, i.e. one-fourth of the actions are chosen randomly.

The learning algorithm used here is a simplified version of the Q-Learning algorithm [30]. The Q-update function for estimating the utility of an action is:

$$Q^t(a) \leftarrow (1 - \alpha) \times Q^{t-1}(a) + \alpha \times reward, \tag{1}$$

where *reward* is the payoff received from the current interaction and $Q^t(a)$ is the utility estimate of action a after selecting it t times. When agents decide not to explore, they will choose the action with the highest Q value. The reward used in the learning process is a proportional reward of that calculated by Algorithm 1.

The simulation process for repeated interactions in the agent society is presented in Algorithm 2.

We have used two different learning modalities: (a) in the *Multi learning* approach both interacting agents use the payoff to update their memory and action estimate, (b) in the *Mono learning* approach, however, only the first agent selected,

```
// First, we select the majority action
TotalAActions = A₁ + A₂;
TotalBActions = B₁ + B₂;
if TotalAActions < TotalBActions then
|   MajorityAction = B;
if TotalBActions < TotalAActions then
|   MajorityAction = A;
if TotalAActions == TotalBActions then
|   MajorityAction = Randomlyselectedbetween AorB;
// Then, we calculate the reward depending on the agents action selection and
    on the majority action
if Action₁ == MajorityAction then
|   reward₁ = MajorityActions₁ / TotalMajorityActions
else
|   reward₁ = 0
end
```

Algorithm 1: Memory based reward function.

```
for timesteps do
    forall agents do
        Select another partner agent from population;
        Selected agents choose an action;
        The joint action from the selected agents and their history determines payoffs;
        Selected agent(s) use payoff received to update action estimates;
    end
end
```

Algorithm 2: Simulation process.

and not the second one, updates its memory and action estimate after an interaction. Each agent interacts exactly once per time step in mono-learning, whereas in the multi-learning mode, different agents interact different times in the same time step because of random partner selection.

4. Experiments

To evaluate the rate and success of norm emergence we ran experiments with different societal configurations by varying the following system and agent properties:

Memory Size: To be able to analyze the effects of memory sizes, we vary the number of past interactions stored by an agent.

Population Size: We study the effects of scale-up by varying population size.

Neighborhood Size: We study how different neighborhood sizes in a one-dimensional lattice affect the process of emergence of conventions.

Underlying Topology: We observe the dynamics of the process of emergence of conventions depending on the underlying network topology.

Learning Modalities: We compare how conventions are reached with different learning modalities, namely, one or both agents learning from an interaction.

Number of Players: We analyze the effect of different number of players participating in each interaction on the speed of convention emergence.

Actions Set: We study the effect of the search space, namely, the number of options available to agents, on the convention emergence process.

Results reported here have been averaged over 25 runs. Agents are initialized with uniformly random memories, and initially are unbiased in their action choice. We conclude that a social convention has been reached when 100% of the population choose the same action. Other researchers have used a convergence threshold of 90% [13, 16]. However we have observed that with certain reward functions and on certain topologies, even after 90% of the society has converged to a convention, it can still switch back to a different convention. Hence, a threshold of 90% can produce misleading or inaccurate results.

4.1. *Effect of neighborhood size*

To observe the effect of neighborhood size, we use a one-dimensional lattice (as scale-free networks and fully connected stars predetermine the neighbors for each node). In these experiments, we have used a memory size of 5. Figure 2(a) shows a comparison of convergence times for different neighborhood sizes, measured as percentages of the population size, in a multi-learning approach.

We can see that when increasing the neighborhood size, the convergence time is steadily reduced until it stabilizes after a certain neighborhood size. This effect is due to the topology of the network. When the one-dimensional lattice has a small
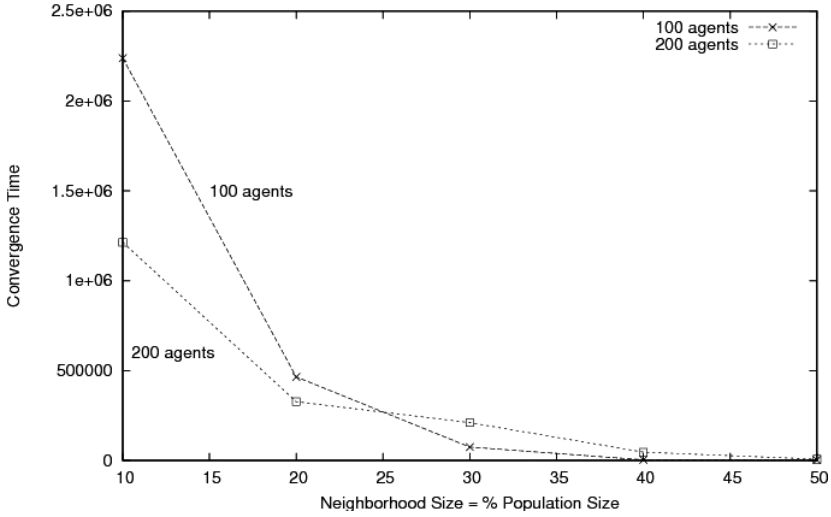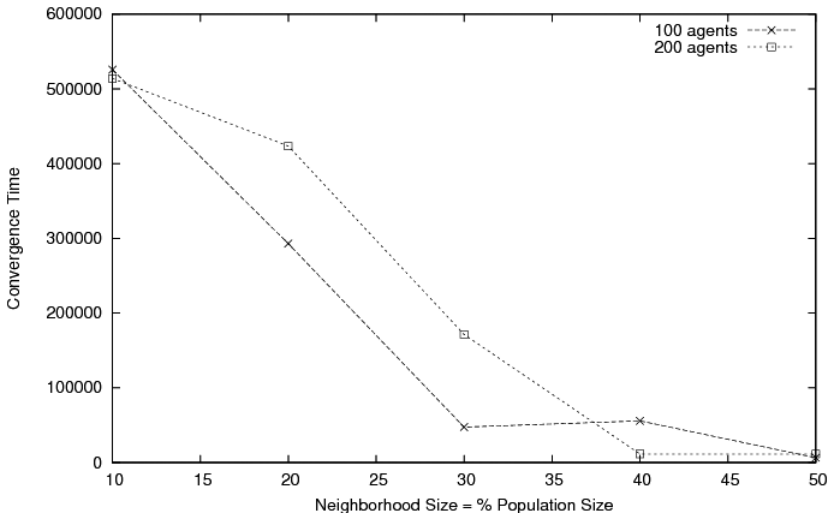
Fig. 2.   Convergence rates with different neighborhood sizes in a one-dimensional lattice. (a) Multi learning. (b) Mono learning.

neighborhood size, on average, the diameter of the graph[a] is high and therefore agents have a relatively higher amount of local interactions. These local interactions might promote different conventions in different parts of the network due to the inherent structure of the convention emergence problem: one convention from the

---

[a]The diameter of a graph is the largest number of vertices which must be traversed in order to travel from one vertex to another.

possible ones has to be adopted by interacting with your neighbors, and as the topology restricts agents to more immediate neighbors (promoting endogamy), they can develop a different convention than agents in another part of the network. These subconventions need to be broken within a certain number of timesteps to achieve the global convention. In the case of larger neighborhood sizes, these endogamy has a softer effect, and therefore the topology does not promote such strong metastable subconventions. It is also interesting to note that for smaller neighborhoods, larger populations exhibits much faster convergence.

