Norm Emergence Under Constrained Interactions in Diverse Societies

Partha Mukherjee
University of Tulsa
partha-mukherjee@utulsa.edu

Sandip Sen
University of Tulsa
sandip@utulsa.edu

Stéphane Airiau
University of Tulsa
stephane@utulsa.edu

ABSTRACT

Effective norms, emerging from sustained individual interactions over time, can complement societal rules and significantly enhance performance of individual agents and agent societies. Researchers have used a model that supports the emergence of social norms via learning from interaction experiences where each interaction is viewed as a stage game. In this social learning model, which is distinct from an agent learning from repeated interactions against the same player, an agent learns a policy to play the game from repeated interactions with multiple learning agents. The key research question is to characterize when and how the entire population of homogeneous learners converge to a consistent norm when multiple action combinations yield the same optimal payoff. In this paper we study two extensions to the social learning model that significantly enhances its applicability. We first explore the effects of heterogeneous populations where different agents may be using different learning algorithms. We also investigate norm emergence when agent interactions are physically constrained. We consider agents located on a grid where an agent is more likely to interact with other agents situated closer to it than those that are situated afar. The key new results include the surprising acceleration in learning with limited interaction ranges. We also study the effects of pure-strategy players, i.e., non-learners in the environment.

Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: Multiagent systems

General Terms

Experimentation

Keywords

Norm, Social dilemma, Coordination game

1. INTRODUCTION

Norms or conventions are key influences on social behavior of humans. Conformity to norms reduces social frictions, relieves cognitive load on humans, and facilitates coordination. "Everyone conforms, everyone expects others to conform, and everyone has good reason to conform because conforming is in each person's best interest when everyone else plans to conform" [11][1]. Conventions in human societies range from fashions to tipping, driving etiquette to interaction protocols. Norms are ingrained in our social milieu and play a pivotal role in all kinds of business, political, social, and personal choices and interactions. They are self-enforcing: "A norm exists in a given social setting to the extent that individuals usually act in a certain way and are often punished when seen not to be acting in this way" [1]. Individual agents in a society can adapt their strategies or behaviors based on feedback from interactions with other agents. Interactions between agents can be formulated as a stage game with simultaneous moves made by the players [10]. Such stage games often have multiple equilibrium [13], which makes the coordination uncertain. While *focal points* [17] can be used to disambiguate such choices, they may not exist in all situations. We are interested in studying the evolution of social conventions or norms that selects one such equilibrium over others based on repeated distributed interactions between agents in a society. Norms can also be thought of as focal points evolved through learning. Hence emergence of norms via learning in agent societies promise to be a productive research area that can improve coordination in agent societies. These aspects of norms or conventions have merited in-depth study of the evolution and economics of norms in social situations [7, 15, 23].

To study the phenomenon of emergence of social norms, we have assumed that the interactions between the computational agents are private, i.e., not observable to the other agents not involved in the interactions. Our experiments involve interactions represented as symmetrical games with the same payoff. We consider a population of agents, where, in each interaction each agent is paired with another agent selected randomly from its neighborhood or from the population in a non-uniform manner. Each agent is learning concurrently over repeated interactions with selected opponents from the society. We also consider the non-uniform selection criterion when physical proximity is used to bias the opponent selection. This kind of learning is referred to in the literature as *Social Learning* to distinguish it from learning from repeated interactions against the same opponent [19].

Cite as: Norm Emergence Under Constrained Interactions in Diverse Societies, P. Mukherjee, S. Sen and S. Airiau, *Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, Padgham, Parkes, Müller and Parsons (eds.), May, 12-16., 2008, Estoril, Portugal, pp. 779-786.

[1]Conventions can therefore be substituted as external correlating signals to promote coordination (all coordination is choosing a solution from a space of possible solutions).

In contrast to prior work on norm emergence [5, 19, 20], this paper investigates the norm emergence phenomenon in more realistic situations where the agents are physically distributed over a grid space and can use different learning algorithms. In physical environments, e.g., real-life physical interactions between humans in the society, agents are much more likely to interact with those in close physical proximity compared to others located further away. Such physical or spatial interaction constraints or biases have been well-recognized in social sciences [12] and, more recently, in the multiagent systems literature [18]. In this paper, we first focus on agents located in a grid world where they interact predominantly with agents in their physical neighborhood. The goal is to evaluate the effects of neighborhood sizes on the rate and pattern of norm emergence. Secondly we evaluate the effects of the following factors on the speed and success of emergence of norms in the agent societies.

1. Homogeneous Vs heterogeneous society of learners.
2. Uniform selection Vs non-uniform selection of opponents in neighborhood.

2. RELATED WORK

The need for effective norms to control agent behaviors is well-recognized in multiagent societies [3, 5]. In particular, norms are key to the efficient functioning of electronic institutions [9]. Most of the work in multiagent systems on norms, however, has centered on logic or rule-based specification and enforcement of norms [6]. Similar to these research, the work on normative, game-theoretic approach to norm derivation and enforcement also assumes centralized authority and knowledge, as well as system level goals [2, 3]. While norms can be established by centralized diktat, norms in real-life often evolve in a bottom-up manner, via "the gradual accretion of precedent" [23]. We find very little work in multiagent systems on the distributed emergence of social norms. We believe that this is an important niche research area and that effective techniques for distributed norm emergence based on local interactions and utilities can bolster the performance of open multiagent systems.

