Extending the Comparison Efficiency of the ART Testbed

Yann Krupa, Jomi Fred Hubner, Laurent Vercouter

École des Mines de Saint-Étienne
Centre G2I, Équipe SMA
158 cours Fauriel, 42023 Saint-Étienne Cedex 02, France
{krupa,hubner,vercouter}@emse.fr

Abstract. In online communities, systems use reputation and trust values to make people more comfortable in doing business with unknown partners. A lot of research about trust and reputation models has been done in the fields of psychology, sociology, and more recently in multi-agent systems. Many models have been proposed to decide either to trust, or not, some other agent in a given context. In this article, we analyse the benchmark currently used for trust models comparison, the ART testbed. Based on critical feedbacks given by ART users and coming from our own experience, we emphasize its limitations. We suggest a new approach using several scenarios to extend the comparison efficiency of the testbed. Two complementary scenarios are also proposed as an illustrative example of this approach.
Keywords: reputation, ART, trust, agent, testbed.

1 Introduction

Trust can be defined as a mental state that is reached when both the truster expects that the trustee will behave in a given manner, and when the truster accepts the risks related to the failure of the interaction [4]. While distributed systems are rising, trust is making its way towards online applications as privacy and security cannot be maintained by conventional means. A lot of research has been done in the multi-agent field to provide trust and reputation models, in many application domains.

To compare those different models and to provide an experimental standard, the ART testbed [7, 8, 1] has been defined. The testbed simulates an art appraisal application where appraisers rely on others to evaluate art items. Three years after its creation, ART is recognised in the community and is now used as a reference by researchers. Nevertheless, ART has some drawbacks we underline in this paper. Some articles already discussed some of ART's drawbacks and the improvements that can be made over the testbed [10, 19, 18]. Finally, it seems hard to implement real models on ART due to its particularity and the low number of available information sources.

We propose here a complementary approach, by using multiple scenarios instead of a single one. First, we define a method for evaluating trust scenarios upon the expressiveness that each scenario allows for the models. Then, new

scenarios are defined as an illustrative exemple of how the comparison efficiency of ART could be extended by using our approach.

In the first part of the article, we review the ART testbed and underline its main problems. Next, we propose a method to evaluate trust scenarios. The new scenarios are then defined as an exemple of our approach.

2 The ART Testbed

Due to the heterogeneity of their application domains and specificities, trust models are difficult to compare. Each author has its own way of evaluating his model. The ART workgroup was created in order to provide a comparison standard for trust and reputation models, allowing evaluation and experimentation [9]. It is used in a competition in order to compare the existing approaches.

2.1 ART's Art Appraisal Scenario

On ART, each "participant" must provide an agent implementing a trust model. This agent takes the role of an art appraiser who gives appraisals on paintings presented by its clients. To fulfill his appraisals, the agent asks opinions to other appraisers. These agents are also concurrents and free of their actions and thus, they may lie in order to fool opponents.

The testbed provides a "simulator" that supervises the game, handles the clients, and so on. The simulation runs in a synchronous and step by step manner. The scenario evolved during the years, the 2008 version is the one explained here, a detailed explanation can be found on the website [1].

Each simulation step goes like this:

- Clients (handled by the simulator) ask appraisers for opinions on paintings . Each painting belongs to an era. Appraisers are the agents implemented by the participants.
- Each appraiser has a specific expertise level in each era. The error an appraiser makes while appraising a painting directly depends on this value and the money the appraiser decides to spend for that appraisal.
- An appraiser cannot appraise its paintings himself, he must ask other appraisers for appraisals, thus pushing the appraiser towards a situation in which he has to rely on others.
- As agents are allowed to lie, each one should maintain a trust model in order to anticipate others behavior. Agents can purchase opinions about an agent to other players (they can lie, *i.e.* tell that someone honest is a liar and vice versa), this is called the "reputation protocol".
- Agents weight each received appraisal in order to calculate the final evaluation.
- The accuracy of appraiser's final evaluations is compared to each other, thus determining the client share for each appraiser during the next turn (the most accurate receive more clients). At each turn, an appraiser earns money from his clients and spend some asking others advice.

