

Notions of Reputation in Multi-Agents Systems: A Review

Lik Mui

MIT Laboratory for Computer Science
200 Technology Square
Cambridge, MA 02139
001-617-253-5860
lmui@lcs.mit.edu

Ari Halberstadt

Magiccookie
9 Whittemore Road
Newton, MA 02458, USA
001-617-332-0960
ari@magiccookie.com

Mojdeh Mohtashemi

MIT Laboratory for Computer Science
200 Technology Square
Cambridge, MA 02139
001-617-253-5860
mojdeh@lcs.mit.edu

ABSTRACT

Reputation has recently received considerable attention within a number of disciplines such as distributed artificial intelligence, economics, evolutionary biology, among others. Most papers about reputation provide an intuitive approach to reputation which appeals to common experiences without clarifying whether their use of reputation is similar or different from those used by others. This paper argues that reputation is not a single notion but one with multiple parts. After a survey of existing works on reputation, an intuitive typology is proposed summarizing existing works on reputation across diverse disciplines. This paper then describes a simple simulation framework based on evolutionary game theory for understanding the relative strength of the different notions of reputation. Whereas these notions of reputation could only be compared qualitatively before, our simulation framework has enabled us to compare them quantitatively.

Categories and Subject Descriptors

I.2.11 [Distributed Artificial Intelligence]: Coherence and Coordination, Intelligent Agent, Multiagent Systems.

General Terms

Algorithms, Measurement, Economics, Experimentation, Theory.

Keywords

Multi-agents systems, reputation, agent modeling, cooperation.

1. INTRODUCTION

Reputation refers to a perception that an agent has of another's intentions and norms. Evolutionary biologists have used reputation to explain why selfish individuals cooperate (e.g., Nowak and Sigmund, 1998). Economists have used reputation to explain "irrational" behavior of players in repeated economic games (e.g., Kreps and Wilson, 1982). Computer scientists have used reputation to model the trustworthiness of individuals and firms in online marketplace (e.g., Zacharia and Maes, 1999). Although an intuitive concept, this paper argues that reputation is not a single notion but one with multiple parts. Several notions of reputation exist in the literature, although their distinction is often not made.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

AAMAS'02, July 15-19, 2002, Bologna, Italy.

Copyright 2002 ACM 1-58113-480-0/02/0007...\$5.00.

Reputation is often confused with concepts related to it, such as trust (e.g., Abdul-Rahman, *et al.*, 2000; Yu, *et al.*, 2001). The trouble with a number of reputation studies lie in their lack of careful understanding based on existing social, biological, and computational literatures regarding reputation. We refer to Ostrom (1998) and Mui, *et al.*, (2002) for a clarification of reputation, trust, and related concepts.

Section 2 reviews the basic notions of reputation as used in several disciplines. Section 3 proposes a typology as a helpful framework to summarize existing notions of reputation. Section 4 discusses a set of simulations aimed at understanding the relative strength of different notions of reputation. The results of these simulations are shown in Section 5. A brief discussion of these results concludes this paper.

2. BACKGROUND

This section provides an overview on the study of reputation across diverse disciplines. The next section unites these studies under a common framework.

Reputation reporting systems have been implemented in e-commerce systems and have been credited with these systems' successes (Resnick, *et al.*, 2000a). Several research reports have found that seller reputation has significant influences on on-line auction prices, especially for high-valued items (Houser and Wooders, 2000; Dewan and Hsu, 2001).

The reputation system in eBay is well studied. *Reputation* in eBay is a function of the cumulative positive and non-positive ratings for a seller or buyer over several recent periods (week, month, 6-months). Resnick and Zeckhauser (2000b) have empirically analyzed this reputation system and conclude that the system does seem to encourage transactions. Houser and Wooders (2000) have used games to study auctions in eBay and describe reputation as the *propensities to default* – for a buyer, it is the probability that if the buyer wins, he will deliver the payment as promised before the close of the auction; for a seller, it is the probability that once payment is received, he will deliver the item auctioned. Their economic analysis shows that reputation has a statistically significant effect on price. Both Lucking-Reilly, *et al.* (1999) and Bajari and Hortacsu (2000) have empirically examined coin auctions in eBay. These economic studies have provided empirical confirmation of reputation effects in internet auctions.

