Predicting Agent Strategy Mix in Heterogeneous Populations

Sabyasachi Saha and Sandip Sen
Department of Math & CS
University of Tulsa
{sabyasachi-saha,sandip-sen}@utulsa.edu

Abstract

Prior research has identified agent behaviors or strategies that can develop and sustain mutually beneficial cooperative relationships with like-minded agents and can resist exploitation from selfish agents. Evolutionary tournaments with different strategies can model scenarios where agents periodically adopt strategies that are outperforming others in the population. However, such experiments can be computationally costly and hence it can be difficult to prescribe a strategy choice for a rational agent given environmental conditions like task distribution, strategy mix in the population, etc. A preferred approach, and the one pursued in this paper, is to analytically capture the dynamics of the strategy mix in the population under an evolutionary tournament. Such an analytical model can be used to predict the long-range winning strategy, which is the strategy to be preferred given the environmental conditions and the initial strategy distribution.

1. Introduction

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 agent based systems (ABSs) of the future would be the capability to interact with other ABSs and humans in different social and role contexts and over extended periods of time. Research in societal aspects of agent behaviors has been relatively scarce [9]. Whereas economic models can provide a basis for structuring agent interactions [14], other non-monetary approaches [1, 2, 3, 4, 5, 8, 10] may provide effective solutions in certain situations. We assume that typical real-world environments abound in *cooperation possibilities*: situations where one agent can help another agent by sharing work such that the helping cost of the helper is less than the cost saving of the helped agent. As agent system designers we can also define rules of interaction to increase the likelihood of cooperation possibilities. We are interested in identifying agent behaviors that allow agents to take advantage of cooperation possibilities in their environments.

Sen *et al.* [10, 11] have presented behaviors that promote cooperation among homogeneous groups and can resist exploitation by malevolent agents in heterogeneous groups. Such behaviors can lead to both improved local performance for individual agents and effective global performance for the entire system. A restrictive assumption in this line of work has been that agents have fixed behaviors. For example, they have assumed that agents with specified behaviors interact repeatedly over a sustained period of time and their effectiveness is calculated as function of the total cost incurred to complete all assigned tasks. The resultant performance reflects cost incurred for local tasks, cost incurred to help other agents with their tasks, and savings obtained from others when help was received.

A more realistic scenario would be to give an agent the freedom of choosing from one of several of these behaviors and to change its behavior as and when it deems appropriate. An agent may be prompted to adopt a behavior if agents using that behavior is seen to be performing better than others. Such a behavior adoption method leads to an evolutionary process with a dynamically changing group composition of agent behaviors [12, 13]. In this paper, we present a mathematical model of the dynamics of the agent population. Our goal is to analytically determine the region of dominance for the different strategies based on the initial population composition, environmental conditions like number of tasks agents need to accomplish, the evolving criteria, etc.

We consider the problem domain where each of the agents are assigned some tasks. The cost of executing a task can be reduced or eliminated if help is obtained from another agent. An agent may be an *expert* of a task type. An *expert* requires less cost to accomplish a task. After all agents have finished processing their assigned tasks, their relative performances are tallied. This comprises one evaluation period, or *generation*, of the behaviors adopted. The behaviors adopted by the agents in the next evaluation period is determined by a performance-proportionate scheme where the probability with which an agent adopts a strategy increases with the average performance of agents employing that strategy in the most recent evaluation period. Thus, it is likely that more agents are produced with behaviors that generated above-average performance. New agent behavior assignments are made as follows: for each agent i, two agents are selected randomly from the population without replacement. Then, of these two selected agents, the behavior of the one with higher performance is adopted by agent i[1]. This leads to a propagation of successful behaviors or traits. As a result, if a behavior produces better performance in one evaluation period compared to other behaviors, we are likely to see more individuals adopting that behavior in the next evaluation period. This generational scheme is semantically equivalent to every agent periodically selecting its behavior based on the current relative performance of the set of available behaviors. This generational approach is akin to work on identifying "evolutionary stable strategy" [6].

The goal of this paper is to identify the dominant strategies under different environmental conditions including initial population composition, the frequencies of the tasks assigned to the agents and the selection criterion used for population evolution. We present a mathematical analysis of the dynamics of the agent population. Using this model we can predict the strategy that will eventually dominate the population given the initial configuration.

