Emergence of Norms Through Social Learning

Sandip Sen and **Stéphane Airiau**
University of Tulsa, USA
{sandip, stephane}@utulsa.edu

Abstract

Behavioral norms are key ingredients that allow agent coordination where societal laws do not sufficiently constrain agent behaviors. Whereas social laws need to be enforced in a top-down manner, norms evolve in a bottom-up manner and are typically more self-enforcing. While effective norms can significantly enhance performance of individual agents and agent societies, there has been little work in multiagent systems on the formation of social norms. We propose a model that supports the emergence of social norms via learning from interaction experiences. In our model, individual agents repeatedly interact with other agents in the society over instances of a given scenario. Each interaction is framed as a stage game. An agent learns its policy to play the game over repeated interactions with multiple agents. We term this mode of learning *social learning*, which is distinct from an agent learning from repeated interactions against the same player. We are particularly interested in situations where multiple action combinations yield the same optimal payoff. The key research question is to find out if the entire population learns to converge to a consistent norm. In addition to studying such emergence of social norms among homogeneous learners via social learning, we study the effects of heterogeneous learners, population size, multiple social groups, etc.

1 Introduction

Norms or conventions routinely guide the choice of behaviors in human societies. 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" [Lewis, 1969][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" [Axelrod, 1997].

While these aspects of norms or conventions have merited in-depth study of the evolution and economics of norms in social situations [Epstein, 2001; Posch, 1995; Young, 1993; 1996], we are particularly interested in the following characterization: "... we may define a convention as an equilibrium that everyone expects in interactions that have more than one equilibrium." [Young, 1996]. This observation has particular significance for the study of norms[2] in the context of computational agents. Computational agents often have to coordinate their actions and such interactions can be formulated as stage games with simultaneous moves made by the players [Genesereth *et al.*, 1986]. Such stage games often have multiple equilibria [Myerson, 1991], which makes coordination uncertain. While *focal points* [Schelling, 1960] can be used to disambiguate such choices, they may not be available in all situations. Norms can also be thought of as focal points evolved through learning [Young, 1996]. Hence, the emergence of norms via learning in agent societies promises to be a productive research area that can improve coordination in and hence functioning of agent societies.

While researchers have studied the emergence of norms in agent populations, they typically assume access to significant amount of global knowledge [Epstein, 2001; Posch, 1995; Young, 1993; 1996]. For example, all of these models assume that individual agents can observe sizable fraction of interactions between other agents in the environment. While these results do provide key insights into the emergence of norms in societies where the assumption of observability holds, it is unclear if and how norms will emerge if all interactions were private, i.e., not observable to any other agent not involved in the interaction.

To study the important phenomenon of emergence of social norms via private interactions, we use the following interaction framework. We consider a population of agents, where, in each interaction, each agent is paired with another agent

[1]Conventions can therefore be substituted as external correlating signals to promote coordination.

[2]Henceforth we use the term norm to refer to social norms and conventions.

randomly selected from the population. Each agent then is learning concurrently over repeated interactions with randomly selected members from the population. We refer to this kind of learning *social learning* to distinguish from learning in iterated games against the same opponent [Fudenberg and Levine, 1998]. Most of our experiments involve symmetrical games with multiple pure-strategy equilibria with the same payoff. In previous work on learning in games, the opponent is fixed but in our work, the opponent is different at each iteration. In addition, the opponent may not use the same learning algorithm. It is unclear, *a priori*, if and how a social norm will emerge from such a social learning framework. Our experimental results and concomitant analysis throws light on the dynamics of the emergence of norm via social learning with private interactions. We also investigate a number of key related issues: the effect of population size, number of choices available, multiple populations with limited inter-population interactions, heterogeneous population with multiple learning algorithms, effect of non-learners in shaping norm adoption, norms for social dilemmas, etc.

