This is a postprint version of the following published document:

Evolutionary-inspired approach to compare trust models in agent simulations

Javier Carbo * and Jose M. Molina

Group of Applied Artificial Intelligence, Computer Science Department, University Carlos III of Madrid, Av. Universidad Carlos III 22, Campus de Colmenarejo, Madrid 28270, Spain

Abstract. In many dynamic open systems, agents have to interact with one another to achieve their goals. These interactions pose challenges in relation to the trust modeling of agents which aim to facilitate an agent's decision making regarding the uncertainty of the behaviour of its peers. A lot of literature has focused on describing trust models, but less on evaluating and comparing them. The most extensive way to evaluate trust models is executing simulations with different conditions and a given combination of different types of agents (honest, altruist, etc.). Trust models are then compared according to efficiency, speed of convergence, adaptability to sudden changes, etc. Our opinion is that such evaluation measures do not represent a complete way to determine the best trust model, since they do not include testing which one is evolutionarily stable.

Our contribution is the definition of a new way to compare trust models observing their ability to become dominant. It consists of finding out the right equilibrium of trust models in a multiagent system that is evolutionarily stable, and then observing which agent became dominant. We propose a sequence of simulations where evolution is implemented assuming that the worst agent in a simulation would replace its trust model with the best one in such simulation. Therefore the ability to become dominant could be an interesting feature for any trust model. Testing this ability through this evolutionary-inspired approach is then useful to compare and evaluate trust models in agent systems.

Specifically we have applied our evaluation method to the Agent Reputation and Trust competitions held at 2006, 2007 and 2008 AAMAS conferences. We observe then that the resulting ranking of comparing the agents ability of becoming dominant is different from the official one where the winner was decided running a game with a representative of all participants several times. Since it is a new evaluation method that, as our application to the ART competition showed, gives additional information on the quality of trust models, it would improve the way they are compared. The application of our proposal is not restricted to the ART domain, we suggest that this kind of evolutionary approach has to be taken into account in any evaluation of trust models in agent systems.

Keywords: Trust models, autonomous agents, evolutionary game theory

1. Introduction

In recent years, the relevance of trust and reputation research has been recognized a lot. The main reason is because they are a key factor to automate electronic commerce. Automation would come from the generalized use of software agents intelligent enough to search and select potential partners without prior interaction or experience [18]. With such intention, the reputation of an agent represents a statistical value about the trust probability computed from previous interactions and recommendations. In this way, reputation-based trust systems provide an incentive mechanism that mitigates the risks of interacting with malicious agents [6].

Many reputation-based trust systems have been proposed [25,28], the most popular ones such as eBay use very simple approaches (average) [27] but many of the academic systems use Bayesian/GameTheoretic [14,34] or belief/cognitive models [8,33] to quantify trust. Performance of these models is often evaluated with ad-hoc implementations and metrics. They consist of simulations of societies with several combinations of agent models (honest, malicious), with potential collusions of providers and recommenders, with differ-ent dynamicity (frequency and intensity of change of agent behaviors). Since it was very difficult to effec-tively compare the evaluation results of these ad-hoc implementations and metrics, a testbed platform for evaluating agent trust and reputation models was developed: the *Agent Reputation and Trust (ART) testbed*

[12].[1] Using this testbed, three international competitions were successfully carried out jointly with the 2006, 2007 and 2008 AAMAS international Conferences (AAMAS is one of the main Agent Conferences nowadays). During these years, the ART testbed was used by dozens of researchers producing relevant publications about the implemented trust models for agents (*sabatini* [15], *iam* [20], *afras* [3], *simplet* [19], *peles* [7] and *uno2008* [22]), while the ART-testbed members have discussed, patched and updated the platform using the feedback from the competitions (see discussion notes on ART web page) and from the agent trust community (through the discussion board of ART). This criticism produced changes in protocols,[2] and outlined new directions of work [16]. The main problem was essentially focused on the scalability of ART games, since with more agents the right use of reputation would play an increasing role to win ART games.

From the extensive use of this testbed, a line of research arose on the ways trust and reputation models should be compared. Authors of [16] suggested distinguishing the way trust is acquired, modeled and updated (trust model) from the way trust is used or applied in decisions (trust strategy). On the other hand, authors of [2] suggested an extension of ART testbed to deal with heterogeneous and different domains and protocols. Finally authors of [19] suggested some predefined scenarios to overcome the limitations of the ART testbed comparing trust models. Although all these publications are related to ART evaluations, defining evaluation metrics for trust and reputation models is a general problem. It is a problem also shared by recommendation algorithms in mobile environments [17] and in e-commerce [29].

This paper belongs to this effort of proposing new ways to compare trust models that provide additional information on determining the quality of trust models. Our desired evaluation method does not intend to replace the classic way to compare reputation models, it just provides a new evaluation criteria to be additionally considered. Instead of defining complex controlled situations to evaluate agents' trust models as [19] did, we aim to define an evaluation metric of trust models different from the usual ones: accuracy, efficiency, adaptability, etc. We pursue a new feature of trust models that will be able to be used to evaluate trust models in any agent simulations including ART testbed. It

does not have an exclusive application into ART competition, we just use this testbed to show the utility of our approach to produce new knowledge on the different abilities or quality of trust models.

In Section 2 we explain how an evolutionary-inspired evaluation method can be applied to compare trust and reputation models. Afterwards, in Section 3, we show an application of this evaluation method on the ART 2006, 2007 and 2008 international competitions. Finally, in Section 4, we outlined some conclusions.

2. Evolutionary approach to trust models in agent simulations

2.1. Evolutionarily stable strategies

Classic Game Theory [11] considers games with discrete actions and time-steps for decision making. In these games as tic-tac-toe and the prisoner's dilemma [26] a particular situation (so called Nash equilibrium)[23] is defined when each player (agent) chooses a strategy that may not be improved for both players (agents) by any other alternative strategy. A Nash equilibrium can be achieved by agents through a process of Darwinian selection of strategies. In this case, the strategies involved in the Nash equilibrium are named evolutionary stable strategies (ESS) [21]. These concepts have shown useful for modeling many problems of social behaviour in animals and humans.