Despite the obvious usefulness of reputation and related concepts for online trading, conceptual gaps exist in current models about them. Resnick and Zeckhauser (2000b) have pointed out the so called *Pollyanna* effect in their study of the eBay reputation reporting system. This effect refers to the disproportionately positive feedbacks from users and rare negative feedbacks. They

have also pointed out that despite the incentives to free ride (for not providing feedbacks), feedbacks by agents are provided in more than half of the transactions. This violates the rational alternative of taking advantage of the system without spending the effort to provide feedback. Moreover, these studies do not model deception and distrust. As shown by Dellarocas (2000), several easy attacks on reputation systems can be staged. These studies also do not examine issues related to the ease of changing one's pseudonym online. As Friedman and Resnick (1998) have pointed out, an easily modified pseudonym system creates the incentive to misbehave without paying reputational consequences.

Economists have extensively studied reputation in game theoretic settings. Much of the economic studies on reputation relates to repeated game. In particular, the Prisoner's Dilemma or the Chain Store stage game is often used in these studies (e.g., Andreoni and Miller, 1993; Selten, 1978). In such repeated games, reputation of players is linked to the existence of cooperative equilibria. Game theorists have postulated the existence of such equilibrium since the 1950's in the so called *Folk Theorem* (Fudenberg and Maskin, 1986). However, the first proof did not come until 1971 in the form of discounted *publicly observable* repeated game between two players (Friedman, 1971). Recent development in game theory has extended this existence result to *imperfect publicly monitored* games and to some extent *privately monitored* games (Kandori, 2002), and to games involving changing partners (Okuno-Fujiwara and Postlewaite, 1995; Kandori, 1992). Economists often interpret the sustenance of cooperation between two players as evidence of "reputation effects" (Fudenberg and Tirole, 1991).

Entry deterrence is often studied by game theorists by using notions of reputation. Kreps and Wilson (1982) borrows Harsanyi (1967)'s theory of imperfect information about players' payoffs to explain "reputation effects" for multi-stage games involving an incumbent firm versus multiple new entrants. They show that equilibria for the repeated game exist (with sufficient discounting) so that an incumbent firm has the incentive to acquire an early reputation for being "tough" in order to decrease the probability for future entries into the industry. Milgrom and Roberts (1982) report similar findings by using asymmetric information to explain the reputation phenomenon. For an incumbent firm, it is rational to seek a "predation" strategy for early entrants even if "it is costly when viewed in isolation, because it yields a reputation which deters other entrants." (*ibid.*) More recently, Tirole (1998) and Tadelis (2000a) have studied reputation at the firm level — firm reputation being a function of the reputation of the individual employees. Tadelis (2000b) has further studied reputation as a tradeable asset, such as the tradename of a firm.

Scientometrics (or bibliometrics) is the study of measuring research outputs such as journal impact factors. Reputation as used by this community usually refers to number of cross citations that a given author or journal has accumulated over a period of time (Garfield, 1955; Baumgartner, *et al.*, 2000). As pointed out by Makino, *et al.*, 1998 and others, cross citation is a reasonable but sometimes confounded measure of one's reputation.

Within **computer science**, Zacharia and Maes (1999) have suggested that reputation in an on-line community can be related to the ratings that an agent receives from others. Their Sporas and Histos systems use the notions of *global* versus *personalized* reputation. Reputation in Sporas is similar to that used in eBay or Amazon, based on average of all ratings given to an agent. Histos

retrieves reputation based on who makes a query and the local environment surrounding the inquirer.

Abdul-Rahman, *et al.*, (2000) have studied reputation as a form of social control in the context of trust propagation — reputation is used to influence agents to cooperate for fear of gaining bad reputation. Although not explicitly described, they have considered reputation as a propagated notion which is passed to other agents "by means of word-of-mouth".

