Learning Trust Strategies in Reputation Exchange Networks

Karen K. Fullam and K. Suzanne Barber
Laboratory for Intelligent Processes and Systems
The University of Texas at Austin
1 University Station Stop C5000, Austin, TX, 78712
+1-512-471-5350
{kfullam, barber}@lips.utexas.edu

ABSTRACT

An agent's *trust decision strategy* consists of the agent's policies for making trust-related decisions, such as who to trust, how trustworthy to be, what reputations to believe, and when to tell truthful reputations. In reputation exchange networks, learning trust decision strategies is complex, compared to non-reputation-communicating systems. When potential partners may exchange reputation information about an agent, the agent's interactions with one partner are no longer independent from interactions with another; partners may tell each other about their experiences with the agent, influencing future behavior. This research enumerates the types of decisions an agent faces in reputation exchange networks, explains the interdependencies between these decisions, and correlates rewards to each decision. Experimental results using the Agent Reputation and Trust (ART) Testbed demonstrate the success of strategy-learning agents over agents employing naive strategies. The variation in performance of reputation-based learning vs. experience-based learning over different opponents illustrates the need to dynamically determine when to utilize reputations vs. experience in making trust decisions.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence – *Intelligent Agents, Multiagent Systems*

General Terms

Algorithms, Experimentation.

Keywords

Trust, Reputation, Multi-Agent Systems, ART Testbed.

1. Introduction

When an agent does not have the resources to accomplish its goals alone, it may seek needed resources from other agents in a multi-agent system. These resources are acquired through transactions, which expose at least one party to risk, since a transaction partner might not follow through on the (implicitly or explicitly) agreed upon contract, either by choice or inability. An agent can protect itself against this risk of transaction failure by learning estimates of partner trustworthiness.

Learning trust from transaction experiences [5, 12, 16] is advantageous when agents have opportunities for numerous repeated interactions. When the outcome of interactions are observable, transaction experiences provide a trust-strategy-learning agent with trustworthiness feedback that is certain. Unfortunately, basing trust on transaction experiences means risk exposure is unavoidable; interactions must take place while partners' trustworthiness characteristics are being learned [7].

Reputation exchange is useful for quickly learning trustworthiness characteristics of other agents [14]. Adapted from [8], a reputation is a (not necessarily truthful) communication from one agent to another about the sender's trust in a third subject-agent. A multi-agent system whose communication protocols permit the exchange of reputations [11, 17, 18, 20, 21] is advantageous in systems with large populations and sparse transactions. Further, reputation exchange reduces an agent's risk exposure; an agent risks only the price of reputations it purchases, rather than the value of resources exchanged in a potential transaction. Agents entering a multi-agent system can quickly learn trust models by requesting reputations from more knowledgeable agents.

An agent's *trust decision strategy* consists of the agent's policies for making trust-related decisions with the aim of achieving greatest benefit. The simplest trust decision strategy, in systems without reputation exchange, encompasses decisions regarding both 1) who to trust (to minimize risk of failed transactions), and 2) how trustworthy to be (to exploit others). But the advantage of agents communicating reputation information adds another dimension to an agent's trust decision strategy; in reputation exchange networks, an agent must make additional decisions regarding 1) which reputations to believe (to better choose who to trust) and 2) how truthful to be in communicating reputations to others (to manipulate what others think about who to trust).

Reinforcement learning has been successfully applied to simple trust decision strategies for deciding who to trust—and how trustworthy to be—in agent transactions [3, 13]. This research seeks to apply similar techniques to help agents learn more

AAMAS'06, May 8-12, 2006, Hakodate, Hokkaido, Japan.

comprehensive trust decision strategies required in networks of agents exchanging reputation information. However, in reputation exchange networks, the complexity of the learning problem increases dramatically, compared to systems in which agents do not communicate reputation information. The agent must make more trust-related decisions, regarding not only who to trust and how trustworthy to be, but also regarding which reputation information to believe and how truthful to be when communicating reputations to others. Compounding the problem, the effects of these decisions may not be immediately observable, since the agent cannot easily gauge the nature and extent of reputation exchange occurring among its potential partners. Further, when potential partners may exchange reputation information about an agent, the agent's interactions with one partner are no longer independent from interactions with another; partners may tell each other about experiences with the agent, influencing how each behaves with the agent in the future.

Toward the goal of enabling agents to learn comprehensive trust decision strategies in reputation exchange networks, this research first enumerates the types of decisions an agent faces in systems with reputation exchange. Next, using the Agent Reputation and Trust (ART) Testbed [10] as a case study, the interdependencies between these decisions are explained. This research employs a learning approach for each decision; by establishing assumptions to ease interdependencies, rewards are correlated to each decision type. Experimental results demonstrate the success of strategy-learning agents over agents employing naive strategies. The variation in performance of reputation-based learning vs. experience-based learning over different opponents illustrates why agents must be able to dynamically determine when to utilize reputations vs. experience in making trust decisions.

