Believing Others: Pros and Cons

Anish Biswas, Sandip Debnath & Sandip Sen
Department of Mathematical & Computer Sciences
University of Tulsa
abiswas,debnath@euler.mcs.utulsa.edu, sandip@kolkata.mcs.utulsa.edu
Phone: 918-631-2985, FAX: 918-631-3077

Abstract

In open environments there is no central control over agent behaviors. On the contrary, agents in such systems can be assumed to be primarily driven by self interests. Under the assumption that agents remain in the system for significant time periods, or that the agent composition changes only slowly, we have previously presented a prescriptive strategy for promoting and sustaining cooperation among self-interested agents. The adaptive, probabilistic policy we have prescribed promotes reciprocative cooperation shown to improve both individual and group performance in the long run. In the short run, however, selfish agents could exploit reciprocative agents. In this paper, we evaluate the hypothesis that the exploitative tendencies of selfish agents can be effectively curbed if reciprocative agents share their "opinions" of other agents. Since the true nature of agents are not known a priori and is learned from experience, believing others can also pose other hazards. We provide a learned trust-based evaluation function that is shown to resist both individual and concerted deception on the part of selfish agents.

Introduction

Recently, agent-based-systems (ABSs) have found increased usage both in the academia and industry (Bradshaw 1997; CACM July 1994 issue 1994; CACM March 1999 issue 1999; Huhns 1997). Agents provide a paradigm for building systems at a higher level of abstraction than the object-oriented paradigm. The component modules of such systems are more complex, more autonomous, and more goal-oriented. Of particular importance is the fact that ABSs can often act proactively to serve the user without explicit guidance. To be successfully adapted as the paradigm of choice, however, ABS technology has to provide more tools and mechanisms to ease the development of such systems, and provide both increased functionality and reliability than can be provided at the current stage.

With the burgeoning of agent based electronic commerce, recommender systems, personal assistant agents, etc. it is becoming increasingly clear that agent systems must interact with a variety of information sources in an open, heterogeneous environment. One of the key factors for successful ABSs of the future would be the capability to interact with other ABSs and humans in different role contexts and over extended periods of time. The ABSs of the future will be situated in a social context, playing a variety of roles in different relationships and problem solving situations. Borrowing on the social cliche leveled at humans, we would like to conjecture the following about the agents of the future: *Agents must be social entities.*

Research in societal aspects of agent behaviors, unfortunately, has been relatively scarce. Whereas economic models may provide a basis for structuring agent interactions (Wellman 1993), other approaches inspired by non-monetary relationships (Armstrong & Durfee 1998; Axelrod 1984) may provide more effective social relationships in certain situations. We have been interested in agent strategies for interactions with other agents that can promote cooperation in groups. our approach is different from other researchers who have tried to design good social laws that can be imposed on agents (Shoham & Tennenholtz 1992). In particular, we have studied environments where agents stand to gain from each other over sustained interactions. The goal of our work is to develop strategies that promote cooperation among homogeneous groups and can resist exploitation by malevolent agents. Such strategies can lead to both improved local performance for individual agents and effective global behavior for the entire system. These are the desirable features for open systems where self-interested agents are required to share resources.

We have developed and analyzed probabilistic reciprocity schemes as strategies to be used by self-interested agents to decide on whether or not to help other agents (Sen 1996). The goal of this work has been to identify procedures and environments under which self-interested agents may find it beneficial to help others. We claim that if the group composition changes only slowly, and there is sustained interaction between the agents. Probabilistic reciprocity strategies are considerably more sophisticated than deterministic reciprocity schemes like tit-for-tat (Axelrod 1984; Cesta & Miceli 1996) and avoids major problems associated with the latter schemes (Sen 1996).

We have experimentally evaluated the probabilistic reciprocity mechanism in multiagent domains where agents can exchange tasks with other agents. An agent decides to help another agent if it does not have too negative a balance with that other agent. The mechanism provides parameters to set the risk tolerance level of the agent designer, i.e., the designer may design agents that are ready to help others even with a large negative balance of help, or design agents that are quick to shun agents with which they have any outstanding balance of help. Our experiments under a variety of environmental conditions, group composition, work estimate difference, etc. have shown that under prolonged interaction the probabilistic reciprocity strategy produces close to optimal individual and group performance. Additionally, this strategy is stable against selfish intruders, i.e., in the long run, selfish agents perform worse than reciprocative agents in a mixed group.

