Multiagent Benevolence as a Societal Norm

Abdulla M. Mohamed
Center for Information Technology
Department of Electrical & Computer Engineering
University of South Carolina
Columbia, SC 29201, USA
(803) 777-9540
mohamed@engr.sc.edu

Michael N. Huhns
Center for Information Technology
Department of Electrical & Computer Engineering
University of South Carolina
Columbia, SC 29201, USA
(803) 777-5921
huhns@sc.edu

Abstract

This paper describes an analysis of benevolent agents in multiagent systems (MAS). We first present a definition and motivation for benevolence that is appropriate for MAS. We then describe requirements for the structure and behavior of benevolent agents and construct a simulator, called Mattress In the Road (MIR), that can analyze and verify such requirements. Using MIR, simulations of benevolence are conducted and the results are analyzed thoroughly. Finally, we suggest some MAS applications that are driven by benevolent agents, and speculate about a more sociable Web due to benevolence as a norm.

Keywords

Organization and social structure; social simulation; benevolence

1 Introduction

Agents can exist and function alone or as part of a society. Each agent has a collection of goals or tasks that it will attempt to accomplish and, if the agent is part of a group within a society, some of the goals might be non-local. That is, each member agent will contribute some effort toward reaching a non-local goal. Each agent's contribution to its group is controlled by its behavioral characteristics, such as cooperation, altruism, friendliness, and benevolence.

A classic example of benevolence is the problem of a *mattress in the road:* an obstacle that can cause a traffic jam, because vehicles will have to slow down to maneuver around it. This results in a delay for everyone. A benevolent agent will stop and move the mattress out of the way so other agents can proceed on their way without any delay. Such an action would cause the benevolent agent more delay than if it just tried to avoid the mattress like everyone else, and the agent receives no immediate reward or compensating "benefit" for its action.

Software agents are unlikely to encounter mattresses, so where might a benevolent agent in an information system have an opportunity to behave benevolently? The agent could clean up stalled or failed transactions, close sockets that were left open by a process that terminated early, or remove locks set by failed or former processes. When it does not have either the authority or ability to take action, it can simply provide notifications to agents or systems that do.

So, what are benevolent agents? What role do they play in a multiagent system (MAS)? Philosophers, sociologists, psychologists, and biologists have studied the concept of benevolence for many years. Recently, researchers in AI have begun considering it, but they have mostly chosen a definition based on the mathematical utility for an individual agent. This definition is incomplete in our view. In the following sections, we argue that benevolence should also have a classical basis that recognizes the moral goodness of an agent and includes social awareness. We present a complete definition of benevolent agents and use it in our testbed, MIR, to demonstrate the role of these agents in an MAS.

2 Background

Nwana and Wooldridge state that agent technology is the most "rapidly" growing area in computer science, but there is no agreement among researchers on what an agent is [Nwana and Wooldridge 1997]. Similarly, there is no agreement on what a *benevolent* agent is. Definitions of benevolence for agents are split into two different strands. Researchers such as Castelfranchi, Conte, Jennings, Wooldridge, d'Inverno, and Luck define benevolent agents as those that accept all other agents' requests for help. For example, d'Inverno and Luck describe a benevolent agent as "an agent for the requesting agents" [Luck and d'Inverno 1996]. Other researchers, such as Rosenschein and Genesereth, define benevolent agents in terms of the similarities of their goals. They believe that benevolent agents have common or nonconflicting goals, and they call this part of the paradigm the *benevolent agent assumption* [Rosenschein and Genesereth 1985].

Goal adoption is classified into three types, namely, terminal, instrumental, and cooperative adoption. Terminal adoption, also called benevolent, occurs when an agent adopts others' goals without any personal advantages in mind, and the goal will not help the agent to achieve any of its own goals. Instrumental adoption occurs when an agent adopts others' goals with some personal advantage for itself. For example, feeding chickens helps them grow (satisfying their goal), and at the same time, it provides us with more food to eat (satisfying our goal). Finally, cooperative adoption happens when an agent adopts a goal because it is shared with another agent [Castelfranchi 1992].

In Castelfranchi's later work, he modified his view of benevolent agents: this work emphasizes the fact that a benevolent agent must adopt other agents' goals and interests without being asked by the recipient agents and even without the recipients' expectations [Conte and Castelfranchi 1995].