2. Trust Decision Strategies

A trust decision strategy describes an agent's choices regarding four trust-related decision types, defined according to agent role (trustee or truster) and transaction (fundamental or reputation), as shown in Figure 1. Fundamental transactions are the basic transactions driving the need for trust modeling in the multi-agent system. Reputation transactions are distinct from fundamental transactions because the outcome of reputation transactions directly impacts decisions regarding fundamental transactions (this impact is shown later). An agent's trust-based decisions are:

1) *Should I trust?* For each potential fundamental transaction, the agent must decide whether to participate in— and expose itself to the risk of— the transaction.
2) *How trustworthy should I be?* For each agreed upon fundamental transaction, the agent must identify the degree to which cheating or honesty yields the greatest benefit.
3) *Should I believe this reputation?* For each potential reputation transaction, the agent must decide whether to conduct the transaction, then use the reputation to determine whether it should trust regarding fundamental transactions.
4) *How accurate should the reputation I tell be?* For each agreed upon reputation transaction, the agent must identify its level of truthfulness in communicating a reputation. The agent can only control the accuracy of reputations it communicates to the extent of its own trust model accuracy.

Agent Role	**Transaction:** Reputation	**Transaction:** Fundamental
Trustee	How accurate should the reputation be?	How trustworthy should I be?
Truster	Should I believe this reputation?	Should I trust?

Figure 1. Trust decision types by transaction and agent role.

How the agent makes these decisions depends, in part, on the characteristics of the transaction. If the transaction protocol calls for sequential action between the two partners (for example, payment then delivery), the first partner to act serves as truster, and the second as trustee. If the transaction protocol dictates simultaneous action, both partners concurrently act as both trustee and truster. Future decisions are also impacted by the observability of a transactions outcome. For example, if an agent receives a payment as the transaction outcome, most likely it can tell if the payment is satisfactory. However, an agent may be unable to tell whether a reputation transaction is successful (the received reputation is accurate) if it has little knowledge about the trustworthiness of the reputation's subject agent.

2.1 The Complexity of Learning Trust Decision Strategies

This research proposes a method by which agents can learn trust decision strategies using a q-learning technique. Q-learning [15] is a well-known tool for learning best decision-making strategies by associating actions with expected rewards. Several researchers [3, 4, 9, 13, 19] have employed reinforcement learning techniques to discover strategies for trust decisions. However, these trust decisions only relate to fundamental transactions, as in the Prisoner's Dilemma [2]. Learning trust decision strategies in reputation-transaction systems is significantly more difficult, since in these systems, fundamental transaction decisions are not independent from each other and from reputation transaction decisions. An agent's fundamental transactions with one partner may influence its fundamental transactions with another, if its two partners are able to communicate reputation information. Learning an agent's trust decision strategy is also complex because multiple decisions can affect common results. For example, falsely spreading negative reputations about a trustworthy partner might cause other agents to trust the lying agent while isolating the falsely accused agent. Similarly, falsely spreading positive reputations about untrustworthy partners might hurt those agents receiving the reputations, thus improving the standing of the lying agent. Or, an agent might perform truthfully in hopes of receiving better treatment from partner agents, yet truthful behavior might benefit the agent's partners too much.

Considering the interdependent nature of these trust-related decisions, an agent attempting to learn the best combination of decisions faces a complexity problem. When making decisions about "should I trust?" an agent faces $O\left(2^{ae}\right)$ decision combinations from which to choose, assuming that choosing to trust is a binary decision, a represents the number of agents in the system, and e represents the number of possible transaction categories for which agents might have different expertise. When deciding "how trustworthy should I be?" an agent has n^p decision combinations, where n describes a number of possible, discrete degrees of fundamental-transaction trustworthiness and p describes the number of fundamental transaction requests received from other agents. Since p can be as large as $(a-1)e$, the complexity of this decision is $\mathrm{O}\left(n^{ae}\right)$.

When deciding "should I believe this reputation?" an agent faces $\mathrm{O}\left(2^{a^2 e}\right)$ decisions, assuming that choosing to believe a reputation is a binary decision, a (signifying the number of agents in the system) represents the number of reputation providers and the number of subject agents about whom reputations may be provided, and e (the number of possible transaction expertise categories) represents the number of categories for which each subject agent's trustworthiness can be questioned. When deciding "should I tell an accurate reputation?" an agent has m^q decision combinations, where m describes a number of possible, discrete degrees of reputation-transaction trustworthiness and q describes the number of reputation transaction requests received from other agents. Since q can be as large as $(a-1)ae$, the complexity of this decision is $\mathrm{O}\left(m^{a^2 e}\right)$.

Each decision of a given type is interdependent, but further, each decision type is interdependent. Therefore, in a given discrete timestep, in which an agent has the opportunity to interact with each other agent, as both truster and trustee, for both fundamental and reputation transactions, the agent's trust decision strategy has a complexity of $\mathrm{O}\left(2^{ae+a^2 e} n^{ae} m^{a^2 e}\right)$. It is not feasible for an agent to learn its trust decision strategy without introducing some assumptions to relax interdependencies between decisions. The following section demonstrates how an agent can introduce these assumptions yet still learn a useful trust decision strategy.

3. Case Study: The ART Testbed

The Agent Reputation and Trust (ART) Testbed [1] is used as a case study for learning trust decision strategies. A brief overview of the ART Testbed game is described here. Readers are referred to [10] for more detail; all ART Testbed notation described here is consistent with that reference.