According to [10], in order to be solved by this evolutionary way to find an equilibrium of strategies, problems have to accomplished three strong assumptions: First, the evolving population of agents is assumed to be infinitely large, second, the payoffs that agents receive are assumed to be without noise, third, each agent is assumed to play against every other agent to determine its fitness. However, other authors demonstrate the dynamics and equilibria of evolutionary stable strategies apply also to finite populations [9].

From the point of view of research in autonomous agents, there is much interest in representing with agents the way how cooperation and reciprocity evolves in human society. Of our particular interest are the typical real-world situations where one agent can help another agent by sharing work/tasks such that the helping cost of the helper is less than the expense of the helped agent [30].

In this situation, the society of agents contains initially representatives of different cooperation strategies

in similar proportions. 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. Agents then 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 the cost incurred for local tasks, the cost incurred to help other agents with their tasks, and the savings obtained from others when help was received.

Although in classical MAS settings, a trust model used by an agent is not open and available to other agents, an interesting alternative scenario (which could be also realistic) would be to give an agent the free-dom of choosing from one to several of these coop-erative strategies and to change its strategy as and when it seems to be appropriate [1,31]. An agent may be prompted to adopt a strategy if agents using that strategy are seen to be performing better than others. Such a strategy adoption method leads to an evolu-tionary process with a dynamic system composed of changing agent strategies by allowing propagation of more successful strategies and elimination of the un-successful ones. In this way, an evolutionary selection method identifies which cooperative strategies can per-form dominantly.

In spite of its very different approach and motiva-tion, we have also to mention that evolutionary game theory has also been applied in agent competitions in another way: Authors of [32] compute a Bayes–Nash equilibria in agent *WhiteBear* to generate a success-ful strategy for the 2005 International Trading Agent Competition.

2.2. Evolutionarily stable trust models

In order to apply an evolutionary approach to trust games, we first have to recognize that trust games do not satisfy completely the three strong assumptions of game theoretic problems:

- the evolving population is assumed to be infinitely large: the population of trust games has to be large enough to allow reputation making sense, since small populations work fine just with direct inter-actions. In spite of this, the number of participant agents in ART competitions is not very large (less than 20) and therefore its small size does not jus-tify this assumption;

- the payoffs that agents receive are assumed to be without noise: Noise is inherent to trust prob-lems [24], since trust is used because it is applied in domains where there is no objective, univer-sal and specific ways to evaluate products or ser-vices, such as the domain of appraising paintings in ART testbed. In fact, ART designers modeled noise as the variance of a normal distribution. But we can state at least that the payoffs that agents re-ceive are assumed to be fair (directly proportional to the quality of the behaviour shown);
- each agent is assumed to play against every other agent to determine its fitness: this assumption may be mostly accepted in trust games since in them it is not often to include groups or coalitions of agents representing the same interests, they represent individuals with different tasks, abilities and goals and act exclusively on behalf of the in-dividuals.

Obviously trust is involved in decisions of classic game theoretic problems of cooperation, but only in an implicit way. The main difference between them and trust problems is the explicit exchange of reputation opinions between agents, but while decisions in classic and evolutionary game theory problems such as tic-tac-toe and prisoners' dilemma are quite simple (choosing a position, cooperate/defect) that can be implemented as a single rule/function, decisions in trust games are rather more complex (choosing several real numbers in a continuous range of possible values) that have to be implemented by a sophisticated set of rules/functions. Trust problems in the agent community involve the next steps for a cooperative game:

- deciding who and how (costs) to request help in the own assigned tasks;
- deciding who and how (costs) to answer to re-quests of help in others assigned tasks;
- deciding how (weights) to aggregate help from different agents in each task.

Furthermore trust games include decisions related to the explicit exchange of opinions about the reputation of third parties:

- deciding who and how (according to potential costs involved) to request reputation on the ability of third parties in performing own assigned tasks;
- deciding who and how (according to potential costs involved) to answer to requests of reputation on the ability of third parties in performing others assigned tasks;

- deciding how (assigning weights) to aggregate reputation from different agents about each ability of a third party in a task.

Additionally, some trust simulations include the possibility of agents promoting (advertising) their own ability performing tasks.

Furthermore, decision making in all these steps is highly interdependent, where these dependencies are an open discussion in psychology and sociology, and therefore implemented in very different ways, particularly the way reputation is updated and aggregated depending on the source of information [4,5].

This complexity avoids the inclusion of mutations (crossing operators) in evolutionary trust games, since trust models are seen as an integrated module that encapsulates all trust decisions represented, addressed and computed in very interdependent way. Combining parts of trust models into an aggregated new trust model (mutation) is then not possible with the current most extended view of trust models. But since the ART testbed unified information exchanged and communication protocols in a particular trust game to compare heterogeneous trust models, it at least allows reproduction (propagation of most successful trust models) and death (elimination of unsuccessful ones). It means that agents may decide proactively to change their trust model. This approach of adopting trust models dynamically conceptually makes sense since the global goal of using trust in agent societies is to establish some kind of social control over malicious or distrustful agents (through the exchange of local and subjective evaluations between partners). So the idea of agents changing trust models is coherent with the final intention of trust decisions to filter out the agents who do not behave properly in such society. Following this line, it seems to be realistic that agents with a failing trust model would replace it and they would adopt a successful trust model in the future.

The absence of mutation, the limited definition of reproduction (just a complete change of trust model is allowed) and the insatisfaction of the three assumptions of game theoretic problems avoid considering our proposal as a 'pure' evolutionary approach, instead of it, our proposal is noted as an evolutionary-inspired approach.

Particularly we suggest implementing these evolutionary-inspired ideas in trust games following the next rule: the loser of a direct confrontation (several turns of a particular game) among single or multiple instances of trust models (not uniform population) would be removed from such society of agents. While the winner

trust model of that game would be reproduced and included as an additional instance of the winning trust model.

