CSE 472 Agent Smiths Seminar Presentation

Believing Others:Pros and cons

Sandip Sen *Department of Mathematical and Computer Sciences, University of Tulsa*

Citation

Sandip, S. (2002). Believing Others: Pros and Cons, Artificial Intelligence, 142:179-203.

Presenters

Daniel Baylog
Huan Nguyen
Agent Smith (substituting for Marco Perches)

Outline

- Background
- Related Work
- Model Structure
- Experiments
- Future Work
- Conclusion
- Questions?

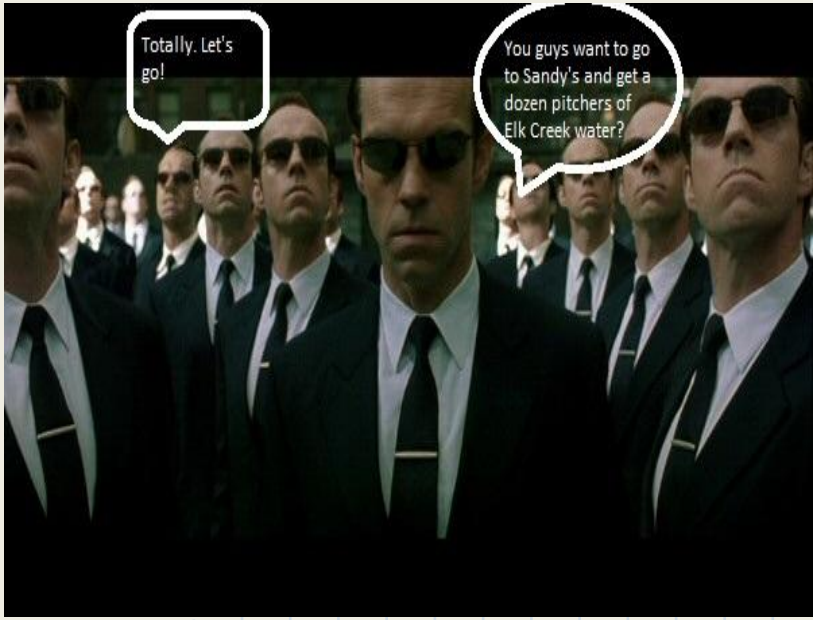
Abstract

- Agents act self interested
- Need a way to promote cooperation among self interested agents
- Use reciprocative agents that share opinions of other agents

Agents of the Future

Agents must be able to interact with other ABS (Agent Based Systems) and humans in different role contexts over extended periods of time

- Agents of the future must be SOCIAL entities
- Agent Interaction is measured in Cooperation Possibilities
  - These exist where an agent can help another agent such that the cost of helping is less than the cost saved by the agent needing help.



Non-Monetary System

Non-Monetary mechanisms provide for better social relationships in certain situations.

- Agents Take advantage of cooperation possibilities by trading “helps”
  - Cost measured in terms of time costs.
    - Time can’t be stored like money
    - Using time causes agents to use unoccupied time to develop mutually beneficial relationships.

Cooperative Relationships

- Not only a benefit to individual agents
  - This can also benefit the condition of the entire society/environment.
    - Individual Agent designers design strategies to make agents profitable.
    - Designers of Agent Systems/Mechanisms want entire system to run smoothly.
    - Can individual rational action lead to max local utility and improve system performance? Not always, but for certain domains, yes!

Related Work

- **Robert Axelrod** - stable cooperative behavior can arise in self-interested agents when they adopt a reciprocative attitude towards each other.
- **Castelfranchi and Falcone** - Argued for the necessity of trust in social interactions between agents to promote cooperation.
- **Cesta, Micelli, and Rizzo** - reveals weaknesses with agents that always helped when asked.
- **Castelfranchi, Conte, and Paolucci** - studied the performance of a group of agents consisting of selfish agents and reciprocative agents that share their opinions of other agents.

Probabilistic Reciprocity

- Assume a multiagent system with N agents, each agent designed to carry out T tasks.
- j^{th} task assigned to i^{th} agent is t_{ij}, and cost is C_{ij}. If Agent K carried out with its own task t_{kl}, cost incurred for task t_{ij} is C^{kl}_{ij}.
- If Agent K can carry out the task of another agent, with a lower cost than the incurred by the assigned agent, the first agent can cooperate with the second.