In the ART Testbed's artwork appraisal domain, agents act as appraisers who are hired by clients to deliver *appraisals* about paintings, each for a fixed client fee f. Agents compete for clients, who are assigned based on agents' past appraisal accuracy. In attempting to produce accurate appraisals, appraiser agents may purchase *opinions* (valuations of the painting by other agents) for a fixed fee, c_p. Opinion providers choose an amount, c_g, to pay (to the system) to generate an opinion, symbolizing the effort to perform a valuation; the amount c_g corresponds to the accuracy of the generated opinion. Appraiser agents may also purchase *reputations* (for a fixed fee c_r) from each other to help assess the trustworthiness of opinion providers. Agents compete to achieve the highest bank balance.

Designed as a tool for experimentation and competition, the ART Testbed has earned the attention of numerous researchers in the international trust research community. The Testbed design allows variation of agent strategies and game parameters for easy experimentation, as well as competition against and experiment replication by other researchers. Transactions are conducted sequentially, permitting the separate examination of truster and trustee decisions. Opinion purchases correlate to fundamental transactions, the outcomes of which (opinion accuracy) are observable at the end of each timestep, when true painting values are revealed. The outcomes of reputation purchases are not observable, though reputation receivers may estimate the accuracy of the reputations they receive against previously built trust models. Most importantly, the game's complexity illustrates the interdependencies of trust-related decisions.

The four trust-related decisions translate to the ART Testbed domain as shown in Figure 2. As an Opinion Requester, an agent must decide which opinions to purchase. As an Opinion Provider, an agent must decide how much payment, c_g, to invest in the opinions it communicates. As a Reputation Requester, an agent must decide which reputations to purchase. As a Reputation Provider, an agent must decide how truthful to be, compared to its own trust models, when communicating reputations.

		Transaction	
		Reputation	Fundamental
Agent Role	Provider	How accurate a reputation should I provide?	How accurate an opinion should I provide?
	Requester	Should I purchase this reputation?	Should I purchase this opinion?

Figure 2. Trust decision types in the ART Testbed domain.

Several assumptions are made to the ART Testbed game to remove details unnecessary to this research purpose. First, agent transaction protocols are simplified: Opinion Providers do not communicate opinion certainties, and Opinion and Reputation Providers always satisfy requests, though provided opinions and reputations may be untruthful. Further, agents do not generate opinions for their own clients, and they weight all received opinions equally when calculating their final appraisals. To eliminate non-agent-controlled variances in opinion accuracy, all agents are assigned the same expertise levels for all painting eras (fundamental transaction types).

3.1 Identifying Decision Interdependencies

Figure 3 shows a diagram of interdependencies between the ART Testbed's four trust-related decisions (darkened diamonds in the figure) and resulting observable feedback (costs and earnings, darkened boxes in the figure), enumerated here:

1) *Opinion Requester:* The Opinion Requester's decision regarding which opinions to purchase determines its opinion costs and the accuracy of its resulting final appraisals, which, in turn, impact its client revenue.
2) *Opinion Provider:* The Opinion Provider's decision regarding the accuracy of appraisals to provide determines its opinion order costs (c_g). This decision also affects Opinion Requesters' accuracy of final appraisals and client revenue. Since appraiser agents compete for client revenue, effects on Opinion Requesters' client revenue also impact the client revenue of Opinion Providers. Finally, the Opinion Provider's opinion accuracy influences the number of future requests from Opinion Requesters, which determines the Opinion Provider's future opinion revenue.
3) *Reputation Requester:* The Reputation Requester's decision regarding which reputations to purchase impacts its

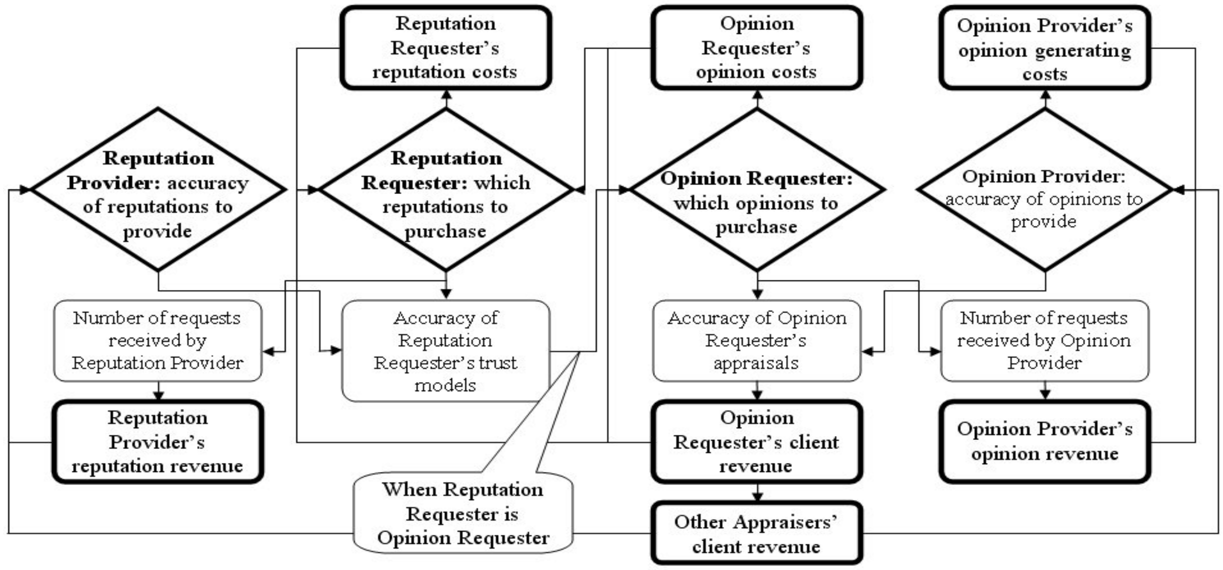

Figure 3. Trust decision interdependencies in the ART Testbed domain.

own reputation costs, as well as its ability to determine which opinions to request, which (as mentioned for the Opinion Requester's role) impacts both the Reputation Requester's opinion costs and client revenue.