We now turn to the focus of the current paper. Even though probabilistic reciprocative agents outperform selfish agents in mixed groups, they still waste some efforts in helping out selfish agents. This is because the reciprocative agents have a bias to initiate help to promote cooperative relationships in the future. A selfish agent can then benefit from this initial cooperative advances from each of the reciprocative agents in a mixed group. This is aided by the fact that reciprocative agents do not share their experiences or impressions of the other agents. In other words, there is no "words of mouth" transmission of the reputation or reliability of the agents in the agent group.

A hypothesis that follows easily from the above observation is the following: *Sharing of experiences about other agents among reciprocative agents will limit the exploitative gains of selfish agents.* Operationalizing this hypothesis, however, requires a closer inspection of the issues at hand. Since it is not clear a priori who is a selfish agent and who is a reciprocative agent (otherwise this whole exercise is moot because accurate identification immediately gives the right strategy to adopt while interacting with others), at the outset it is not possible to limit sharing of experiences only between selfish individuals. When an agent Z decides to use information supplied by an agent X to decide whether or not to help agent Y, then believing X can be advantageous or disadvantageous to Z based on the true nature of X. If X is selfish, it might find it useful to taint Y's reputation, and that of other agents, so that Z will consider X to be a relatively trustworthy agent. As such, we need to augment the reciprocative agents' strategy to believe only the agents who are trustworthy. In this paper, we evaluate the effectiveness of these strategies in mixed groups.

Reciprocity as an adaptive mechanism

The evolution of cooperative behavior among a group of self-interested agents have received considerable attention among researchers in the social sciences and economics community. Researchers in the social sciences have focused on the nature of altruism and the cause for its evolution and sustenance in groups of animals (Krebs 1970; Schmitz 1993; Trivers 1972). Mathematical biologist and economists have tried to explain the rationality of altruistic behavior in groups of self-interested agents by proposing various fitness models that analyze the success of altruistic individuals and more importantly the evolution of genetic traits supporting altruistic behavior (Dugatkin *et al.* 1994; Nee 1989; Nowak, May, & Sigmund 1995). Our goal in this paper is not to model altruistic behavior in animals; so we do not address the issues raised in the social science literature on this topic. Our purpose is to propose mechanisms by which cooperation can be encouraged and established in groups of self-interested agents. To this end, we have to compare and contrast and build upon the work reported by game theorists and economists on this topic. Space limitations do not permit a thorough review of the literature. Hence, we first identify a common trait in most of this body of work that we have surveyed, identify some underlying problems with the common trait, and then motivate how our proposed approach addresses these problems.

Most of the work by mathematical biologists or economists on the evolution of altruistic behavior deals with the idealized problem called Prisoner's dilemma (Rapoport 1989) or some other repetitive, symmetrical, and identical 'games'. Some objections have already been raised to using such sanitized, abstract games for understanding the evolution of complex phenomena like reciprocal altruism (Boyd 1988). In the following we analyze in some detail one of the often-cited work that share the typical assumptions made by economists and mathematical biologists, and then present our own set of suggestions for relaxing the restrictive assumptions made in that work.

In a seminal piece of work Robert Axelrod has shown how stable cooperative behavior can arise in self-interested agents when they adopt a reciprocative attitude towards each other (Axelrod 1984). The basic assumptions in this work include the following: agents are interested in maximizing individual utilities and are not predisposed to help each other; agents in a group repeatedly interact over an extended period of time; all interactions are identical (they are playing the same "game" again and again); agents can individually identify other agents and maintain a history of interactions with other agents; individual agents do not change their behavioral strategy over time; composition of agent groups change infrequently and the changes are minimal (only a few agent leaves and joins a group at a time). Using primarily simulated games, and, to a lesser extent, theoretical analysis, Axelrod convincingly argues for the effectiveness of simple behavioral rules for a variety of agent interactions. Specifically, he shows that a simple, deterministic reciprocal scheme of cooperating with another agent who has cooperated in the previous interaction (this strategy, for obvious reasons, is referred to as the *tit-for-tat* strategy), is quite robust and efficient in maximizing local utility. Whereas such a behavioral strategy can be exploited by strategies designed for that purpose, in general, the tit-for-tat strategy fairs well against a wide variety of other strategies. Two properties of the tit-for-tat strategy deserve special mention:

