Social norm emergence in virtual agent societies (Short Paper)

Bastin Tony Roy Savarimuthu, Maryam Purvis and Martin Purvis
Department of Information Science
University of Otago
P O Box 56, Dunedin, New Zealand
(tonyr, tehrany, mpurvis)@infoscience.otago.ac.nz

ABSTRACT

The advent of virtual environments such as SecondLife call for a distributed approach for norm emergence and spreading. In open virtual environments, monitoring various interacting agents (avatars), using a centralized authority might be computationally expensive. The number of possible states and actions of an agent could be huge. An approach for sustaining order and smoother functioning of these environments can be facilitated through norms. Agents can generate norms based on interactions. In particular, those social norms that incur certain cost to an individual agent but benefit the whole society are more interesting than those benefit both the agent and the society. The problem is that the selfish agents might not be willing to share the norm adherence cost. In this work, we experiment with notion proposed by Axelrod that social norms are best at preventing small defections where the cost of enforcement is low. We also study how common knowledge can be used to faciliate the overall benefit of the society. We believe our work can be used to facilitate norm emergence in virtual online societies.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Multiagent systems*; J.4 [**Computer Applications**]: Social and Behavioral Sciences—*Sociology*

General Terms

Design, Experimentation

Keywords

Norms, Agents, Societies, Emergence, Punishment, Common Knowledge

1. INTRODUCTION

Norms are expectations of an agent about the behaviour of other agents in the society. The human society follows norms such as tipping in restaurants and exchange of gifts at Christmas. Norms are of interest to researchers because they help to improve the predictability of the society. They also reduce the computations required by an agent to make a decision. Norms have been of interest in different areas of research such as Sociology, Economics, Psychology and Computer science [9]. Norms have been shown to facilitate co-ordination and co-operation among the members of the society [2, 17]. Some of the well established norms may become laws.

While the discussion on how norms emerge and spread remains a research issue among scientists in Sociology, the advent of new ways of human interactions proxied through software agents in virtual 3D worlds such as SecondLife [1] have created interest among researchers in MAS to work on the applicability of the concept of norms in these digital societies.

We believe that software agents that operate autonomously or on behalf of human users in these virtual worlds cannot be effectively monitored and controlled through centralized policing mechanisms. The explosion of possible action states for an agent in a open environment is huge. It would be computationally expensive and taxing for a centralized monitor to enforce behavioural regularities to ensure smoother functioning of these systems. We believe that an alternative approach based on norms could be effectively used in such scenarios where norms can be derived and built using a bottom-up approach through interactions between the agents.

Researchers have categoried norms into several categories [9]. One such categorization is based on the the cost-benefit analysis of a norm from a perspective of an individual agent and how that relates to the society as the whole, proposed by Horne [11]. There are four categories of norms according to Horne. There are norms which benefit both the individual agents and the society and some norms incur cost to both the agent and the society. There are some norms that benefit the agent but cost the society. Some norms cost the agent but are beneficial to the society. We are interested in the last category of norms where norms cost the agent but the whole society is benefitted. For these norms to be established there should be agents in the society that will help in the enforcement of these norms. In this paper we are interested in how one such norm, the norm against littering might spread within an agent society. In this context we also examine how the concept of common knowledge can be used to facilitate norm emergence.

2. BACKGROUND

Due to multi-disciplinary interest in norms, several definitions for norms exist. Habermas [10], a renowned sociologist, identified norm regulated actions as one of the four action patterns in human behaviour. A norm to him means *fulfilling a generalized expectation of behaviour*, which is a widely accepted definition for social norms. When members of a society violate the societal norms, they may be punished. Many social scientists have studied why norms are adhered. Some of the reasons for norm adherence include a)

Cite as: Social norm emergence in virtual agent societies (Short Paper), Bastin Tony Roy Savarimuthu, Maryam Purvis and Martin Purvis, *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. 1521-1524.

fear of authority b) rational appeal of the norms and c) feelings such as shame, embarrassment and guilt that arise because of non-adherence.

2.1 Normative multi-agent systems

Norms have been of interest to multi-agent system researchers for over a decade [4, 8, 17]. Norms in multi-agent systems are treated as constraints on behaviour, goals to be achieved or as obligations [6]. There are two main research branches in normative multi-agent systems. The first branch focuses on normative system architectures, norm representations, norm adherence and the associated punitive or incentive measures [3, 12]. The second branch deals with the emergence of norms.