4) *Reputation Provider:* The Reputation Provider's decision regarding the accuracy of reputations to provide influences the Reputation Requesters' ability to determine which opinions to request, which affects the Reputation Requester's, and ultimately the Reputation Provider's, client revenue. Also, the Reputation Provider's reputation accuracy influences the number of future requests from Reputation Requesters, which determines the Reputation Provider's future reputation revenue.

Multiple decisions, over all four agent roles, influence multiple feedback elements, making the correlation of specific decisions to their resulting rewards difficult. The following section introduces dependency-eliminating assumptions and proposes methods for calculating rewards associated with each decision.

3.2 Trust Decision Rewards

To conduct q-learning for each trust-related decision, a reward must be assigned to each decision. Since decisions are interdependent, and observable costs and revenues are aggregated, two assumptions are presented to ease these interdependencies and divide observable earnings into per-decision rewards:

Assumption 1: Individual decisions of a single decision type are assumed independent of each other (i.e. all of an agent's opinion-requesting decisions are independent of each other, regardless of opinion provider or era).

Assumption 2: Client revenue is only influenced by the Opinion Requesting decision (thus agents' client revenues are independent of each other).

3.2.1 Opinion Requester Rewards

The Opinion Requester's choice regarding the purchase of a single opinion is binary: either request or not request the opinion. The choice to not request an opinion is assumed to yield a reward of zero, but calculating a reward ($Q_{OR}(true)$) associated with choosing to request an opinion is more complicated. Utilizing the assumptions above, a reward associated with each choice to request an opinion can be estimated by attributing a portion of the Opinion Requester's resulting client revenue first to each computed appraisal, then to each opinion composing that appraisal. First, the total client revenue, v_{total}, attributed to all client appraisals, is calculated as the number of Opinion Requester's clients in the next timestep, n_c, times the earned, per-client fee, f:

$$v_{total} = n_c f \ .$$

Next, the estimated client revenue portion, v_c, attributed to a single appraisal, for client c, is calculated as:

$$v_c = v_{total}\left(\frac{\frac{1}{\varepsilon_{p_c^*}}}{\sum_{i\in C}\left(\frac{1}{\varepsilon_{p_i^*}}\right)}\right),$$

where C represents the set of the Opinion Requester's client appraisals and $\varepsilon_{p_c^*}$ represents the normalized error of the client's painting appraisal, p_c^*, from the painting's true value, t_c:

$$\varepsilon_{p_c^*} = \left|\frac{p_c^* - t_c}{t_c}\right|.$$

An estimated client revenue portion, $v_{c,a}$, is attributed to each opinion (by opinion provider a for client c's painting):

$$v_{c,a} = v_c\left(\frac{\frac{1}{\varepsilon_{p_{c,a}}}}{\sum_{i\in A}\left(\frac{1}{\varepsilon_{p_{c,i}}}\right)}\right),$$

where A represents the set of Opinion Providers for painting c and $\varepsilon_{p_{c,a}}$ represents the normalized error of a single opinion, $p_{c,a}$, from the painting's true value, t_c:

$$\varepsilon_{p_{c,a}} = \left| \frac{p_{c,a} - t_c}{t_c} \right| .$$

Finally, the reward, $w_{OR,c,a}$, associated with requesting the opinion from Opinion Provider a about client c's painting is calculated as the estimated client revenue portion, $v_{c,a}$, minus the cost, c_p, of requesting the opinion:

$$w_{OR,c,a} = v_{c,a} - c_p \ .$$

This reward is averaged for all requested opinions from a given Opinion Provider a and era e, to which paintings c belong.

3.2.2 Opinion Provider Rewards

The Opinion Provider must choose the accuracy of the opinion it provides to an Opinion Requester. More precisely, the Opinion Provider must choose the amount, c_g, to pay to generate an opinion, where c_g is related to opinion accuracy. First, it is assumed that the Opinion Provider makes the same opinion accuracy decision for all opinion requests from the same Opinion Requester about paintings from the same era. Thus, using the independence assumptions, the reward, $w_{OP,a,e}$, associated with choosing an amount, c_g, to pay to generate an opinion, is calculated as:

$$w_{OP,a,e} = n_{p,a,e}\left(c_p - c_{g,a,e}\right),$$

Where $n_{p,a,e}$ represents the number of opinion requests received in the next timestep from agent a about paintings from era e, c_p represents the per-opinion cost received for providing opinions, and $c_{g,a,e}$ represents the chosen amount paid to generate each opinion for requester a about paintings from era e.