Now, it is well-known that if each agent was to perform only one task, i.e., the number of interactions between two agents were at most one, the selfish strategy will dominate the reciprocative strategy. On the other hand, if the group of agents were completely stable, i.e., the agents interacted with each other infinitely often, the reciprocative strategy will dominate the selfish strategy. The switch in dominance happens at an intermediate value of the number of tasks per agent, and is dependent on other environment factors like initial group composition. The goal of our mathematical analysis and predictive model is to identify this switch over point. Using such a predictive model, then, it is possible to generate a decision surface on the number of tasks required for the reciprocative strategy to dominate the other strategies given the initial strategy distribution in the population. So, if an agent knows the population configuration and number of tasks then from this analysis it can predict the population configuration after any time period and it can decide which strategy will be dominant.

1 Selection of the best candidate from a set of randomly selected candidates is known as *tournament selection* in the genetic algorithms literature [7].

2. Adaptation via Reciprocity

A significant body of work by mathematical biologists or economists on the evolution of altruistic behavior deals with the idealized problem called Prisoner's dilemma [2] or some other repetitive, symmetrical, and identical 'games'. To consider a well-known study in this area, Axelrod demonstrates that a simple, deterministic reciprocal scheme or the *tit-for-tat* strategy is quite robust and efficient in maximizing local utility [2]. Sen criticizes that 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 [10].

The evaluation framework used by Axelrod considers an evolving population composition by allowing propagation of more successful behaviors and elimination of unsuccessful ones. In this paper, we evaluate the variants of exploitative and reciprocative behaviors suggested by Sen *et al.* [11] in a generational framework as used by Axelrod [2]. This allows us to see what behaviors emerge to be dominant or are evolutionarily stable.

3. Probabilistic reciprocity

We now present our probabilistic reciprocity framework for deciding whether or not to help another agent. Each agent is assigned to carry out T tasks. The mth task assigned to the ith agent, t_{im}, will cost it C_{ij} if the mth task is of type j. However, if agent k carried out this task together with its own tasks, the cost incurred for task j by agent k is C_j^{kl} (no cost is incurred by agent i), where agent k is doing tasks of type l. If $C_{ij} > C_j^{kl}$, there exists a cooperation possibility as agent k can help agent i save C_{ij} by incurring a cost of only C_j^{kl}.

We define S_{ik} and W_{ik} respectively as the cumulative savings obtained from and extra cost incurred by agent i from agent k over all of their previous exchanges. Also, $B_{ik} = S_{ik} - W_{ik}$ is the balance of these exchanges (note that, in general, $B_{ik} \neq -B_{ki}$).

Sen [10] proposes a probabilistic decision mechanism that satisfies a set of criteria for choosing when to honor a request for help that was described at the end of the previous section. 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:

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

where C_{avg}^k is the average cost of tasks performed by agent k, and β and τ are constants. This is a sigmoidal probability function (not a probability distribution) where the probability of helping increases as the balance increases and is more for less costly tasks.

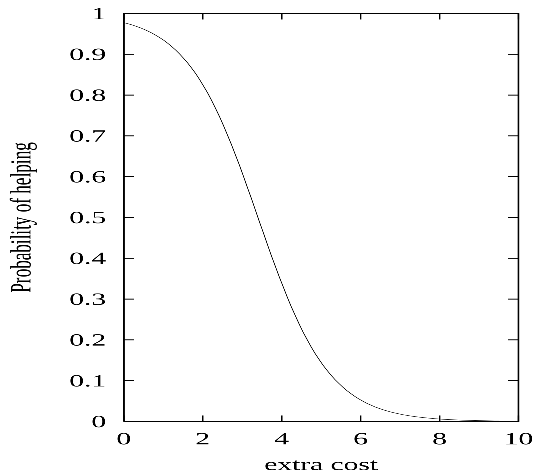

Figure 1. Probability distribution for accepting request for cooperation.

A sample probability distribution is presented in Figure 1. β can be set to a low value to move the probability curve left (less inclined to cooperate) or to a high value to move the curve to the right (more inclined to cooperate). Initially, $B_{ki} = 0$ for all i and k. At this point the probability that an agent will help another agent by incurring an extra cost of $\beta * C^k_{avg}$ is 0.5. 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 $\beta * C^k_{avg}$, but will rarely accept cooperation requests with an extra cost greater than that value.

These are the only two "parameters" in the equation 1 that can be fine tuned to adjust the level of cooperation. The other equation variables determine the actual dynamics of the agent behaviors. The level of cooperation or the inclination to help another agent is dynamically adapted based on past interactions with that agent. Note that the sigmoid is one of several function classes that can be used to represent a probabilistic reciprocity behavior.