- if all agents use this strategy, system performance is optimal,
- it is stable against invasion by selfish agents (i.e., if an agent who never returns help is introduced into a group of tit-for-tat agents, the former cannot obtain greater utility than that obtained by tit-for-tat agents).

Though Axelrod's work is interesting and convincing, we believe that the assumptions used in his work makes the results inapplicable in a number of domains of practical interest. We now analyze some of this critical assumptions, identifying how they are violated in domains of practical interest, and motivate the need for an alternative framework for reciprocal behavior (we believe the term *reciprocal behavior*, as compared to the term *altruistic behavior*, more appropriately summarizes the motivation and mechanism that we use) that avoids these unrealistic assumptions:

Initial decision: If the first decision is to defect, rather than cooperate, tit-for-tat produces completely selfish behavior in homogeneous groups!

Symmetrical interactions: Axelrod assumes that every interaction is perfectly symmetrical, and the payoff from cooperation is identical to both parties. More frequently in real-life interactions, a cooperating agent incurs a cost to save some work of another agent. While individual interactions are asymmetrical, averaging over an ensemble of interactions can put one agent as many times in the position of the benefactor as in the position of the beneficiary.

Repetition of identical scenarios: It is unlikely that identical situations will recur in real life.

Lack of a measure of work: Since all interactions are assumed to be identical, there is no need to measure the cost of cooperation. Real life scenarios present differing circumstances which need to be compared based on some common metric.

Hence, the simple reciprocative strategy is not the most appropriate strategy to use in most real-life situations because most of the underlying assumptions that motivate its use are violated in these situations. Our proposal is for agents to use a reciprocity-based interaction scheme that is based on more realistic assumptions. More specifically, we believe that a probabilistic, rather than a deterministic reciprocity scheme is more suitable for real-life problems. Such a scheme should have at least the following properties:

- allow agents to initiate cooperative relationships (this implies that it should be able to handle asymmetrical interactions),
- use a mechanism to compare cooperation costs,
- allow agents to be inclined to help someone with whom it has a favorable balance of help (have received more help than have offered help),
- be able to flexibly adjust inclination to cooperate based on current work-load (e.g., more inclined to cooperate when less busy, etc.).

Probabilistic reciprocity

We assume a multiagent system with N agents. Each agent is assigned to carry out T tasks. The jth task assigned to the ith agent is t_{ij}, and if agent k carried out this task independently of other tasks, the cost incurred is C^k_{ij}. However, if agent k carried out this task together with its own task t_{kl}, the cost incurred for task t_{ij} is C^{kl}_{ij}. Also, the cost incurred by agent k to carry out its own task t_{kl} while carrying out task t_{ij} for agent i is C^{kij}_{kl}. In this paper, we allow an agent to carry out a task for another agent only in conjunction with another of its own tasks.

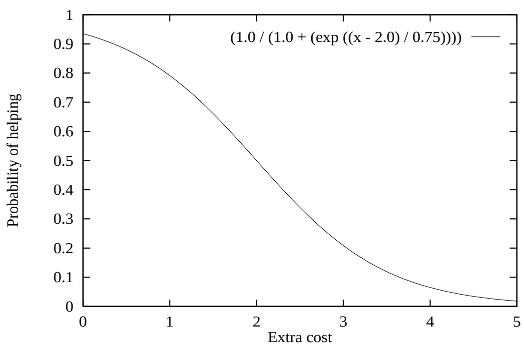

Figure 1: Probability distribution for accepting request for cooperation.