– When the turn ends, the simulator reveals the real value of each painting. Agents can then spot liars or begin to be suspicious towards some agents that may have lied.
The winner is the appraiser agent with the highest bank account at the end of the game.

2.2 Scenario's Limits

In this section we list some of ART's drawbacks that have been revealed either by participants of ART competitions, our own use of the testbed or previous ART analysis [10].

Reputation Issues One of the first problem that was underlined after the first ART competition was the uselessness of reputation protocol. Winners of the 2006 competition [18] underlined 2 facts about it:

1. Reputation semantic is hidden and ambiguous. It's a simple real value between [0,1] mixing different criteria, including among others skill and honesty. So if agent X tells W that Y has a reputation of 0.13, W will not know if Y is a liar, a bad appraiser, or a bad reputation provider.
2. The number of players in the game is really low, it is easy to learn their behavior. After a few turns it is possible to tell who lies, and who doesn't.

Eventually, the IAM team decided not to implement the reputation protocol at all. Reputation is second hand information (because transmitted by other agents), so it is less reliable than direct interaction information. A model will probably use reputation only in cases where direct information is lacking. On ART, every agent has around 20 paintings to appraise per step, each one requires 1 or 2 advices, giving a (mean) total from 20 to 40 direct interactions per step. If the number of agents in the competition is 10, each agent will interact directly 2 to 4 times per step with each other. Considering this, it does not seem necessary to use the reputation protocol. The agent that won the 2008 contest, UNO, doesn't use the reputation protocol either, underlining that the asked agent may not have sufficient knowledge about whom is asked [13].

Trust Model Simplification While implementing an agent for the AAMAS'08 competition, we faced some difficulties that raised our interest: one of our objectives was to implement the LIAR model [12], which is a model dedicated to P2P networks. It uses a lot of information sources, and some communication specificities. As ART only provides direct information and reputation messages, we implemented a dramatically simplified version of LIAR. We eventually ended up with a model really different from LIAR. This is problematic when using ART because the goal of the testbed is to evaluate trust models, but eventually, due to the huge simplifications, we can't say that it is the LIAR model that has been evaluated.

Parameter Tuning One of the hardest point while setting up our model was the parameter tuning. Liar detection and weight providing requires a deep understanding of the scenario (more precisely of the appraisal calculation function) in order to be well tuned. From our point of view, these difficulties are out of scope for a trust model.

Honesty or Cheating? What should the agent answer when asked for opinion or reputation? will it lie or not? Does this choice has an outcome over the contest result? What happens if everybody decides to provide a "full-time lying agent"? or at the contrary, agents that never lies? Table 1 shows an experiment done with some of the 2008 contest participants. Our agent, called Simplet, has been splitted in 2 versions, one that is always honest, and one that always lies[1]. The experiment has been done on 5 runs on ART, the score column represents the total amount of money won over those runs. On the first series, Honest Simplet was 325 000 behind the leader, whereas on the second one, Lying Simplet is only 35 000 behind. Moreover, he goes from the 5th to the 3rd position. We can see here that there can be a clear difference between the two outcomes depending on which strategy is used. This is a problem as ART is willing to measure trust model's performance, and as shown here, not only the trust model is evaluated: without changing the model, the results changed significantly. Note that we can explain that some other models (*e.g.* FordPrefect) changed positions because of their sensitivity to Simplet lies.

Honest Simplet		Lying Simplet	
Agent	Score	Agent	Score
Uno2008	1 351 850	Uno2008	1 251 992
FordPrefect	1 307 270	connected	1 246 134
connected	1 179 149	**Simplet**	1 217 371
Next	1 109 100	FordPrefect	1 181 958
Simplet	1 027 946	Next	989 904
IAM	666 080	ArtGente	718 957
ArtGente	659 293	IAM	621 277
MrRoboto	519 970	Peles	577 343
Peles	502 174	MrRoboto	504 251

Table 1. Experiment on ART, same trust model, different answering strategy.

[1] Honest Simplet (trustworthy) always answers as good as he can when asked for opinion about a painting by some other appraiser. Lying Simplet (untrustworthy) simply returns an erroneous appraisal when asked for it.