Similar convergence results are also obtained with the mono learning approach shown in Fig. 2(b). When the neighborhood size crosses about 30% of the population size, the convergence time does not significantly decrease anymore. The relation of the neighborhood size and the diameter follows a geometric distribution and is shown in Fig. 3. We see that when neighborhood sizes cross 30% of the population size the diameter of the network is no longer significantly reduced, and hence the convergence times are also not significantly reduced any further.

4.2. *Effect of memory size*

In this set of experiments, we want to observe the effect of different memory sizes on convention emergence for different network topologies. We fix the population size at 100 agents. For the one-dimensional lattice, we use a fully connected network. We present the convergence times for different memory sizes in Fig. 4.

The results show that larger memory sizes increase convergence times for the scale-free and fully connected networks. This phenomenon is due to the configuration of the reward function and the learning algorithm. Each action in memory gets
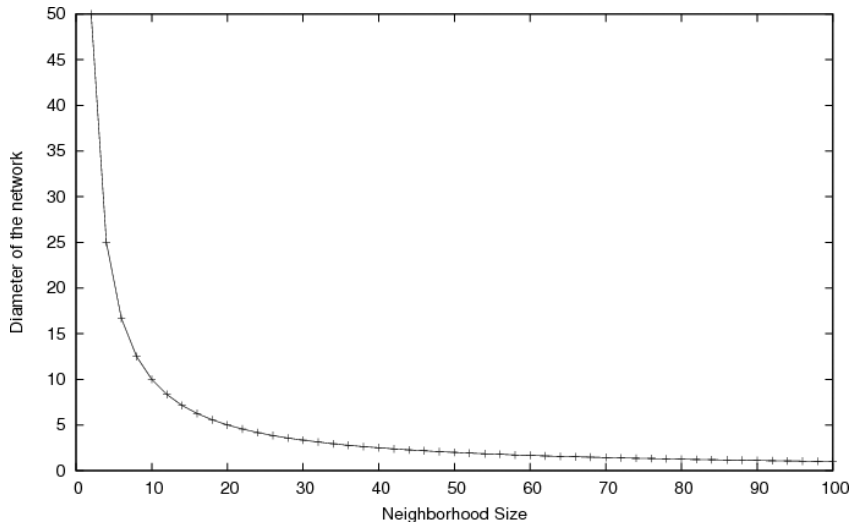

Fig. 3. Diameter relation with neighborhood size in a one-dimensional lattice with population = 100.
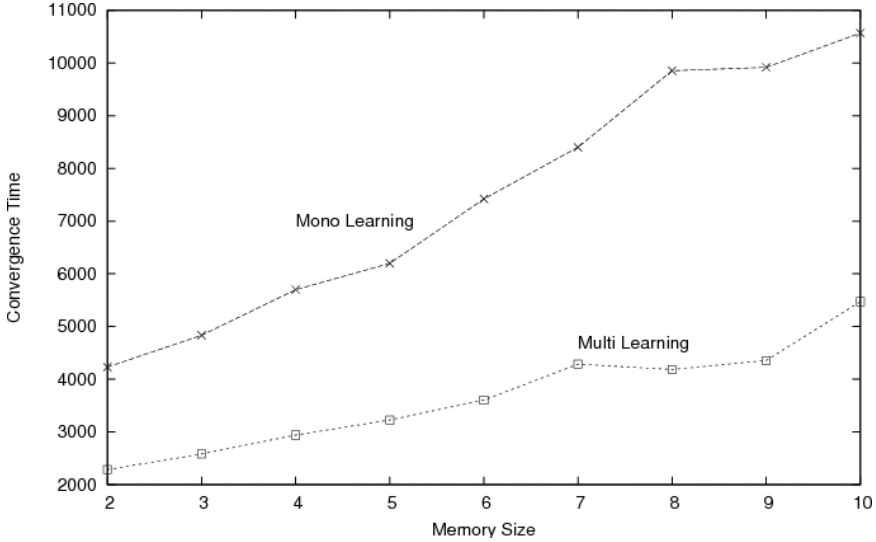
Fig. 4.   Effect of memory size in convergence time on a fully connected network (100 agents).

a relatively high reward for smaller compared to larger memory sizes (refer to the reward function defined in Algorithm 1). The learning algorithm, therefore, receives larger reinforcements for the actions performed for smaller memory sizes, resulting in faster convergence. Convergence is accelerated in this situation because higher rewards have a larger impact on the Q value updated by the learning Algorithm 1. On the other hand, when dealing with higher memory windows, the proportional reward is much smaller, and therefore, the reinforcement will be smaller. Due to this smaller reinforcement, a higher number of interactions, and hence higher number of timesteps, will be needed to reinforce that action to same degree, thereby increasing convergence time.

We also note from Fig. 4, that the mono-learning approach takes longer to converge than the multi-learning approach. A part of this difference is explained by the fact that the average number of learning interactions in a multi-learning approach is twice that of the mono-learning approach for the same number of time steps. There is, however, an additional clear trend of accelerated learning when both agents are learning from the same interaction.

In Fig. 5 (note that the y-axis is in a logarithmic scale), we can observe the relative performance of different topologies for different memory sizes with the mono-learning approach. During this experiment, we limited the execution of the simulations to one million timesteps. We observe that the Fully Connected Stars network takes the most time to converge, followed by the Scale-Free network. For both the Scale Free and the Fully Connected Networks we can observe that the convergence time increases with increasing memory size. These inefficiencies are largely due to more time taken to break or resolve conflicting subconventions
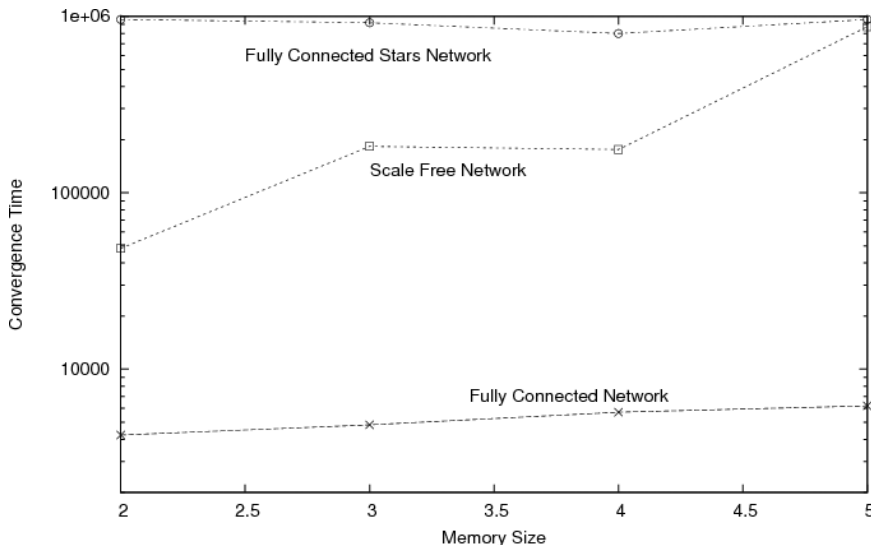
Fig. 5. Topologies comparison with different memory sizes with mono learning approach (100 agents).

that form with scale-free and fully connected stars networks but not for fully connected networks (see following sections for an explanation). Accordingly, the fully connected network, scales up much better with increasing memory size.