Jennings and Wooldridge define a benevolent agent as one that helps another agent whenever it is asked [Jennings and Wooldridge 1995]. Similarly, Jennings and Campos term benevolent agents as those that perform all goals that they are capable of on a first-come first-serve basis and accept all requests [Jennings and Campos 1997]. Moreover, Jennings and Kalenka, while describing a good decision-making function, select benevolence. The function of a benevolent decision is to "accept all requests made" [Jennings and Kalenka 1998].

Rosenschein defines benevolent agents as those that "hold common goals" [Rosenschein 1985]. In addition, he and Genesereth state that previous DAI studies assumed that all agents have nonconflicting goals. Researchers had focused on how agents could help each other achieve their common goals or how they could use common resources without interfering with each other. In reality, not all agents are benevolent; they don't all have common goals or help each other benevolently. Each agent has its own goals and intentions that it would like to achieve [Rosenschein and Genesereth 1985].

Others [Durfee et al. 1987] think that Rosenschein and Genesereth miscalled the agent that shares some goals a benevolent agent. In contrast, they think that these agents are selfish, because they only take actions that will help them achieve their own "interpretation" of the goals.

Sen investigated the circumstances in which one agent should help another agent perform a given task when the other agent requests help. The decision criterion is that this action should enable the agent who is conducting the help action to perform more effectively in the long run. For his experiment, Sen uses the principle of reciprocity, which means agents only help those agents who helped them in the past or can help them in the future. Sen's analysis and experiments show that reciprocal behavior improves the individual agent's performance in the long run over purely selfish behavior [Sen 1996].

3 Definition and Motivations for Benevolent Agents

Benevolent agents have been defined, characterized, and analyzed by a number of researchers, primarily computational and social psychologists. But other fields of science, such as philosophy and biology, addressed the concept of benevolence much earlier. Some used the term benevolence, whereas others used altruism to describe the same phenomenal behavior.

In 1871, Darwin suggested that a man helps other fellow men hoping to be helped back by others in the future [Darwin 1871]. One hundred years later, Trivers converted Darwin's idea to the *Theory of Altruism* [Trivers 1971]. Philosophers and biologists approach and describe benevolence as a pure concept of virtue, compassion, and moral sentiments. They describe the benevolent action as the doing of a kind action to another from mere good will and without any obligation; it is a moral duty only.

Computational scientists instead analyzed and measured benevolence in terms of individual costs and benefits. Most researchers ignored the origin of benevolence, whose long history in philosophy and biology explores virtue and moral duty. They thought that benevolence should not be taken for granted, but should be considered an important "phenomenon" that develops in societies of autonomous agents from exploration of agent emotions. Also, they think that in the present MAS theories, the description of benevolence is missing the emotional components [Bazzan et al. 1998].

So, what is the right approach to define and study benevolence? Should it be a pure moral or a pure individual benefits approach? A combination of both is what we are using. In other words, we take the concept of benevolence from where it originated—philosophy and biology—and apply it to computational agents. We are studying benevolence as a concept of goodness, social duty, and utility function.

An agent is benevolent if:

1. The agent *voluntarily* helps other agents without being commanded to do so.
2. The agent's benevolent actions are intended to benefit the society to which the agent belongs.
3. The agent should not expect an *immediate* reward or benefit for its benevolent actions. If it did, then the agent is instrumental, not benevolent [Conte and Castelfranchi 1995].
4. The agent's benevolent action is taken while the agent is pursuing one of its own goals in such a way that it should neither prevent nor help the agent accomplish its goal.

According to our definition of benevolent agents, benevolent actions should benefit the benevolent agents' society and will not stop them from reaching their goals. This will benefit the benevolent agent in the long run, i.e.; it is an indirect benefit. In other words, if the society is doing well, then all its members, including the benevolent agent, must be doing well too. Another motivation is the belief that the agent's benevolent actions may encourage others to act benevolently in the future, thereby providing compensation in the longer term. This relates to Blackmore's work on memes, where she states that altruism spreads altruism (meme-fountain) [Blackmore 1999]. It is important to understand that a benevolent entity can exist only in an environment with other entities, never alone.