3.2.3 Reputation Provider Rewards

The Reputation Provider must choose the accuracy of the reputation it provides to a Reputation Requester. For this research, it is assumed that all agents have a mutual understanding that reputation values correspond to q-values ($Q_{OR}(true)$) for deciding in the affirmative to request opinions. Since the ART Testbed only permits the exchange of reputations as values between zero and one, q-values are converted from a scale between $-f$-c_p (the assumed worst-case penalty for requesting an opinion) and f-c_p (the assumed best-case reward for requesting an opinion). Therefore, the conversion from $Q_{OR,s,e}(true)$ (the Reputation Provider's actual q-value for requesting opinions from opinion provider (subject-agent) s for expertise era e) to $r_{s,e}$ (the scaled, believed reputation about s for e) is given by:

$$r_{s,e} = \frac{Q_{OR,s,e}\left(true\right) + f + c_p}{2f} \ .$$

Reputation accuracy is parameterized using the δ variable to describe the relationship between the Reputation Provider's believed reputation, $r_{s,e}$, and the actual reputation, $r^*_{a,s,e}$, it provides to Reputation Requester a. Intuitively, $\delta_{a,s,e}$ represents the percent difference between the believed reputation $r_{s,e}$ and the limits of permissible reputation values, where negative δ values imply pessimistic reputations and positive δ values imply optimistic reputations. Formally, $r^*_{a,s,e}$ is calculated as:

$$r^*_{a,s,e} = \begin{cases} r_{s,e}\left(\delta_{a,s,e}+1\right), \text{if } -1 \le \delta_{a,s,e} < 0 \\ r_{s,e}\left(1-\delta_{a,s,e}\right)+\delta_{a,s,e}, \text{if } 0 \le \delta_{a,s,e} \le 1 \end{cases} .$$

Using the independence assumptions, the reward, $w_{RP,a,s,e}$, associated with providing a reputation whose accuracy correlates to $\delta_{a,s,e}$ is given by:

$$w_{RP,a,s,e} = n_{r,s,a,e} c_r \ ,$$

where c_r is the fixed reputation purchase cost and $n_{r,s,a,e}$ is the number of reputations requested in the next timestep by a about s for era e. Since the maximum possible value of $n_{r,s,a,e}$ is 1 (in a single timestep, a Reputation Requester can only request one reputation from a given provider about a given subject-agent and expertise era), an $n_{r,s,a,e}$ value of zero or one is recorded in each timestep for a chosen δ value, until the timestep in which a new reputation provider decision must be made for the given requester, subject-agent, and era.

3.2.4 Reputation Requester Rewards

The Reputation Requester's decision to purchase reputations is binary. As in the case of the opinion-requesting decision, the reward of not requesting a reputation is assumed to be zero. The reward associated with requesting a reputation is based on the indirect benefit of making savvy opinion requests based on quality reputations, and is estimated as follows. The error of a reputation, $\varepsilon_{r^*_{a',s,e}}$, is estimated by averaging the error between the reputation and computed rewards for requested opinions in C_e (the set of opinions whose paintings belong to this reputation's era e):

$$\varepsilon_{r^*_{a',s,e}} = \frac{\sum_{i \in C_e} \left| r^*_{a',s,e} - w_{OR,i,s} \right|}{\left| C_e \right|} \ ,$$

where a' represents the provider of the reputation about opinion provider (subject-agent) s.

An estimated client revenue portion, $v_{r,a',s,e}$, is attributed to each reputation (by reputation provider a' about subject-agent s and era e) in this way: a portion of the client revenue ($w_{OR,c,s}$, attributed to requesting an opinion c) is attributed to each requested reputation according to the reputation's estimated error (A' is the set of reputation providers about s for e). These client revenue portions are averaged over all opinions whose paintings belong to era e:

$$v_{r,a',s,e} = \frac{\sum_{i \in C_e} \left(w_{OR,i,s} \left(\frac{\frac{1}{\varepsilon_{r^*_{a',s,e}}}}{\sum_{j \in A'} \left(\frac{1}{\varepsilon_{r^*_{j,s,e}}} \right)} \right) \right)}{\left| C_e \right|} .$$

The reward, $w_{RR,a',s,e}$, attributed to requesting a given reputation is simply its calculated client revenue portion minus its cost:

$$w_{RR,a',s,e} = v_{r,a',s,e} - c_r \ .$$

The Reputation Requester must make a second decision which brings to light a significant problem [6]: how should reputation information influence the agent's existing experience-based trust model (the Opinion Requester's $Q_{OR}(true)$ values)? Since a Reputation Requester purchases reputations for the purposes of making better opinion requesting decisions, an agent in the reputation requesting role should be jointly considered in the opinion requesting role. This research employs a weighting parameter λ to perform a weighted average between aggregated reputations and the existing $Q_{OR,s,e}(true)$ value. When λ is 0, no weight is given to reputations, and opinion request decisions are made using only previous experience according to Section 3.2.1. Conversely, when λ is 1, the Opinion Requester's $Q_{OR}(true)$ values are given no weight, instead, the Opinion Requester makes opinion requests based on an average of q-values received as reputations. Reputation request rewards are multiplied by λ since reputations are only responsible for λ fraction of the opinion requesting decision. Note that when an agent serves as a Reputation Provider, it provides reputations from its experience-based $Q_{OR}(true)$ values, excluding any reputations it may have received.

4. Experiments

Once rewards are associated with each trust-related decision as in Section 3.2, q-learning is performed. Experiments are conducted to compare the resulting trust-strategy-learning technique against several standard strategies.