Probabilistic Reciprocity II

- If Agent K helps agent I, it incurs the extra cost of C_{kl}, but agent I saves cost of C_{ij}.
- Since cost of helping to helper agent is less than saving of the helped agent, cooperation possibility exists.

Probabilistic Reciprocity III

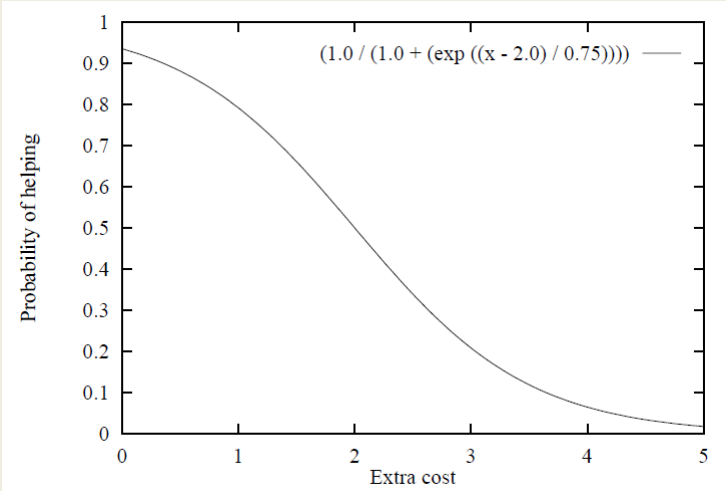
- Proposition for probabilistic decision mechanism to satisfy criteria for determining to help an agent.
  - S_{ik} and W_{ik} = savings obtained from and extra cost incurred by agent i from agent k over all previous exchanges.
  - $B_{ik} = S_{ik} - W_{ik}$, the balance of exchanges.
- Probability that agent k will carry out task t_{ij} for agent i while it is carrying out task t_{kl} is given by:

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

- Sigmoidal probability function where the probability of helping increases as the balance increases and is more for less costly tasks.

Probabilistic Reciprocity IV

- C_k is the average costs of tasks performed by agent K, and *Beta* and *Tau* are constants.
  - *Beta* moves the probability curve left or right.
  - At the beginning B_{ki} is 0 and there is a .5 probability that one agent helps anyway
  - *Beta* * C_k is the extra cost
  - *Tau* controls the steepness of the curve.



Assumptions

- This section outlines Assumptions for the agents according to this design principle:
  - An agent does not change strategy in the course of an experiment. IE: Selfish agent does not become reciprocative.

Assumptions II

- The motivation for reciprocity comes from self interested agents interacting in open environments that are abound in cooperation possibilities. However, The Agents could all be using different strategies, and choose to cooperate on a task-by-task basis, leaving the environment very volatile.

- Helpful agents are honest, while exploitative agents can lie and deceive. In reality a helpful agent could lie too but for the purposes of this paper and trust based relationships, here that will not be the case.

Assumptions III

- Selfish Agents are not utility maximizers. They are being used to study the effects of disruptive agents in the population.

Assumptions IV

- The composition of the agent group is stable for some amount of time as measured by the number of tasks executed or the number of interactions between agents.

- There exists sufficient number of cooperation possibilities with roughly symmetrical possibilities so that sometimes Agent Smith helps Agent K and vice versa.

Agent Strategies

Two base strategies that this paper expands upon:

- **Selfish Agents:** Agents always ask for cooperation but never accept when cooperation is requested of them. Benefits by exploiting philanthropic agents.

- **Reciprocative agents:** Deterministically use cost and savings to determine whether or not to cooperate when requested.