While the first game would be run over a system with single instances of all trust models (uniform population), through this analogy we can run several games where trust models are being adopted and dropped by agents, defining a repeated game set that would allow us to evaluate the ability of trust models to be dominant or an evolutionarily stable trust model in case we would reach an equilibrium.

Additionally, we can by analogy then define as an evolutionarily stable strategy a trust model which, if applying this reproduction schema in as many games as needed it becomes dominant (adopted by a majority of agents) and it cannot be defeated by any alternative strategy. We can even determine the level of dominance by the percentage of agents implementing such strategy in the equilibrium reached, and by the number of games needed to reach such equilibrium. Equilibrium would be reached when no more changes of trust models take place.

This evolutionary-inspired approach to evaluate trust models implies the next assumptions that could be seen as limitations since not always can be satisfied in real-world trust problems:

- The payoffs of all the agents have to be publicly known.
- Games have to be run in an independent way. Information acquired in a game is not used for the next games.
- Trust models have to be available to be reproduced.
- Each game has to be long-enough to decide and differences between agents are large enough to change trust model.

## 3. Application of an evaluation of trust models based on an evolutionary-inspired approach into ART competitions

3.1. Agent Reputation and Trust (ART) testbed, players and competitions

3.1.1. Agent Reputation and Trust (ART) testbed

The ART testbed [12][4] allows the comparison of different trust models using reputation models in the art appraisal domain. In this domain, agents are play-

4
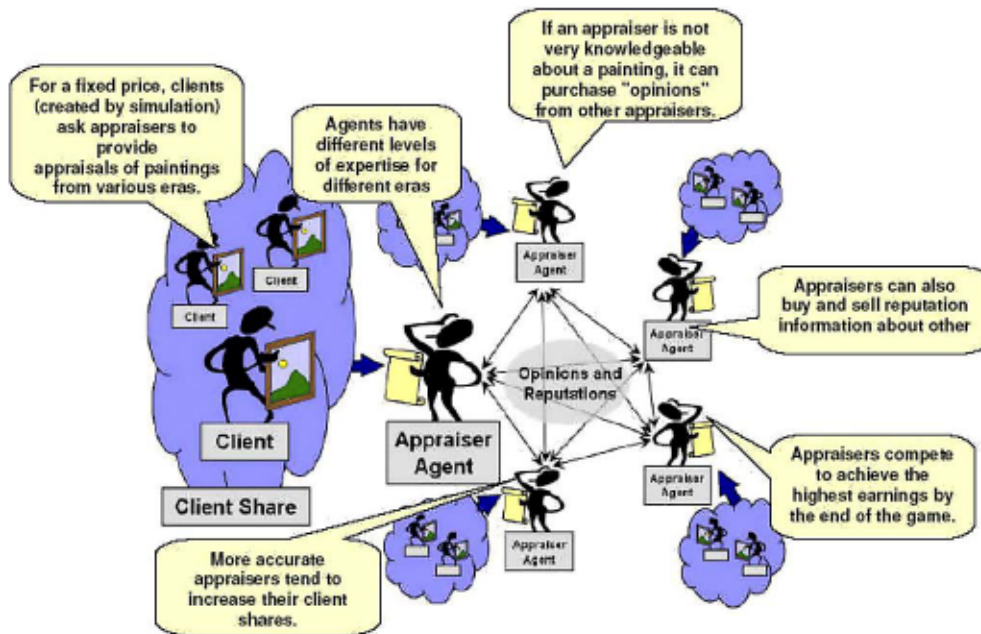
ers/competitors that appraise paintings and implement trust models. Figure 1 shows an outline of the ART domain.

At each timestep, the simulator engine presents each appraiser agent with paintings (generated by the simulation engine) to be appraised, paying a fi ed fee for each appraisal request. Very close valuations of paintings to the real value would lead to more future clients, and therefore to more earnings to win the competition. The corresponding steps of a turn in ART games is shown in Fig. 2.

Each painting belongs to an era from among a set of artistic eras, while agents have different levels of expertise (ability to appraise) in each era. An agent can appraise its own paintings and may request opinions (at a fi ed cost) from other appraisers to get its valuation of the painting close to the real value (specially useful in the eras where the agent has low expertise). An agent can act also as provider of appraisals in response to opinion (about paintings) requests from other agents. Additionally, an agent can similarly request reputation information about other appraisers (at a fi ed and much lower cost than opinions). The winner of an ART game is the agent who earned more money over the number of iterations that were run in the game. Such earnings come from different sources: paintings appraised to the own clients (Client Fee), paintings appraised to other appraiser agents (Opinion Cost) and reputations shared
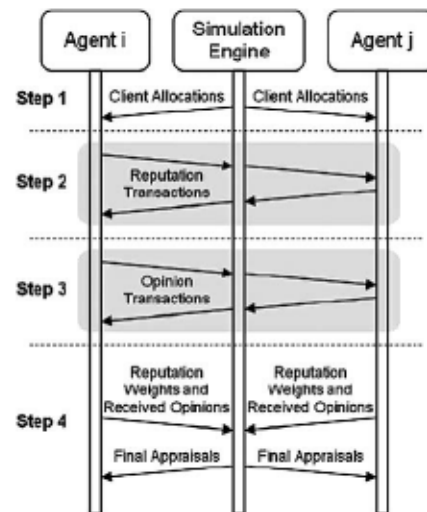
Fig. 2. Steps of a gameturn in ART domain. Source: [13].

with other appraiser agents (Reputation Cost), where Client Fees are the main source of income since: Client Fee $\gg$ Opinion Cost $\gg$ Reputation Cost. Trust models implemented in agents have to implement decisions involved in these protocols (aggregating opinion and reputation values, considering who to ask and answer, weighting opinions, etc.).

In 2008, a new version of the testbed included the possibility of expertise of agents about eras changing over time (expertise dynamicity). Every time step,

a number of eras was randomly selected for each agent to change the corresponding expertise agents have about these eras. For every positive change in expertise of agent about an era, a negative change in expertise of the same amount was applied to another era, so that the average total expertise of the agent was not modified.