2.2 Related work on emergence of norms

The work on norm emergence focuses on two main issues. The first issue is on norm propagation within a particular society. According to Boyd and Richerson [5], there are three ways by which a social norm can be propagated from one member of the society to another. They are a) Vertical transmission (from parents to offspring), b) Oblique transmission (from a leader of a society to the followers) and c) Horizontal transmission (from peer to peer interactions). Norm propagation is achieved by spreading and internalization of norms. Boman and Verhagen [4, 18] have used the concept of normative advice (advice from the leader of a society) as one of the mechanisms for spreading and internalizing norms in an agent society. The work done by Savarimuthu et al. [15] uses a distributed approach for normative advise based on the notion of leadership. Another recent development is the consideration of the role of network topologies on norm emergence [13, 14]. Sen et al. [16] have experimented with the emergence of traffic norm using social learning.

In his well known work, Axelrod [2] has shown the role of meta-norms to be effective in norm emergence. He also discusses several other approaches that might be useful for norm establishment which include the role of power, reputation, internalization and punishment. The contribution of this paper to this area are two fold. Firstly, we investigate Axelrod's statement that social norms are better suited for preventing smaller defections when the enforcement costs are low using social simulations in the context of norm emergence. Secondly, we introduce the notion of common knowledge that can help sustaining norms in agent societies.

3. EXPERIMENTAL SETUP AND PARAMETERS

The agents in the virtual online society are conceived as particles moving in a 2 dimensional space of linear size L. This virtual environment can be considered as a communal region such as a park. The agents explore and enjoy the park by moving around. Collisions of these particles in the virtual space represent interactions between agents in a social space. Each collision corresponds to two agents observing each other's actions. When two agents interact (when they meet each other within certain area of the park) they can observe each other performing one of the two actions, Litter (L) or Not Litter (NL), i.e. keep the environment clean. The payoff matrix that corresponds to the littering scenario is given in table 1.

Table 1: Payoff matrix

	L	NL
L	0.5, 0.5	0.5,-0.5
NL	-0.5, 0.5	-0.5,-0.5

An agent in our society starts with a score (*s*) of 100. When an agent litters, it gets a payoff of 0.5 while the cost associated with non littering is -0.5. These are the payoffs to an individual agent. When an agent litters, it pollutes the area that belongs to the commons. So, this can be considered defecting the entire society. To model this aspect, each agent receives a negative payoff of 1/N (where N is the total number of agents in the society) for every agent that litters. So an agent's final payoff value is the sum of the individual payoff and the negative partial payoff to the society (1/N) obtained in case of a defection.

Let us now assume that the society does not have a law against littering and hence there is no centralized policing mechanism. In this scenario, any agent that believes that there should be no littering in the society, might choose to punish the other agent whom it observes littering. A punishment cost (P_{cost}) is incurred by the non-litterer when punishing a littering agent. Every agent in the society is initialized with an autonomy value from 0 to 9 based on a uniform distribution. Autonomy refers to the stubbornness of the agent. This value governs the number of punishments required by an agent to move from L to NL (change of strategies).

Another parameter that we have defined in the system is the minimum Survival score (S_{score}). When an agent's score *s* goes below S_{score} it changes its strategy (moves from L to NL or vice versa). S_{score} is set to 50 in our experiments. We define the Litter Level of the park (LL_{park}) at any point of time as the cumulative number of littering actions which is reset to 0 at certain intervals of time (at the end of every 1000 iterations in our experimental set up).

4. EXPERIMENTS AND RESULTS

4.1 Role of punishments with low enforcement cost

In the first experiment there are 100 agents, 50 agents of type L and 50 of type NL. In every society there will be certain percentage of agents that are vengeful enough to punish another agent when they observe certain behaviour that they consider to be inappropriate. Let us assume that there are certain percentage of non-littering agents that are punishers (*p*) (*p*=0.1, 0.25 and 0.5). P_{cost} is kept low (0.01) in these experiments.

In each iteration two agents are randomly chosen to interact. We conducted experiments upto 18000 iterations. At the end of the simulation, we observe whether a littering or non-littering norm emerges. In our experiments we consider a norm to have emerged if all the agents are either of type L or NL (100% norm emergence). In other works, the value for norm emergence has varied from 70% to 100%.