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

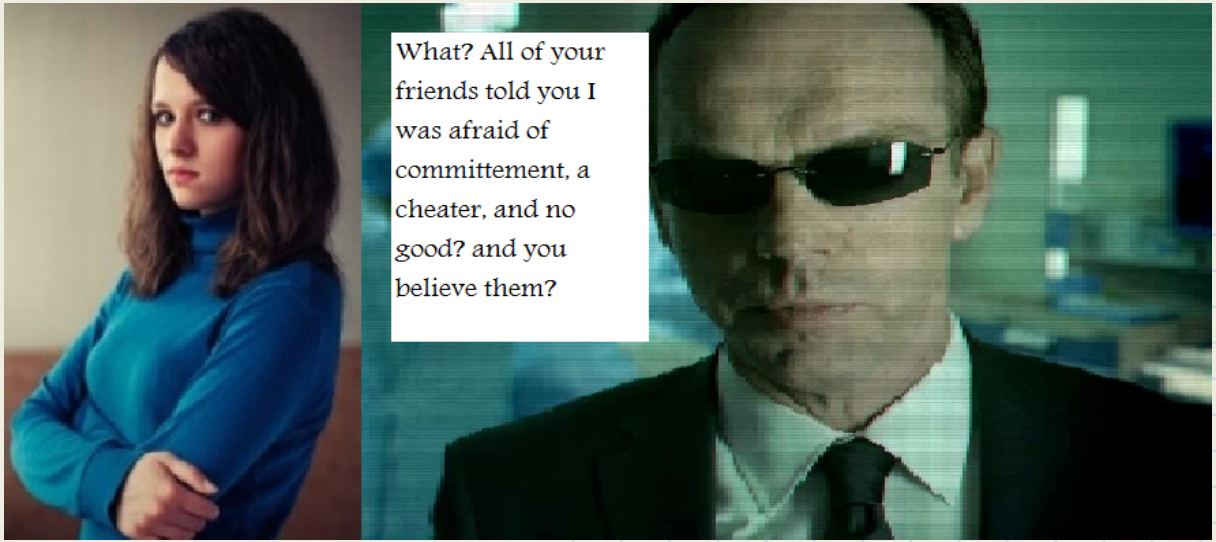
Agent Strategies II

An expansion upon the aforementioned two agent strategies are:

- **Believing Agent:** Doesn't use own balance in making decisions, but the balance as reported by all other agents. Instead of Bki from the earlier equation, a believing agent uses:

$$\sum_{j \neq i} B_{ji}$$

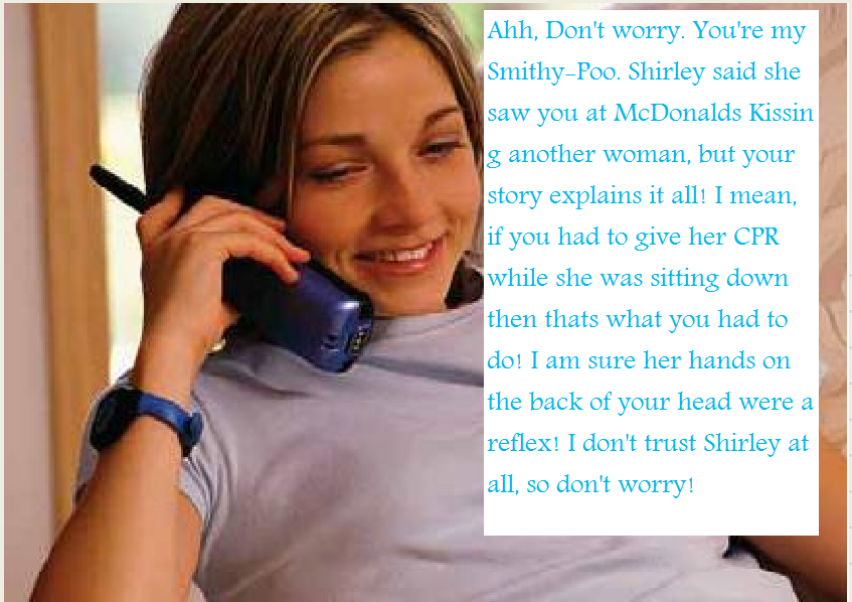
agent i^4.