Benevolent agents will not take a benevolent action if they will be harmed, i.e., if the action will prevent the agent from reaching its goals. In the *mattress in the road* example, an agent will pickup the mattress if the agent is not in a hurry and simply exploring the region, but will not pickup the mattress if one of its passengers is having a heart attack and needs to be rushed to a hospital.

4 Analysis of Benevolent Agents

The basic question we would like to answer is "When is benevolence useful or harmful for the agent and its society?" We expect that benevolence is beneficial to a society as a whole, and thus to each of its members, when it leads to an overall improvement in efficiency or results. However, we show below that it can be individually harmful if an agent spends all of its time performing benevolent actions and never makes any progress towards its own goals. It can also be harmful if only a small proportion of a society's members is willing to take any benevolent action and the rest of the members are not. But if all or many of a society's members are willing to undertake benevolent actions for the goodwill of their society, then there are situations where benevolent behavior will definitely be useful. Benevolence then becomes a societal norm.

For an individual agent, depending on its goals, benevolence might or might not be the appropriate behavior. For example, a "business" agent that needs to make the best deal on a contract will not take a benevolent action that will help its competitor agents. On the other hand, a search agent might perform the benevolent action of updating search engine results, which over time might reduce the overall traffic on the Internet, thereby benefiting all of the Internet's users.

A benevolent agent will have a list of goals that it needs to accomplish. At the same time, it will also work on achieving some of its societal goals. The actions taken by the agent should not in any way have negative impacts on its society. While the agent is working toward a goal, it might encounter a situation where a benevolent action is needed for the good of the others, but it is not part of the goal that the agent is striving toward. For example, an agent's main goal might be to clean a nuclear facility by picking nuclear waste up from the floor and dropping it at safe dumping areas. While the agent is carrying some waste and moving towards the dump area, it might encounter some obstacles. The agent will take a benevolent action by moving the obstacle out of the way. Such an action will cost the agent some time delay, because it could simply just avoid the obstacle. But this benevolent action will clear the way for other agents, so they do not have to waste their time trying to avoid it. Thus, this benevolent action helps all of the society's members as well as the benevolent agent itself in the long run.

The reward for a benevolent action is not immediate, and the results on the society will be observed over the long term. Measuring the effects of such actions is not straightforward. But in general, benevolent actions should assist the society of agents to accomplish their objectives, and at the same time not prevent individual agents from reaching their own goals.

5 Mattress In the Road Simulator

The Mattress In the Road (MIR) simulator is a tool we have constructed to simulate agents' benevolent behavior. MIR consists of agents and mattresses. Each agent enters the road with some entrance probability (default is 25 %) and moves along the road to the end. Some agents will drop mattresses accidentally according to some probability (default is 10%). Agents can be benevolent or non-benevolent. Benevolent agents will remove the mattress from the road with some delay cost (default is 10 clock cycles). Non-benevolent agents will avoid the mattress with less cost (default is 5 clock cycles) (see Figure 1).

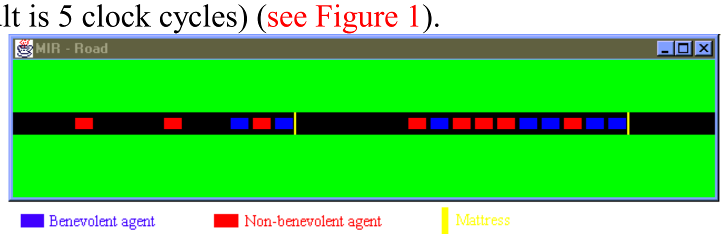

Figure 1: MIR road environment

Once an agent enters the road, it gets assigned a minimum and maximum time limit to complete the road. These time limits are selected randomly and are measured in terms of road length (default is 32 units in length). For example, when the minimum is 1 and the maximum is 3 (default values), it means that the agent will have a random time limit range from 32 to 96 clock cycles. If an agent uses more than its time limit to complete the road, than it is late and it increases the percentage of late arriving agents. A benevolent agent recomputes its time limit based on how long it has traveled and how long it has to go to complete the road. Once its time limit becomes equal to or less than the time needed to complete the road, the benevolent agent will not remove any mattresses from the road in order to reach its goal, which is to complete the road within its time limit. This demonstrates that such benevolent agents are indeed rational, since they do not take benevolent actions that harm or stop them from reaching their main goals.