In our formulation, norms evolve as agents learn from their interactions with other agents in the society using multiagent reinforcement learning algorithms [14]. Most multiagent reinforcement learning literature involve two agents iteratively playing a stage game and the goal is to learn policies to reach preferred equilibrium [16]. Another line of research considers a large population of agents learning to play a cooperative game where the reward of each individual agent depends on the joint action of all the agents in the population [21]. The goal of the learning agent is to maximize an objective function for the entire population, the world utility.

The social learning framework we use to study norm emergence in a population [19] is somewhat different from both of these lines of research. This framework considers a potentially large population of learning agents. At each time step, however, each agent interacts with a single opponent agent chosen from the population, and the opponent changes at each interaction. The payoff received by an agent for a time step depends only on this interaction as is the case when two agents are learning to play a game. In the two-agent case, a learner can adapt and respond to the opponent's policy. In our framework, however, the opponent changes at each interaction. It is not clear *a priori* if the learners will converge to useful policies in this situation. Other work with similar interaction assumptions either use deterministic adaptation schemes or assume knowledge of local state of other agents [5].

3. SOCIAL LEARNING FRAMEWORK

The specific social learning situation for norm evolution that we consider is that of learning "rules of the road". In particular, we will consider the problem of which side of the road to drive in and who yields if two drivers arrive at an intersection at the same time from neighboring roads [2]. We will represent each interaction between two drivers as a n-person, m-action stage game. These stage games typically have multiple pure strategy equillibria. In each time period, each agent is paired with another agent from the population to interact according to some interaction bias. An agent is randomly assigned to be the row or column player in any interaction. We assume that the stage game payoff matrix is know to both players, but agents cannot distinguish between other players in the population. Hence, each agent can only develop a single pair of policies, one as a row player and the other as a column player, to play against any other player from the agent population. The learning algorithm used by an agent is fixed, i.e., an intrinsic property of an agent.

When two cars arrive at an intersection, a driver will sometimes have another car on its left and sometimes on its right. These two experiences can be mapped to two different roles an agent can assume in this social dilemma scenario and corresponds to an agent playing as the row and column player respectively. Consequently, an agent has a private bimatrix: a matrix for when it is the row player, one matrix for when it is the column player. Each agent has a learning algorithm and learns independently to play. An agent is randomly assigned as the row or the column player in every interaction. Each agent develops a pair of policies, one for its role as a row player and another for its role as a column player. An agent does not know the identity of its opponent, nor its opponent's payoff, but it can observe the action taken by the opponent (perfect but incomplete information).

We consider the agents are distributed over space where each agent is located at a grid point (see Figure 1). Each agent has a fixed location on the grid and hence a static set of neighbors. In this grid world, an agent can interact only with agents located within its neighborhood. The neighborhood of an agent is composed of all agents within a distance D of its grid location. We have used the Manhattan distance metric, i.e., $|x_1 - x_2| + |y_1 - y_2|$ is the distance between grid locations (x_1, y_1) and (x_2, y_2). Different D values are used to represent different neighborhood sizes.

In each time period, each agent interacts with another agent in the society. The selection of opponents follow either of two modes:

Uniform Selection: Agents are randomly selected from the neighborhood of the learner. So every agent within the neighborhood is selected with uniform probability for the interaction.

[2]It might seem to the modern reader that "rules of the road" are always fixed by authority, but historical records show that "Society often converges on a convention first by an informal process of accretion; later it is codified into law." [23].

Algorithm 1: Non-uniform selection of learners

initialization : neighborhood distance $= D$;
for *Each player* $i \leftarrow 1$ ***to*** $|G|$ **do**
  $Sum\_dist^i = 0$;
  for *Each neighbor* j *with dist* $d_j^i < D$ **do**
    $Sum\_dist^i = sum_{j=1}^{|Nb_i|} \frac{1}{d_j^i}$;
  for *Each neighbor* j *with dist* $d_j^i < D$ **do**
    $Pr_j^i = \frac{\frac{1}{d_j^i}}{\sum_{j=1}^{|Nb_i|} \frac{1}{d_j^i}}$;

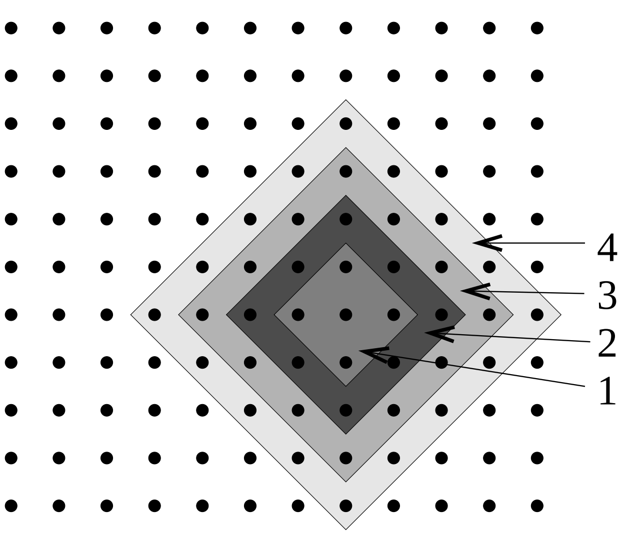


Figure 1: Agents located on a grid and allowed to interact only in a limited neighborhood.