4.1 Experimental Setup

The ART Testbed parameters [10] for these experiments include client fee (f) of 100, opinion cost (c_p) of 10, and reputation cost (c_r) of 1. The accuracy of generated opinions is governed by the parameters s^* (related to appraiser expertise, the standard deviation of opinion error distribution) and α (determines the impact of the opinion-generating cost, c_g). For these experiments, s^* equals 0.1 for all agents regarding all eras, and α is set to 5.0. In each experiment, the simulation assigns client paintings from three eras. The parameter q, which determines the influence of previous client shares on current client shares, equals 0.5. Agents are permitted to select c_g values in one-unit increments between zero and ten (since ten is the opinion purchase price c_p, investments larger than that amount are impractical).

Table 1 displays the seven strategy variations compared. The Naïve-Honest strategy requests opinions from all other agents in the system and provides very accurate opinions, investing $c_g = 10$ to generate each opinion. This strategy is naïve in that reputations are not utilized, therefore none are requested, and default, optimistic reputations of 1.0 are provided to requesters. The Naïve-Cheating strategy is similar to the Naïve-Honest strategy except that it provides very inaccurate opinions (investing $c_g = 0$ in the opinions it provides) and pessimistic reputations (0.0). The Op-Learning-Honest strategy learns opinion requesting decisions from experience (Section 3.2.1) and opinion providing decisions as in Section 3.2.2, while requesting reputations from all agents and providing truthful reputations. The Op-Learning-Cheating strategy behaves similarly to Op-Learning-Honest, but consistently provides pessimistic reputations of 0.0. The Random strategy randomly makes decisions for each of the four decision types. The two Learning strategies demonstrate the learning techniques described in this paper, one with a λ value of 0.0 (learning opinion requesting decisions based on experience only), and the other using λ of 1.0 (learning opinion requesting decisions based on reputations only). All Learning agents employ the same q-learning parameters for each learned decision: learning rate of 0.5, discount factor of 0.0, and temperature, or degree of exploration, of 0.5.

Table 1. Agent strategies.

Strategy Name	Opinion Req.	Opinio n Prov.	Reputatio n Req.	Reputatio n Prov.
Naïve Honest	request from all	accurate (c_g=10)	none	1.0 (δ = 1)
Naïve Cheating	request from all	inacc. (c_g=0)	none	0 (δ = -1)
Op-Learning Honest	learn by experience	learn	request from all	truthful (δ = 0)
Op-Learning Cheating	learn by experience	learn	request from all	0 (δ = -1)
Random	random	random	random	random
Learning (λ = 0)	learn by experience	learn	learn	learn
Learning (λ = 1)	learn from reputation s	learn	learn	learn

In each of ten experiments, a single Learning agent (λ=0 or λ=1) competes against five agents, all of a single strategy: Naïve-Honest, Naïve-Cheating, Op-Learning-Honest, Op-Learning-Cheating, or Random. Each experiment of six agents is run for ten thousand timesteps (for three independent runs) to demonstrate strategy convergence. Each agent's performance is gauged in terms of converged, per-timestep earnings.

4.2 Experimental Results

Results from all ten experiments are shown in Figure 4. In each experiment, per-timestep earnings for each agent converged within approximately the first one hundred timesteps. In all cases, the Learning agent's five opponents (all of like strategy) achieved per-timestep earnings which were statistically similar to each other; these earnings were averaged to arrive at the opponent's average per-timestep earnings (white bars) shown in Figure 4. The opponents' earnings are compared against the Learning agent's earnings (black bars). All differences between Learning agents and their opponents are statistically significant (α = 0.1).

In the first pair of experiments, a Learning agent (λ=0 or λ=1) competes in a system with five Naïve-Honest agents. The Learning agent increases its bank balance at nearly twice the rate of each Naïve-Honest agent. Because the Learning agent purchases fewer reputations when λ=0 than when λ=1, the λ=0 agent achieves slightly higher per-timestep earnings than the λ=1 case, and its opponents' earnings are slightly lower (failing to utilize reputations does not penalize the Learning agent, because its opponents are honest). The Learning agent's behavior converges to always requesting opinions from the Naïve-Honest agents, while investing little in the opinions it provides and, in the λ=0 case, rarely requesting reputations.