In addition, we simulate an MAS that consists of both benevolent and non-benevolent agents. The mix percentage represents the ratio of benevolent agents to non-benevolent agents entering the road. For example, 10% means that 10 out of 100 agents will be benevolent. The control panel (see Figure 2) is used to enter all other environment and agent properties, such as clock cycle time (milliseconds), car entrance probability (%), mattress probability (%), number of clock cycles needed to remove the mattress (cycles), number of clock cycles needed to avoid the mattress (cycles), type of agents (non-benevolent, benevolent, or a mixture of both), and minimum and maximum time limits of agents.

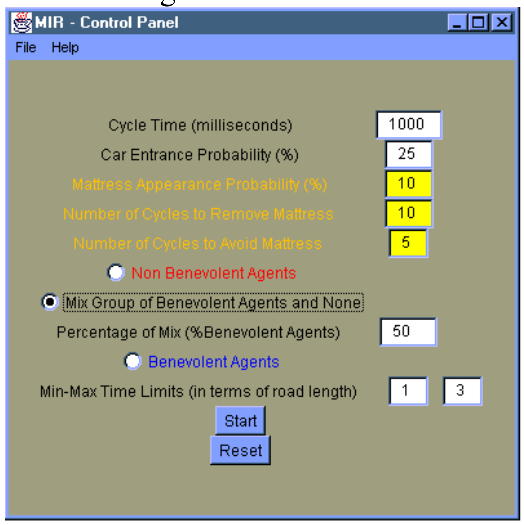

Figure 2: MIR control panel

Figure 3 shows the simulation result window. In this window, we display the current simulation time (cycles) that indicates the elapsed time. We also display the total number of cars that have completed the road. The average time to complete the road and the percentage of cars that arrive late are updated every time an agent completes the road. Figure 4 is a dynamic graph of the average time to complete the road as simulation time elapses.

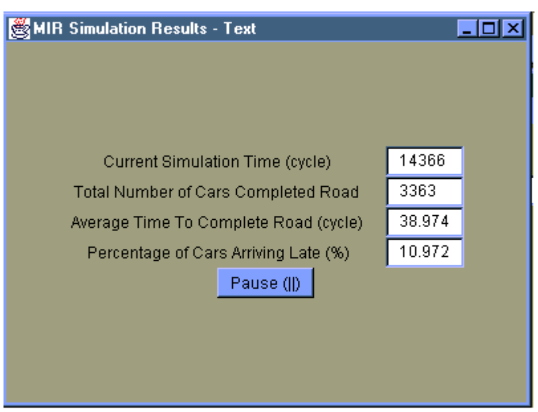

Figure 3: MIR simulation results text window

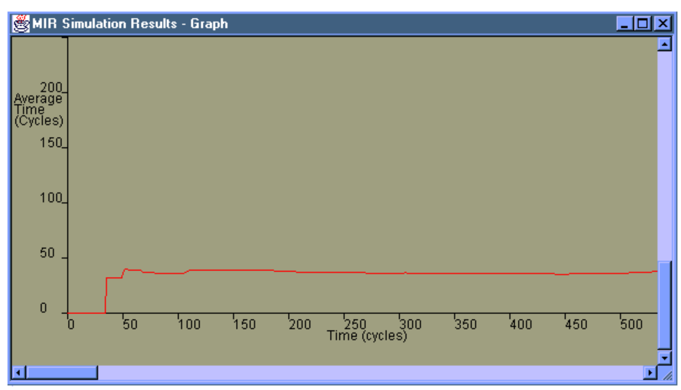

Figure 4: MIR simulation results graph window

6 Simulation

Is benevolence good or bad? Do we design our agents to be benevolent or non-benevolent? Is it necessary for all agents to be benevolent? The answers are strongly related to the application and the environment where the agents exist. For our MIR Testbed, to determine the benefits of an agent being benevolent or non-benevolent, a full study of all factors that could influence the dynamics of the agents and their environment was conducted. These factors include the traffic density (car appearance probability), mattress probability, and the percentage of benevolent to non-benevolent agents on the road. To measure the performance of the MIR, two values are used, namely, the average time required for cars to complete the road and the percentage of cars arriving late (see Figures 3 and 4).