Non-uniform Selection: Agents located closer to the learners are selected for an interaction with higher probability. The probability of selection Pr_j^i is computed from Algorithm 1, where d_j^i is the distance between agent i and agent j and $|G|$ is the grid size. The physical proximity acts as a bias in the selection process.

4. RESULTS

We present experiments with a society of N agents placed in a $\sqrt{N}$x$\sqrt{N}$ grid. For the experiments in this paper, we use 225 agents placed on a 15 by 15 grid. We run experiments using the two-action coordination game, where agents receive high payoff for using the same action and otherwise receive a low-payoff (see Table 1). It can model the situation where agents are deciding which side of the road to drive on. Note that either action combinations (0,0) or (1,1) would work equally well. The goal is then for all agent to develop a norm of choosing the same action consistently.

	left	right
left	4, 4	-1, -1
right	-1, -1	4, 4

Table 1: Payoff in a coordination game.

A payoff of 1.5 is achieved when the agents use a uniform distribution over their actions when playing the game. The maximum reachable payoff for this game is is 4 and is obtained when the players play joint actions (L,L) or (R,R). However, as the learners use ε-greedy scheme, the maximum payoff value cannot be reached. We recognize that a norm has emerged when the average payoff reaches 3.5.

Though some aspects of results from our simulated agent society can be transferred to human situations (with additional mechanisms such as empowering agents with sanction schemes), our results are targeted towards a better understanding of how to develop self-adaptive agent societies. Accordingly, we make no claims about using our results to predict human social behavior.

4.1 Learning Algorithms Used

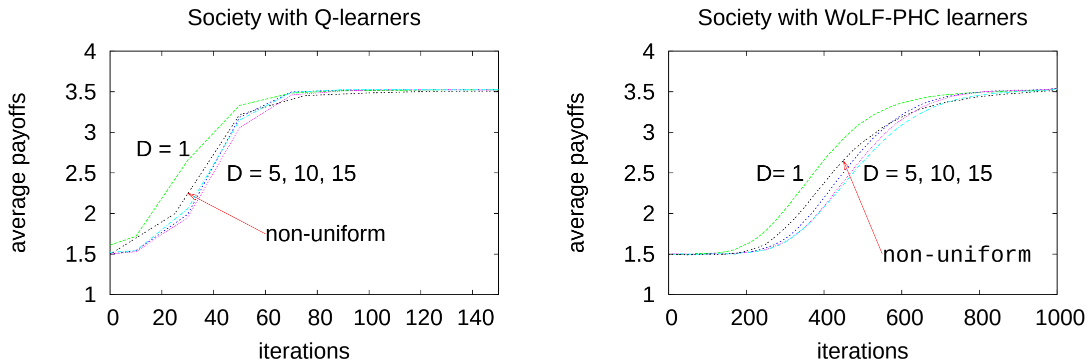
We use four different learning algorithms for norm emergence : Q-Learning [22] with ε-greedy exploration with learning rate $\alpha = 0.1$ and probability of exploration $\varepsilon = 0.1$, WoLF-PHC (Win or Learn Fast-Policy Hill Climbing) [4] with learning rate $\alpha = 0.1$, Fictitious Play (FP) [8] with rate of learning 0.1 and Highest Cumulative Reward (HCR) [5, 20]. Q-Learning is well suited for repeated games against unknown opponents and is widely used in multiagent systems. WoLF-PHC can learn mixed strategies and is guaranteed to converge to a Nash equilibrium of the repeated game in 2-person 2-action games. Fictitious Play (FP) is the basic learning approach widely studied in the game theory literature. An FP player uses the historical frequency count of its opponents' past actions and tries to maximize expected payoff by playing the best response to the observed mixed strategy. HCR is a deterministic scheme that uses finite memory of size M and chooses the action that fetched the maximum cumulative value over the last M interactions. We will also present some experiments when a small minority of the agent population are non-learners, i.e., they play fixed strategies.

4.2 Effect of neighborhood size

For the first set of experiments, all agents use the WoLF-PHC learning algorithm. We have experimented by varying the neighborhood size and observed the corresponding effects on the rate of convergence of the learning agents. We present results from experiments with both uniform and non-uniform selection to understand the effect of the neighborhood size on learning of agents is observed (see Figure 2). We have tested with four neighborhood distances, D (1, 5, 10, and 15), for each agent. When $D = 1$ only an adjacent agent is a neighbor (there are 4 neighbors in that case). The computation of number of neighbors should follow the recurrence relation $D_i = 4 \cdot i + D_{i-1}$, where $D_1 = 4$. When the distance is 15, every agent is a neighbor of every agent.

We present in Figures 2(a) and 2(b) the dynamics of the average payoff over a run of populations of Q-learning and WOLF-PHC learners respectively when all agents are learning concurrently. We observe that the smaller the neighborhood distance, the faster the emergence of a norm. It is also interesting to note from Figure 2(a) and 2(b) that the learning rate for non-uniform opponent selection falls in between the smallest ($D = 1$) and the larger group ($D = 5$, 10, 15) of neighborhood sizes. Norm emergence in society with Non-uniform selection does not depend on neighborhood D.