Figure 1 shows 6 different lines. For each *p* value, there are two lines, one respresenting the number of litterers and the other representing the number of non-litterers. As the number of litterers decrease, the number of non-litterers increase (and vice versa). For this reason, these lines for a given *p* value, are the mirror images of each other. It is of interest to observe whether the littering or non-littering group reaches the value of 100. All the 6 lines start from a value of 50. Note that non-litterers are represented using hollow symbols while the litterers are shown using solid symbols.

It can be observed from figure 1 that as the number of punishers increase, the norm against littering is established. When the values of *p* are 0.25 and 0.5, the system converges to a norm against littering. It is interesting to note that when *p*=0.1, the system initially moves towards a NL norm, but when the number of punishers are less, the punishers' score reaches the minimum threshold (S_{score}) and hence to ensure their survival, they become litterers. So, this results in a whole society of litterers.

Once there are adequate number of punishers, the cost of punishment is spread across the non-littering punishers, hence their individual scores do not reach the minimum threshold and they are successful in converting the litterers to become non-litterers. So, the important characteristics that governs this change are the autonomy of the individual agents (littering agents) and the minimum threshold for survival (S_{score}). If a society has large autonomy values and high threshold for survival, the system will end up with litterers.

This result can be observed in many social interactions. For example, when you go to a restaurant, if you were the first ones, you might be polite and keep your voice low when interacting with your friends. As the restaurant becomes crowded, you might observe the noise levels rising. As the noise level increases, there is no incentive for you to keep your voice down. Moreover, you might be forced to speak out loud as that is the only way you might be heard by others. This case is similar to the non-litterers becoming litterers after certain threshold is surpassed.

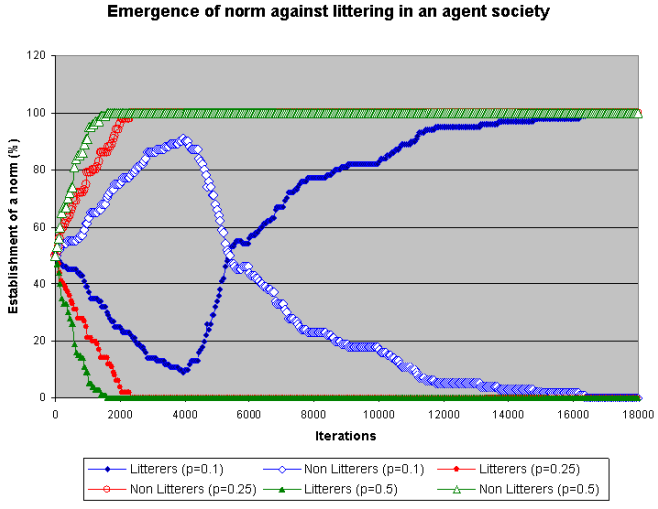

Figure 1: Emergence of norm against littering in an agent society

4.2 Role of punishments with high enforcement cost

This experiment shows that when the enforcement cost increases the system drives all the agents to litter. This experiment shows the behaviour of the system for three different values of punishment cost (P_{cost} = 0.1, 1, 10). There are 25% punishers in a society for all the three experiments. Again, note that non-litterers are represented using hollow symbols while the litterers are shown using solid symbols.

It can be observed from figure 2 that apart from the lower enforcement costs (P_{cost} = 0.1), the other higher values (P_{cost} =1, 10) result in littering. When P_{cost}=0.5 (not shown in the figure), the system oscillates between NL and L for different runs and for larger values of P_{cost}, the result is a littering society.

From this experiment it can be inferred that social norms can successfully be established and sustained against smaller defections when the costs of enforcements are low. But for norms that require larger costs of punishment (e.g. honour killing), social norms might not be very useful. In those cases, institutionalized mechanisms such as laws would be best suited.