A full multidimensional statistical analysis of the effects of these factors on benevolence is underway to decide whether benevolence is beneficial or not for an application. We are varying some of the factors and fixing others in order to conduct a statistical regression analysis. This type of analysis will assist us in evaluating and describing the performance of the MAS as a function of all factors. Using MIR, we will be able to write the following equation:

```
avg. completion time = f(traffic, mattress probability, % benevolent agents).
```

A complete study using MIR will be addressed in future papers.

For this paper, a number of simulations were run to delineate the scope of our ongoing research. The average completion time is plotted for two MAS's made of just benevolent or just non-benevolent agents vs. traffic density (section 7.1) and mattress probability (section 7.2). In addition, the percentage of benevolent to non-benevolent agents is varied and average completion time is measured and plotted (section 7.3).

7 Discussion of Simulation Results

Car entrances and mattress appearances are random values. If there is no car blocking the start of the road, then car entrance set to 25% means that a new car enters the road on the average of once every four clock cycles. Similarly, a mattress probability of 10% means that once every ten clock cycles a random location on the road is selected. If that location does not have any other mattresses, and it is behind a car, than a mattress is dropped at that location. This randomness is why some of the simulation results are not smooth curves.

In all simulations to date, only one factor of the environment is varied while the rest remain fixed. This provides an easy way to study how each factor effects the performance of the MAS. The factors under investigation are traffic density, mattress probability, and percentage of benevolent to non-benevolent agents. The performance measures is the average time needed for a car to complete the road.

7.1 Benevolence vs. Traffic Density

During the study of the effect of traffic density, the following values were fixed:

- Mattress probability = 10%
- Number of clock cycles to remove a mattress = 10 cycles
- Number of clock cycles to avoid a mattress = 5 cycles
- Minimum time limits = 1 and maximum time limits = 3 (unit length)
- Simulation time = 2000 clock cycles

The car appearance was varied from 0 to 100 %, and the result average completion time is plotted in Figure 5. From these graphs, we can see that the benevolent agents' performance is better than that of the non-benevolent agents regardless of traffic density. After 80% traffic density, there is a difference between the performances of the agent types, but it is not as significant as the difference in the case where the density is under 80%. Once traffic is high, the benevolent action of moving the mattresses from the street will help ease the flow of cars. But due to the high traffic density, cars will still be delayed while waiting for others to move. Thus, benevolent agents will always perform better than non-benevolent agents' will whether the traffic is low or high, but in high traffic; the difference is not very significant.

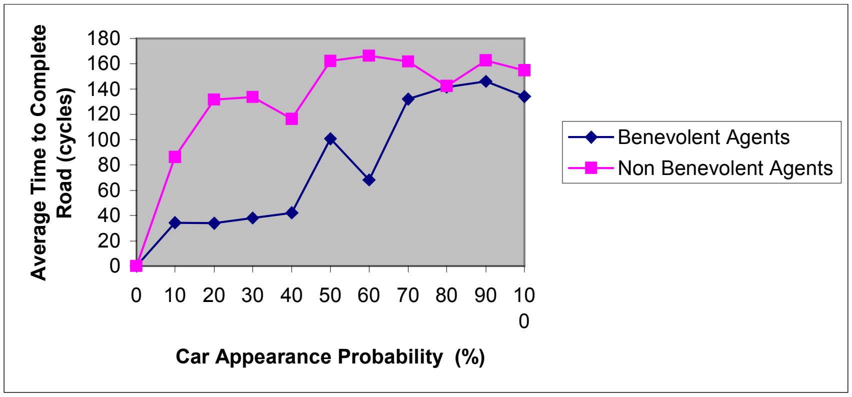

Figure 5: Average completion time vs. Traffic density

7.2 Benevolence vs. Mattress Probability

During the study of the effect of mattress probability, the following values were fixed:

- Car entrance probability = 25%
- Number of clock cycles to remove a mattress = 10 cycles
- Number of clock cycles to avoid a mattress = 5 cycles
- Minimum time limits = 1 and maximum time limits = 3 (unit length)
- Simulation time = 2000 clock cycles

The mattress probability was varied from 0 to 100 %, and the resultant average completion time is plotted in Figure 6. From this graph, we can see that for mattress probability up to 40%, the benevolent agents' performance is better than that of the non-benevolent agents. After 40%, there is no significant difference between the performance of benevolent and non-benevolent agents. In other words, as the mattress probability increases, benevolent agents spend as much time as non-benevolent agents on the road. This makes sense for two reasons. The first reason is that the time that benevolent agents will save for the others by removing the mattress from the road will only benefit a few others, since other mattresses will appear quickly due to the high mattress probability. For example, a benevolent agent will spend 10 clock cycles to remove a mattress that might benefit only one other agent before another mattress is dropped in front of it, so total delay time is 10 cycles. On the other hand, if the agents are non-benevolent, each will spend 5 clock cycles to avoid the same mattress, and the total delay time is equal to that of the benevolent agents (10 cycles).