When an agent has four neighbors ($D = 1$), the agents



(a) Learners : Q-Learning agents

(b) Learner : WoLF-PHC Players

Figure 2: Learning of homogeneous agents.

learn to coordinate faster by driving on the same side of the road than when it has 35 or 99 neighbors ($D = 5$ and 10 respectively). For a given number of iterations, the agents interact more often with a particular neighbors for smaller neighborhoods. This means that the impact an agent has on another agent is larger when the neighborhood size is small. In addition, an agent with few neighbors will encounter few different behaviors from its neighbors, and it is *a priori* easier to coordinate with a small set of agents rather than a larger one. As the neighborhood distance increases, an agent has to coordinate with many other agents, and in addition, interactions between two particular neighbors in the network become less frequent. This decreasing interaction frequency between pairs of learners increases the time for exploration of the behavior space and thereby influences the learning patterns of the agents in the network. This problem is exaggerated when every agent is everyone's neighbor ($D = 15$) which further reduces the rate of learning.

When the entire population uses the same learning algorithm, from Figures 3(a), 3(b) and 3(c) it is clearly observed that population of Q-Learners is fastest to evolve a convention($\approx$ 100 iterations), followed by the society of WoLF ($\approx$ 1000 iterations) and FP ($\approx$ 50000 iterations) for selected values of the neighborhood distance D.

Figure 4 represents, for largest ($D = 15$) and smallest ($D = 1$) neighborhoods, the policy of each agent in the population at different iterations in a single run. Each cell represents the policy of an agent: the darker it is, the higher the probability of driving on the left, whereas lighter colors denote higher probability of driving on the right. When a cell is completely dark, or white, it means that the learning algorithm of the agent has converged. In the particular runs we present, the norm of "driving on the right" emerges (over different runs "driving on the left" and "driving on the right" norms were evolved in roughly the same number of runs). At iteration 145, the agents are exploring and are receiving low payoff (see corresponding payoff dynamics in Figure 2). At iteration 355, for $D = 1$, we are close to the inflection point for the curve of the payoff dynamics: the agents start to favor one norm over the other. For $D = 15$, however, there is a lesser bias favoring one action. We can see that, on the average, the snapshot for $D = 1$ is lighter than that with $D = 15$. At iteration 480, we can see that many more agents have converged for the smallest compared to the largest neighborhood. So smaller neighborhoods induce faster learning among agents on a grid.

The above effect of agent neighborhood size on learning rate was somewhat surprising. A priori, it was unclear whether smaller neighborhoods will engender divergent norms to initially form over the agent space, which would subsequently delay the convergence of the population to a consistent norm. Such effects, however, were overshadowed by the effects of increased interaction frequencies between neighbors in our framework.

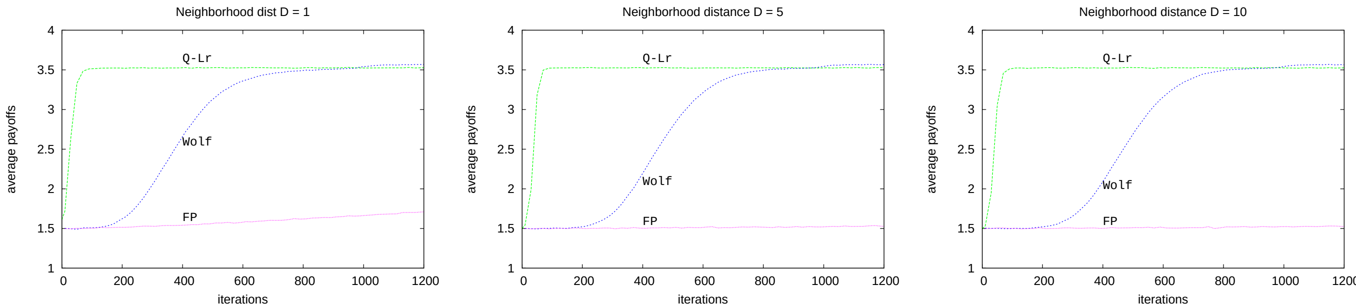
4.3 Influence of non-learning agents

So far, we have observed that all norms with equal payoffs were evolved roughly with the same frequency over multiple runs. This is expected because the payoff matrix for the coordination game (Table 1) has no preference for one norm over the other. Extraneous effects, however, can bias a society of learners towards a particular norm. For example, some agents may not have learning capabilities and always choose a predetermined action. We now study the influence of agents playing a fixed pure strategy (FPS agent) on the emergence of a norm. We are interested in the effect of multiple pure strategy players with the same or different fixed strategies.

We do not preclude the possibility of multiple coexistent norms in sufficiently isolated populations [19]. Without sufficient isolation, stochastic biases introduce enough differential to lead to norm conformance. The norm adopted with larger number of FPS agents is more likely to emerge. Even with a few FPS agents, for a given agent, most of its neighbors are learners and influences this agent's eventual norm selection.