4. Set of agent behaviors

In this paper we plan to include the following philanthropic, selfish and reciprocative agent types [10]:

Philanthropic agents: Agents that always honor a cooperation request irrespective of past experience.

Selfish agents: Agents who ask for help but never return favors. Selfish agents can thrive on the benevolence of philanthropic agents.

Reciprocative agents: Agents that use the probabilistic reciprocity scheme described above.

Another variant of the reciprocative strategy that we have considered here is as follows [11]:

Earned-Trust based reciprocative agents: While evaluating a request for help, these agents consider balances of only those agents with whom they themselves have favorable balances. 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. This behavior is an augmentation of the believing reciprocative agent and was required to counter false balance reporting by exploitative agents.

5. Decision surface

5.1. Decision surface formation

In this subsection, our objective is to discuss the mathematical analysis of the derivation of decision surface such that an agent can predict the dominant strategy given the environmental configuration available to it and without doing any experimentation or exploration in the domain. We consider the *proportion of selfish* and *proportion of philanthropic* agents as the independent variables. These two variables completely determine the initial strategy distribution in the population, because the sum of the proportion of the three different agents must be one. We want to find out, for each pair of independent variables, the minimum number of tasks required for the reciprocative strategy to dominate the other strategies.

Let there be N agents in the environment. In the initial population, $\langle p_s, p_p, p_r \rangle$ is the proportion of the selfish, philanthropic and reciprocative agents. In this paper, we have considered two types of tasks: *type1* and *type2*. The proportion of task types $\langle tp_1, tp_2 \rangle$ is assumed to be equal, *i.e.* $tp_1 = tp_2 = 0.5$.

Given the proportion of the initial population, one can find out the number of different types of agents. $N_{r,l}$, $N_{p,l}$ and $N_{s,l}$ are the number of *reciprocative, philanthropic* and *selfish* agents respectively, which are expert in task type $l = 1, 2$, where, $N_{r,l} = N * p_r * tp_l$, $N_{p,l} = N * p_p * tp_l$ and $N_{s,l} = N * p_s * tp_l$. We have to predict the evolutionarily dominant strategy given this initial configuration of the agent population and the total number of tasks per agent. From such a predictive mechanism, we will compute the minimum number of tasks for which the reciprocative strategy will become the evolutionarily dominant strategy.

$P(i, k, j, l)$ is the probability that agent k, if it is asked, will help agent i for a particular task t of type j when agent k is expert in tasks type l. This $P(i, k, j, l)$ is defined as,

$$P(i,k,j,l) = \frac{1}{1 + \exp^{\frac{C^{kl}_j - \beta * C^k_{avg} - B_{ki}}{\tau}}}, \text{if, k is reci and} j = l$$

$$
\begin{aligned}
&= 1, \text{if, k is phil. and} j = l \\
&= 0, \text{otherwise,}
\end{aligned}
$$

The first expression is defined in Equation 1. Since, B_{ki} is defined as $B_{ki} = S_{ki} - W_{ki}$ for all i, k, where S_{ki} and W_{ki} are cumulative savings from and extra cost incurred by agent k from agent i. Initially, $S_{ki} = 0$ and $W_{ki} = 0$ and hence $B_{ki} = 0$.

We need to calculate $P_1(i, k, j, l)$, which is the probability that for a task of type j in the task distribution of agent i, agent k, an expert in task type l, will be the one to help agent i. This event corresponds to the situation that all the agents asked before agent k will refuse to help and agent k will help.

$$
\begin{aligned}
P_1(i,k,j,l) &= \sum_{a=1,\neq i}^{N_{r,l}} [Pr(L_{k,a} \cap R(i,a,j,l)) * P(i,k,j,l)], \\
&\quad \text{if k is reciprocative} \\
&= \sum_{a=1,\neq i}^{N_{r,l}} [Pr(L^p_{k,a} \cap R^p(i,a,j,l))], \\
&\quad \text{if k is philanthropic} \\
&= 0, \text{ if k is selfish}
\end{aligned} \tag{2}
$$

where $L_{k,a}$ is the event that k is selected as the ath among the $N_{r,l}$ reciprocative agents that are expert in tasks of type l, i.e., after $a-1$ reciprocative agents expert in task type l. $R(i, a, j, l)$ is the event that all those $a - 1$ agents refuse to help agent i for tasks of type j. So,