The second reason is that benevolent agents are rational agents. They start by removing the mattresses from the road, but because of the high number of mattresses, they run out of spare time. Once their time limits become equal to or less than the time needed to complete the remaining distance of the road, they stop removing any mattresses from the road (become non-benevolent) in order for them to complete the road in time.

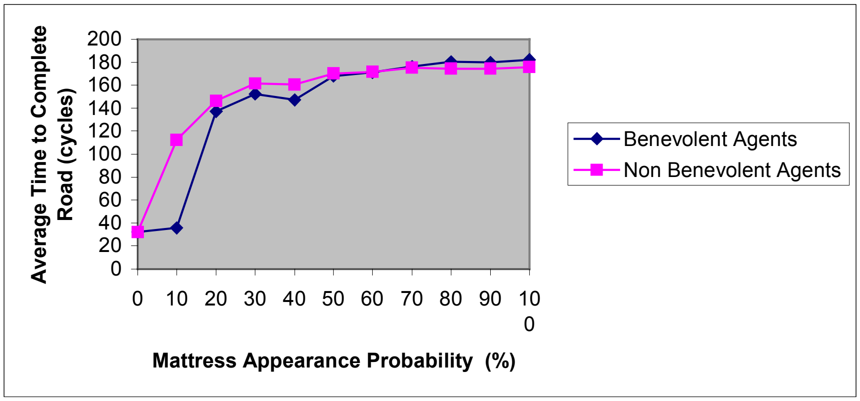

Figure 6: Average completion time vs. Mattress probability

7.3 Percentage of Benevolence in MAS

During the study of the effect of the percentage of benevolent to non-benevolent agents, the following values were fixed:

- Mattress probability = 10%
- Car entrance probability = 25%
- Number of clock cycles to remove mattress = 10 cycles
- Number of clock cycles to avoid mattress = 5 cycles
- Minimum time limits = 1 and maximum time limits = 3 (unit length)
- Simulation time = 2000 clock cycles

The percentage of benevolent to non-benevolent agents was varied from 0% (all non-benevolent agents) to 100% (all benevolent agents), and the result average completion time is plotted in Figure 7. From this graph, we can see that once the percentage of benevolent agents exceeds 20%, the MAS's performance increases dramatically. But another interesting point is that as the percentage of benevolent agents increases beyond 50%, the MAS's performance does not change significantly. Thus, benevolent agents help the MAS to perform better, and their existence is important. But we do not need every agent to be benevolent; we only need about half. In real life, we all wish to be benevolent, but sometimes circumstances force us to avoid taking benevolent actions.

This also supports our definition of benevolent agents and the way we modeled them in MIR. As we defined benevolent agents, we emphasized the fact that benevolent agents are rational and thus do not take any action that will stop them from reaching their goals. As a result, some benevolent agents will not take benevolent action if it will cause some consequences to them. In MIR, if you select to simulate an MAS that consists of pure benevolent agents, some of them will run out of time and will decide not to pickup mattresses from the road, becoming non-benevolent, in order to complete the road without being late (their main goal).

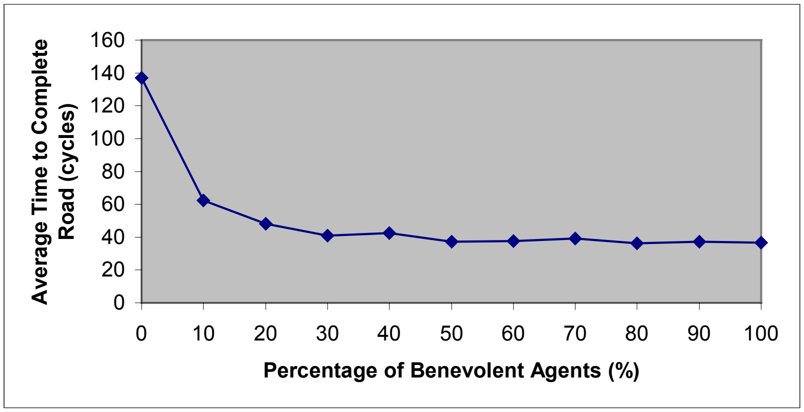

Figure 7: Average completion time vs. Percentage of benevolent to non-benevolent agents