4.3.1 Non-learners use same strategy

In the first experiment, we replace some learning agents by FPS agents and we study the effect of the speed of emergence of a norm. When there are no FPS agents, as the learners explore early in the run, they should encounter each joint ac-



(a) Neighborhood Dist : 1 (b) Neighborhood Dist : 5 (c) Neighborhood Dist : 10

Figure 3: Learning of homogeneous agents.

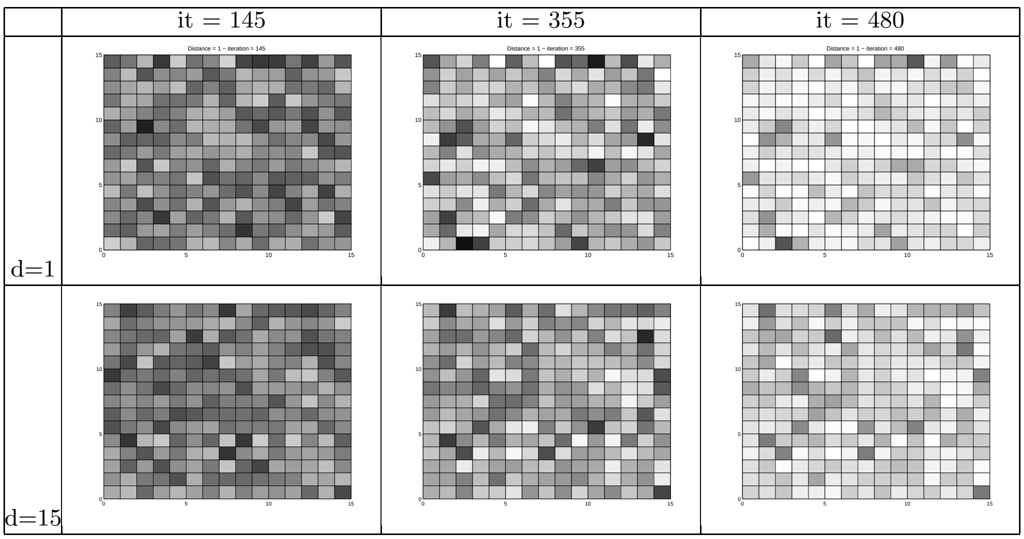


Figure 4: All agents are learning. The whiter cells represents probabilities close to one and darker ones close to probability zero.

tion in the same proportion on average. When FPS agents are present, however, learners that have an FPS agent in their neighborhood should observe a bias towards one strategy which the FPS agent always chooses. As agents start to exploit, a learner i that has an FPS agent f in its neighborhood should exploit this bias and consequently, it is more likely to play the action played by f. This bias should also be boosted by i's neighbors which are also in the neighborhood of f. Our hypothesis is that with more FPS agents that play the same action, e.g., all FPS agents want to drive on the right, the corresponding norm would emerge faster in the population. In Figure 6, we compare the results when there are no FPS agents and either 1, 2, 3, or 4 FPS agents in the population of WoLF-PHC learners[3]. For these experiments, we used $D = 5$. Note that all the FPS agents play the same action (driving on the right).

[3]When there are multiple FPS agents, we located them as far as possible from each other. When there are two FPS agents, they are located at (4,4) and (12,12). When there are three FPS, they are located at (4,4),(7,8) and (12,12). When there are four, they are located at (4,4),(12,12),(4,12) and (12,4)

The first observation from Figure 6 is that norms do not emerge any faster with only one FPS: the local effect of a single FPS agent is insufficient to expedite convergence to a norm. When there are two or more FPS agents, however, we observed the expected faster norm emergence. With our choice of locations for the two FPS agents, no learner has both FPS agents as neighbors. However, the speed of emergence is faster than with one FPS agent in the population. When there are three FPS agents, some agents have two FPS agents in their neighborhood, which could help them to converge faster. However, this is not the case as we observe a minor effect on the speed of emergence. When there are four FPS agents, more learners have two FPS agents in their neighborhood, and we do observe a positive impact on the speed of emergence. As we had expected, the speed of emergence increases with the number of FPS agents. However, we cannot yet accurately predict the variation of the speed of emergence with number of FPS agents, and we plan to further investigate this issue.