Agent Strategies III

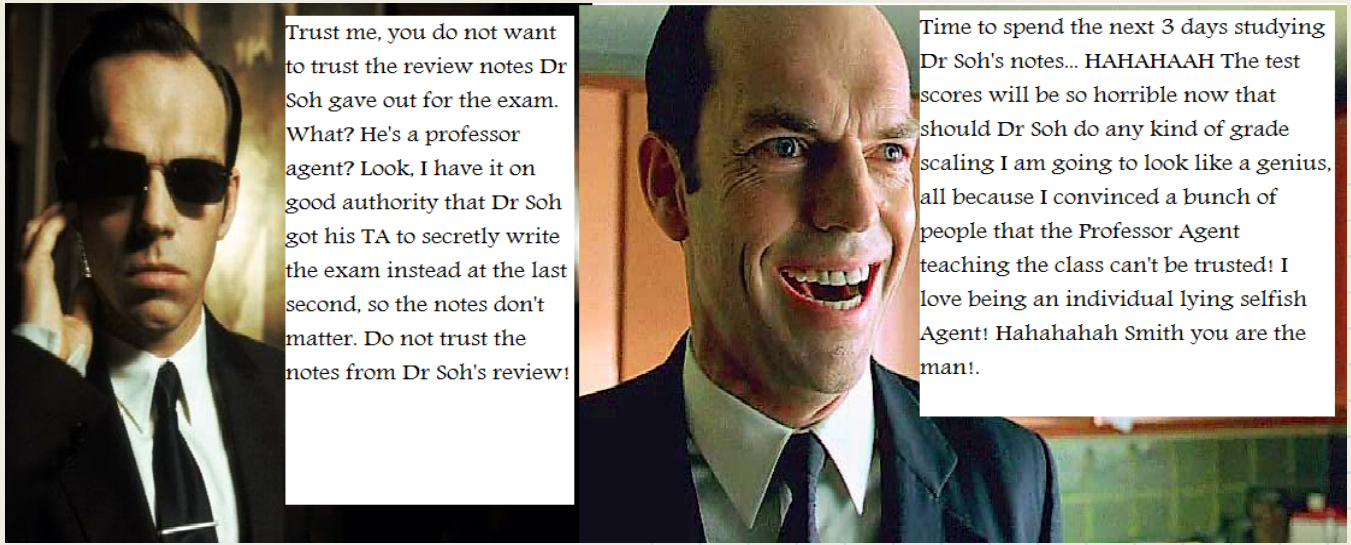
Learned Trust Reciprocative Agents: The same thing as a believing agent, however it will only consider input from an agent it holds in favorable regard. Instead of B_{ik}, for calculating the probability of trusting agent i:

$$\sum_{j \neq i \wedge B_{kj} > 0} B_{ji}$$



Agent Strategies IV

Individual Lying Selfish Agents: These agents realize that other agents are using a balance system to decide whether to trust agents and will reveal false impressions about said agents, either to leave other agents to help itself more often due to lack of trust in other agents, or to cause other agents to have less gainful interactions.



Agent Strategies V

Collaborative Lying Selfish Agents: Not only spoil the reputation of helpful agents but try to band together to boost the the reputation of selfish agents and agents with whom it has zero balance.

Experiment Description I

- There are N agents
- each agent is assigned to deliver T packets
- All the packets are located at a centralized depot
- packet destinations are located at one of R radial fins
- The distance is between 1 and D
- Agents movement is limited (depot to fin, fin to depot)
- Agents are assigned the next packet when returning to the depot
- The agents will check to see if there are other agents in the depot

Experiment Description II

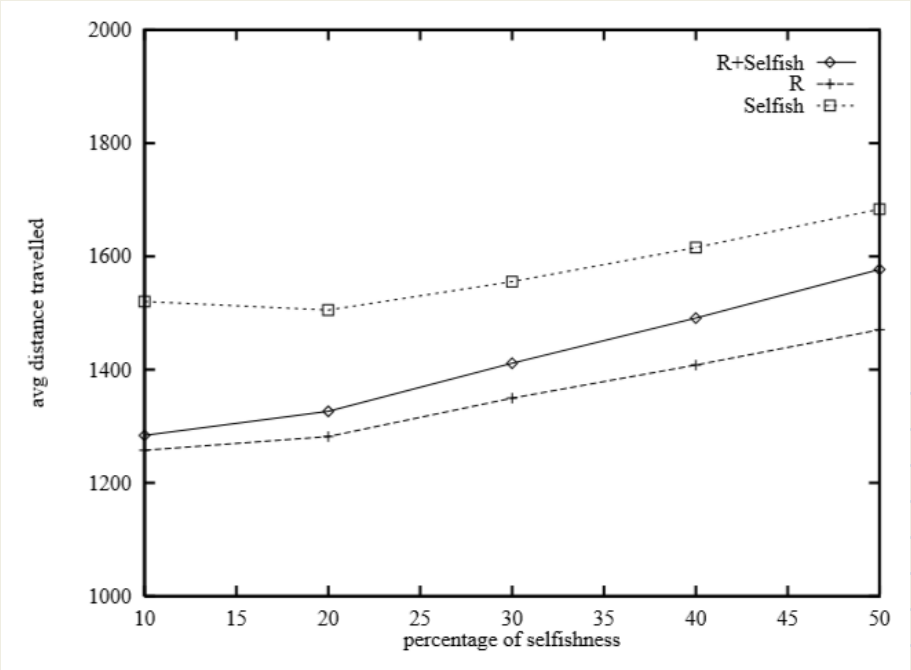
- Values used by authors for their experiments
  - 100 agents
  - 500 packets for each agent to deliver
  - 4 radial fins
  - max distance from the depot is 3
  - Beta = 0.5
  - Tau = 0.75