8 Future Applications of Benevolent Agents

8.1 Collective Store Database

Parisi, Pedone, and Cecconi discuss the ideas of individual survival strategies (ISS) and social survival strategies (SSS). Social survival strategy employs a collective store (CS) to which all individuals in a group contribute some of their resources. The collective store in turn redistributes the resources to group members by some allocation criteria or converts the resources into something new. Resources may include essential provisions, money, or knowledge—or CPU time and data storage space. Through simulations, the researchers concluded that a group using a collective store could survive severe environmental conditions, while individuals without a collective store would perish. In addition, the raw resources that individuals contributed could be transformed into new resources that no single individual could produce [Cecconi and Parisi 1998] [Pedone and Parisi 1997].

The concept of a collective store strategy implies benevolent behavior of the agents. By examining our definition of benevolence, we clearly can see that all benevolence criteria are met. Each agent contributes its resources (or its surplus) to the collective store willingly (autonomy) and without any guarantee that it will receive the appropriate amount, or even anything at all, from the store. In addition, the store will decide who needs the resource the most and provide them with it, which in turn will benefit the group (rationality).

The collective store could be implemented as a large database of query results and information (see Figure 11). And benevolent query agents contribute their search result to this collective store database. When heavy Internet traffic degrades the search environment, the collective store database could help those agents seeking information on the Web. This is the basis for Internet search services such as Excite, Lycos, and AltaVista, except that users do not have to contribute anything in exchange for using these services. However, agents making greater contributions to a collective store might be given higher priorities in the subsequent use of the store. The collective store could refine the data submitted by different agents and derive new results through data mining techniques. Moreover, a collective store can gather data from agents that have better Web access capabilities and redistribute them to those with poorer capabilities, such as low-bandwidth PDAs [Huhns and Mohamed 1999].

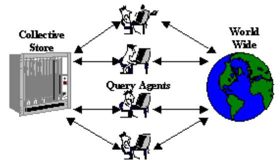

Figure 11: Collective Store Database

8.2 Benevolent Query Agents

One of the most common Web agents is a query agent. A query agent searches the Web to find an answer to a user's request, and in so doing it may visit many sites and databases. When asked, a benevolent query agent would freely share its query results with other agents on the Web, even though it may have consumed substantial resources to get this knowledge and might have to consume more to share it. Through one agent's benevolence, other agents charged with similar queries would not have to explore all the sites or databases the first explored: they can simply use its results. Thus, benevolent agents can help reduce Internet traffic, leading to faster Web processing for all [Huhns and Mohamed 1999].

9 Conclusion

There has been a lot of research on cooperation among agents, but benevolence has not been addressed comprehensively. Based on our model of benevolence, the society of a benevolent agent will benefit from its benevolent actions, and the benevolent agent will benefit in the long run by being a member of such a society. This societal compensation an agent receives through membership is why benevolence only makes sense within a society of agents that has goals shared by its members. On the other hand, benevolence might not be suitable for multiagent systems where there is competition for the same goal, such as money or power.

Moreover, as we begin spending more time on the Web, the demand will rise for agents that can perform Web tasks for us. Each agent will represent its owner, serving as the owner's surrogate for Web tasks and transactions. To be an effective surrogate, agents will have to be imbued with their owners' preferences and characteristics, such as cooperation, friendliness, sociability, and benevolence. Then the Web will be a friendlier and more productive environment for work, learning, and recreation. Social behavior such as benevolence will find its way very soon into Internet applications as we move toward a more sociable web.

References

[Bazzan et al. 1998] Ana L. Bazzan, Rafael H. Bordini and John A.Campbell, "Moral Sentiments in Multi-Agent Systems," *Pre-proceeding of the 5th International Workshop on Agents Theories, Architectures, and Languages (ATAL-98),* Research Note UCL-CS [RN/98/29].