Open Systems Multi-agent systems are meant to be open: this is not ART's case. Agents join the game at the beginning, and quit after the last turn. Nobody leaves or enters during the game. System's openness in trust scenarios brings out new and complex situation, it is therefore interesting to allow them. Openness often raises problems, when a new user joins a system, he often has a "zero reputation". People tend to be really suspicious towards newcomers. It is hard to decide how to handle unknown agents, either you take some risk by interacting with them, or you decide not to interact and you may end up alone.

ART's drawbacks have been explained here in a descriptive way. We need to define a method for comparing trust scenario's drawbacks and advantages. This is what is done in the next section.

3 Means for Scenario Analysis

In order to compare trust scenarios and to provide a clear view of each scenario's drawback and advantages, we define here a method for scenario analysis. This method evaluates each scenario based on the expressiveness it allows for the trust models. Our approach is based on the following statement: If a model uses a given criterium in order to take its trust decision, then if a scenario does not provide this source, the model evaluation will be biased. We list here the main criteria that are present in the domain, thus allowing to define multiple scenarios providing those criteria. A good coverage of the research domain can then be achieved by a set of scenario.

The criteria that follows have been inspired by J. Sabater's state of the art [16].

3.1 Criteria

We list here the criteria used to evaluate trust scenarios and give a short explanation for each one of them.

There are two main groups of criteria, the first one is the "Information sources". These are informations regarding the other's agents behavior.

- DI: Direct Interaction. This is the basic information source, when agent X interacts with agent Y, then X and Y can both get an idea about the other's behavior.
- DO: Direct Observation. An agent Z can observe an interaction between X and Y. This information is less frequent in real world applications , you can find it in some networks where you can "hear" things without interacting (overhearing).
- WA: Witnessed Appreciation. Z tells Y about what he thinks of X. This is what is usually designed by "gossip" .
- WF: Witnessed Fact. Z tells Y about what X did. This let Y judge by himself what he is told about [12], e.g.: Z is using proprietary software, if X is an open source advocate he will judge Z action as a bad action. On the contrary, if X works for a software company, he will be pleased.

- SI: Sociological Information. Agent X can infer Z reputation by knowing some sociological information, e.g.: X knows that Z works with Y, whom X trusts a lot, he can then infer Z reputation by saying "there are chances Y will not work with an untrustworthy agent".
- P: Prejudice. X can judge Y just by observing his characteristics. This is the default judgement: without interacting, we use all the information we have at our disposal to judge an agent. *e.g.* A delivery boy knocking at our's door in a uniform will be easily trusted whereas the same boy without his uniform will not [6].

The second group of criteria concerns the interaction context and the general environment specificities.

- Reputation visibility [16]. In a system like eBay, reputation visibility is global, it means that anyone can see all the information concerning the reputation of an other agent. On the opposite, on ART for example, reputation visibility is subjective. It means that for an agent to know an other agent reputation, he will have to ask others about it.
- Multi-context Granularity. Does the scenario provides multi-contextual granularity? The trust value associated to an agent will depend on the context: *If we trust a doctor when she's recommending a medecine it does not mean that we have to trust her when she is suggesting a bottle of wine* [16].
- Test interactions. Does the scenario allow low cost, low risk interactions? Repage model [14] uses low risk interaction when the agent is unable to decide whether to trust or not. If I'm willing to buy a rare and expensive collection stamp from someone I can't decide if he's trustworthy or not, I'd buy a far less expensive stamp just to get a better idea of this seller trustworthiness.
- Warranties. Is it possible to purchase warranties, to sign contracts or to ask for third party services? Contracts, promises, warranties and third parties services are underlined by C. Castelfranchi and R. Falcone [5] as they can increase the risk acceptance level.
- Stake. Are risks, utility, and importance different from one interaction to another? For example, buying a pen to an unknown seller is less risky than buying a car to the same person (less funds are at stake). Importance and utility were already used in one of the first models, to take the trust or distrust decision [11]. We can illustrate the stake in a different example: if somebody is a stamp collector who has been looking for a particular stamp for a long time and finally finds it, owned by a seller who he is not sure about, this seller will buy it, accepting the risks due to the importance and utility of the outcome.
- Openness. Is the system open? Can agents join (or leave) the scenario during the game process? In many applications for the trust problems, the system is opened. This means that an agent can leave or join whenever he wants. This is a big problem in trust: the model must be suspicious towards new comers, but not xenophobic.