$$
Pr(L_{k,a} \cap R(i,a,j,l)) = Pr(L_{k,a}) * Pr(R(i,a,j,l)|L_{k,a})
$$

where,

$$
Pr(L_{k,a}) = \frac{\binom{N_{r,l}-1}{a-1}}{\binom{N_{r,l}+N_{p,l}}{a}}
$$

and as agent decisions are independent,

$$
Pr(R(i,a,j,l)|L_{k,a}) = \prod_{t=1}^{a-1} (1 - P(i, a_t, j, l)),
$$

where a_t is the tth agent selected for asking for help. Now, since all the agents are starting with the same balance and probabilities to help the other guys it is immaterial in which order or who exactly are the agents that are selected before agent j is selected.

The probabilities $Pr(L^p_{k,a})$ are similar to the probability $Pr(L_{k,a}$ except that the numerator in the expression for the former contains $N_{r,l}$ instead of $N_{r,l} - 1$ (this is because for a philanthrop, all the reciprocatives may have already been asked). Also, the probability $Pr(R(i,a,j,l)|L_{k,a})$ is similar to the probability $Pr(R^p(i,a,j,l)|L_{k,a})$ except that the expression for the former uses a instead of $a - 1$ range of the product (for similar reasons as above).

Let us now consider the expected change of balance between two reciprocative agents i and k for a particular task of type l. One can compute the expected savings and expected spending of agent i for agent k by

$$
S_{ik} = S_{ik} + \sum_{l=1}^{2} P_1(i,l,k,l) * C_{i,l} * pt_l
$$

$$
W_{ik} = W_{ik} + \sum_{l=1}^{2} P_1(k,l,i,l) * C_{i,l} * pt_l
$$

and hence,

$$
B_{ik} = S_{ik} - W_{ik}.
$$

Using these B_{ik} values we can again find out the probability of helping for the next task. So, we can calculate the performance of an agent as the expected net wealth it will generate after processing all the assigned tasks. The expected net wealth generated by an agent is the total of the expected balance with the other agents.

As we have discussed earlier, at the end of each evaluation period, i.e., after every agent completes all the tasks they are assigned (may be with the help of other agents), the performances are tallied. At this point, agents will choose their strategies by the performance-proportionate scheme discussed in Section 1. This will determine our new expected population with $\langle p_s, p_p, p_r \rangle$ as the proportion tuple. This proportion will be approximated as the probability of an agent choosing the corresponding strategies.

To illustrate this point we now calculate the expected probability that an agent will adopt reciprocative strategy *i.e.* p_r. It is defined as,

$$
\begin{aligned}
p_r &= Pr(reci, reci) + Pr(reci, self) * Pr(reci \geq self) \\
&\quad + Pr(reci, philan) * P(reci \geq philan)
\end{aligned}
$$

where,

$$
Pr(reci, reci) = \frac{\binom{\lfloor N * p_r \rfloor}{2}}{\binom{N}{2}}
$$

and,

$$
Pr(reci, oth) = \frac{\binom{\lfloor N * p_r \rfloor}{1} * \binom{\lfloor N * p_{oth} \rfloor}{1}}{\binom{N}{2}}
$$

where, oth = selfish or philanthropic. Then the expected value of $Pr(reci \geq oth)$ can be found from the previously calculated expected performance. Similarly, we can find out p_p and p_s values and determine the expected new strategy distribution in the population in the next generation. We repeat this process until the agent population becomes homogeneous, i.e., all agents use the same strategy.

So, for a particular initial agent population and number of tasks available to each agent an expected dominant strategy is found. Using this information, in turn, we identify the region of dominance by the different strategies and the decision surface separating them, given environmental factors like number of tasks per agent and the starting strategy distribution in the population.

5.2. Analysis of the derived decision surface

In this subsection, we will identify and discuss the decision surface that separates the region of dominance of the different strategies. We consider, *proportion of selfish* and *proportion of philanthrop* agents are the independent variables. We vary the values of the *proportion of selfish* and *proportion of philanthrop* agents from 0.1 to 0.8. We set the other parameters in the environment as follows: $N = 60$, $\beta = 0.8$, $\tau = 0.9$. We have obtained the decision surface which we will discuss and analyze in this subsection.