If an agent, k, can carry out the task of another agent, i, with a lower cost than the cost incurred by the agent who has been assigned that task ($C^i_{ij} > C^{kl}_{ij}$), the first agent can cooperate with the second agent by carrying out this task. If agent k decides to help agent i, then it incurs an extra cost of C^{kl}_{ij} but agent i saves a cost of C^i_{ij}. The obvious question is why should one agent incur any extra cost for another agent. If we consider only one such decision, cooperation makes little sense. If, however, we look at a collection of such decisions, then reciprocal cooperation can more than compensate for the immediate cost incurred in helping the other agent in the current interaction.

We now propose a probabilistic decision mechanism that satisfies the set of criteria for choosing when to honor a request for help that we described at the end of the previous section. We will define S_{ik} and W_{ik} as respectively the savings obtained from and extra cost incurred by agent i from agent k over all of their previous exchanges. Also, let $B_{ik} = S_{ik} - W_{ik}$ be the balance of these exchanges (note that, in general, $B_{ik} \neq -B_{ki}$). The probability that agent k will carry out task t_{ij} for agent i while it is carrying out its task t_{kl} is given by:

$$Pr(i,k,j,l) = \frac{1}{1 + \exp \frac{C_{ij}^{kl} - \beta * C_{avg}^{k} - B_{ki}}{\tau}}, \qquad (1)$$

where C_{avg}^{k} is the average cost of tasks performed by agent k, and β and τ are constants. This gives a sigmoidal probability distribution in which the probability of helping increases as the balance increases and is more for less costly tasks. We include the C_{avg} term because while calculating the probability of helping, relative cost should be more important than absolute cost.

We present a sample probability distribution in Figure 1. The constant β can be used to move the probability curve left (more inclined to cooperate) or right (less inclined to cooperate). At the onset of the experiments B_{ki} is 0 for all i and k. At this point there is a 0.5 probability that an agent will help another agent by incurring an extra cost of $\beta * C_{avg}^{k}$. The constant τ can be used to control the steepness of the curve. For a very steep curve approximating a step function, an agent will almost always accept cooperation requests with extra cost less than $\beta * C_{avg}^{k}$, but will rarely accept cooperation requests with an extra cost greater than that value. Similar analyses of the effects of β and τ can be made for any cooperation decision after agents have experienced a number of exchanges. In essence, β and τ can be used to choose a cooperation level (Goldman & Rosenschein 1994) for the agents. The level of cooperation or the inclination to help another agent is dynamically adapted with problem solving experience. Over time, an agent will adapt to have different cooperation levels for different agents.

Agent strategies

There are two types agents that we have used in our previous work on which we will expand on in this paper:

Selfish agents: Agents who will request for cooperation but never accept a cooperation request. Selfish agents can benefit in the presence of philanthropic agents by exploiting their benevolence.

Reciprocative agents: Agents that uses the balance of cost and savings to stochastically decide whether to accept a given request for cooperation.

The augmentations on these strategies are as follows:

Believing reciprocative agents: These are agents who use not only their own balance with another agent, but also the balances as reported by all other agents when deciding whether or not to provide help. More precisely, in place of using B_{ki} in Equation 1, a believing reciprocative agent k uses $\sum_{j \neq i} B_{ji}$ while calculating the probability of helping agent i[1].

[1] We assume that while k is deciding to help i it finds out the balances that everyone else has with i, but does not ask i itself about it. If k were to ask i about its balance with others, lying agents would be able to easily exploit k.

Earned-Trust based reciprocative agents: These agents also use combined balances, but includes balances of only those agents with whom it has a favorable balance. More precisely, in place of using B_{ki} in Equation 1, a conservatively trusting reciprocative agent k uses $\sum_{j \neq i \wedge B_{kj} > 0} B_{ji}$ while calculating the probability of helping agent i.

Individual lying selfish agents: These agents are designed to exploit the fact that believing or trusting reciprocative agents use balances provided by other agents. These agents reveal false impressions about other helpful agents to ruin their reputation. More precisely, when such an agent, j is asked for its balance with another agent i, it reveals B'_{ji} given by:

$$\begin{aligned} B'_{ji} &= C * (-B_{ji}), \text{ when } B_{ji} > 0 \\ &= B_{ji}, \text{ otherwise,} \end{aligned}$$

where C is a positive constant. This means that the more an agent i helps it, the larger the negative balance an individual selfish agent will report about agent i to other agents.