4.3. *Effect of learning modalities*

In this set of experiments, we want to observe the difference in convergence times with the two learning modalities for different topologies. We first compare results of the two learning modalities in a one-dimensional lattice with 100 agents (see Fig. 6, where the y-axis is drawn on a logarithmic scale). For smaller neighborhood sizes, i.e. when the network diameter is high, multi-learning takes longer to converge than mono-learning. After reaching the point where the diameter is no longer affected by the neighborhood size (as discussed before, this happens when the neighborhood size is about 30% of the population size), the multi-learning performs better. The reason for this interesting phenomenon is the creation of local subconventions with multi-learning when the neighborhood size is small.

When agents have a small neighborhood size, they will interact often with their neighbors, resulting in diverse subconventions forming at different regions of the network. With the multi-learning approach, agents reinforce each other in each interaction. Such divergent subconventions conflict in overlapping regions. To resolve these conflicts, relatively more interactions between the agents in the overlap area between regions adopting conflicting subconventions is necessary. Unfortunately, agents in the overlapping regions may have more connections in their own subconvention region and hence will be reinforced more often by their subconventions, which makes
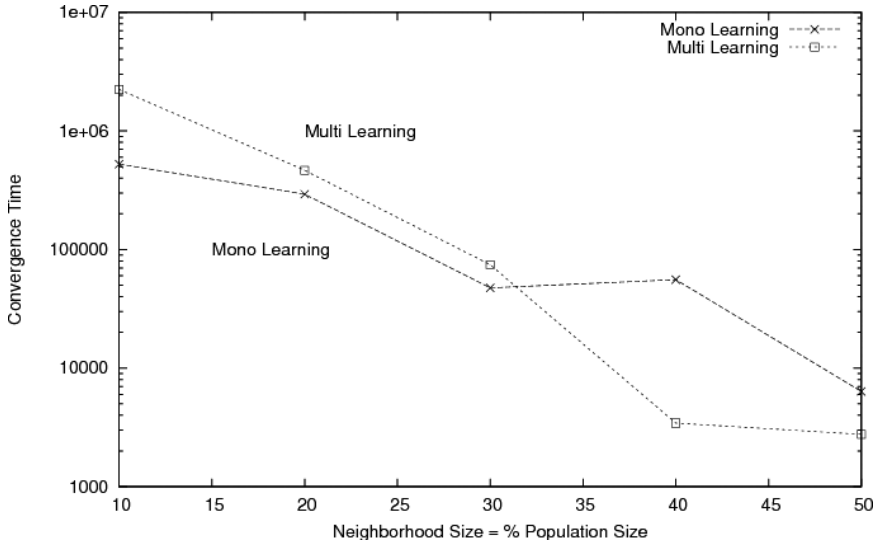
Fig. 6.   Different learning approaches in one-dimensional lattices with different neighborhood sizes.

it harder to break subconventions and arrive at a consistent, uniform convention over the entire society. In the case of the mono-learning approach, the agents in the overlapping region will not be disproportionately reinforced by the other agents sharing its subconvention, making it easier to break those subconventions.

On the other hand, when neighborhood sizes are large, and hence network diameters are small, agents interact with a larger portion of the population. This makes it more difficult to create or sustain subconventions. In addition, this large neighborhood size is more effectively utilized by the multi-learning as agents will be learning from all the interactions they are involved in, and not only from the interactions initiated by them.

For the scale-free and fully-connected stars, systematic variation of neighborhood size is not possible in general. We do observe an interesting phenomenon for these kind of networks. When the multi-learning approach is used in scale-free networks and fully connected stars, subconventions are persistent and the entire population does not converge to a single convention! This is the first time in all of our research on norm emergence that we observed the coexistence of stable subconventions.

The explanation of this rather interesting phenomena can be found in the combination of the memory-based reward function and the inherent topologies of such networks. We present, in Fig. 7, a portion of a representative Scale Free or Fully Connected Stars network where subconventions have formed. We see that agent 1 (hub node 1) and its connected leave nodes (nodes 10, 11, and 12) have converged to one subconvention (represented by the color of the nodes) that is different from the subconvention reached by agent 2 (hub node 2) and its connected leave nodes
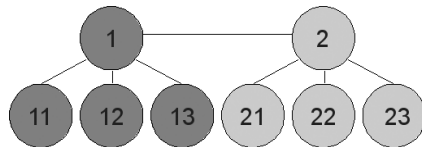
Fig. 7.   Subnetwork topology resistent to subconventions in a multi-learning approach.

(nodes 21, 22, and 23). As an agent has equal probability of interaction with any of its neighbors, both agents 1 and 2 interact more frequently with their associated leave nodes that share their subconvention.

Also note that when two agents interact and both actions have been used equally often in their combined memories, as will be the case when agents 1 and 2 interact, the majority action will be selected randomly, giving advantage to one of the agents. For the subconventions to be broken in this scenario, it is needed that for one of the hub nodes the following holds true: (1) the agent's q-value for its preferred action decreases, and (2) the q-value for the action preferred by the other agent increases. In order for *the agent's q-value for its preferred action to decrease*, a number of repeated interactions (proportional to the memory size) between the hub nodes (in our examples 1 and 2) have to occur, and as there will be no clear majority action, the preference has to be given to the same action, e.g. that preferred by agent 2, in all those interactions. As the reward for the agent 1's action will then be 0, its q-value will start decreasing. In order for the agent 1's q-value for its non-preferred action to increase, a number of interactions (also proportional to the memory size) between it and agent 2 has to occur and agent 1 has to explore in that interaction and try agent 2's preferred action. This will result in agent 1's estimate of agent 2's preferred action to increase, albeit slowly. Only when both these fortuitous events follow each other, and without the intervention of another interaction with the leaves associated with agent 1 (which would reinforce the subconvention), can the subconvention be ultimately broken. The likelihood of these sequence of events happening is exceedingly small and hence subconventions routinely arise with the multi-learning approach. Viewed another way, the leaf nodes can only interact with their hubs and each of them will reinforce the subconvention action for their associated hub node in every time step, making it very unlikely that conflicting subconventions will be resolved in situations such as in Fig. 7.