Sabater, *et al.* (2001) have defined reputation as the "opinion or view of one about something" and have modeled 3 notions of reputation: individual, social, and ontological. Individual reputation refers to how a single individual's impressions are judged by others. Social reputation refers to impression about individuals based on the reputation of the social group they belong to. Ontological refers to the multifaceted nature of reputation — depending on the specific context.

Mui, *et al.*, (2001) and Yu, *et al.*, (2001) have proposed probabilistic models for reputation. The former uses Bayesian statistics while the latter uses Dempster Shafer evidence theory. Reputation for an agent is inferred in both cases based on propagated ratings from an evaluating agent's neighbors. These propagated ratings are in turn weighted by the reputation of the neighbors themselves.

In the field of **evolutionary biology**, Pollock and Dugatkin (1992) have studied reputation in the context of iterated prisoners' dilemma games (Axelrod, 1982). They have introduced a new interaction strategy (named *Observer Tit For Tat*) which determines whether to cooperate or defect based on the opponent's reputation. Reputation here is inferred from the ratio of cooperation over defection. Nowak and Sigmund (1998, 2000) use the term *image* to denote the total points gained by a player by reciprocation. The implication is that image is equal to reputation. Image score is accumulated (or decremented) in every direct interaction among agents. Following the studies by Pollock and Dugatkin (1992), Nowak and Sigmund (1998) have also studied the effects of third party observers of interactions on image scores. Observers have a positive effect on the development of cooperation by facilitating the propagation of observed behavior (image) across a population. Castelfranchi, *et al.* (1998) explicitly have reported that communication about "Cheaters"'s bad reputation in a simulated society is vital to the fitness of agents who prefer to cooperate with others.

Among **sociologists** studying social networks, reputation as a quantitative concept is often studied as a network parameter associated with a society of agents (Freeman, 1979; Krackhardt, *et al.*, 1993; Wasserman and Faust, 1994). Reputation or prestige is often measured by various centrality measures. An example is a measure proposed by Katz (1953) based on a stochastic coincidence matrix where entries record social linkages among agents. Because the matrix is stochastic, the right eigenvector associated with the eigenvalue of 1 is the stationary distribution associated with the stochastic matrix (Strang, 1988). The values in the eigenvector represent the reputation (or *prestige*) of the individuals in the society. Unfortunately, each individual is often modeled with only one score, lacking context dependence.

In her Presidential Speech to the American Political Science Society, Ostrom (1998) has argued for a holistic approach to study reputation based on how reputation, trust, and reciprocity interrelate. Based on her qualitative model, a computational model for these related concepts has been proposed by Mui, *et al.* (2002).

3. REPUTATION TYPOLOGY

3.1 Contextualization

Reputation is clearly a context-dependent quantity. For example, one's reputation as a computer scientist should have no influence on his or her reputation as cook. Formal models for context-dependent reputation have been proposed by Mui, *et al.*, (2001), Sabater, *et al.*, (2001), among others. Existing commercial reputation systems in eBay or Amazon provide only 1 reputation rating per trader or per book reviewer. Context-dependent reputation system (e.g., based on value of items) might help mitigate cybercrimes involving self-rating on small value items among a cartel of users for gaining reputation points (*c.f.*, US Dept of Justice, 2001).

3.2 Personalization

Reputation can be viewed as a *global* or *personalized* quantity. For social network researchers (Katz, 1953; Freeman, 1979; Marsden, *et al.*, 1982; Krackhardt, *et al.*, 1993), prestige or reputation is a quantity derived from the underlying social network. An agent's reputation is globally visible to all agents in a social network. In the same way, scientometricians who use citation analysis to measure journal or author impact factors (*i.e.*, reputation) also rely on the underlying network formed by the cross citations among the articles

studied (Garfield, 1955; Baumgartner, *et al.*, 2000). Many reputation systems rely on global reputation. In the case of Amazon or eBay, reputation is a function of the cumulative ratings on users by others. Global reputation is often assumed in research systems such as those in Zacharia and Maes, (1999)'s Sporas, Nowak and Sigmund (1998)'s image score without observers, Rouchier, *et al.*, (2001)'s gift exchange system, among others.