4.3.2 *Non-learners use different strategies*

In the previous experiment, all FPS agents were playing

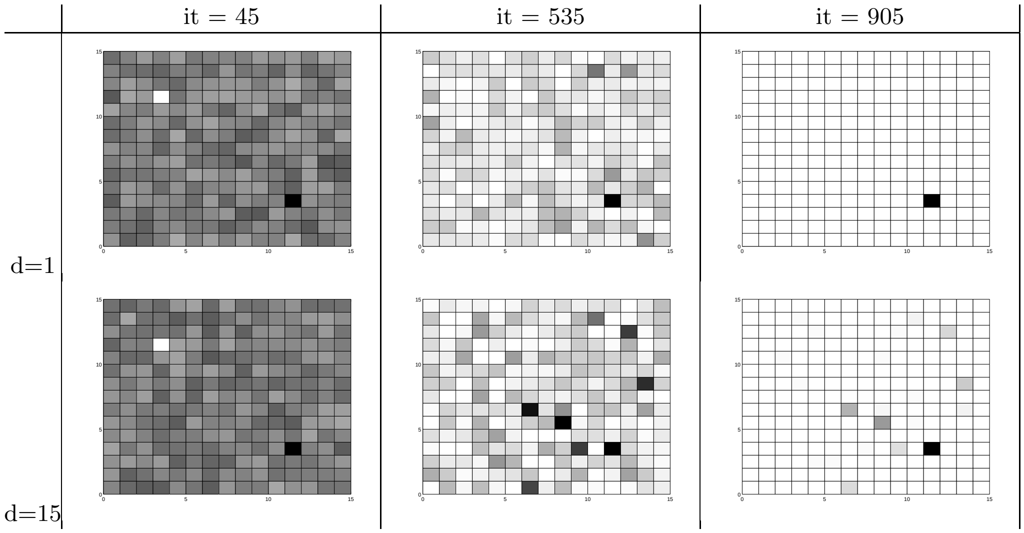


Figure 5: Probabilities of agents driving on the left. Two FPS players play different fixed strategies.

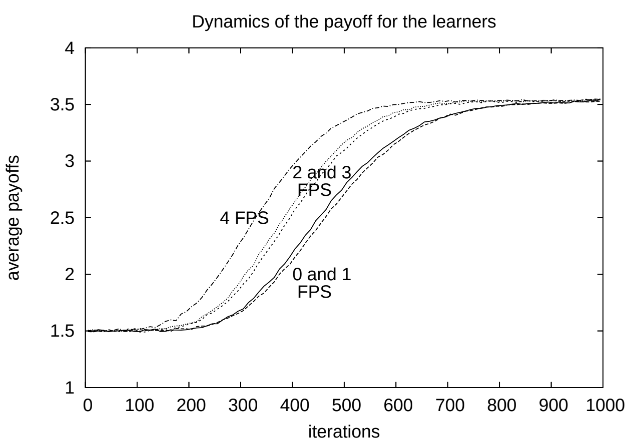


Figure 6: Influence of of non-learners, using identical strategy, on learning rate ($D = 5$).

the same fixed strategy (driving on the right), and they are able to speed up the emergence of a norm. But FPS agents in practice may be unrelated and adopt conflicting behavior, e.g., some agents alway "drive on the right" and some others always "drive on the left". In this case, they are likely to decrease the speed of emergence, or even prevent the convergence of a norm in the entire population. In [19], we have observed that two populations that interact infrequently can develop different norms. Hence, it may be possible that FPS agents influence other agents in their neighborhood, hence, different norms emerge in different neighborhoods. In the next set of experiments, we used two FPS agents playing different strategy R (for driving on right) and L (driving on left).

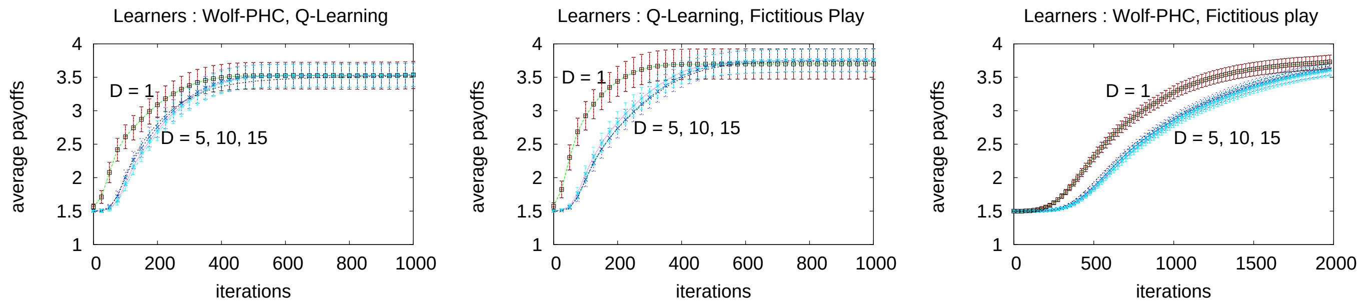
In Figure 5 we present snapshots representing the state of the policy of the agents in the population at different stages of the simulation. The two FPS agents are located at locations (4,12) for R and (12,4) for L. In the two runs, for $D = 1$ and $D = 15$, presented in Figure 5, "driving on the right" is the norm that emerges. We notice that the emergence is faster when the size of the neighborhood is smaller. When the simulation is at iteration 45, the agents are exploring, and the policies of the agents are close to $< 0.5, 0.5 >$. When the simulation is at 535, the population starts to learn and a norm starts to be preferred by a majority of agents. We were expecting that neighbors of the FPS agents will converge to the policy of the near-by FPS agent. But we do not observe this phenomenon, even when the size of the neighborhood is equal to one (for example the agent that is just below the agent choosing L has converged to the norm of R). This may be due to the fact that even with $D = 1$, three of the neighbors are learners, who might ultimately lead the neighbor of L to choose R. We plan to run further experiments to explain this phenomenon. When we ran multiple runs, we observe that each time, the entire population of learning agents converges to a norm: the norms driving on the right and driving on the left emerges with equal frequency. Hence, we did not observe the establishment of multiple norms in these population. This is particularly significant since, with the payoffs we chose (see Table 1), using a single norm in the population maximizes social welfare[4]. Hence, social learning is able to produce social welfare maximizing outcomes even in the presence of non-learners.