2 Related work

The need for effective norms to control agent behaviors is well-recognized in multiagent societies [Boella and van der Torre, 2003; Vázquez-Salceda *et al.*, 2005]. In particular, norms are key to the efficient functioning of electronic institutions [Garcia-Camino *et al.*, 2006]. Most of the work in multiagent systems on norms, however, has centered on logic or rule-based specification and enforcement of norms [Dignum *et al.*, 2002; Vázquez-Salceda *et al.*, 2005]. 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 [Boella and Lesmo, 2002; Boella and van der Torre, 2003]. While norms can be established by centralized dictat, a number of real-life norms evolve in a bottom-up manner, via "the gradual accretion of precedent" [Young, 1996]. 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. We focus on the importance for electronic agents solving a social dilemma efficiently by quickly adopting a norm. Centralized social laws and norms are not sufficient, in general, to resolve all agent conflicts and ensure smooth coordination. The gradual emergence of norms from individual learning can facilitate coordination in such situations and make individuals and societies more efficient.

In our formulation, norms evolve as agents learn from their interactions with other agents in the society using multiagent reinforcement learning algorithms [Panait and Luke, 2005; Tuyls and Nowé, 2006]. Most multiagent reinforcement learning literature involve two agents iteratively playing a stage game and the goal is to learn policies to reach preferred equilibrium [Powers and Shoham, 2005]. 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 [Tumer and Wolpert, 2000]. 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 is somewhat different from both of these lines of research. We are considering a potentially large population of learning agents. At each time step, however, each agent interacts with a single agent, chosen at random, from the population. 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.

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 interaction at the same time from neighboring roads [3]. We will represent each interaction between two drivers as a n-person, m-action stage game. These stage games typically have multiple pure strategy equilibria. In each time period each agent is paired with a randomly selected agent from the population to interact. 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 when it is the row player, one matrix when it is the column player. Each agent has a learning algorithm to play as a row player and as a column player and learns independently to play as a row and 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). The protocol of interaction is presented in Algorithm 1.

[3] 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." [Young, 1996].

```
for a fixed number of epoch do
    repeat
        remove randomly agents p_row and p_col from the
        population ask each agent to select an action;
        send the joint action to p_row and p_col for policy
        update;
    until all agents have been selected during the epoch ;
```

Algorithm 1: Interaction protocol.

We use three different learning algorithms for learning norms: Q-Learning [Watkins and Dayan, 1992] with ϵ-greedy exploration, WoLF-PHC [Bowling and Veloso, 2002] and Fictitious Play (FP). Q-learning has been widely used in multiagent systems, but converges only to pure strategies. WoLF-PHC (Win or Learn Fast - policy hill climbing) can learn mixed strategies. Though WoLF is guaranteed to converge to a Nash equilibrium of the repeated game in a 2-person, 2-actions game against a given opponent, it is not clear whether it is guaranteed to converge in social learning. Finally, FP is the basic learning approach widely studied in the game theory literature [Fudenberg and Levine, 1998]. An FP player uses the historical frequency count of its opponent's past actions and tries to maximizing expected payoff by playing a best response to that mixed strategy, represented by this frequency distribution.

4 Results

4.1 Example of a social dilemma

One typical example of the use of norms or convention is to resolve social dilemmas. A straightforward example of this is when two drivers arrive at an intersection simultaneously from neighboring streets. While each player has the incentive of not yielding, myopic decisions by both can lead to undesirable accidents. Both drivers yielding, however, also creates inefficiency. Ideally, we would like norms like "yield to the driver on right", which serves all drivers in the long run. Hence, the dilemma is resolved if each member of the population learns to "yield" as a row (column) player and "go" as a column (row) player. The player that yields gets a lesser payoff since it is losing some time compared to the other player. The players know whether they are playing as a row or a column player: the row player sees a car on its right, and the column player sees a car on its left. The action choices for the row player are to go (G) or yield to the car on the right (Y_L), and they are go (G) or yield to the car on the left (Y_L) for the column player. We model this game using the payoffs presented in Table 1(a). Note that for a social norm to evolve, all agents in the population has to learn any one of the following policy pairs: (a) (row:G, col:Y_L), i.e., yield to the car on the left , or (b) (row:Y_R, col:G), i.e., yield to the car on the right. We say a norm has emerged in the population when all learners make the corresponding choice except for infrequent random exploration.