[Blackmore 1999] Susan Blackmore, "*The Meme Machine*", Oxford University Press Inc., New York, 1999.

[Castelfranchi 1992] Cristiano Castelfranchi, "No More Cooperation, Please! In Search of the Social Structure of Verbal Interaction," *Communication from an Artificial Intelligence Prospective*, Ortony, A., Slack, J., and Stock, O. (eds.), Springer-Verlag, Berlin, 1992.

[Cecconi and Parisi 1998] Federico Cecconi and Domenico Parisi, "Individual versus social survival strategies," *Journal of Artificial Societies and Social Simulation*, vol. 1, no.2, The SimSoc Consortium, Dept. of Sociology, University of Surrey, Guildford, UK, 1998.

[Conte and Castelfranchi 1995] Rosaria Conte, and Cristiano Castelfranchi, "*Cognitive and Social Action,*" UCL press, London, UK, 1995.

[Darwin 1871] C. Darwin, "*The Descent of Man and Selection in Relation to Sex*", John Murray, London, 1871.

[Durfee et al. 1987] E. H. Durfee, V. R. Lesser, and D. D. Corkill, "Cooperation Through Communication in Distributed Problem Solving Network", *Distributed Artificial Intelligence,* M. N. Huhns (ed.), Kaufmann, San Mateo, California, 1987, pages 29-58.

[Huhns and Mohamed 1999] M. N. Huhns and A. M. Mohamed, "Benevolent Agents", *IEEE Internet Computing*, Vol.3, No. 2, Mar.-Apr. 1999.

[Jennings and Wooldridge 1995] N. Jennings and M. Wooldridge, "Agent Theories, Architectures, and Languages: A Survey", *Intelligent Agent: ECAI-94 Workshop on Agent Theories, Architectures, and Languages*, Nicholas Jennings and Michael Wooldridge (eds.), Springer-Verlag, Berlin, 1995.

[Jennings and Campos 1997] N. Jennings and J. Campos, "Towards a Social Level Characterization of Socially Responsible Agents," *IEEE Proceedings on Software Engineering*, 1997, pages 11-25.

[Jennings and Kalenka 1998] N. Jennings and S. Kalenka, "Socially Responsible Decision Making by Autonomous Agents", *Proceeding of Fifth Int. Colloq. on Cognitive Science*, San Sebastian, Spain, 1998. (Invited paper to appear.)

[Luck and d'Inverno 1996] M. Luck and M. d'Inverno, "Engagement and Cooperation in Motivated Agent Modelling", *Distributed Artificial Intelligence Architecture and Modelling: Proceedings of the First Australian Workshop on Distributed Artificial Intelligence*, Zhang and Lukose (eds.), Lecture Notes in Artificial Intelligence, 1087, 70-84, Springer-Verlag, 1996.

[Nwana and Wooldridge 1997] H. S. Nwana and M. Wooldridge, "Software Agent Technology", *Software Agents and Soft Computing*, H S Nwana and N Azarmi (Eds.), Springer-Verlag, Berlin, Germany, 1997.

[Pedone and Parisi 1997] Roberto Pedone and Domenico Parisi, "In What Kinds of Social Groups can Altruistic Behavior Evolve?" *Simulating Social Phenomena*, Conte, R., Hegselmann, R., and Terno, P. (eds.), Springer-Verlag, Berlin, 1997.

[Rosenschein 1985] J. Rosenschein, "*Rational Interaction: Cooperation Among Intelligent Agents*", Ph.D. Dissertation, Computer Science Department, Stanford University, Stanford, CA 94305, 1985.

[Rosenschein and Genesereth 1985] J. S. Rosenschein and M. R. Genesereth, "Deals Among Rational Agents," *IJCAI-85*, pp. 91-95, Aravind Joshi (eds.), Morgan Kaufmann Publishers Inc., California, 1985.

[Sen 1996] Sandip Sen, "Reciprocity: a foundational principle for promoting cooperative behavior among self-interested agents," *Proceedings ICMAS-96*, Kyoto, Japan, December 1996, pp. 322-329.

[Trivers 1971] R. L. Trivers, "The Evolution of Reciprocal Altruism*", Quarterly review of Biology*, Vol.46, 1971, pages 35-56.