Collaborative lying selfish agents: These agents not only try to spoil the reputation of helping agents, but also collaboratively bolsters the reputation of other selfish agents or agents with whom it has zero balance. More precisely, when such an agent, j is asked for its balance with another agent i, it reveals B'_{ji} given by:

$$\begin{aligned} B'_{ji} &= C * (-B_{ji}), \text{ when } B_{ji} > 0 \\ &= \mathcal{P}, \text{ otherwise} \end{aligned}$$

where C is a positive constant as above and $\mathcal{P}$ is a large positive constant. Note that we assume that since the selfish agent never helps anyone, other agents with whom it has 0 balance is to be treated as selfish agents. This means, initially it treats all agents equivalently. Only when the reciprocative agents start helping it do these collaborative lying selfish agents turn against them!

Experimental results

In the simple package delivery problem that we have used for experimentally evaluating strategies, we assume there are N agents, each of which is assigned to deliver T packets. All the packets are located in a centralized depot. The packet destinations are located on one of R different radial fins, and at a distance between 1 and D from the depot. Agents can only move towards or away from the depot following one of the fins; they cannot move directly between fins. On arriving at the depot, an agent is assigned the next packet it is to deliver. At this point, it checks if any other agents are currently located in the depot. If so, it can ask those agents to deliver this packet.

The cost of an agent to deliver one of its packets individually is double the distance of the delivery point from the depot. If it carries another package to help another agent, it incurs one unit of extra cost per unit distance traveled when it is carrying its own packet and this extra packet. In addition, if it is overshooting its own destination to help the other agent, an additional cost measured as double the distance between the destination of its packet and the destination of the other agent's packet is incurred.

In this section, we present experimental results on the package delivery problem with agents using the reciprocity mechanism described in Section to decide whether or not to honor a request for cooperation from another agent. The number of agents and the number of packets to be delivered by each agent are chosen to be 100 and 500 respectively. The other parameters for the experiments are as follows: $R = 4$, $D = 3$, $\tau = 0.75$, and $\beta = 0.5$. Each of our experiments are run on 10 different randomly generated data sets, where a data set consist of an ordered assignment of package deliveries to agents. All the agents are assigned the same number of deliveries. The evaluation metric is the average cost incurred by the agents to complete all the deliveries.

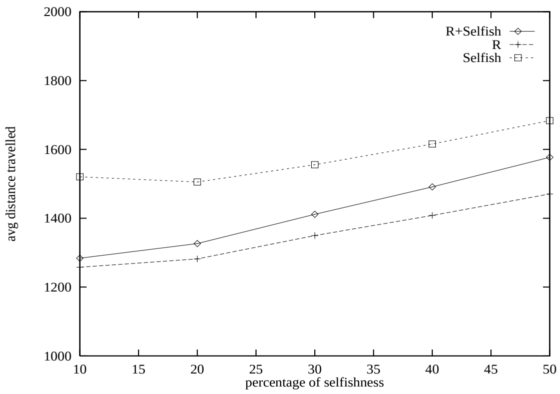

Figure 2: Performance of Reciprocative (R) and Selfish agents in mixed groups.

The first set of experiments we report is from our previous work where reciprocative and selfish agents are evaluated in mixed groups while varying the percentage of selfish agents. From the corresponding results presented in Figure 2 we see that though the selfish agents are able to exploit the reciprocative agents somewhat (if they had to deliver all of their packets by themselves, their average distance traveled would be approximately 2000), they still cannot outperform the reciprocative agents for a wide range of group mix. Since exploitation by the selfish agents adversely affects the performance of the reciprocative agents, we conjectured that if the reciprocative agents could share their balances, an agent that receives help from others but never helps back will be identified early by everyone. Such early identification will severely limit the exploitative potential of these selfish agents and also enable the reciprocative agents to perform better by eliminating cost incurred in helping these selfish agents.