On the other hand, as an agent is reinforced only once each time-step in the mono-learning approach, the processes required to break the subconventions are more likely, even though it sill has a relatively small probability. This probability decreases with larger memory sizes, and hence subconventions are more likely to emerge with larger memory sizes when using mono-learning. Therefore, with larger memory sizes, subconventions will be harder to break and this phenomenon caused the significant increase in convergence time for scale-free networks and full-connected star networks (we discussed this in the previous section with reference to results displayed in Fig. 5).

4.4. *Weighted reward*

To facilitate the reconciliation of subconventions produced by the topological structure, we decide to investigate a reasonable modification of the reward function. The current reward function only takes into account the previous actions chosen by both agents. In particular, the identity, or more specifically, the social position of the interacting agents did not influence the rewards calculated. We can, however, easily imagine scenarios where the position or social status of an agent can influence the payoff calculation. A straightforward way to incorporate social status in reward calculation would be to use a multiplicative weight, depending on the degree of the interacting node,[b] in Algorithm 1 presented in Sec. 3. As a result, interactions with central, better connected nodes will produce higher rewards than those with relatively isolated nodes on the fringe of the network. By using this weighted reward we are allowing the hub agents to have a larger influence on other agents. The new reward function is shown in Algorithm 3. Here A_x and B_x are the number of A and B actions in memory that agent x has taken, $Action_x$ is the last action taken by agent x, $MajorityActions_x$ is the number of actions equal to the majority action that agent x has previously taken, $TotalMajorityActions$ is the number of actions of the majority action, and $Degree_x$ is the degree of agent x in the network. Note that this modified reward function will not produce different results for the one-dimensional lattice networks, as all nodes in such networks have the same degree and hence will have the same multiplicative factor in the reward function.

The results for the Scale Free network using the new Weighted Reward function and for the two learning modalities are shown in Fig. 8. When we compare the results with this new Weighted reward function with those with the unweighted (uniform) reward function for Scale-Free networks with multi-learning (see Fig. 9), we observe that the weighted reward function results in faster convergence.

```
// First, we select the majority action
TotalAActions = A₁ + A₂;
TotalBActions = B₁ + B₂;
Weight = Degree₂;
if TotalAActions < TotalBActions then
   │   MajorityAction = B;
if TotalBActions < TotalAActions then
   │   MajorityAction = A;
if TotalAActions == TotalBActions then
   │   MajorityAction = RandomlyselectedbetweenAorB;
// Then, we calculate the reward depending on the agents action selection and on the
   majority action
if Action₁ == MajorityAction then
   │   reward₁ = Weight × (MajorityActions₁ / TotalMajorityActions);
else
   │   reward₁ = 0;
end
```

Algorithm 3: Memory and social position based reward function.

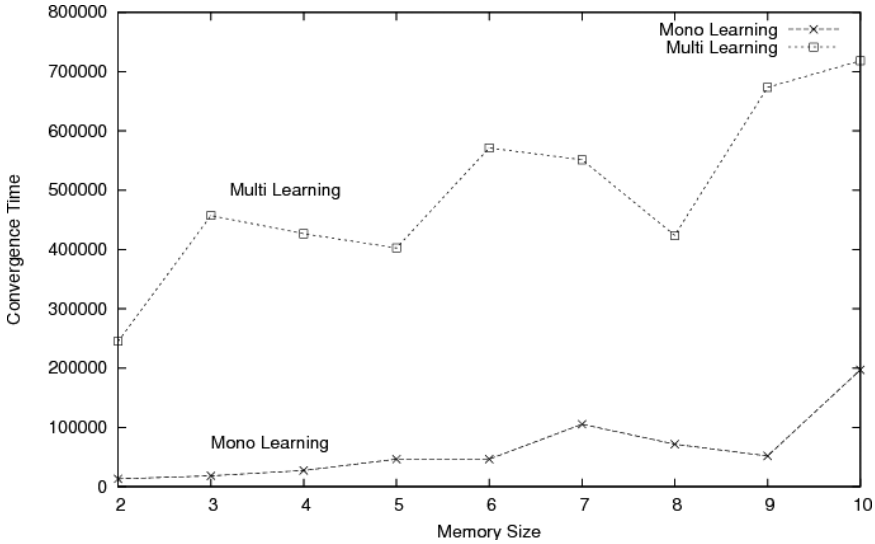[b]The degree of a vertex in a graph is number of edges connected to that vertex.

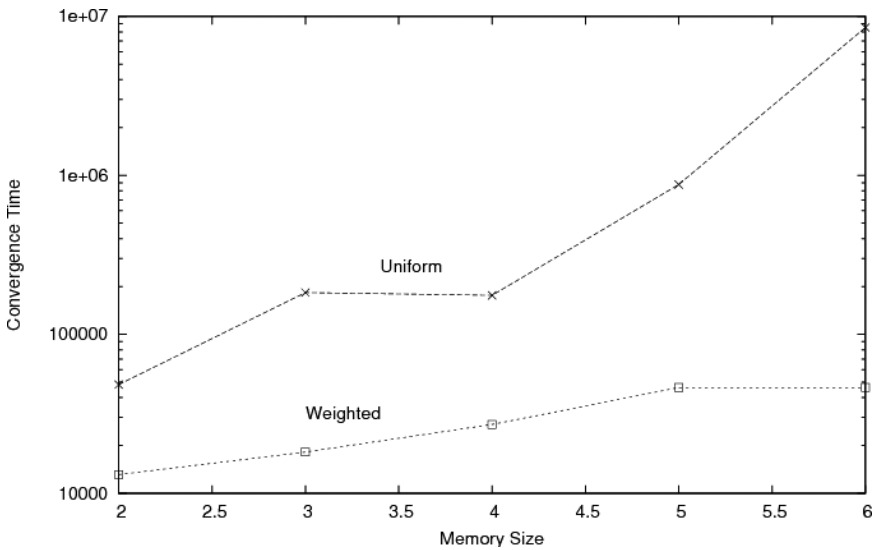Fig. 8. Learning approached comparison in scale free network with weighted reward function.

Fig. 9. Reward functions comparison in scale free network with multi-learning.

The main reason for this is that the weighted reward function allow the hub nodes with more connections to have a larger influence on other nodes and allows them to resolve subconventions, and thereby producing faster convergence.