4.4 Effect of Heterogeneous Learners

In heterogeneous populations the learning algorithms used by different agents vary. We first performed experiments to evaluate the influence on the norm emergence for each possible subset of the first three learning algorithms. In a given setting agents choose randomly from the available set of learning algorithms. We have also examined the influence on the norm convergence when all four algorithms are used in equal numbers by the agents in the society.

First we consider populations where the agents play with one of two learning algorithms. There are three such hybrid societies given three learning schemes. Figure 7(a), 7(b), and 7(c) show the convergence of social learning in such

[4]If two regions of the population were to adopt distinct norms, the agents at the border and their neighbors would suffer a loss of payoff. When a single norm emerges, only the neighbor of the FPS agents suffer a loss of payoff.



(a) WoLF and Q-Learners (b) Q and FP Learners (c) WoLF and FP Learners

Figure 7: Agents behaving in different two learner societies.

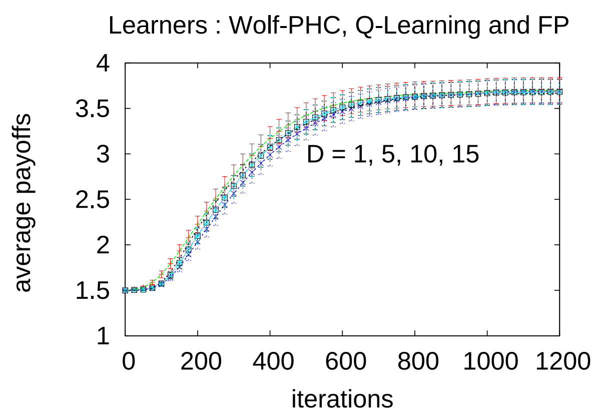


Figure 8: Hybrid Agent Society : WoLF, Q-Learners and FP players with varying neighborhood.

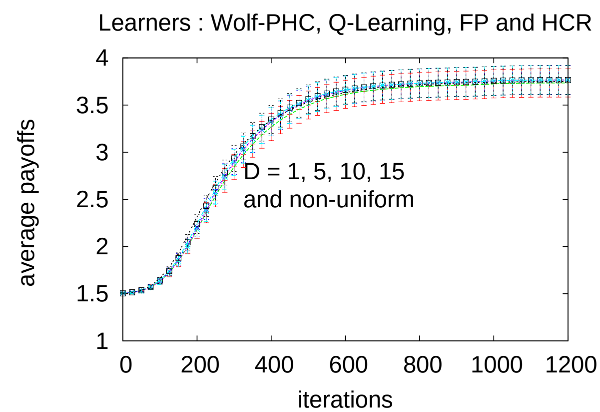


Figure 9: Hybrid Agent Society : WoLF, Q-Learners, FP players, and HCR group with varying neighborhood.

situations.

For the learning algorithms in the population, relative norm emergence rates (see Figure 7) are similar to that in homogeneous population of learners. We observe that societies when Q-Learners are present converge faster. In Figure 7(a) and 7(b) norms emerge within $\approx$ 500 iterations but when agents are using only WoLF and FP, around 2000 iterations are required to evolve a norm (Figure 7(c)).

Now if the degree of learning heterogeneity increases and all three learner groups are incorporated into the population, we find an interesting effect of the neighborhood size on learning rate. The learning curves show no statistical difference(see Figure 8). The effect of the neighborhood size reduces in this case as the probability of getting a neighbor with same algorithm decreases. From figures 2 and 7, we observe that the time taken to evolve norms by heterogeneous groups, with a pair of learners is in between the time taken by the corresponding homogeneous groups. Similarly in a society with three learners, shown in Figure 8, time taken for emergence of norm falls in between that taken by the two slowest mixed groups containing different combinations of the three different learning schemes.

We also run experiments where 25% of the agent societies uses HCR, the deterministic adaptation technique, with M = 25 and the rest is equally divided between the three learning schemes. The results obtained from figure 9 corroborates the patterns observed in the society with three types of learners.

5. CONCLUSIONS

We investigated a bottom-up process for the evolution of social norms that depends exclusively on individual experiences rather than observations or hearsay. This social learning framework requires each agent to learn from repeated interaction with anonymous members of the society. This is in contrast to most results in multiagent learning where two or more agents learn from repeatedly interacting with the same or different group. 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 and reasoning capabilities of members of the society. Our primary goal in this paper was to evaluate the effect of heterogeneous learning populations and spatial interaction constraints on the speed and nature of norms that emerges through social learning. We surmised that limiting interactions may isolate sub-populations, thus allowing for different norms to evolve in different parts of the space. Resolving such emerging conflicts and producing a consensus norm could have been time-consuming and challenge to the learners. Non-uniform exploration captures a certain kind of mobility: because of mobility biases and constraints, the interactions between two agents decrease with increasing distance separating them.

Experimental results, however, demonstrate that agent populations where interactions are restricted to immediate neighbor produce faster convergence to social norms! This is very likely due to the increased number of interactions between immediate neighbors which allow them to quickly identify mutually-agreed behavior. This neighbor agreement speed is found to overshadow the effect of time taken to resolve divergent norms. The learning rate for non-uniform opponent's selection, where the likelihood of selecting on opponent is inversely proportional to distance, is similar to larger neighborhood sizes. Mobility biases, reflected by neighborhood sizes, do have real implications for socio-cultural-religious norms in human societies. We do not focus on modeling human social phenomena, but given interaction biases between agents in electronic societies, it is important to understand corresponding effects.