In the next set of experiments we evaluated mixed

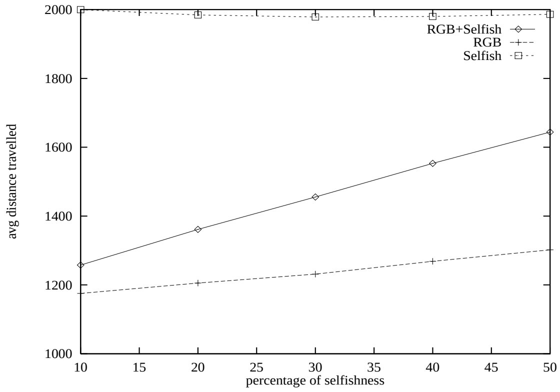

Figure 3: Performance of believing Reciprocative (RGB) and Selfish agents in mixed groups.

groups of believing reciprocative agents and selfish agents. As we see from the results presented in Figure 3, the sharing of balances does indeed severely restrict the exploitative edge of the selfish agents. In groups where they are a small minority, they have to do almost all of their work by themselves. In groups where they are a larger percentage of the group size, they get some leverage out of the fact that only few reciprocative agents are present to share their balances. As expected, the early identification of selfish agents also enable the reciprocative agents to improve their performance significantly. The problem with this approach is that since a reciprocative agent gets balances from everyone else (since it does not know a priori which of the others is selfish or cooperative), the selfish agents has the incentive to undermine the process by giving false balances about other agents.

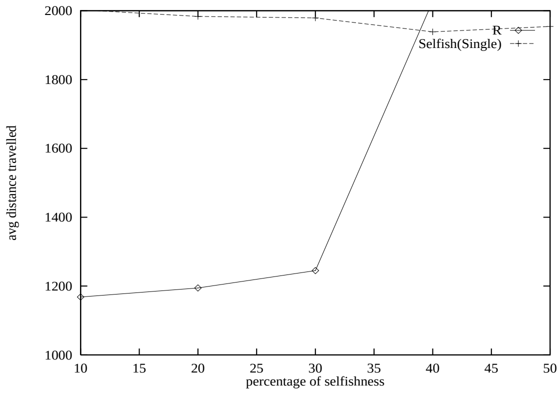

Figure 4: Performance of believing Reciprocative and Individual lying Selfish agents in mixed groups.

In the next set of experiments, we experiment with mixed groups of believing reciprocative agents and individual lying selfish agents. From Figure 4 we observe that when there are few selfish agents, their lying behavior does not noticeably affect the performance of believing reciprocative agents. But as the the percentage of such lying agents increases above a threshold of about 35%, critical mass of negative information surmounts the positive impression created by mutual help between reciprocative agents. At this point the reciprocative agents stops helping each other, and since they do not receive any help from selfish agents, they end up doing all of their work by themselves. Interestingly enough, the lying agents still appear to be able to get some help from the reciprocative agents. The other, more sinister, form of lying can occur when selfish agents collude to not only vilify the reputation of reciprocative agents, but falsely tout the helpful nature of themselves. The believing reciprocative agent will be gullible enough to sway by this false group impression which will even override any negative balance it might have with those agents. This is actually the other extreme of the effect of group balances: instead of rightly identifying "bad guys", now one will wrongly identify the bad guys as "good guys."

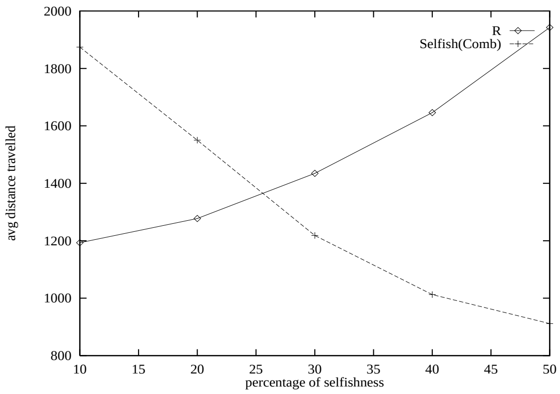

Figure 5: Performance of believing Reciprocative and Collaborative lying Selfish agents in mixed groups.