For the Fully Connected Stars networks, we observe that the Weighted Reward function produces faster convergence when using the mono-learning approach (see Fig. 10). However, subconventions continue to persist with multi-learning
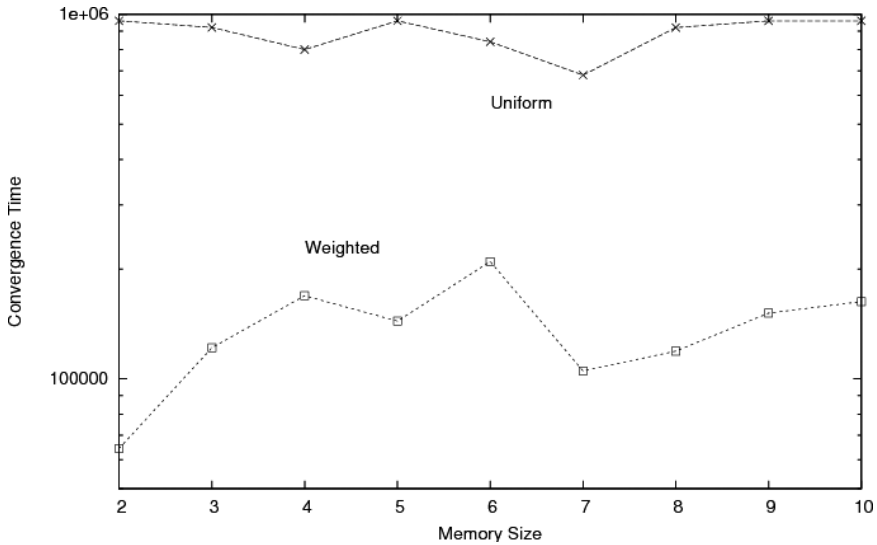
Fig. 10.   Reward functions comparison in fully connected stars network with mono-learning.

approaches. The reason for this effect is due to the uniform degree distribution of the hub nodes in the network and the design of the reward function. The Fully Connected Stars networks engender a three phase convention emergence process: (1) first the leaf nodes drive the hubs, then (2) the hubs have to coordinate, and (3) finally the leaf nodes will have to coordinate with their hub. The second of these phases still takes significant time and is fragile, as explained in Sec. 4.3.

5. Effect of Number of Players

To investigate the effect of the number of players per interaction on the convention emergence process we designed the following experiment set. Agents were situated in a fully connected network. As the actual design of the does not consider situations with more than two players. Also, we modify the simple Memory Based Reward Function (shown in Algorithm 1) to produce a new reward function (shown in Algorithm 4) that reinforces the convention taken by the majority of the players.

```
// First, we select the majority action
Choose MajorityAction as the most frequent action over memories of all interacting players;
In case of tie, select randomly amongst the possible ones;
// Then, we calculate the reward depending on the agents action selection and on the
    majority action
if Action_1 == MajorityAction then
|   reward_1 = 1
else
|   reward_1 = 0
end
```

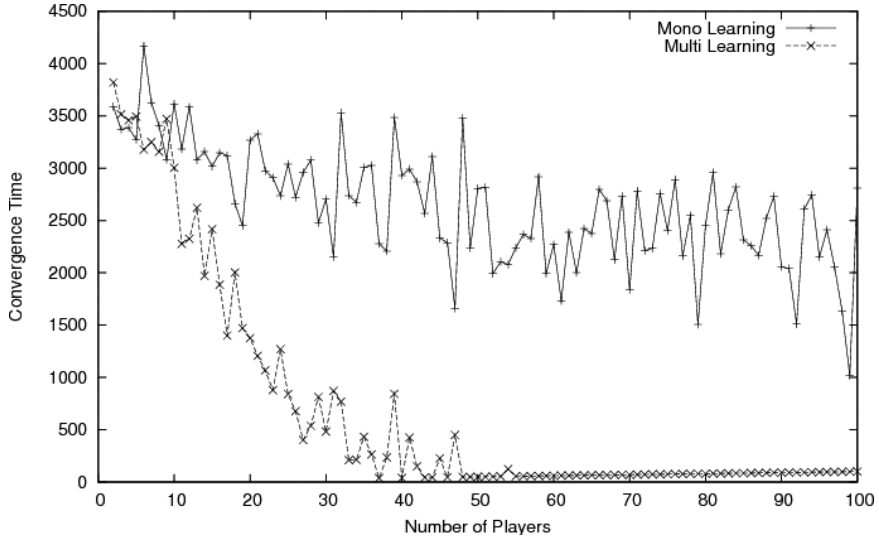**Algorithm 4**: MultiPlayer reward function.

Fig. 11.   Different players in fully connected networks with different learning approaches.

Results from this experiment are presented in Fig. 11. We observe that the convergence time seems to decrease when the number of players in each interaction increases. The reason of this phenomena is found on the design of the reward function: agents will reinforce the action played by the majority. In situations with smaller number of players, fewer agents are consulted while deciding the majority action. This leads to the emergence of more subconventions, and these agents reinforce each other. On the other hand, in scenarios with larger number of players per interaction, the majority action will be calculated by consulting the history of more agents, and therefore, these subconventions will be less likely to appear. In essence, more consistent rewards are provided to a larger number of agents, hence promoting convergence of choices and enabling the emergence of a global convention. Information sharing between a larger number of individuals then have a similar effect on global convention emergence as the likelihood of interacting with a larger proportion of the society. Both facilitate uniform action adoption.

In Fig. 12 we can observe how the neighborhood size affects (as identified in Sec. 4.1) the emergence of multiplayer conventions. We previously observed that the emergence of conventions in one-dimensional lattices was strongly affected by the neighborhood size: the convergence times are drastically reduced when the diameter of the network (inversely proportional to the neighborhood size) is reduced. We can now observe similar effects in multiplayer situations. We also observe that the convergence time also increases when increasing the number of players. This phenomena occurs due to the learning modality used by the agents (a multi-learning approach) and the relation between the number of players and the neighborhood size. When a small number of players interact (relative to the neighborhood size),
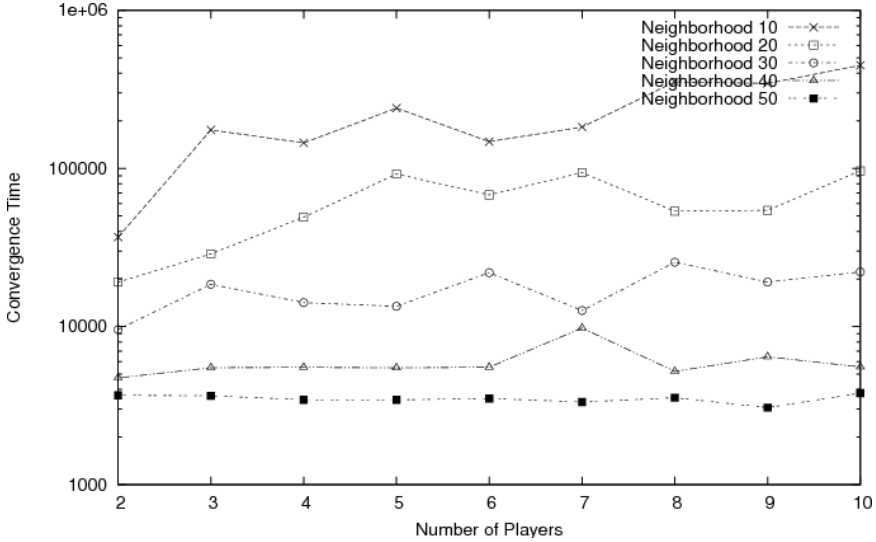
Fig. 12.   Neighborhood sizes comparison with different players with multi-learning approach.

subconventions are less likely to be created. However, when the number of players per interaction is larger (and closer to the neighborhood size), agents will always interact with the same agents, creating and reinforcing subconventions.