For clearly describing the decision surface, we state the following two theorems. We are not including the proofs in the paper. For the proofs please refer to the following URL: `www.mcs.utulsa.edu/~sahasa/proofs.pdf`:

Theorem 1: *If both philanthropic and selfish agents are present in the population, the philanthropic agents will always be dominated by the selfish agents.*

Theorem 2: *In the presence of philanthropic agents, reciprocative agents cannot dominate the selfish agents.*

Since the philanthropic agents help others without considering their nature it is impossible for them to dominate the other strategies, as we have shown in Theorem 1. But their presence help the exploitative agents to dominate the environment. Since, the selfish agents always exploit the philanthropic agents, in the presence of selfish agents the performance of the philanthropic agents deteriorate sharply and the agents adopt other strategies. Also, as stated in Theorem 2, in the presence of philanthropic agents of different expertise, a reciprocative agent cannot perform better than the selfish agents. Only after the philanthropic strategy become *extinct*, does the reciprocatives have any chance of dominating the selfish agents. Within each evaluation period, the reciprocatives initially help the selfish agents and expects help from them. It takes some time for them to recognize the selfish agents and stop helping them. Once the reciprocative agents stop helping the selfish agents, the selfish needs to do all of the work by themselves and incur more cost compared to the reciprocatives who will be better off by exchanging help among themselves. But, if the number

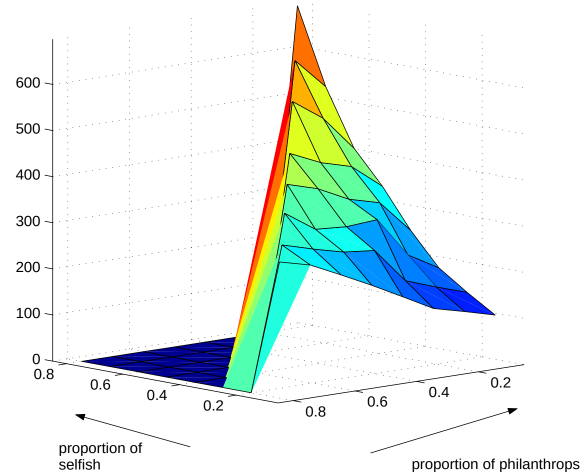

Figure 2. Decision surface separating the regions of dominance of basic reciprocative and selfish strategies.

of tasks within an evaluation period is small, the reciprocatives will not be able to compensate the initial cost it incurred by helping the selfish agents. So, for each pair of initial proportion of selfish and philanthrops, there will be a task number above which the reciprocatives are expected to be the dominant strategy.

In the first scenario, we model the evolution of a mix of basic reciprocative, selfish and philanthropic agents. In Figure 2, we show the decision surface separating the region of dominance of the reciprocative and the selfish strategies. For any point above the surface, reciprocative is the dominant strategy. Note that the dark region signifies the emptiness of the point *e.g.* the point $\langle 0.8, 0.5, * \rangle$ does not exist as the sum of proportions has to be less than or equal to one. We observe that as the proportion of selfish agent in the initial population is increased, while keeping the proportion of philanthrops constant (this implies a decrease in the initial proportion of reciprocative agents), it takes more tasks for the reciprocative to ultimately dominate the population. A similar trend is observed when the philanthrop proportion is increased keeping the selfish proportion constant.

To demonstrate the expected evolution of the population over a single evolutionary run, we plot the proportion of the three different strategies over different evaluation periods in Figure 3. Here the proportion of philanthrops are 0.2 with the rest of the population equally divided among selfish and reciprocative agents, and the number of tasks is 400. After two evaluation periods the floor value of the expected number of philanthrops become zero. In the presence of philanthrop agents, the expected performance of the selfish agents was better than that of reciprocative agents as they were exploiting the philanthrops. But after the phi-

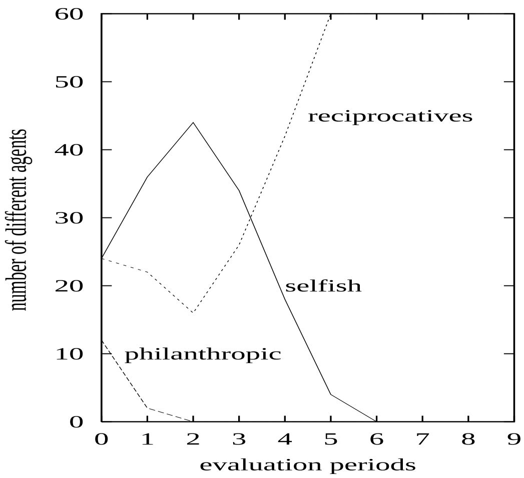

Figure 3. Expected number of agents of different types over evaluation periods. Initial proportion of selfish and philanthrops are 0.4 and 0.2 respectively and number of tasks is 400.

lanthrops die off, expected performance of the reciprocatives becomes dominant and more agents are expected to adopt reciprocative strategy. After about six evaluation periods the expected number of selfish agents drop down to zero leaving reciprocatives as the dominant strategy.