3.1.2. Agent Reputation and Trust (ART) players

For instance, *Uno2008* agent [22] classifies agents into four categories according to the level of knowledge about them regarding its role (opinion and reputation provider). These categories are used to decide between applying exploration or exploitation strategies with them. In this way authors of *Uno2008* define the amount of effort dedicated to exploitation given a quality threshold QT, so that agents with a higher trust value will be used as appraisers, with an adaptive maximum number of requests to perform for each painting, T. This maximum is computed from: $T = (\frac{clientFee \times questionPercentage}{opinionCost})$ where values of QT and *questionPercentage* were empirically set to 0.7 and 0.4, respectively. *Uno2008* assumes a general policy of being honest with every request it receives.

Another example of an ART trust model is *IAM* [20]. It computes the estimated benefit of having an opin-ion from a particular agent to justify the decision of re-questing an opinion from it (it calculates the expected variance of the final appraisal). Then the *IAM* agent se-lects the most profitable providers to ask for opinions. Furthermore, *IAM* agent assumes a general strategy of being honest with its partners although it looks for a minimum level of reciprocity. It classifies some agents as cheaters based on their previous interactions, and it then generates random responses for cheating agents. An additional important issue is how much effort *IAM* uses to generate opinions, it computes an empirical study to decide the compromise value (of $Ci = 4$) of investment according to reverse engineering applied to the equations that ART designers used to generate opinions of paintings from their real value. In *IAM*, the opinion certainty provided to other agents, cv, de-pends on its expertise in the era of the particular re-quest, Si, and its intended spending to generate an opin-ion ($Ci = 4$), resulting: $cv = 1 - \frac{(1 + \alpha/Ci) \times Si}{1.5}$ where α is a constant used in the generation of real value of paintings shared by all agents during the competition (it was set to 0.5 by organizers), and 1.5 is the maximum deviation of the Gaussian distribution generating an agent's appraisal value.

3.1.3. Agent Reputation and Trust (ART) competitions

Using the ART testbed, three international competitions were successfully carried out jointly with the last AAMAS international Conferences. The corresponding way to define the winner of competitions was slightly different each year.

- In 2006, for scalability reasons, games could involve at most 5 agents, so the 13 participants were grouped in 14 different preliminary rounds. The 5 agents with better average score in such preliminary rounds played the final round. Each round was played 10 times with games of 60 minutes length whatever timesteps they took.[5]
- In 2007, games involved all the participants (16) plus some dummy agents (9) of 60 minutes length whatever timesteps they took. Games were run 10 times to avoid noise due to initial conditions.[6]
- In 2008, 9 games were played with all the participants (11). Three of them with low, medium and high expertise dynamicity respectively and with 90 timesteps per game.[7]

As publications on ART testbed and ART competitions stated, order of interactions is not relevant since all of them are implemented at once by the simulation engine, and standard deviation was low enough with the repetition values established in competitions [12,16]. Although three alternative competition games were applied, we think that quality of trust models cannot be determined completely, it has to be shown facing different situations. Performance of agents will depend on the opponents they face in games and on the particular game setup. Our point is that such games based on direct competitions of an instance of each agent per game give some information about the quality of trust models, but more information could be obtained to determine the best trust model with other different types of games. Therefore, new game setups have to be defined to evaluate performance of agents facing many different situations. Specifically we suggest a game that shows which participant is implementing an evolutionarily stable trust model.

3.2. Finding a ESS among 2006 ART competitors

Once the type of game to be run has been defined in Section 2.2, we have applied it to the participant agents

of 2006 ART competition (the parameter setup and the code of 2006 participants are public and therefore these experiments are repeatable[8]). The participants and the resulting earnings of such competition are also public.[9]

As the ART testbed in 2006 had an important limitation of scalability, games could involve at most 5 agents. Because of this scalability limitation of the 2006 ART testbed the 13 participants were grouped in 14 different preliminary rounds, that leaded to a fina game with the 5 agents implementing the best trust models in the preliminary rounds. Therefore we considered the application of our evolutionary-inspired games to these best 5 agents who played the fina round in the 2006 competition. As we consider these participants (a single instance of each competitor) as the firs game in our evolutionary-inspired simulation, then the results become equal to those of the officia ranking: *IAM* wins and *sabatini* loses. So second game includes 2 *IAM* agents, no *sabatini* agent and the other 3 agents as participants of the second game. We proceed in same way including an extra winner agent and excluding the loser agent in consecutive games. Next we show in Table 1 the agents that win and lose each consecutive game with the corresponding earnings. The earnings shown in Table 1 are computed in the same way as the competition, as a sum of all the bank balances, and game length and game repetition as define in the same way (60 minutes maximum, 10 times) so all the experiments shown here can be easily repeated.

Specificall with this experiment we have shown that although the strategy of the winner of the 2006 international competitions spreads in the firs game (with 2 agents implementing *IAM* trust model out of 5 participant agents), it never becomes dominant (there is never a majority of iam agents). In fact it is defeated by other trust model, *joey*, which becomes totally dominant (5 *joey* agents out of 5). Therefore *IAM* is not an evolutionarily stable trust model, so its superiority to the other agents is, at least, arguable. We also found out that the right equilibrium of trust models that forms an evolutionarily stable society is composed by 5 *joey* agents. Finally, from the order in which agents are excluded from the society, we can propose an alternative ranking of trust models in Table 2 which is slightly different from the competition ranking.

Since games in the 2006 version of ART testbed just involved 5 agents, we assume that conclusions of our evolutionary approach to trust comparisons would be more significan in the 2007 and 2008 games, since 5 agents is a small population. 2007 and 2008 competitions had more diversity of agents in the games. But even with 5 agents we can see how the apparent best quality of *IAM* agent strongly depends on the opponents. On the other hand, the evolutionarily stable trust model, *joey*, is not able to win the initial game, but it wins and it avoids to be excluded in all the next games becoming totally dominant of the game.