The neighborhood size also plays an important role when dealing with multi-player interactions. Assuming that all the players in an interaction must be neighbors, we can infer that the amount of neighbors an agent has directly affects to the convergence time. We have accordingly designed two methods that will adapt the neighborhood size of the network, depending on the number of players per interaction: the reduced and the super-reduced. The *super-reduced* scheme assigns the minimum amount of neighbors needed for the number of players specified. Recall that the neighborhood size should be an even number N, representing $N/2$ of neighboring agents on each side of the agent. Therefore, in the super-reduced method, for either a 2 or 3 players games, the neighborhood size will be $N = 2$. We have observed that using the *super-reduced* method, in the case of the upper limit of players (3 players with a neighborhood size $N = 2$), makes agents interact repeatedly with the same agents every timestep. Depending on the type of reward function, this repeated interaction with the same players would lead to metastable subconventions. Therefore, we relaxed this method creating the *reduced* method. The *reduced* method assigns the minimum amount of neighbors needed for the amount of players specified *plus 2*. This addition of two extra neighbors introduces variety into the games (reducing the endogamy), allowing agents to play with different agents in different interactions.

From the results presented in Fig. 13 we can observe the convergence times of both methods (reduced and super-reduced) with different number of players
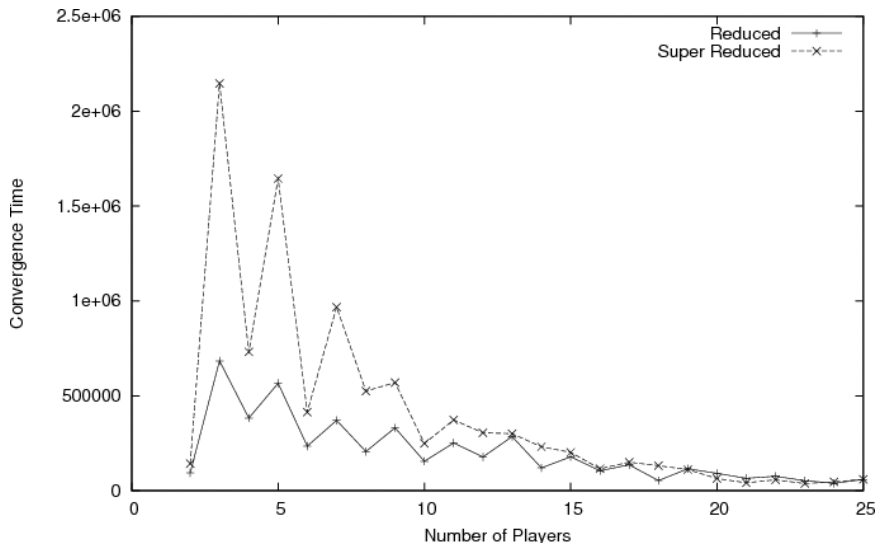
Fig. 13. Reduced and super reduced neighborhood sizes with different players using the multi-learning modality.

per interaction. We can observe how the *super-reduced* method produced larger convergence times and more pronounced changes between the even and odd value of each neighborhood size.[c] The reason why the *super-reduced* takes longer to converge is because of the interacting topology. When agents are not allowed to interact with other agents than their direct neighbors, no variance is introduced (complete endogamy), promoting the appearance of a "frontier effect". This frontier effect affects the agent located in the middle of two regions of clear preference: this agent in the "frontier" will be doubtful about its preference (one example of this situation can be seen in Fig. 14). If the agents that the frontier agent interact with are always the same, it will be harder to break the frontier effect this agent is under the influence of. Until one of the involved agents explores a different action, this frontier effect will not be broken.

Moreover, in Fig. 13 we observed another phenomenon that has not been explained yet. We observe that the scenarios with an odd number of players games take longer to converge than those with an even number of players. We can find an explanation for this phenomenon in the design of the reward function and the frontier effect: the majority action is chosen by observing the actions of the interacting agents, and, if those are even, there might be a tie. In case of a tie, the majority action is chosen randomly, giving a clear advantage to one of the actions in the "frontier" region. In the case of an odd amount of players, the frontier agent will be affected by the same amount of non-frontier agents. We can observe an example

---

[c]For $N = 2$, 2 and 3 players games can be played. For $N = 4$, 4 and 5 players games can be played. For $N = m$, m (even) and $m + 1$ (odd) players games can be played.
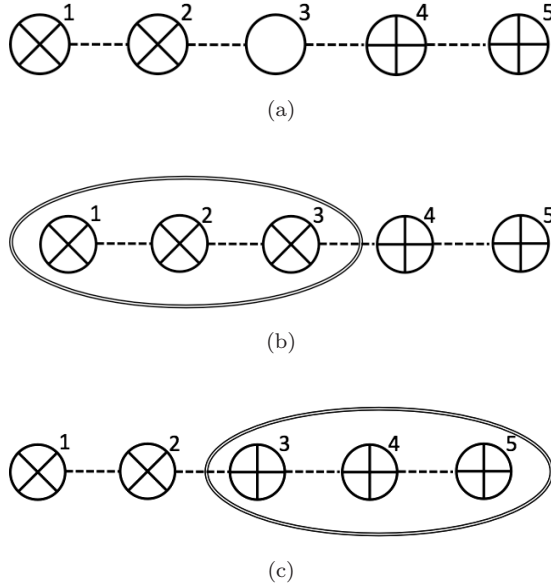
Fig. 14.   Frontier effect in 3-players game with two competing conventions. (a) Agent 3 doubtful. (b) Agent 2 interacts and affects agent 3. (c) Agent 4 interacts and affects agent 3.

scenario in Fig. 14. Initially, in Fig. 14(a), Agent 3 is doubtful about its preference. When Agent 2 interacts, in Fig. 14(b), it affects (together with agent 1) agent 3, biasing their action preference to one convention. On the other hand, when agent 4 interacts, it will affect (together with agent 5) agent 3, making it change its prefer- ence to the other convention. The random selection of the majority action in case of a tie, possible only for even number of players, speeds up the process of con- vergence, although producing a larger number of preference change per agent. On the other hand, when having an odd number of players, we obtain the previously explained "frontier" effect. This effect ensures a longer convergence time (because the frontier agent needs to explore in order to break the frontier) with a smaller number of preference changes. Experimental results are shown in Fig. 15 confirming our hypotheses.