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

- Each experiment is run on 10 different randomly generated data sets, where a data set consist of an order assignment of package deliveries to agents.
- All agents are assigned the same number of deliveries.
- Evaluation metric is the average cost incurred by the agents to complete all the deliveries.

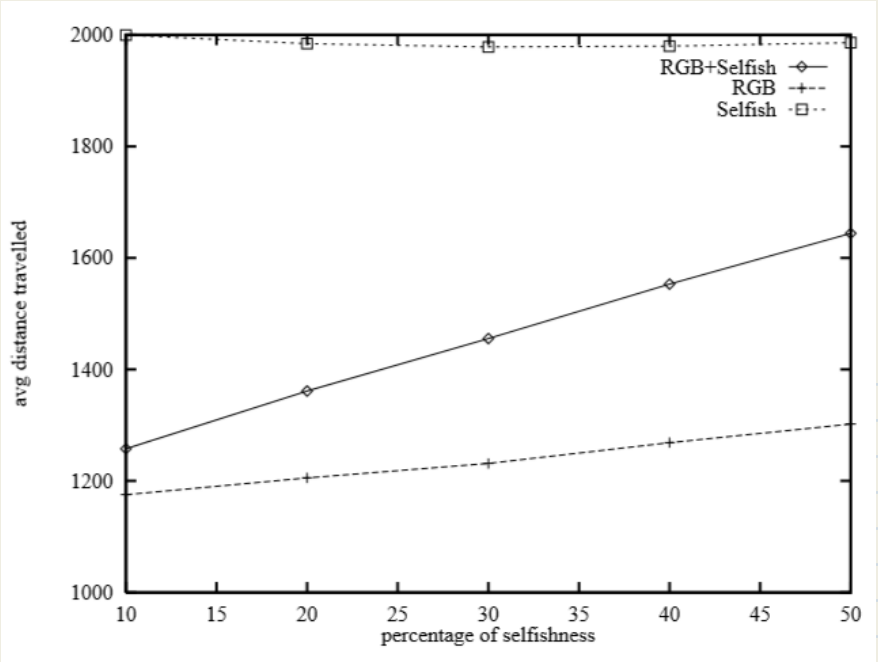
Experimental Results I

- Reciprocative and Selfish agents in mixed groups
  - Selfish agents are somewhat able to exploit the reciprocative agents.