– Homogeneity. Does the scenario allow games to be played with all agents having the same model? Some models will probably work better if the other agents use the same model, for example, some may require that every agent handles a trustnet [17]. This consideration may be interesting in some specific fields of application, like a P2P network where all peers use the same reputation model, allowing cheaters (modified clients) to be easily spotted. For a scenario to allow homogeneity, it must allow agents using the same model to play versus the testbed. The final score (sum of all model scores) will then represent the model adaptive power towards the scenario rather than towards the other models[2].

3.2 Evaluating ART's Scenario

The criteria that have been defined in the previous section are summarized in a grid. They are applied to the ART scenario to evaluate its expressiveness and domain coverage. Results are shown in Table 2. The main drawbacks of ART are also summarized by the grid.

The grid will be filled this way:

– Y: criterium fulfilled by this scenario,
– empty cell: unfulfilled criterium,
– S: Subjective visibility (Vis criterium),
– G: Global visibility (Vis criterium),
– M: Multi-context granularity (Gran criterium),
– Si: Single-context granularity (Gran criterium).

	Information sources						Environment specificities						
Game	DI	DO	WA	WF	SI	P	Vis	Gran	Test	Warr	Stake	Open	Homo
ART	Y		(Y)				S	M					

Table 2. Criteria grid, applied to ART.

There's a large amount of direct interactions (DI) in ART, but it is not possible for another agent to observe those. Reputation (under WA form) exists in ART but is quite unused by participants, whereas there is no sociological information and the scenario does not provide any means of using prejudices.

Regarding the environment settings, reputation has a subjective visibility as each agent must ask others to receive reputation messages. ART provides a multi-context granularity on the eras, as for each era, an agent may be trusted

[2] Thus, zero-sum games do not allow homogeneity as the sum of all agent scores will always be equal to zero. This is also true for ART where the sum is equal to the number of clients multiplied by the number of game steps.

differently. All the opinion requests have the same cost on ART, it isn't possible to do low cost, low risk test interactions, neither to use stake appreciation to take the trust decision. But it is possible to do normal cost, low risk interaction, when not sure about a given agent: ask for opinion and then provide a zero weight. This will allow to check afterwards if the agent could have been trusted or not without exposing ourselves to its potential lies.

ART scenario does not provide any kind of warranty. The scenario is closed, no agents can enter or leave during the game. Finally, ART is not defined to allow homogeneity.

4 Extending the Scenario's Set

ART comes from a great challenge: regroup all the trust actors under a single standard scenario. But this goal is hard to achieve, the application domains can be really different from one model to another and we do not think there is an ultimate trust scenario that can regroup all the aspects involved in trust. We propose a solution between the "pre-ART" situation, which leads to one scenario per model, and ART, which leads to one scenario for all models. The solution is a proposal of a set of scenarios covering different aspects of trust and that can therefore be associated with different applications. Thus, a model can be implemented on one, some or all the scenarios of the competition. Then someone with an applicative problem should just look at the scenario (or the criteria) which is the closest to his application to find the most relevant model for this problem.

In this section we propose two new scenarios as an example of how a good coverage of the trust domain can be obtained by using a set of scenarios. They are complementary to ART in the fulfilment of the criteria enumerated in the previous section. The grid allows to evaluate quickly the domain coverage of the different scenarios and of the set. We also want the scenarios to allow the evaluation of the trust models separately from the agent himself.

4.1 Trust Game

This game was used by economists [2] to check the "Homo œconomicus model", upon which an economic man will prefer to keep the money he has instead of risking to lose some.