5.1. *Effect of action set*

This last experiment was designed to evaluate how the topology plays an important role in the emergence of conventions in environments where there are more than two conventions to choose from. We can observe in Fig. 15, that the larger the search space in the number of possible conventions, the longer it takes for the conventions to emerge. Specially pronounced is the effect with smaller neighborhood sizes. The explanation for this phenomena is again the creation of metastable subconventions
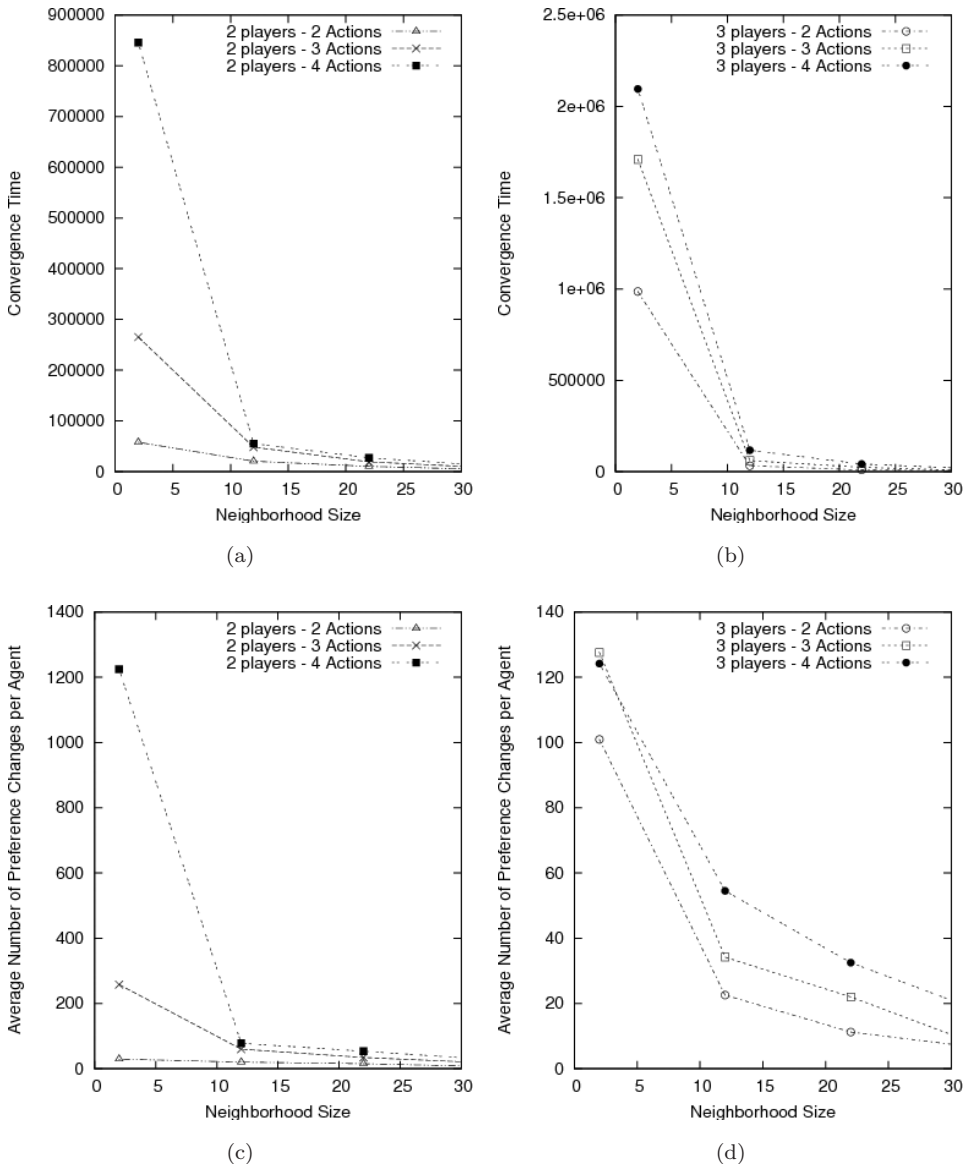
Fig. 15. Multi player and multi action games. (a) Convergence time for 2 player games. (b) Convergence time for 3 player games. (c) Average number of preference change for 2 player games. (d) Average number of preference change for 3 player games.

promoted by the endogamy produced by the topological structure as explained in Sec. 4.1. However in the previous experiments, agent only had two options where to choose from. Having a larger set of options generate a larger number of different subconventions that need to be broken.

6. Conclusion

Our primary goal in this paper was to study how norm emergence in real environments are likely to be influenced by both physical neighborhood effects imposed by mobility restrictions and biases as well as diverse learning, memory and reasoning capabilities of members of the society. We have presented a set of experiments to study the emergence of social conventions based not only on direct interactions but also on the memory (and previous history) of each of the agents under different interconnection topologies between agents. In this social learning framework requires that each agent learns from repeated interaction with anonymous members of the society. In particular, agents learn to adopt conventions that are good for interaction with arbitrary population members rather than learning policies targeted for interacting with particular individuals.

Our initial hypotheses were that different characteristics of the topology in which agents are located would produce different convergence times for reaching a social convention. Experimental results confirm this hypotheses. We have shown that the emergence of transitory subconventions are the cause of the delay of the emergence of conventions, as identified also by other authors [4]. Specific topology structures promote endogamy between a certain subgroup of agent that might produce a different subconvention than the rest. We also observe that memory size have a pronounced affect on the emergence of conventions in all topologies studied with agents having larger memory sizes taking longer to reach conventions. This is due to the fact that the reward amount for a given action is inversely proportional to the memory size. As a result, reward sizes are smaller for larger memory sizes, requiring a higher number of interactions for a convention to be reached.

Moreover, we have observed how the learning modality does directly affect in the performance of the convention emergence process. We observe that subconventions are more likely to appear and are more resistant when using the multi-learning approach, and might not be resolved for scale-free and fully-connected star networks. To aid in breaking such stalemates, we introduced a new, plausible reward function which allows socially important nodes (those with more connections to other agents) to have more influence in the reward function. This new reward function accelerates the emergence of conventions in scale-free networks but subconventions persist in fully-connected star networks.

Finally, we have observed how, in games with more than two players, convergence times are affected differently by the structure of the topology: more clustered societies converge faster with larger player games; on the other hand, less clustered societies take longer to converge when the number of players increase.

7. Future Work

One question that we plan to answer in future work is under what circumstances and configuration of parameters the one-dimensional lattice behaves similarly to the scale-free network for large population sizes. We have observed that when

the population size increases, the convergence times in the one-dimensional lattice increases at a much faster rate compared to scale-free networks. We believe that a dynamic adjustment of the neighborhood size on a one-dimensional lattice, and hence the network diameter, will produce similar dynamics to those obtained with scale-free networks.

We also want to experiment with heterogeneous populations, as done by Mukherjee *et al.* [20]. In the current paper, all the agents are initialized with the same parameters and with the same distribution of initial memory. We want to observe the resulting dynamics of different types of populations. For example, in a scale-free network, we can initialize the hubs with a specific bias towards a certain action, and observe the speed of convergence of the rest of the population. Another interesting experiment to be carried out is when agents in the same population are initialized with different memory sizes.