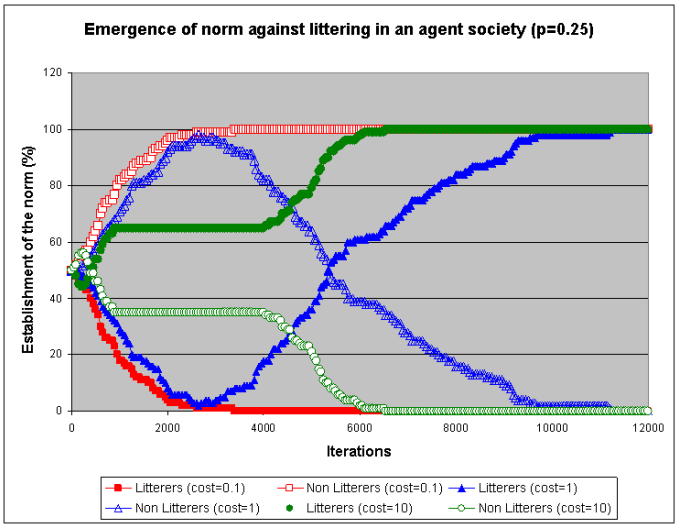

Figure 2: Emergence of norm against littering in an agent society (p=0.25)

4.3 Conditional punishment based on common knowledge

In human societies we tend to gather information about the state of the world through certain common knowledge sources such as newspapers, television, radio and even from some influential, well connected people. At any point of time, not everyone in the society might know about the current state of an issue or a problem. But, once some information is available through the common knowledge sources, it can be assumed that there would be an increase in the awareness of the situation in the society. Young Chwe [7] describes the role of common knowledge in solving coordination problems. In this section we describe our experiment on how common knowledge helps norm emergence in the context of social littering.

Let us assume that a common knowledge source is available (e.g. a newspaper). This common knowledge source periodically informs the agents about the state of the park. The agents in the society can choose to look at the information available from the knowledge source periodically. Based on the information available, the agents can choose to react. For example, whenever the park's litter level is greater than certain value ($LL_{park} > 50$ in our experiment), the non littering agents can choose to punish. For example, say there were only 10% of the non-litterers were punishers originally. After the information is known to all the other non litterers (remaining 40% of non-litterers), can choose to punish the litterers based on a conditional probability which is their vengefulness value. Each non-littering punisher agent has a vengefulness value (V) which is similar to the autonomy value and this is initialized at the start of the experiments. A non-littering punisher with vengencefulness value of 8 will punish a litterer 8 out of 10 times.

Figure 3 shows a comparison of punishment mechanisms with and without the use of common knowledge keeping P_{cost}=0.1 as a constant. The figure shows four lines that correspond to the establishment or fading of the non-littering norm. We have omitted the lines that show the trendlines of the littering norm for the sake of clarity of the diagram.

When p=0.1, the punishment mechanism that makes use of common knowledge results in a non-littering norm (hollow triangles) while the mechanism that does not use this mechanism results in

a littering norm (solid trianges). When p=0.25, the punishment mechanism that uses common knowledge (hollow squares) converges faster than the one that does not use it (solid squares). So, it can be inferred that the availability of common knowledge has increased the rate of establishment of a norm against littering in one case (p=0.25) and has resulted in the emergence of a new norm in another (p=0.1).

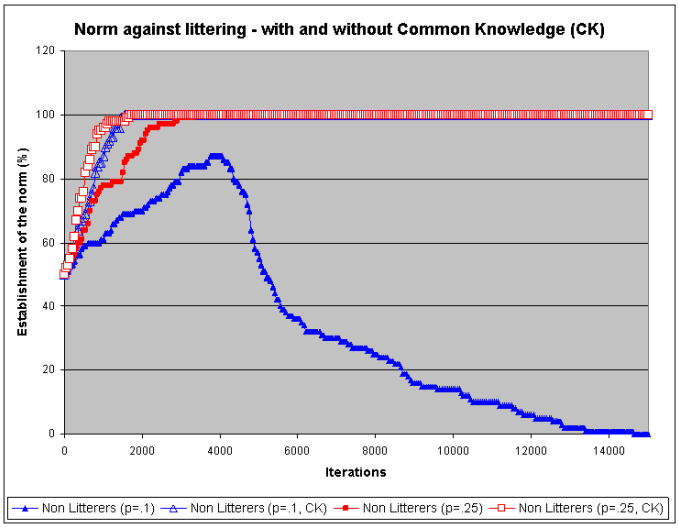

Figure 3: Comparison of emergence of norm against littering - with and without Common Knowledge (CK)

5. DISCUSSION AND FUTURE WORK

The results presented in this paper demonstrate that norms can be established through a bottom-up process based on a distributed, peer to peer punishment mechanism. Members of Wikipedia have been successful in establishing a norm of collaboration using a peer to peer mechanism based on careful scrutiny of the content. Wikipedia example makes it clear that more number of enforcers are needed for a system to work effectively.