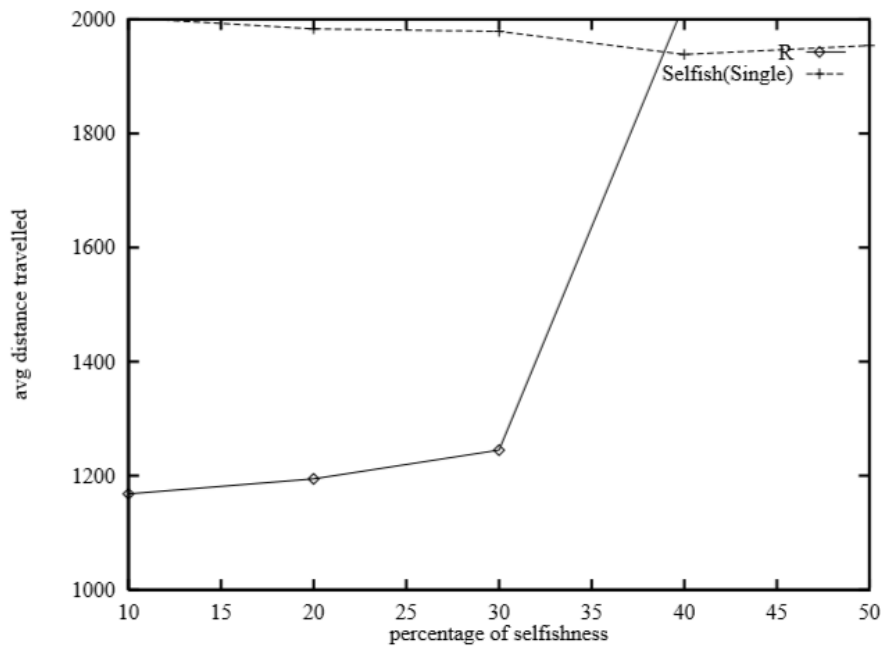
Experimental Results II

- Believing Reciprocative and Selfish agents in mixed groups
  - Sharing of balances severely restrict the exploitative edge of the selfish agents



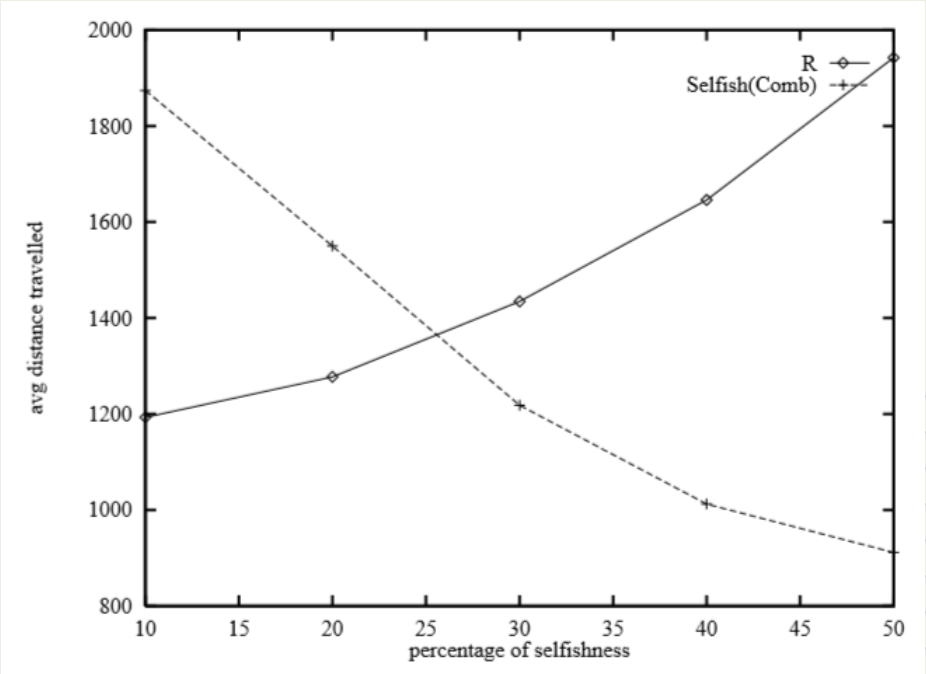
Experimental Results III

- Believing Reciprocative and Individual Lying Selfish agents in mixed groups
  - few selfish agents = the lying behavior does not noticeably affect the believing reciprocative agents.
  - a lot of selfish agents = negative information greatly affects performance.