Finally, we want to analyze, more formally and in a more detailed manner, the effect of subconventions. This work has provides us with a basic understanding that subconventions are facilitated by the topological configuration of the environment (isolated areas of the graph which promote endogamy) or by the agent reward function (concordance with previous history, promoting cultural maintenance). Consequently, we need to develop and test mechanisms for agents to proactively discover and dissolve these subconventions in real-life scenarios.

Acknowledgments

References

[1] Adamic, L. A., Lukose, R. M., Puniyani, A. R. and Huberman, B. A., Search in power-law networks, *Phys. Rev. E* **64** (2001) 046135.

[2] Albert, R. and Barabasi, A.-L., Statistical mechanics of complex networks, *Rev. Mod. Phys.* **74** (2002) 47–97.

[3] Andrighetto, G., Villatoro, D. and Conte, R., Norm internalization in artificial societies, *AI Communications* **23** (2010) 325–339.

[4] Araújo, T. and Aubyn, M. S., Education, neighborhood effects and growth: An agent-based model approach, *Adv. Complex Syst.* **11** (2008) 99–117.

[5] Axelrod, R., An evolutionary approach to norms, *Am. Polit. Sci. Rev.* **80** (1986) 1095–1111.

[6] Barabasi, A. and Bonabeau, E., Scale-free networks, *Sci. Am.* **288** (2003) 60–69.

[7]   Boella, G., Norm governed multiagent systems: The delegation of control to autonomous agents, in *Proc. of IEEE/WIC Intelligent Agent Technology Conference* (IEEE Press, 2003), pp. 329–335.

[8]   Boella, G., Torre, L. and Verhagen, H., Introduction to the special issue on normative multiagent systems, *Auton. Agent. Multi-Ag.* **17** (2008) 1–10.

[9]   Boissier, O. and Gâteau, B., Normative multi-agent organizations: Modeling, support and control, draft version, in *Normative Multi-agent Systems*, eds. Boella, G., van der Torre, L. and Verhagen, H., number 07122 in Dagstuhl Seminar Proceedings (2007), Internationales Begegnungs — und Forschungszentrum fuer Informatik (IBFI), Schloss Dagstuhl, Germany, http://drops.dagstuhl.de/opus/volltexte/2007/902/pdf/07122.BoissierOlivier.Paper.902.pdf.

[10] Castelfranchi, C. and Tummolini, L., Positive and negative expectations and the deontic nature of social conventions, in *ICAIL'03: Proceedings of the 9th International Conference on Artificial Intelligence and Law* (ACM, New York, NY, USA, 2003), ISBN 1-58113-747-8, pp. 119–125, http://doi.acm.org/10.1145/1047788.1047819.

[11] Coleman, J., Foundations of social theory (1998).

[12] Delgado, J., Emergence of social conventions in complex networks, *Artif. Intel.* **141** (2002) 171–185.

[13] Delgado, J., Pujol, J. M. and Sangüesa, R., Emergence of coordination in scale-free networks, *Web Intel. Agent Syst.* **1** (2003) 131–138.

[14] Garcia-Camino, A., Noriega, P. and Rodriguez-Aguilar, J., Implementing norms in electronic institutions, in *AAMAS'05: Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems* (ACM, New York, NY, USA, 2005), ISBN 1-59593-093-0, pp. 667–673, http://doi.acm.org/10.1145/1082473.1082575.

[15] Huberman, B. A. and Glance, N. S., Evolutionary games and computer simulations, *Proc. Natl. Acad. Sci. USA* **90** (1993) 7716–7718.

[16] Kittock, J. E., Emergent conventions and the structure of multi-agent systems, in *Lectures in Complex Systems: The Proceedings of the 1993 Complex Systems Summer School, Santa Fe Institute Studies in the Sciences of Complexity Lecture Volume VI, Santa Fe Institute* (Addison-Wesley, 1993), pp. 507–521.

[17] Leskovec, J., Backstrom, L. and Kleinberg, J., Meme-tracking and the dynamics of the news cycle, in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD'09 (ACM, New York, NY, USA, 2009), ISBN 978-1-60558-495-9, pp. 497–506, http://doi.acm.org/10.1145/1557019.1557077, http://doi.acm.org/10.1145/1557019.1557077.

[18] López y López, F., Luck, M. and d'Inverno, M., A normative framework for agent-based systems, in *Normative Multi-Agent Systems*, eds. Boella, G., van der Torre, L. and Verhagen, H., number 07122 in Dagstuhl Seminar Proceedings (Internationales Begegnungs- und Forschungszentrum fuer Informatik (IBFI), Schloss Dagstuhl, Germany, 2007), http://drops.dagstuhl.de/opus/volltexte/2007/933/pdf/07122.LopezyLopezFabiola.Paper.933.pdf.

[19] Mukherjee, P., Sen, S. and Airiau, S., Norm emergence in spatially contrained interactions, in *Proceedings of ALAg-07* (Honolulu, Hawaii, USA, 2007).

[20] Mukherjee, P., Sen, S. and Airiau, S., Norm emergence with biased agents, *International Journal of Agent Technologies and Systems* (*IJATS*) **1** (2009) 71–84.

[21] Newman, M. E. J., The structure and function of complex networks, *SIAM Rev.* **45** (2003) 167–256.

[22] Saam, N. J. and Harrer, A., Simulating norms, social inequality, and functional change in artificial societies, *Journal of Artificial Societies and Social Simulation* **2** (1999).

[23] Sen, O. and Sen, S., Effects of social network topology and options on norm emergence, in *COIN@AAMAS&IJCAI&MALLOW* (2009), pp. 211–222.

[24] Sen, S. and Airiau, S., Emergence of norms through social learning, *Proceedings of IJCAI-07* (2007) 1507–1512.

[25] Shoham, Y. and Tennenholtz, M., On social laws for artificial agent societies: Off-line design, *Artif. Intel.* **73** (1995) 231–252.

[26] Shoham, Y. and Tennenholtz, M., On the emergence of social conventions: Modeling, analysis, and simulations, *Artif. Intel.* **94** (1997) 139–166.

[27] Urbano, P., Balsa, J., Antunes, L. and Moniz, L., Force versus majority: A comparison in convention emergence efficiency (2009), pp. 48–63.

[28] Villatoro, D., Sen, S. and Sabater-Mir, J., Topology and memory effect on convention emergence, in *Proceedings of the International Conference of Intelligent Agent Technology (IAT)* (IEEE Press, 2009).

[29] Walker, A. and Wooldridge, M., Understanding the emergence of conventions in multi-agent systems, in *Proceedings of the First International Conference on Multi-Agent Systems*, ed. Lesser, V. (MIT Press, San Francisco, CA, 1995), pp. 384–389, citeseer.ist.psu.edu/walker95understanding.html.

[30] Watkins, C. J. C. H. and Dayan, P., Q-learning, *Mach. Learn.* **8** (1992) 279–292.