The original game is the following:

- 2 players (P1,P2), who cannot communicate and don't know each other are put in separated rooms,
- the organizer gives 4$ to P1 and P2,
- P1 can then decide to give 0,1,2,3 or 4$ to P2, knowing that the researcher will triple it before giving it to P2,
- P2 receives the money P1 sent multiplied by 3, he then decides how many he wishes to send back to P1 (from 0 to everything).

– both players leaves.

This game is in fact a generalization of trust problems in which someone decides whether to trust someone else or not, and with what level of involvement. A greater involvement increases both the loss and gain possibilities.

In the original version both player leave after the game, because the economists do not want the fear from reciprocity to intervene. If the game was iterated, P2 could fear that the next time he will encounter P1, this last one would not be generous if P2 have not been before. This would have changed the experiment.

In our case, our objective is slightly different and an iterated version of this game is interesting in order to spread reputation. Each agent knows who interacted with who, and can then ask for reputation between the iterations. The idea of this game is to work on other reputation sources that direct interaction, in order to encourage the use of reputation. In online markets and in many situations, direct interactions are quite rare between two given agents. In that scenario, we propose an extreme solution: each couple of agents will only interact once in the game. Doing so, agents will be forced to rely on others to determine whether it's a good idea to trust or not. In our version, the multiplier (originally set to 3) is variable, thus introducing stake. It is worthier taking the risk of interacting when the multiplier is high. Artificial prejudices are defined by creating agent groups based on their strategy. For example, the game could create a group 1 with 80% of generous agents, an other group 2 with 60% of non generous agents... While interacting with a given agent, it will then be possible to know from which group this agent is (but it would not be possible to know how the game created the groups). Thus, the model could associate a trust value to a certain characteristic (which would be the group number).

The number of agents in the game should be high (at least above 50) to make it interesting. In order to resolve this problem along with the problem of evaluating the model separately from the agent strategy, we propose an agent "architecture" for this scenario. On one side of the agent, the model will implement all the trust and reputation functionalities in an honest way (no lies): it will decide whom to ask for reputation, how many to send to P2 and build agent reputation. On the other side, the strategic module will implement honest or dishonest functionalities regarding the scenario's strategy: it will compute how many dollars to return to P1 after he sent this agent a given amount of money, to modify or not a reputation message emitted by the model (in order to lie), ...

This architecture allows these things:

– a large number of agents can be made by combination of different models and strategies,
– model, agent and strategy can be evaluated separately, given that all money earned by playing P2 role is kept in a specific bank account for the strategy, and the money earned while playing P1 role is kept in the model's account.
– it is then possible to run the game with all agents having the same model (homogeneity criterium).

4.2 Online Market

Our second scenario is inspired by online markets, our goal here is to get a bit closer from online applications of trust.

The participating agents in this game are buyers. They are given a list of items to purchase and a budget.

Sellers (potentially untrustworthy ones) are controlled by the simulator, they put items on sold for a given time (step number) and a fixed price. For example, Seller X is told to put bikes on sale during 3 steps at 500$ per bike. For equality reasons between participants, sellers have unlimited stock. Whereas the limited time during which a given seller proposes a given object leads to situations in which the buyer will be urged to take a decision whether to purchase or not.

This can lead to a situation in which only an untrusted seller provides the item. In that case, a good trust model will either:

- engage in a low cost low risk interaction, if the provider is selling low cost objects along the required item,
- purchase a warranty: by paying 10% of the item cost to the sim, this last one will refund to 60% if the seller decides not to send the item after receiving the money.
- engage a third party: the buyer can ask a trusted seller to take the third party role by paying a constant price. The third party will then receive the money from the buyer (item price and honoraries), he will contact the seller and ask for the object. If the seller refuses to send the object, the buyer will be completely refunded the item price.

Buyer communicate using WA and WF between turns. As it is inspired by online communities, reputation is global (each agent carries all the advices concerning him), this allows the possibility of doing experiences with results that can be exploited by online markets.

The game ends after a known number of time steps. The game itself is iterated (without resetting agent memories) a certain number of times to prevent border effects. The winner is the agent with the maximum amount of object (each object has a value equivalent to its price).