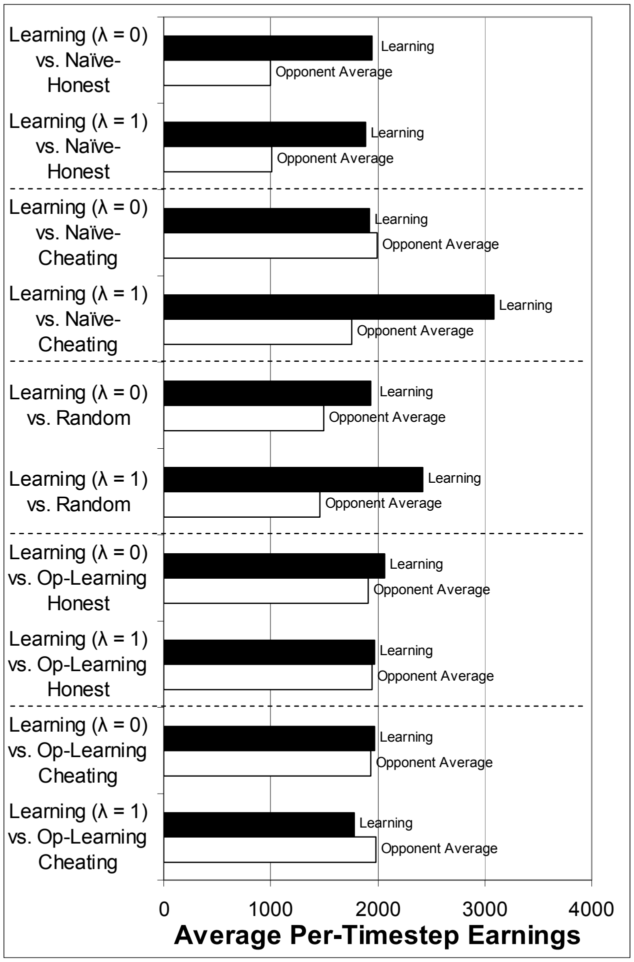

Figure 4. Average, per-timestep earnings for learning strategies (λ=0 and λ=1) versus opponents.

In the second pair of experiments, a Learning agent (λ=0 or λ=1) competes against five Naïve-Cheating agents. The Learning agent with λ=0 maintains per-timestep earnings slightly less than those of the Naïve-Cheating agents. The Learning agent with λ=1 maintains much higher earnings than the Naïve-Cheating agents; it makes use of its opponents' pessimistic reputations, an accurate reflection of the opponents' tendency to cheat.

The third pair of experiments demonstrate that the Learning agent achieves higher earnings than its five Random strategy opponents in both the λ=0 and λ=1 cases. The λ=1 Learning agent achieves higher per-timestep earnings than the λ=0 agent, but Random agents' earnings are the same between the two cases. The Learning agent succeeds by employing a mixed strategy for each of its four decision types.

The fourth pair of experiments compares the Learning agent against five Opinion-Learning-Honest agents. In both cases (λ=0 and λ=1), the Learning agent maintains greater earnings than its opponents. All agents learn to always request opinions and to provide opinions with accuracy that is both beneficial to competitors (to ensure repeated transactions) yet not too costly to generate. Because their opponents naively always request reputations, the Learning agents provide randomly-accurate reputations. As in the Naïve-Honest experiments, the Learning agent yields higher earnings when λ=0 than when λ=1 because the agent purchases fewer reputations in the λ=0 case.

In the last pair of experiments, the Learning agent competes against five Opinion-Learning-Cheating agents. When λ=1, the Learning agent earns lower earnings than its opponents, because it is misled by its opponents' inaccurate reputations. When λ=0, the Learning agent achieves earnings higher than its opponents, relying on the experience-based learning, rather than reputations, to decide which opinions to purchase.

The experiments demonstrate that in only two cases (Learning λ=0 vs. Naïve-Cheating and Learning λ=1 vs. Op-Learning Cheating), the Learning agent earns less than its opponents. However, in both cases, adjusting the λ value (to λ=1 against Naïve-Cheating and to λ=0 against Op-Learning Cheating) allows the Learning agent to outperform its opponents. Further, generalizations can be made about when certain λ values are beneficial: learning from experience (λ=0) achieves higher earnings when opponents tend to be trustworthy, while learning from reputations (λ=1) is beneficial when reputations accurately reflect opponent trustworthiness.

5. Conclusions

Experimental results demonstrate that learning trust decision strategies, either from experience or reputations, yields higher earnings than several static strategies in the ART Testbed domain Further, the variation in earnings of reputation-based learning (λ=1) vs. experience-based learning (λ=0) over different opponents demonstrates successful agents must be able to dynamically determine when to utilize reputations vs. experience in making trust decisions.

This research yields a domain independent contribution by enumerating decisions—and their resulting complexity—that make up a trust decision strategy (who to trust, how trustworthy to be, what reputations to believe, and how truthful to be in telling reputations). Other contributions of this work are adaptable to various types of problems in which trust must be evaluated and reputations may be exchanged. The work identifies interdependencies (which exist since potential partners can influence each other by exchanging reputations about an agent) between those decisions. By introducing assumptions to ease these interdependencies, rewards are attributed to each decision to facilitate reinforcement learning of trust decision strategies. Reward allocation can be adapted to other domains by adjusting transaction costs and earnings associated with domain-introduced interdependencies.

Approaching the complexity of learning trust decision strategies requires the introduction of numerous assumptions. However, these assumptions pave the way for future investigation, which will seek to relax the independence assumptions of Section 3.2 to improve the overall rewards a learning agent achieves. The authors plan to use learning techniques to help an agent learn the best λ, the degree of reliance on experience- versus reputation-based trust, relative to each opinion-requesting decision. The performance of the strategy-learning techniques described in this research requires further analysis against more advanced opponent agents with dynamic strategies. Also, more agent combinations must be tested—including agent self-play—since trust strategies may yield different results as the composition of the system (strategies and numbers of opponents) change. Since, at the time of writing, the ART Testbed is still awaiting its first competition, sophisticated strategies have yet to be uncovered for comparison. However, the experiments conducted in this research will serve as a benchmark for future progress.

6. Acknowledgment

This research is sponsored in part by the Defense Advanced Research Project Agency (DARPA) Taskable Agent Software Kit (TASK) program, F30602-00-2-0588 and by a subcontract from The University of Texas Applied Research Labs Office of Naval Research Grant N00014-05-1-0857, entitled Improvised Explosive Device. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

7. References

[1] ART Testbed Team. Agent Reputation and Trust Testbed. http://www.lips.utexas.edu/art-testbed/, 2005.