We first note that in iterated play between two players, i.e., if the population consisted of only two agents, other policy combinations may also emerge. For example, in addition to the above two possible norms, it can be the case that one of the two learners learn to "go" both as row and column player and the other player learns to yield in both situations. Although not "fair", this situation is possible in our framework since each agent independently learns to play as a row and a column player.

	G	Y_L
G	-1, -1	3, 2
Y_R	2, 3	1, 1

(a) social dilemma game

	0	1
0	4, 4	-1, -1
1	-1, -1	4, 4

(b) coordination game

Table 1: Stage games corresponding to social interactions.

When a third agent is introduced, as the agents do not know the identity of the opponents, no agent can any longer benefit from always choosing "go". This is because all other agents must always "yield" to the "go" agent, and then those agents will receive relatively poor utility when playing each other. As a result, they will also learn to "go". To optimize performance they will have to learn to settle to a norm which everyone else also follows. Though we only hoped that this would happen via social learning in a large population, our experimental results show that a uniform norm always emerges in a population of three or more agents. For example, in a population of 200 agents using WoLF, we ran $1,000$ runs, and we observed that the population converged to the "yield to the left" norm 482 times, and "yield to the right" norm 518 times. We present the averaged dynamics of the payoffs and the frequency of the joint action during learning in Figure 1. From the dynamics we can see that at first the agents avoid the collision and prefer to yield. Then, one agent notice that it can exploit this situation by choosing to "go" as the other one is yielding. Depending on who notices this first, the population converges to one norm or the other. Note that the plot in Figure 1 is averaged over all the runs, which explains why the (G, Y_L) and (Y_R, G) appear almost 50% of the time. The presence of the other joint-actions is due to exploration. 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 [Epstein, 2001; Posch, 1995; Young, 1993].

4.2 Influence of population size, number of actions and learning algorithm

The time required for the emergence of a norm in a society of interacting agents, measured by the number of interaction periods before most agents adopt the norm, depends on several factors. Here we study the influence of the size of the population, the learning algorithm used, and the number of actions available to the agents.

First we consider the effect of population size. With a larger population, the likelihood that two particular agents interact decreases. Hence the variety of opponents as well as the diversity of personal interaction history increases with the population size and the population takes more time to evolve a norm. In Figure 2, we present the dynamics of the aver-

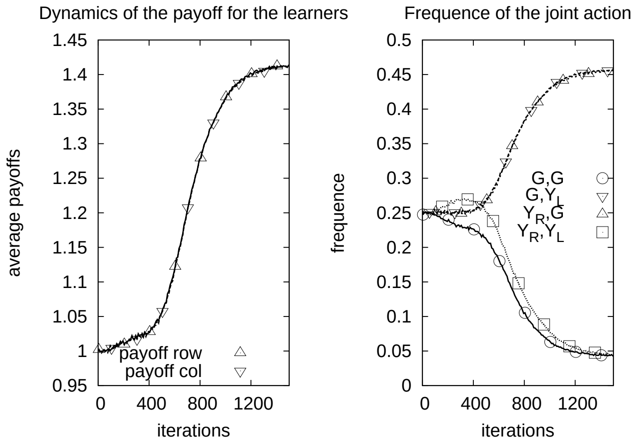

Figure 1: Social dilemma game with 200 agents using WoLF, averaged over 1,000 runs. The population converges 482 times to (G, Y_L) and 518 times to (Y_R, G).

age agent reward for the social dilemma game in a population of agents using WoLF with different population sizes: with more agents, it takes longer for the entire population to converge on a particular norm. It is well-known that tight-knit, small societies, groups, clans develop eclectic norms that are often not found in larger, open societies.