In the second scenario, we find out the decision surface separating the dominant regions of the *earned trust based* reciprocatives and the selfish agents (see Figure 4). In this case, the decision surface is lower than that in Figure 2, i.e., it takes much fewer tasks for the *earned trust based* reciprocative agents to dominate the selfish agents compared to that required by the basic reciprocative agents for the same initial proportions of selfish and philanthropic agents.

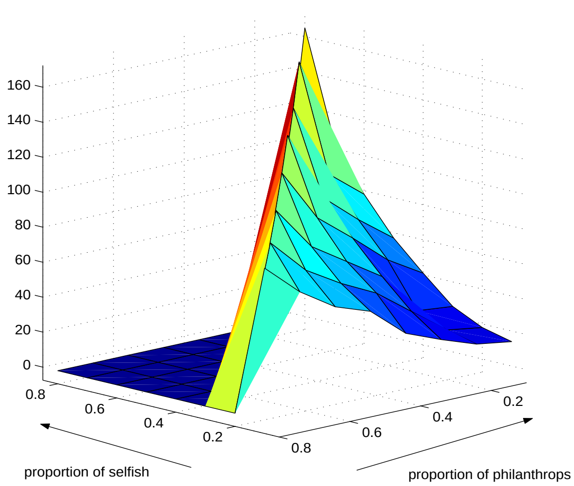

Figure 4. Decision surface separating the regions of dominance of *earned trust based* reciprocative and selfish strategies.

6. Conclusion and future work

In this paper, we have presented an analytical model for predicting the mix of different strategy distributions under an evolutionary scheme. Such an evolutionary scenario captures the dynamic of agents periodically adopting strategies that have been providing larger payoff in the current environment. Our goal has been to identify the evolutionarily dominant strategy given the starting strategy distribution and the number of tasks to be performed per iteration before agents reconsider changing their strategies. Our analytical model helps us to predict the dominant strategy given this information. More importantly such predictive analysis allows us to construct a decision surface separating regions of dominance of selfish and reciprocative strategies. As a result, a rational agent can choose the most beneficial strategy for the long run given the initial strategy profile in the population and the assigned task load.

As an extension of this work, we have planned to run a set of simulations to evaluate the accuracy of our predicted population mix and the tasks required before reciprocative strategies dominate.

References

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

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

[3] R. Azoulay-Schwartz and S. Kraus. Stable strategies for sharing information among agents. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence*, pages 1128–1134, 2001.

[4] C. Castelfranchi, R. Conte, and M. Paolucci. Normative reputation and the costs of compliance. *Journal of Artificial Societies and Social Simulation*, 1(3), 1998.

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

[6] T. Day and P. D. Taylor. The evolution of temporal patterns of selfishness, altruism, and group cohesion. *The American Naturalist*, 152(1):102–113, July 1998.

[7] K. Deb and D. Goldberg. A comparative analysis of selection schemes used in genetic algorithms. In G. J. Rawlins,

editor, *Foundations of Genetic Algorithms*, pages 69–93, San Mateo, CA, 1991. Morgan Kaufman.

[8] A. Glass and B. Grosz. Socially conscious decision-making. In *Proceedings of the Fourth International Conference on Autonomous Agents*, pages 217–224, New York, NY, 2000. ACM Press.

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

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

[11] S. Sen, A. Biswas, and S. Debnath. Believing others: Pros and cons. In *Proceedings of the Fourth International Conference on Multiagent Systems*, pages 279–286, 2000.

[12] S. Sen and P. S. Dutta. The evolution and stability of cooperative traits. In *Proceedings of the First Intenational Joint Conference on Autonomous Agents and Multiagent Systems*, pages 1114–1120, New York, NY, 2002. ACM Pres.

[13] K. Tuyls, K. Verbeeck, and T. Lenaerts. A selection-mutation model for q-learning in multi-agent systems. In *Proceedings of the Second Intenational Joint Conference on Autonomous Agents and Multiagent Systems*, pages 693–700, 2003.

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