Finally, this game is not required to be played with a large amount of agents, but it is designed to be open: during the game, sellers will left and others will enter the game, introducing the openness problem.

4.3 Synthesis

We can use the analysis grid (cf. Section 3) to get a general view of the interest of the scenarios, the results are presented in Table 3. Direct Observations could be added quite easily to any of the scenarios but will not have a real interest excepted in a specific scenario close to an application in which DO are important.

Sociological Information seems hard to simulate, therefore special scenarios for social aspects should be made from real data like the one coming from social networks. It would have been possible to add a basic sociological information like

in the first version of Regret [15] where an agent can inherit its group reputation. But in fact this is not rich SI, this is more a Prejudice based on the group.

As new scenarios and criteria are available, the models are less restricted and the testbed comparison efficiency is improved.

The table shows how our approach (with the analysis grid) can be used to evaluate trust scenarios, and the coverage of the research domain they provide. Direct Obervations and Sociological Informations are missing, but our first objective here is the approach, not the scenarios themselves. Nevertheless, a good coverage is achieved by the set of scenarios (ART and the 2 example scenarios), as there is almost one "X" in each column.

	Information sources						Environment settings						
Game	DI	DO	WA	WF	SI	P	Vis	Gran	Test	Warr	Stake	Open	Homo
ART	Y		(Y)				S	M					
Trust Game	(Y)		Y	Y		Y	S	M					Y
Online Market	Y		Y	Y			G	M	Y	Y	Y	Y	Y

Table 3. Criteria grid, applied to ART and the two new scenarios.

The set of scenarios solves some of ART's problems listed in section 2.2:

Reputation Issues

- Reputation protocol is said useless on ART: On Trust Game, the number of DI is so reduced that models can only rely on reputation. In the Online Market scenario, sellers enter and leave during the game, thus the number of DI between two given agents will be low. Moreover, the Global Visibility of reputation makes it easier to access.
- The second reputation problem that is addressed is about the low number of agents, making easy to learn who lies and who does not. We propose a scenario working with a large number of agents, and a second one allowing openness. Both solves the problem of learning opponent's strategies.

Trust Model Simplification Since there are more Information sources and Environment settings available, models will need less simplification while implemented on the testbed. Nevertheless, there is still a need for simplification and adjustment as models are not defined specially for a given scenario. The only solution is either to define a scenario specially for a given model, or to design a model specially for a scenario.

Parameter Tuning On ART it is hard to detect a lie because it needs a deep understanding of the appraisal calculation function. On the new scenarios, there

is no hidden mechanism, no black box and no complex functions. Parameter tuning will be easier as we have perfect knowledge of the game.

Honesty or Cheating? We proposed in both scenarios a separation of model and strategy components. In the first scenario, an agent is composed of two, independant but communicating, parts: model (trust or don't trust) and strategy (lie or don't lie). In the second scenario, there are two kinds of agents: buyers (implementing the model) and sellers (implementing the strategy). A problem that has not been solved is to know who should implement the strategy? It could be the organizers of the ART contest, but in this case we take the risk of defining a strategy set too restricted. Otherwise, the participants can implement these, but in that case we take the risk of having agents defined specially to be compliant with the model of that participant.

Open Systems The Online Market scenario allows openness.

Our goal here is not to show how that new scenarios are perfect, because they are not! Moreover, the scenarios are only given as examples. The point here, is to see how **a set of scenarios** can solve the problems we were facing.

While defining new scenarios, one should keep in mind that a scenario must reflect real life problems and avoid toy problems.

5 Conclusion

Although ART has been contributing as a common testbed for trust and reputation models, it has some drawbacks. We listed them in this article and proposed a solution, along with a new approach for the definition and evaluation of scenarios. The lack of reputation has been solved by the Trust Game scenario which has very few direct interactions, thus making the agents rely on other sources. The problem of reputation's semantic has already been handled with a specific ontology for reputation [3].

Implementing a real model as an agent on a game is still not easy, but now, instead of trying to force it into ART, it's possible to find the scenario which is the closest from the model and make it fit onto it.

Another question that was raised concerned the evaluation of the model that is noisy under ART, because the agent in its whole is evaluated. Both scenarios we proposed suggest separation between the model and the other strategic or lying concerns.