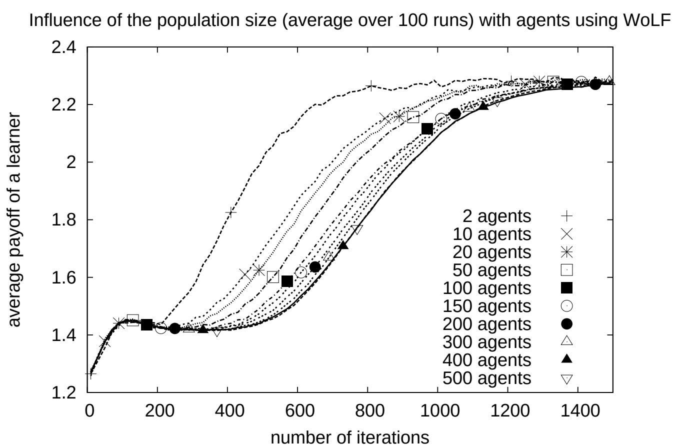

Figure 2: Dynamics of the average payoff of learners using WoLF with different population sizes (average over 100 runs).

Next, we consider the effect of the number of actions available to each agent. For the rest of paper, we use the coordination game presented in Table 1(b). This stage game models the situation where agents need to agree on one of several equally desirable alternatives. For example, for the two-action case, this game can represent the situation where agents choose on what side of the road to drive. When both agents drive on their left, or on their right, there is no collision, else there is a penalty. The societal norms that we would want to evolve are either driving on the left or driving on the right. The stylized game, representing other, non-driving scenarios, can be expanded to n-actions: the agents receive a payoff of 4 when they choose the same action and a payoff of -1 when their actions differ. In Figure 3, we show the dynamics of the probability of an agent choosing action 1 for each learning algorithm in a population of 20 agents for $n = 2$. Then, we ran this experiment with $n \in \{2, 3, 4\}$ in a population of 200 agents using WoLF. The results are presented in Figure 4. When the number of actions increase, the proportion of joint actions with high payoff decreases. When the agents explore at the beginning, the expected utility is less with a larger game. Over time a norm emerges, with the average payoff of the population approaching 4. It takes longer to evolve norms for larger action sets as the space of joint actions increases quadratically.

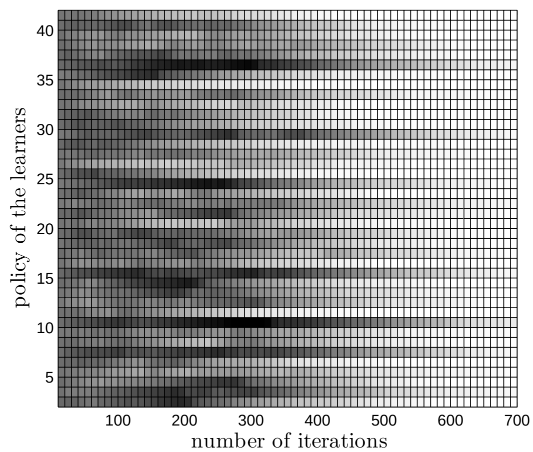

Figure 3: Dynamics of the probability to play action 0 (each agent is represented by two lines: policy to play as a row and as a column player; darker color represents probability closer to 1): all agents converge to a probability close to 0, i.e., chooses action 1.

Finally, we consider the effect of the learning algorithm used by the agents. Since there is no clear choice of learn-

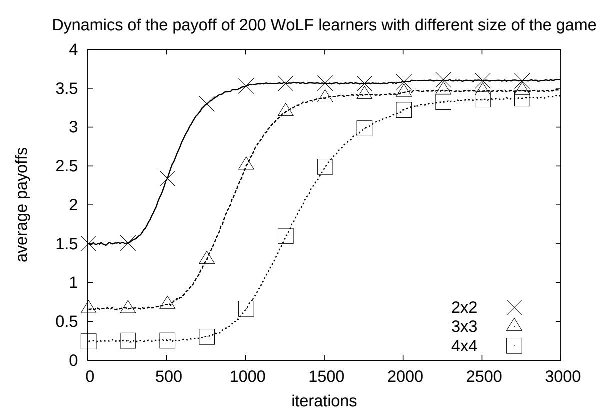

Figure 4: Dynamics of the payoff of learners using WoLF with different game sizes (average over 100 runs).