In this set of experiments, we experimented with mixed groups of believing reciprocative agents and collaborative lying selfish agents. From Figure 5 we observe that the collaborative lying agents are able to exploit the reciprocative agents quite effectively and overwhelms them when their percentage in the group is more than about 25%. In contrast to the individually lying agents, the collaborative lying agents not only cause poor performance of reciprocative agents, but saves itself a lot of problem solving cost by receiving help from the reciprocative agents. It is clear that collaborative lying is a threat which if not countered will make the believing reciprocative strategy unstable. One can always revert to using the base reciprocative agent, which does not believe others, and hence is not susceptible to either individual or group lying. But then we have to be happy to concede some non-trivial exploitation by even non-lying selfish agents. Our conjecture was to alter the believing reciprocative agent strategy by believing only those agents who have proven to be trustworthy based on past experience. That is, if someone has consistently been of help, it is reasonable to believe its opinion. Whereas it is unwise to believe someone who has not returned help-giving behaviors. We believed that such a learned-trust based reciprocative agent strategy may withstand both individual and collaborative lying by selfish agents.

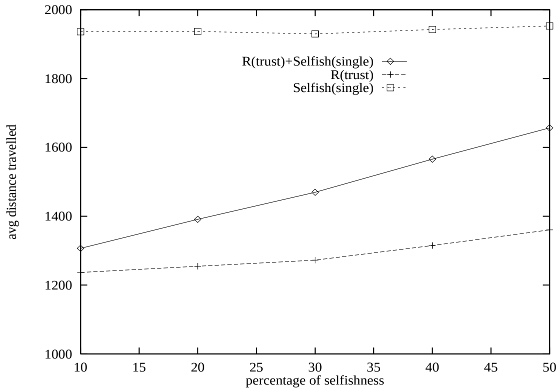

Figure 6: Performance of learned-Trust based Reciprocative, R(Trust), and Individual lying Selfish, Selfish(Single), agents in mixed groups.

In this set of experiments, we evaluated mixed groups of learned-Trust based Reciprocative (RGB) and Individual lying Selfish agents. Results presented in Figure 6 show a clear improvement in performance of reciprocative agents. When compared with Figure 4, we see that selfish agents get some help from the learned-trust based reciprocative agents compared to believing reciprocative agents. The amount of help received by the lying selfish agents is still much less than what the selfish agents received from reciprocative agents in our previous work (see Figure 2). An interesting observation is the level of exploitation and hence the performance of selfish and reciprocative agents vary only by a small amount over different group mixes. This set of experiments clearly demonstrated that learned-trust based reciprocative agents can effectively handle lying selfish agents (this also means they will be able to handle selfish agents who do not lie).

In the last set of experiments, we evaluated mixed groups of earned-trust based reciprocative and collaborative lying selfish agents. From the results in Figure 7 we see that as in the previous case, the learned-trust based reciprocative agents are able to distinguish between themselves and the lying selfish agents. It is interesting to note that comparing figures 6 and 7 we find that the collaborative lying agents perform even worse than individual lying agent when pitted against the learned-trust based reciprocative agents. Thus, it is convincingly demonstrated that the learned-trust based reciprocative limit exploitation of all the different kinds of selfish agents we have studied.

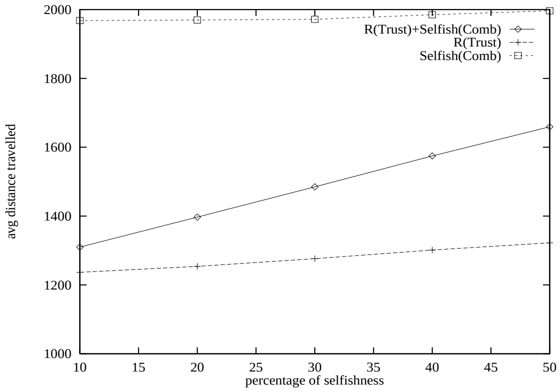

Figure 7: Performance of learned-Trust based Reciprocative, R(trust) and Collaborative lying Selfish, Selfish(Comb), agents in mixed groups.