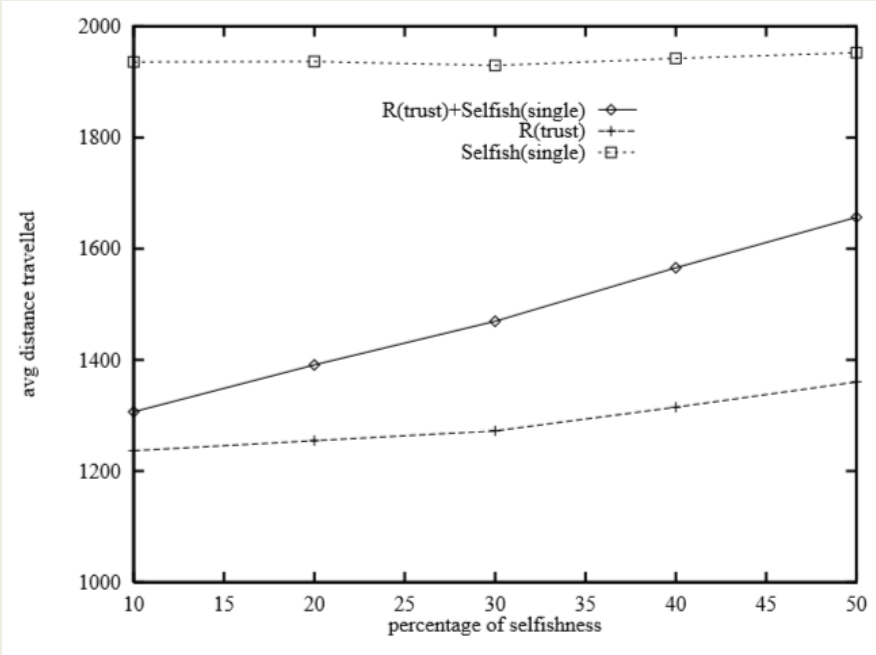
Experimental Results IV

- Believing Reciprocative and Collaborative Lying Selfish agents in mixed groups
  - lying agents are able to exploit the reciprocative agents effectively and overwhelm them when their percentage in the group is more that about 25%



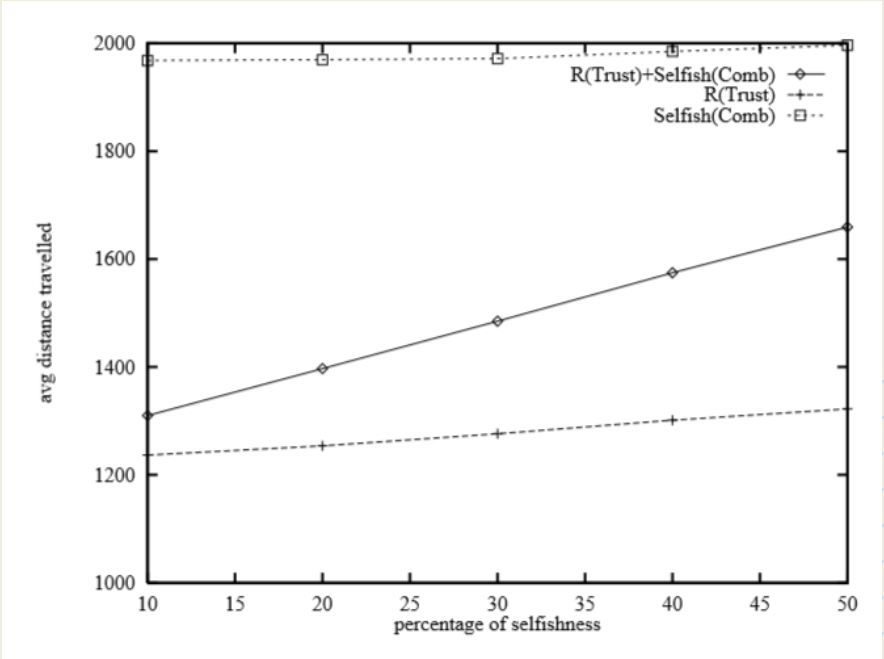
Experimental Results V

- Learn-Trust based Reciprocative and Individual Lying Selfish agents in mixed groups
  - Learn-trust based reciprocative agents can effectively handle lying selfish agents



Experimental Results VI

- Learned-Trust based Reciprocative and Collaborative Lying Selfish agents in mixed groups
  - learn-trust based reciprocative agents are able to distinguish between themselves and the lying selfish agents.



Conclusion

- In general, Reciprocative agents outperform selfish agents in low selfish percentages
- Believing Reciprocative agents with individual reputation weighting fares well against purely selfish agents, but are susceptible to lying/collusion
- Learned-Trust Reciprocative agents avoid these pitfalls even in very hostile environments
  - Realizes benefits of others' opinions while maintaining security of individual reputations!

Future Work

- “Analytically capture the dynamics of the evolution of balance of helps in homogeneous and heterogeneous groups.”
  - More analysis on the actual helps (cooperation efforts) themselves: do they increase as agents trust eachother more? How does group composition affect this?

- How does composition of selfish and reciprocative agents change as a function of time (if agents are free to choose)?

- Using only utility-maximizing agents, not special agents just to make things interesting.

Questions?