3.3. Finding a ESS among 2007 ART competitors

Next, we have applied the same type of game to the participant agents of 2007 ART competition with the parameter setup and code of 2007 competition participants.[10]

In this case we consider as participants of the firs game of our evolutionary-inspired simulation a single instance of each competitor, without the dummy agents included in the competition games. The results of such firs game become similar to those of the officia ranking: *IAM2* wins and *xerxes* loses. So second game includes 2 *IAM2* agents, no *xerxes* agent and the other 14 agents as participants of the second game. Next we show in Table 3 the agents that win and lose each consecutive game with the corresponding earnings. The earnings shown in Table 3 are computed in same way as the competition, as a normalized bank balance (bank balance divided by number of timesteps in the game),

Table 1
Evolution simulation 2006 results

Game number	Winner	Earnings	Loser	Earnings
1	IAM	114,753	sabatini	90,699
2	joey	130,797	IAM	85,299
3	joey	128,388	frost	74,730
4	joey	135,390	neil	88,812
5	joey	126,573	IAM	90,192

Table 2
Comparison of rankings of 2006 agents

Competition rank	Evolution rank	Agent name	Excluded in game number
2	1	joey	–
1	2	iam	5
3	3	neil	4
4	4	frost	3
5	5	sabatini	1

Table 3

Evolution simulation 2007 results

Game number	Winner	Earnings	Loser	Earnings
1	IAM2	17,377	xerxes	−8610
2	IAM2	14,321	lesmes	−13,700
3	IAM2	10,360	reneil	−14,757
4	IAM2	10,447	blizzard	−7093
5	agentevicente	8975	Rex	−5495
6	IAM2	8512	alatriste	−999
7	artgente	8994	agentevicente	2011
8	artgente	10,611	agentevicente	1322
9	artgente	8932	novel	424
10	IAM2	9017	IMM	1392
11	artgente	7715	marmota	1445
12	artgente	8722	spartan	2083
13	artgente	8966	zecariocales	1324
14	artgente	7285	IAM2	2599
15	artgente	7475	IAM2	2298
16	artgente	8384	UNO	2719
17	artgente	7639	IAM2	2878
18	IAM2	6279	JAM	3486
19	IAM2	14,674	artgente	2811
20	artgente	8035	IAM2	3395

Table 4

Comparison of rankings of 2007 agents

Competition rank	Evolution rank	Agent name	Excluded in game number
6	1	artgente	–
1	2	IAM2	–
2	3	JAM	18
7	4	UNO	16
4	5	zecariocales	13
5	6	spartan	12
9	7	marmota	11
13	8	IMM	10
10	9	novel	9
15	10	agentevicente	8
11	11	alatriste	6
12	12	Rex	5
3	13	blizzard	4
8	14	reneil	3
14	15	lesmes	2
16	16	xerxes	1

and game length and game repetition define in the same way (60 minutes maximum, 10 times) so all the experiments shown here can be easily repeated. As it was expected, since there are no dummies to easily cheat, the earnings of agents are much lower than those of the competition and the differences between winners and losers becomes closer in the last games.

Specificall with this simulation, we have shown that although the trust model of the winner of the 2007 international competition, *IAM2*, spreads in the society of agents (until 6 agents implementing *IAM2* trust model out of 16 participant agents), it never becomes dominant (there is never a majority of *IAM2* agents). In fact it is defeated by another trust model of *artgente* agent, which becomes finall dominant (11 *artgente* agents out of 16). Therefore *IAM2* is not an evolutionarily stable trust model, so its superiority as winner of 2007 competition is, at least, relative. It seems that *IAM2* agent performs rather well when there is enough diversity in the society, but, when it has to play against clone agents implementing the same trust model, its performance becomes heavily affected. On the other hand, *artgente* agent shows no ability to win initial games, but it avoids to be excluded in all the games and when diversity of society is reduced it becomes an almost unbeatable opponent of *IAM2* agent. We also found out that the right equilibrium of trust models that

form an evolutionarily stable society is composed by 10–11 *artgente* agents and 6–5 *IAM2* agents. Finally, from the order in which agents are excluded from the society, we can generate an alternative ranking of trust models in Table 4 which is very different from the competition ranking. Jointly with the alternative ranking we can see the game number in which the corresponding agents were excluded. From it, we can also comment that *JAM* agent lasts until game number 18, which is an excellent result. This shows that it is clearly better than other agents not present in the equilibrium society since *JAM* agent avoids exclusion 1 games (number 17) competing with a society of *IAM2* and *artgente* agents. So we can remark that *JAM* agent has more relevance than just the 3rd place in the alternative ranking (in fact it was the 2nd best agent in the officia competition).

3.4. Finding a ESS among 2008 ART competitors

Finally we have applied this type of game to the participant agents of the 2008 ART competition.[11]

We consider as participants of the firs game in our evolutionary-inspired simulation a single instance of each competitor, without the dummy agents included in the competition games. Again the results become similar to those of the officia ranking: *uno2008* wins and *hailstorm* loses. So second game includes

Game number	Winner	Earnings	Loser	Earnings
1	uno2008	2,690,281	hailstorm	1,012,622
2	uno2008	2,332,164	olpagent	964,783
3	uno2008	2,133,745	peles	944,896
4	uno2008	2,111,883	artgente2	1,609,998
5	uno2008	2,013,963	IAM	967,961
6	connected	1,840,376	mrroboto	1,340,291
7	uno2008	1,809,956	nextagent	1,344,296
8	uno2008	1,790,527	simplet	1,226,474
9	connected	1,683,182	uno2008	1,518,626
10	connected	1,743,356	uno2008	1,543,274
11	connected	1,710,364	fordprefect	1,440,956
12	connected	1,751,602	uno2008	1,499,951
13	connected	1,677,487	uno2008	1,516,218
14	connected	1,756,878	uno2008	1,397,317
15	uno2008	1,964,339	connected	1,372,826

Competition rank	Evolution rank	Agent name	Excluded in game number
2	1	connected	–
1	2	uno2008	–
3	3	fordprefect	11
5	4	simplet	8
4	5	nextagent	7
6	6	mrroboto	6
8	7	iam	5
7	8	artgente2	4
9	9	peles	3
10	10	olpagent	2
11	11	hailstorm	1

2 *uno2008* agents, no *hailstorm* agent and the other 9 agents. Next we show in Table 5 the agents that win and lose each consecutive game with the corresponding earnings. The earnings shown in Table 5 are computed in the same way than the competition, as a sum of all the bank balances, and game length and game repetition define in the same way (90 iterations, 3 times each with 0.05, 0.1 and 0.3 amount expertise change) so all the experiments shown here can be repeated.