Conclusions and Future Work

In this paper, we consider the effects of believing other agents' opinions when deciding to help an agent. We evaluate the effects of lying selfish agents, where both individual and group level exploitative schemes may be used. We study the probabilistic reciprocity based strategy to come up with individual and group based exploitative strategies. These schemes are shown to be able to "invade" a homogeneous group of reciprocative agents, thus making that strategy non-stable. While the reciprocity based strategy can be augmented by information received from other agents to counter individual exploitation by lying selfish agents, this augmentation is particularly susceptible to group exploitation. We introduce an experience based trusting mechanism for reciprocative agents that is able to successfully withstand invasion by both individual and group level exploitative schemes. The addition of the trust mechanism then restores the stability of the probabilistic reciprocity based strategy.

One of our future goals is to analytically capture the dynamics of the evolution of balance of helps in homogeneous and heterogeneous groups. For example, given a particular group composition and random interactions between members, how do the balances of selfish and reciprocative agents change as a function of time. Either difference or differential equation models can be constructed to represent the dynamics of these societies. In addition to identifying the ascendancy of exploitative or cooperative relationships, such models can also allow us to identify the formation of demes or working coalitions based on interaction histories.

References

Armstrong, A. A., and Durfee, E. H. 1998. Mixing and memory: Emergent cooperation in an information marketplace. In *Proceedings of the Third International Conference on Multiagent Systems*, 34–41. Los Alamitos, CA: IEEE Computer Society.

Axelrod, R. 1984. *The Evolution of Cooperation*. Basic Books.

Boyd, R. 1988. Is the repeated prisoner's dilemma a good model of reciprocal altruism? *Ethol. Sociobiology* 9:211–222.

Bradshaw, J. M. 1997. *Software Agents*. Menlo Park, CA: AAAI Press/The MIT Press.

1994. Communications of the ACM, july 1994, volume 37, number 7. Special Issue on Intelligent Agents.

1999. Communications of the ACM, march 1999, volume 42, number 3. Special Issue on Multiagent Systems on the Net and Agents in E-commerce.

Cesta, A., and Miceli, M. 1996. Help under risky conditions: Robustness of the social attitude and system performance. In *Proceedings of the Second International Conference on Multiagent Systems*, 18–25. Menlo Park, CA: AAAI Press.

Dugatkin, L.; Wilson, D.; III, L. F.; and Wilkens, R. 1994. Altruism tit for tat and 'outlaw' genes. *Evolutionary Ecology* 8:431–437.

Goldman, C., and Rosenschein, J. S. 1994. Emergent coordination through the use of cooperative state-changing rules. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, 408–413.

Huhns, M. N. 1997. *Readings in Agents*. San Francisco, CA: Morgan Kaufmann.

Krebs, D. 1970. Altruism – an examination of the concept and a review of the literature. *Psychological Bulletin* 73(4):258–302.

Nee, S. 1989. Does hamilton's rule describe the evolution of reciprocal altruism? *Journal of Theoretical Biology* 141:81–91.

Nowak, M.; May, R.; and Sigmund, K. 1995. The arithmetics of mutual help. *Scientific American* 272(6):76–81.

Rapoport, A. 1989. Prisoner's dilemma. In Eatwell, J.; Milgate, M.; and Newman, P., eds., *The New Palgrave: Game Theory*. London: Macmillan. 199–204.

Schmitz, D. 1993. Reasons for altruism. *Social Philosophy & Policy* 10(1):52–68.

Sen, S. 1996. Reciprocity: a foundational principle for promoting cooperative behavior among self-interested agents. In *Proceedings of the Second International Conference on Multiagent Systems*, 315–321. Menlo Park, CA: AAAI Press.

Shoham, Y., and Tennenholtz, M. 1992. On the synthesis of useful social laws for artificial agent societies (preliminary report). In *Proceedings of the National Conference on Artificial Intelligence*, 276–281.

Trivers, R. 1972. The evolution of reciprocal altruism. *Quarterly Review of Biology* 46:35–57.

Wellman, M. P. 1993. A market-oriented programming environment and its application to distributed multicommodity flow problems. *Journal of Artificial Intelligence Research* 1(1):1–22.