We investigate the effects of varying neighborhood sizes, selection criterion, different learning strategies and heterogeneous societies on the speed and stability of norm evaluation. Our experimental results confirm that such distributed, individual learning is indeed a robust mechanism for evolving stable social norms. These results confirm that only private experience is sufficient for the emergence of a norm in a society of learning agents. This is in contrast with prior work on norm evolution which requires agents to have knowledge about non-local interactions between other agents and their strategies [5, 7]. The effects of heterogeneity, in terms of learning algorithms used, is more interesting. Increasing diversity obscures the effect of different neighborhood sizes.

We have also studied the influence of agents playing fixed strategy on the emergence of norms. In particular, few agents playing the same strategy are able to bias the choice of the norm adopted. We plan to investigate the influence of fixed strategy players in non-uniform mode. In this paper, we considered that agents were located in a grid, but real world communication and social experiences have more complex topologies. We view this work as a base for exploring more complex environments and norms. In the future, we plan to investigate the emergence of norms in more realistic scenarios, and in particular in scale free networks, small world networks and social networks. In addition, we want to evaluate the framework on different games that exhibit social dilemma. We would also like to study other intriguing phenomena like punctuated equilibria in social norm evolution [23] within our framework.

Acknowledgment: A DOD-ARO Grant #W911NF-05-1-0285 partially supported this work.

6. REFERENCES

[1] R. Axelrod. *The complexity of cooperation: Agent-based models of conflict and cooperation*. Princeton University Press, Princeton, NJ, 1997.

[2] G. Boella and L. Lesmo. A game theoretic approach to norms. *Cognitive Science Quarterly*, 2(3–4):492–512, 2002.

[3] G. Boella and L. van der Torre. Norm governed multiagent systems: The delegation of control to autonomous agents. In *Proceedings of IEEE/WIC IAT Conference*, pages 329–335. IEEE Press, 2003.

[4] M. Bowling and M. Veloso. Multiagent learning using a variable learning rate. *Artificial Intelligence*, 136:215–250, 2002.

[5] J. Delgado, J. M. Pujol, and R. Sangüesa. Emergence of coordination in scale-free networks. *Web Intelligence and Agent Systems*, 1(2):131–138, 2003.

[6] F. Dignum, D. Kinny, and L. Sonenberg. From desires, obligations and norms to goals. *Cognitive Science Quarterly*, 2(3–4):407–430, 2002.

[7] J. M. Epstein. Learning to be thoughtless: Social norms and individual computation. *Computational Economics*, 18:9–24, 2001.

[8] D. Fudenberg and K. Levine. *The Theory of Learning in Games*. MIT Press, Cambridge, MA, 1998.

[9] A. Garcia-Camino, J. Rodriguez-Aguilar, and C. Sierra. A rule-based approach to norm-oriented programming of electronic institutions. *ACM SIGecom Exchanges*, 5(5):33–41, 2006.

[10] M. Genesereth, M. Ginsberg, and J. Rosenschein. Cooperation without communications. In *Proceedings of the National Conference on Artificial Intelligence*, pages 51–57, Philadelphia, Pennsylvania, 1986.

[11] D. Lewis. *Convention: A Philosophical Study*. Harvard University Press, 1969.

[12] N. M. and M. R. M. Evolutionary games and spatial chaos. *Nature(London)*, 359:826–829, 1992.

[13] R. B. Myerson. *Game Theory: Analysis of Conflict*. Harvard University Press, 1991.

[14] L. Panait and S. Luke. Cooperative multi-agent learning: The state of the art. *Autonomous Agents and Multi-Agent Systems*, 11(3):387–434, 2005.

[15] M. Posch. Evolution of equilibria in the long run: A general theory and applications. *Journal of Economic Theory*, 65:383–415, 1995.

[16] R. Powers and Y. Shoham. New criteria and a new algorithm for learning in multi-agent systems. In *Proceedings of NIPS*, 2005.

[17] T. C. Schelling. *The Strategy of Conflict*. Havard University Press, 1960.

[18] F. Schweitzer, J. Zimmermann, and H. MÂÍuhlenbein. Coordination of decisions in a spatial agent modela. *Physica A*, 303(1–2):189–216, 2001.

[19] S. Sen and S. Airiau. Emergence of norms through social learning. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence(IJCAI'07)*, pages 1507–1512, January 2007.

[20] Y. Shoham and M. Tennenholtz. On the emergence of social conventions: Modeling, analysis and simulations. *Artificial Intelligence*, 94(1–2):139–166, 1997.

[21] K. Tumer and D. H. Wolpert. Collective intelligence and Braess' paradox. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence*, pages 104–109, Menlo Park, CA, 2000. AAAI Press.

[22] C. J. C. H. Watkins and P. D. Dayan. Q-learning. *Machine Learning*, 3:279 – 292, 1992.

[23] P. H. Young. The economics of convention. *The Jounal of Economic Perspectives*, 10(2):105–122, Spring 1996.