The results obtained in our experiments are in agreement with Axelrod's statement that the norms are best in preventing smaller defections when the cost of enforcement is low. Also, it was shown that common knowledge can be used as a mechanism for improving norm establishment when used in conjunction with the punishment mechanism.

We agree that our results are preliminary. However, this work is aimed towards experimenting with mechanisms that might be suitable for generating norms in a bottom-up approach than a prescriptive top-down approach. In particular, our work is relevant for virtual online societies where behavioural norms should be derived by the agents themselves rather than adhering to an enforced law.

We are currently extending our simulation scenario to include agents of different personality types. Another extension to our system is to experiment with the emergence of different norms among different sub-groups within an agent society. To achieve that we need an application domain that has more states than a simple coordination game. Another important extension is to test the model on network topologies as agents evolve norms based on the influence from agents that they are connected to. The concept of distributed norm emergence is applicable to many applications in agent societies (e.g. buyer-seller scenarios in supply chain management, file sharing). A fertile ground for the study and experimentation of new mechanisms for norm emergence is the social networking applications.

6. REFERENCES

[1] Second Life. http://secondlife.com/.

[2] R. Axelrod. An evolutionary approach to norms. *The American Political Science Review*, 80(4):1095–1111, 1986.

[3] G. Boella and L. van der Torre. An architecture of a normative system: counts-as conditionals, obligations and permissions. In *Proceedings of The Fifth International Joint Conference on Autonomous Agents and Multi Agent Systems, AAMAS*, pages 229–231, New York, NY, USA, 2006. ACM Press.

[4] M. Boman. Norms in artificial decision making. *Artificial Intelligence and Law*, 7(1):17–35, 1999.

[5] R. Boyd and P. J. Richerson. *Culture and the evolutionary process*. University of Chicago Press, Chicago, 1985.

[6] C. Castelfranchi and R. Conte. *Cognitive and social action*. UCL Press, London, 1995.

[7] M. Chwe. *Rational Ritual: Culture, Coordination and Common Knowledge*. Princeton University Press, 2001.

[8] R. Conte, R. Falcone, and G. Sartor. Agents and norms: How to fill the gap? *Artificial Intelligence and Law*, 7(1):1–15, 1999.

[9] J. Elster. Social norms and economic theory. *The Journal of Economic Perspectives*, 3(4):99–117, 1989.

[10] J. Habermas. *The Theory of Communicative Action : Reason and the Rationalization of Society*, volume 1. Beacon Press, 1985.

[11] C. Horne. Sociological perspectives on the emergence of norms. *Social Norms (Hechter, M. and Opp, KD, eds)*, pages 3–34, 2001.

[12] F. López y López and A. A. Márquez. An architecture for autonomous normative agents. In *Fifth Mexican International Conference in Computer Science (ENC'04)*, pages 96–103, Los Alamitos, CA, USA, 2004. IEEE Computer Society.

[13] J. M. Pujol. *Structure in Artificial Societies*. PhD thesis, Software Department, Universitat Politénica de Catalunya, 2006.

[14] B. T. R. Savarimuthu, S. Cranefield, M. Purvis, and M. Purvis. Role model based mechanism for norm emergence in artificial agent societies. In *Proceeding of the AAMAS 2007 workshop on Coordination, Organization, Institutions and Norms in agent systems (COIN)*, pages 1–12, 2007.

[15] B. T. R. Savarimuthu, M. Purvis, S. Cranefield, and M. Purvis. Mechanisms for norm emergence in multi-agent societies. In *Sixth International Joint Conference on Autonomous Agents and Multi-agent Systems (AAMAS'2007)*, pages 1097–1099, 2007.

[16] S. Sen and S. Airiau. Emergence of norms through social learning. In *Proceedings of Twentieth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1507–1512, Hyderabad, India, 2006. MIT Press.

[17] Y. Shoham and M. Tennenholtz. On social laws for artificial agent societies: Off-line design. *Artificial Intelligence*, 73(1-2):231–252, 1995.

[18] H. Verhagen. *Norm Autonomous Agents*. PhD thesis, Department of Computer Science, Stockholm University, 2000.