ing algorithms to use in general, we wanted to evaluate a few representative learning algorithms. We study the influence of the learning algorithms on a population of 200 agents playing the two-action game. When the entire population uses the same learning algorithm, the population of Q-Learners are the quickest to evolve a norm (≈ 100 iterations), followed by a population of WoLF (≈ 1000 iterations), and the population of agents using FP ($\approx 40,000$ iterations). The payoff reached at convergence is different for different algorithms due to different exploration schemes. We also show results of hybrid population using equal proportions of any two or all three of these algorithms. The time taken by mixed groups to evolve norms are in between the time taken by the corresponding homogeneous groups.

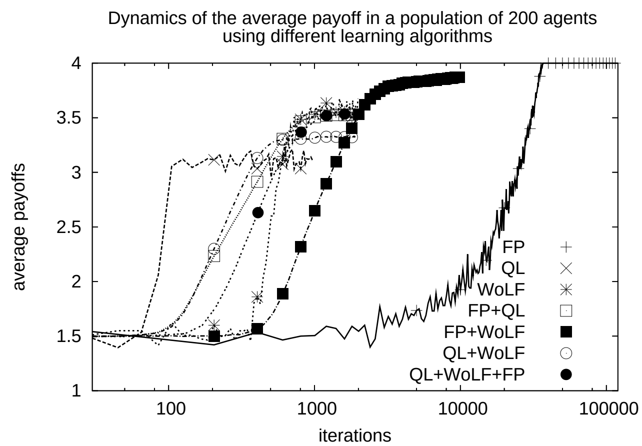

Figure 5: Dynamics of the payoff of learners using different learning algorithms (population of 200 agents, average over 100 runs).

4.3 Influence of fixed agents

So far, we have observed that all norms with equal payoffs were evolved roughly with the same frequency over multiple runs. This is understandable because the payoff matrix in Table 1(b) does not support any 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 repeat a predetermined action. We study the influence of agents playing a fixed pure strategy on the emergence of a norm. For this study, we use the coordination game of Table 1(b) and consider a population with $3,000$ learners, $n_f = 30$ agents playing the fixed strategy 0 (driving on the left), and n_f agents playing strategy 1 (driving on the right). We ran experiments where we add additional agents playing the pure strategy 1. Figure 6 presents the percentage of time that the norm $(0, 0)$, i.e., everyone driving on the left, and $(1, 1)$, i.e., everyone driving on the right, emerges. Note that when there are equal number of agents playing fixed strategy 0 and fixed strategy 1, one of the two norms emerges with almost equal frequency. It is interesting to note that with only 4 additional agents choosing to drive on the right, the entire population of 3000 agents almost always converges to driving on the right! There just

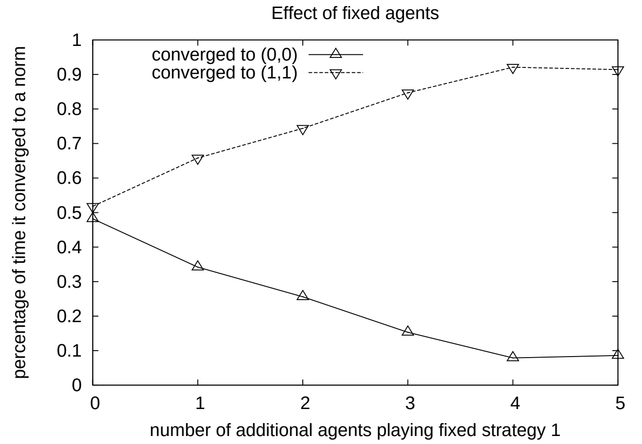

Figure 6: Number of times each norm emerges (average over 100 runs): a small imbalance in the number of agents using a pure strategy is enough to influence an entire population

might therefore be some truth to the adage that most fashion trends are decided by a handful of trend setters in Paris!