As it was expected, since there are no dummies, the value of the earnings of the winning agent in the firs game is lower than that of the competition, but they are in general very similar. We can also observe that the differences between winners and losers become closer in the last games. From game 6, the differences are halved, so we can conclude that there is a significan gap between the agents excluded before such game number (*IAM*, *artgente2*, *peles*, *olpagent* and *hailstorm*) and the rest of them (*uno2008*, *connected*, *fordprefect*, *simplet*, *nextagent* and *mrroboto*). This division into two groups is different according to the greatest difference in the earnings of agents in the competition results. There *uno2008*, *connected*, *fordprefect* and *nextagent* had much score than the rest of agents.

Equilibrium was reached in game 15 because connected and *uno2008* would next be continuously alternating winners, and therefore 16th game would be an exact repetition of 14th game.

Specificall with this experiment we have shown that although the strategy of the winner of the 2008

international competition, *uno2008*, spreads in the society of agents (until 7 agents implementing *uno2008* trust models out of 11 participant agents), even becoming dominant (there is a majority of *uno2008* agents), it is finall defeated by other trust model (*connected*), which becomes dominant (8 *connected* agents out of 11) in the fina equilibrium (evolutionarily stable society) form by 7–8 *connected* agents and 3–4 *uno2008* agents. Therefore *uno2008* is not an evolutionarily stable trust model, so its superiority as winner of 2008 competition is, at least, relative. It seems that *uno2008* agent performs rather well when there is enough diversity in the society, but, when it has to play against clone agents implementing the same trust model, its performance becomes heavily affected. On the other hand, connected agent shows no ability to win initial games, but it avoids to be excluded in all the games and when diversity of society is reduced it becomes a unbeatable opponent of *uno2008* agent. Finally, from the order in which agents are excluded from the society, we can generate an alternative ranking of trust models in Table 6 which is slightly different from the competition ranking. Jointly with the alternative ranking we can see the game number in which the corresponding agents were excluded.

We can also comment that other abilities of agents different from dominance can be observed using our evolutionary-inspired approach. For instance, *fordprefect* agent lasts until game number 11, which it is an excellent result. This shows that it is clearly better than the other agents not present in the equilibrium society. *Fordprefect* agent avoids exclusion 2 games (number 9 and 10) competing with a society of *uno2008* and *connected* agents. So we could also realize that *fordprefect* agent has more relevance than just the 3rd place in both (officia and alternative) rankings.

Table 7
Evolution simulation 2006 results with accumulative earnings

Game number	Winner	Earnings	Loser	Earnings
1	IAM	114,753	sabatini	90,699
2	joey	230,624	frost	184,135
3	joey	361,018	IAM	279,718
4	joey	485,884	neil	371,052
5	joey	610,736	IAM	461,174

Table 8
Evolution simulation 2007 results with accumulative earnings

Game number	Winner	Earnings	Loser	Earnings
1	IAM2	17,377	xerxes	−8610
2	IAM2	31,698	lesmes	−20,372
3	IAM2	42,058	reneil	−23,013
4	IAM2	52,505	blizzard	−16,089
5	IAM2	60,312	Rex	−9470
6	IAM2	68,824	alatriste	780
7	IAM2	77,515	agentevicente	4530
8	artgente	88,126	novel	9093
9	artgente	97,058	IMM	13,601
10	artgente	106,075	marmota	18,347
11	artgente	113,790	zecarioles	25,642
12	artgente	122,512	spartan	37,918
13	artgente	131,478	UNO	57,032
14	artgente	138,763	JAM	69,856
15	artgente	146,238	IAM2	138,521
16	artgente	154,622	IAM2	145,082
17	artgente	162,261	IAM2	151,286
18	artgente	168,540	IAM2	158,931
19	IAM2	183,214	artgente	176,207
20	IAM2	191,249	artgente	185,320

3.5. Using accumulated instead of local earnings

In the previous games, winner was decided according to the earnings obtained in each game, but we could alternatively consider the total earnings of each agent to decide winners according to the sum of the individual earnings of each evolutionary game. In that way a great advantage obtained in previous games can help an agent to avoid exclusion in future games. As you can see in Tables 7, 8 and 9, fina equilibrium has not changed, but a few number of agents are excluded in different order. It seems that better agents are more difficul to remove from earlier games than in the previous tables, but results do not change significat vely. We consider this alternative also as a reasonable way to compute our evolutionary approach for evaluating trust models with ART testbed.

Table 9
Evolution simulation 2008 results with accumulative earnings

Game number	Winner	Earnings	Loser	Earnings
1	uno2008	2,690,281	hailstorm	1,012,622
2	uno2008	5,022,445	olpagent	2,073,883
3	uno2008	7,156,190	peles	3,113,269
4	uno2008	9,268,073	artgente2	4,884,267
5	uno2008	11,282,036	IAM	5,949,024
6	uno2008	13,082,734	mrroboto	7,423,344
7	uno2008	14,892,690	nextagent	8,902,070
8	connected	16,528,034	simplet	10,251,191
9	connected	18,211,216	fordperfect	11,921,680
10	connected	19,954,572	uno2008	19,455,708
11	connected	21,664,936	uno2008	20,896,664
12	connected	23,416,538	uno2008	22,396,615
13	connected	25,094,025	uno2008	23,912,833
14	connected	26,850,903	uno2008	25,310,150
15	uno2008	28,795,242	connected	28,075,361

4. Conclusions

Due to the relative success of the trust and reputation research, a good design foundation of fair comparisons among trust models will spread the inclusion of reputation and trust communications into more general service-oriented systems that would be truly distributed. According to this intention, we have define what an evolutionarily stable strategy would be in trust domain, and how it can be shown through a repeated game with a simulation of evolution.