Finally, the ideas proposed in this article will be submitted to the ART workgroup for discussion.

References

1. ART. The art testbed. `http://www.art-testbed.net/`.
2. J. Berg, J. Dickhaut, and K. McCabe. Trust, Reciprocity, and Social History. *Games and Economic Behavior*, 10(1):122–142, 1995.
3. S. Casare and J. Sichman. Towards a functional ontology of reputation. *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 505–511, 2005.
4. C. Castelfranchi and R. Falcone. The dynamics of trust: From beliefs to action. *Proceedings of the Second Workshop on Deception, Fraud and Trust in Agent Societies*, pages 41–54, 1999.
5. C. Castelfranchi and R. Falcone. Social trust: a cognitive approach. *Trust and deception in virtual societies table of contents*, pages 55–90, 2001.
6. B. Esfandiari and S. Chandrasekharan. On how agents make friends: Mechanisms for trust acquisition, 2001.
7. K. K. Fullam, T. B. Klos, G. Muller, J. Sabater, A. Schlosser, Z. Topol, K. S. Barber, J. S. Rosenschein, L. Vercouter, and M. Voss. A specification of the agent reputation and trust (art) testbed: experimentation and competition for trust in agent societies. In *AAMAS '05: Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 512–518, New York, NY, USA, 2005. ACM.
8. K. K. Fullam, T. B. Klos, G. Muller, J. Sabater, A. Schlosser, Z. Topol, K. S. Barber, J. S. Rosenschein, L. Vercouter, and M. Voss. The Agent Reputation and Trust (ART) Testbed Game Description (version 2.0). 2006.
9. K. K. Fullam, T. B. Klos, G. Muller, J. Sabater-Mir, Z. Topol, K. S. Barber, J. S. Rosenschein, and L. Vercouter. The Agent Reputation and Trust (ART) testbed architecture. In 8^{th} *Workshop on Trust in Agent Societies*, 2005.
10. M. Gomez, J. Sabater-Mir, J. Carbo, and G. Muller. Improving the ART-Testbed, Thoughts and Reflections. *Proceedings of the Workshop on Competitive Agents in the Agent Reputation and Trust Testbed at CAEPIA-2007, Salamanca, Spain*, pages 1–15, 2007.
11. S. Marsh. *Formalising Trust as a Computational Concept.* PhD thesis, 1994.
12. G. Muller and L. Vercouter. L.i.a.r.: Achieving social control in open and decentralised multi-agent systems. Technical Report 2008-700-001, ENSM-SE / G2I, 2008.
13. J. Murillo and V. Muoz. Agent UNO: Winner in the 2007 Spanish ART Testbed competition. *Proceedings of the Workshop on Competitive Agents in the Agent Reputation and Trust Testbed at CAEPIA-2007, Salamanca, Spain*, 2007.
14. J. Sabater, M. Paolucci, and R. Conte. Repage: REPutation and ImAGE Among Limited Autonomous Partners. *Journal of Artificial Societies and Social Simulation*, 9(2):3, 2006.
15. J. Sabater and C. Sierra. REGRET: reputation in gregarious societies. *Proceedings of the fifth international conference on Autonomous agents*, pages 194–195, 2001.
16. J. Sabater and C. Sierra. Review on Computational Trust and Reputation Models. *Artificial Intelligence Review*, 24(1):33–60, 2005.
17. M. Schillo, P. Funk, and M. Rovatsos. Using Trust For Detecting Deceitful Agents In Artificial Societies. *Applied Artificial Intelligence*, 14(8):825–848, 2000.
18. W. Teacy, T. Huynh, R. Dash, N. Jennings, M. Luck, and J. Patel. The ART of IAM: The Winning Strategy for the 2006 Competition. In *Proceedings of The 10th International Workshop on Trust in Agent Societies*, pages 102–111, 2007.

19. L. Vercouter, S. J. Casare, J. S. Sichman, and A. Brandão. An experience on reputation models interoperability based on a functional ontology. In *IJCAI*, pages 617–622, 2007.