4.4 Emergence of norms in isolated subpopulations

It is well-documented in societies that isolated populations can be using contradictory norms, e.g., driving on the “right” or the “wrong” side of the road. We wanted to replicate this phenomenon using our social learning framework. When two groups of agents interact only infrequently, it is possible that a different norm emerges in each group. In particular, we are interested in studying the degree of isolation required for divergent norms to emerge in different groups. For our experiments, we consider two groups of equal size and a probability p that agents of different groups interact.

Results from this set of experiments are presented in Figure 7. We observe that when the probability of interaction is at least 0.3, a single norm pervades the entire population. In roughly half of the runs all agents learn to drive on the left and for the other half they learn to drive on the right. But for interaction probabilities of 0.2 and less there are runs where divergent norms emerge in the two groups (corresponding to the white space above the shaded bars in Figure 7). This is a very interesting observation and we are surprised by the relatively high interaction probabilities that could still sustain divergent norms.

5 Conclusions

We investigated a bottom-up process for the evolution of social norm that depends exclusively on individual experiences rather than observations or hearsay. Our proposed social learning framework requires each agent to learn from repeated interaction with anonymous members of the society. The goal of this work was to evaluate whether such social learning can successfully evolve and sustain a useful social norm that resolves conflicts and facilitates coordination between population members. Our experimental results confirm

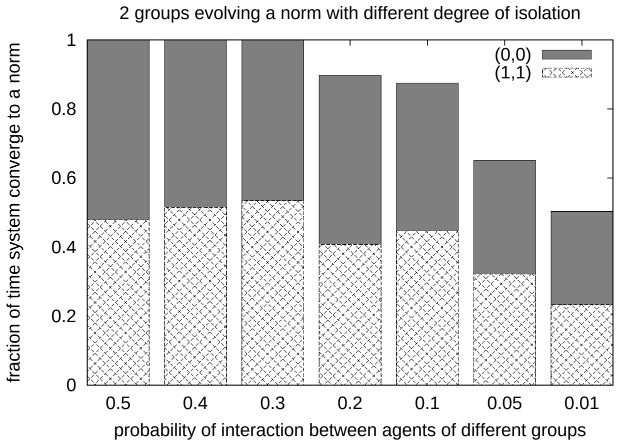

Figure 7: Two groups of 100 agents each evolve norms with different interactions frequencies (average over 1,000 runs). When the probability of interaction is low, the groups can evolve different norms.

that such distributed, individual learning is indeed a robust mechanism for evolving stable social norms.

We investigate the effects of population size, number of actions, different learning strategies, non-learning agents, and multiple relatively isolated populations on the speed and stability of norm evolution. We confirmed that even thorny problems like social dilemmas can be successfully addressed by the social learning framework.

We would like to study other intriguing phenomena like punctuated equilibria in social norm evolution [Young, 1996] within our framework. Other interesting experiments include study of spatial distribution of agents and corresponding effects on rate and divergence of norms.

Acknowledgment: US National Science Foundation award IIS- 0209208 partially supported this work.

References

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

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

[Boella and van der Torre, 2003] 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.

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

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

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

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

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

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

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

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

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

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

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

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

[Tumer and Wolpert, 2000] Kagan Tumer and David 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.

[Tuyls and Nowé, 2006] K. Tuyls and A. Nowé. Evolutionary game theory and multi-agent reinforcement learning. *The Knowledge Engineering Review*, 20(1):63–90, March 2006.

[Vázquez-Salceda *et al.*, 2005] Javier Vázquez-Salceda, Huib Aldewereld, and Frank Dignum. Norms in multi-agent systems: From theory to practice. *International Journal of Computer Systems & Engineering*, 20(4):225–236, 2005.

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

[Young, 1993] Peyton H. Young. The evolution of conventions. *Econometrica*, 61:57–84, January 1993.

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