We have applied such game definitio to the participant agents of 2006, 2007 and 2008 ART competitions and we found out relevant differences with the officia ranking. For instance the winners of 2007 and 2008 competition were not implementing an evolutionarily stable strategy. It seems that both *IAM2* and *uno2008* agents perform rather well when there is enough diversity in the society, but when they have to face clone agents implementing the same trust model their performance becomes heavily affected. Additionally some other minor conclusions were outlined such as: the remarkably long life of *fordprefect* agent in 2008 and *JAM* agent in 2007, the very different ranking of 2007 and the unexpected relevant gap between two groups of agents in 2008. Therefore, the application of evolutionary game theory in ART competitions shows that this evaluation method for trust models is not redundant, it provides new information valuable to compare the quality of trust models. In this way, we have shown that quality of trust models cannot be determined completely with games based on direct competitions of an instance of each agent per game. More information

could be obtained to determine the best trust model with other different types of games. Therefore, performance of agents depends on the opponents they face in games and on the particular game setup. In this way, new game setups have to be define to evaluate performance of agents facing very different situations.

In our simulation we decided to apply a low frequency of changes of trust model (just the agent who lost the game may change of trust model) since there was a low number of participants in ART competitions. Therefore we reach an equilibrium in a small number of games. Other higher frequencies could be applied (particularly in simulations with very large populations of agents). Furthermore other different game conditions could fir the change of trust model, for instance, a very big difference between the loser and the winner after a given number of game iterations (avoiding changes in the case of even games). But if the possibility of changing trust model was unlimited (a very high frequency of changes), then the agents' ability of becoming dominant would lose part of its desirability.

Regarding the results, the only big differences appear in the 2007 competition, while in 2006 and 2008 there are minimal differences. So it would be interesting (but not easy to do) to implement a deeper analysis on the differences between 2006, 2007 and 2008 competitions, since the competitors cannot be run in the testbed of another year. So we have to consider the possibility that the differences between officia ART ranking and the evolutionary ranking might also be in the 2007 testbed design.

As a conclusion we stated that this kind of repeated game has to be taken into account in the evaluation of trust models, since new knowledge on the quality and abilities of trust models can be concluded from them. We do not claim that our evaluation method would be fairer or more efficient Our approach is just another way to compare them. It is not better or worst than others, it is complementary. Any additional way to compare them provides more or less weight to the decision of adopting one instead the others. Specificall , our method is focused on how much advantage an agent loses when its trust model is replicated by other agents.

The application of our proposal would increase the information about the quality of trust models, and then it is a valuable contribution to Trust in Agent Societies literature. It is still open the problem of integrating this new information about the ability of becoming dominant with the classic direct comparison of games composed by one instance of each trust model. None of them is better than the other, and considering an average or weighted sum of rankings seems to have

not enough justification Since they measure the ability of trust models facing two different type of games, weights would depend on the type of games we expect these trust models to face in real world.

However, our approach is very focused on ART testbed, and therefore our approach accepts its assumptions (decided after an open discussion in AAMAS conferences) and our conclusions inherit its limitations, overall two of them: the low number of trust models involved in the games, which affects to trust and evolutionary justification of use; and the use of shared utility functions, which generally had to be, in a more realistic approach, different for each agent. These are the ART-derived problems to generalize our conclusions. In spite of this ART-dependence of our work, we have to remark that although the ART competitors implemented their trust models in an ad hoc way to win ART competition with the previously established winning conditions, our evolutionary competition is run with ART rules, and each run was executed as ART competition did. So the only change included by us, is what the other competitors are, which was unknown for everyone before the games, also in ART competition. So, they had to be designed to beat any other trust model, including those which were very similar to them (or even equal to them as we do in our evolutionary computation).

We state that future work in the trust community would be not just thinking about the testbed design improvements, but also about the games definitio and evaluation metrics. This consideration is not specifi to ART testbed issues. Evolutionarily-inspired games have to be considered in any comparison of trust models. It is a metric that provides additional knowledge about the different abilities/quality of trust models and it can be applied in any of the simulations of agent societies that have been proposed until now for comparing trust models.

Acknowledgements

References

[1] A. Biswas, S. Debnath, S. Sen, I. Debnath and I. Sen, Believing others: Pros and cons, *Artificia Intelligence* **142** (2000), 279–

[2] A.A. Brandao, L. Vercouter, S.J. Casare and J.S. Sichman, Extending the art testbed to deal with heterogeneous agent reputation models, in: *10th Workshop on Trust in Agent Societies*, Honolulu, Hawai, 2007.

[3] J. Carbo and J.M. Molina, An extension of a fuzzy reputation agent trust model (afras) in the art testbed, *Soft Computing* **14**(8) (2010), 821–831.

[4] C. Castelfranchi and R. Falcone, Principles of trust for mas: Cognitive anatomy, social importance, and quantification in: *3rd Int. Conf. on MultiAgent Systems (ICMAS 98)*, 1998, pp. 72–79.

[5] R. Conte and M. Paolucci, *Reputation in Artificia Societies*, Kluwer Academic Publishers, 2002.

[6] C. Dellarocas, The digitization of word of mouth: Promise and challenges of online feedback mechanisms, *Management Science* **49**(10) (2003), 1407–1424.

[7] A. Diniz Da Costa, C.J. Lucena, V. Torres Da Silva, S.C. Azevedo and F.A. Soares, Art competition: Agent designs to handle negotiation challenges, in: *Trust in Agent Societies*, Springer, Heidelberg, 2008, pp. 244–272.