[2] Axelrod, R. The Evolution of Cooperation. New York: Basic Books, 1984.

[3] Banerjee, B., R. Mukherjee, and S. Sen. "Learning Mutual Trust," Proceedings of the Workshop on Deception, Fraud and Trust in Agent Societies at AGENTS-00, pp. 9-14, 2000.

[4] Banerjee, B. and J. Peng. "Countering Deception in Multiagent Reinforcement Learning," Proceedings of the Workshop on Trust, Privacy, Deception and Fraud in Agent Societies at AAMAS-03, Melbourne, Australia, pp. 1-5, 2003.

[5] Barber, K. S. and J. Kim. "Belief Revision Process based on Trust: Agent Evaluating Reputation of Information Sources," Trust in Cyber-societies: Integrating the Human and Artificial Perspectives, vol. 2246, Lecture Notes in Computer Science, Falcone, R., Singh, M., and Tan, Y.-H., Eds.: Springer, pp. 73-82, 2002.

[6] Barber, K. S. and J. Kim. "Soft Security: Isolating Unreliable Agents," Proceedings of The Workshop on Deception, Fraud, and Trust in Agent Societies at AAMAS-02, Bologna, Italy, pp. 8-17, 2002.

[7] Barber, K. S., K. Fullam, and J. Kim. "Challenges for Trust, Fraud, and Deception Research in Multi-agent Systems," Trust, Reputation, and Security: Theories and Practice, R. Falcone, K. S. Barber, L. Korba and M. Singh, Eds., Springer: pp. 8-14, 2003.

[8] Barber, K. S. and K. Fullam. "Applying Reputation Models to Continuous Belief Revision," Proceedings of The Workshop on Deception, Fraud and Trust in Agent Societies at AAMAS-03, Melbourne, Australia, pp. 6-15, 2003.

[9] Crandall, J.W. and M.A. Goodrich. "Establishing Reputation Using Social Commitment in Repeated Games," Proceedings of the Workshop on Learning and Evolution in Agent Based Systems at AAMAS-04, New York, pp. 12-17, 2004.

[10] Fullam, K., T. Klos, G. Muller, J. Sabater, A. Schlosser, Z. Topol, K. S. Barber, J. Rosenschein, L. Vercouter, and M. Voss. "A Specification of the Agent Reputation and Trust (ART) Testbed: Experimentation and Competition for Trust in Agent Societies," Proceedings of The Fourth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2005), Utrecht, pp. 512-518, 2005.

[11] Huynh, T.D., N.R. Jennings, and N. Shadbolt. "On Handling Inaccurate Witness Reports," Proceedings of 8th International Workshop on Trust in Agent Societies at AAMAS-05, Utrecht, pp. 63-77, 2005.

[12] Jonker, C.M. and J. Treur. "Formal Analysis of Models for the Dynamics of Trust Based on Experiences," Proceedings of The 9th European Workshop on Modeling Autonomous Agents in a Multi-Agent World: Multi-Agent System Engineering (MAAMAW-99), pp. 221-231, 1999.

[13] Littman, M. and P. Stone. "Leading Best-Response Strategies in Repeated Games," Proceedings of the Workshop on Economic Agents, Models, and Mechanisms, at IJCAI-01, Seattle, Washington, 2001.

[14] Mui, L., A. Halberstadt, and M. Mohtashemi. "Notions of Reputation in Multi-Agents Systems: A Review," Proceedings of The First International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2002), Bologna, Italy, pp. 280-287, 2002.

[15] Sandholm, T. and R. Crites. "On Multi-agent Q-Learning in a Semi-Competitive Domain," Proceedings of the Workshop on Adaptation and Learning in Multiagent Systems at IJCAI-95, Montreal, Canada, pp. 71-77, 1995.

[16] Schillo, M., P. Funk, and M. Rovatsos. "Using Trust for Detecting Deceitful Agents in Artificial Societies," Proceedings of The Applied Artificial Intelligence Journal, Special Issue on Deception, Fraud and Trust in Agent Societies, pp. 825-848, 2000.

[17] Shi, J., G. Bochmann, and C. Adams. "Dealing with Recommendations in a Statistical Trust Model," Proceedings of the Workshop on Trust in Agent Societies at AAMAS-05, Utrecht, pp. 144-155, 2005.

[18] Teacy, L., J. Patel, N.R. Jennings, and M. Luck. "Coping with Inaccurate Reputation Sources: Experimental Analysis of a Probabilistic Trust Model," Proceedings of The Fourth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2005), Utrecht, pp. 997-1004, 2005.

[19] Weinberg, M. and J. Rosenschein. "Best-Response Multiagent Learning in Non-Stationary Environments," Proceedings of The Third International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2004), New York, pp. 506-513, 2004.

[20] Yolum, P. and M.P. Singh. "Self-Organizing Referral Networks: A Process View of Trust and Authority," Engineering Self-organizing Systems, Lecture Notes in Artificial Intelligence, pp. 195-211, 2003.

[21] Yu, B. and M.P. Singh. "An Evidential Model of Distributed Reputation Management," Proceedings of The First International Joint Conference on Autonomous Agents and Multiagent Systems, Bologna, Italy, pp. 294-301, 2002.