[8] B. Esfandiari and S. Chandrasekharan, On how agents make friends: Mechanisms for trust acquisition, in: *Proceedings of the Fourth Workshop on Deception, Fraud and Trust in Agent Societies*, Montreal, Canada, 2001, pp. 27–34.

[9] S.G. Ficici and J.B. Pollack, Effects of finit populations on evolutionary stable strategies, in: *Proceedings of the 2000 Genetic and Evolutionary Computation Conference*, Las Vegas, Morgan Kaufmann, 2000, pp. 927–934.

[10] D.B. Fogel, G.B. Fogel and P.C. Andrews, On the instability of evolutionary stable strategies, *BioSystems* **44** (1997), 135–152.

[11] D. Fudenberg and J. Tirole, *Game Theory*, MIT Press, 1991.

[12] K. Fullam, T. Klos, G. Muller, J. Sabater, A. Schlosser, Z. Topol, K.S. Barber, J. Rosenschein, L. Vercouter and M. Voss, A specificatio of the agent reputation and trust (art) testbed: Experimentation and competition for trust in agent societies, in: *The Fourth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2005)*, 2005, pp. 512–518.

[13] K. Fullam, T. Klos, G. Muller, J. Sabater, Z. Topol, K.S. Barber, J. Rosenschein and L. Vercouter, The agent reputation and trust (art) testbed architecture, in: *Workshop on Trust in Agent Societies at the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2005)*, 2005, pp. 50–62.

[14] D. Gambetta, Can we trust trust?, in: *Trust: Making and Breaking Cooperative Relations*, Basil Blackwell, 1988, pp. 213–237.

[15] M. Gomez, J. Carbo and C. Benac, Honesty and trust revisited: The advantages of being neutral about other's cognitive models, *Journal Autonomous Agents and Multi-Agent Systems (JAAMAS)* **15**(3) (2007), 313–335.

[16] M. Gomez, J. Sabater-Mir, J. Carbo and G. Muller, Improving the art testbed, thoughts and reflections in: *Proceedings of 12th CAEPIA Conference*, 2007, pp. 1–15.

[17] K. Hegelich and D. Jannach, Effectiveness of different recommender algorithms in the mobile Internet: A case study, in: *International Joint Conference on Artificia Intelligence*, 2009.

[18] A. Josang, R. Ismail and C. Boyd, A survey of trust and reputation systems for online service provision, *Decision Support Systems* **43**(2) (2007), 618–644.

[19] Y. Krupa, J.F. Hubner and L. Vercouter, Extending the comparison efficien y of the ART testbed, in: *Proceedings of the First International Conference on Reputation: Theory and Technology – ICORE 09*, M. Paolucci, ed., Gargonza, Italy, 2009.

[20] W. Luke Teacy, T. Huynh, R. Dash, N. Jennings, J. Patel and M. Luck, The ART of IAM: The winning strategy for the 2006 competition, in: *Proceedings of Trust in Agent Societies, WS Proceedings, AAMAS 2007*, 2007.

[21] J. Maynard-Smith, *Evolution and the Theory of Games*, Cambridge Univ. Press, 1982.

[22] V. Munoz, J. Murillo, B. Lopez and D. Busquets, Strategies for exploiting trust models in competitive multi-agent systems, in: *Multiagent System Technologies*, L. Braubach, W. van der Hoek, P. Petta and A. Pokahr, eds, Lecture Notes in Computer Science, Vol. 5774, Springer, Heidelberg, 2009, pp. 79–90.

[23] J. Nash, Equilibrium points in *n*-person games, in: *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 36, 1950, pp. 48–49.

[24] M.P. O'Mahony, N. Hurley and G. Silvestre, Detecting noise in recommender systems databases, in: *Proceedings of 11th International Conference on Intelligent User Interfaces (IUI 2006)*, 2006, pp. 327–331.

[25] S.D. Ramchurn, D. Huynh and N.R. Jennings, Trust in multi-agent systems, *Knowl. Eng. Rev.* **19**(1) (2004), 1–25.

[26] A. Rapoport, Prisoner's dilemma, in: *The New Palgrave: Game Theory*, M.M.J. Eatwell and P. Newman, eds, Macmillan, London, UK, 1989, pp. 199–204.

[27] P. Resnick, R. Zeckhauser, J. Swanson and K. Lockwood, The value of reputation on eBay: A controlled experiment, *Experimental Economics* **9** (2003), 79–101.

[28] J. Sabater and C. Sierra, Review on computational trust and reputation models, *Artificia Intelligence Review* **24**(1) (2005), 33–60.

[29] B. Sarwar, G. Karypis, J. Konstan and J. Riedl, Analysis of recommendation algorithms for e-commerce, in: *ACM Conference on Electronic Commerce*, 2000, pp. 158–167.

[30] S. Sen, Reciprocity: A foundational principle for promoting cooperative behaviour among self-interested agents, in: *Proceedings of the 2nd Int. Conf. on MultiAgent Systems*, AAAI Press, Menlo Park, CA, 1996, pp. 315–321.

[31] S. Sen and P.S. Dutta, The evolution and stability of cooperative traits, in: *Proceedings of 1st International Joint Conference on Multi-Agent Systems (AAMAS 02)*, ACM Press, 2002, pp. 1114–1120.

[32] I.A. Vetsikas and N.R. Jennings, Generating Bayes–Nash equilibria to design autonomous trading agents, in: *Proceedings of the Twentieth International Joint Conference on Artificia Intelligence*, 2007, pp. 1543–1550.

[33] Y. Wang and M.P. Singh, Trust representation and aggregation in a distributed agent system, in: *Int. Conference on Artificia Intelligence (AAAI)*, AAAI Press, Boston, US, 2006, pp. 1425–1430.

[34] A. Whitby, A. Jøsang and J. Indulska, Filtering out unfair ratings in Bayesian reputation systems, in: *Int. Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, New York, US, 2004.