Personalized reputation has been studied by Zacharia and Maes (1999), Sabater, *et al.*, (2001), Yu, *et al.* (2001), among others. As argued by Mui, *et al.* (2002), an agent is likely to have different reputations in the eyes of others, relative to the *embedded social network*. The argument is based on sociological studies of human behavior (*c.f.*, Granovetter, 1985; Raub and Weesie, 1990; C. Castelfranchi, *et al.*, 1998). Depending on factors such as environmental uncertainties, agents' reputation in the same embedded social network often varies (Kollock, 1994).

How many notions of reputation have been studied? Based on the reviewed literature, an intuitive typology of reputation is proposed as shown in Figure 1. This typology tree is to be discussed one level at a time in the rest of this section. Each sub-section reviews reputation literatures that are relevant to that part of the tree.

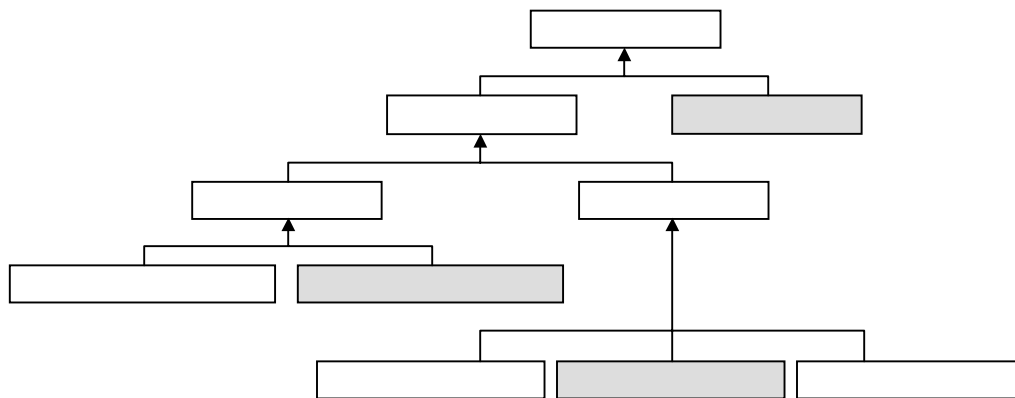
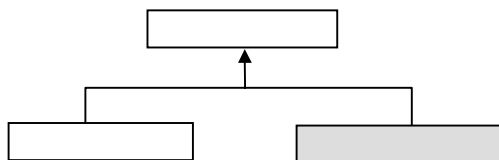


Figure 1. Reputation typology used in the paper. It is assumed that reputation is context dependent. Shaded boxes indicate notions that are likely to be modeled as social (or “global”) reputation as opposed to being personalized to the inquiring agent (see text).

3.3 Individual and Group Reputation



At the topmost level, reputation can be used to describe an individual or a group of individuals. Existing reputation systems such as those in eBay, Amazon, Free Haven, or Slashdot (*c.f.*, Resnick, *et al.* 2000b; Houser and Wooders, 2001; Dingleline, *et al.*, 2001) concentrate on reputation of the individuals.

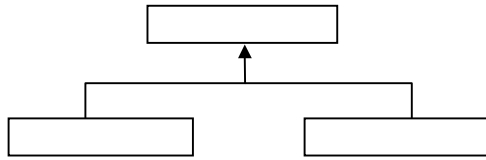
Economists have studied group reputation from the perspective of the firm (Kreps and Wilson, 1982; Tirole, 1996; Tadelis, 2000). A firm's (group) reputation can be modeled as the average of all its members' individual reputation. Among computer scientists, Sabater and Sierra (2001) have studied the *social* dimension of reputation, which is inferred from a group reputation in their model.

Halberstadt and Mui (2001) have proposed a hierarchical group model and have studied group reputation based on simulations using the hierarchical model. Their group model allows agents to belong to multiple overlapping groups and permits reputation inferences across group memberships.

Commercial groups such as Reputation.com and OpenRatings¹ are applying their proprietary secret sauces to manage buyer-supplier company relationships based on individual transactions. Inherent in these models is the distinction between individual and group reputation.

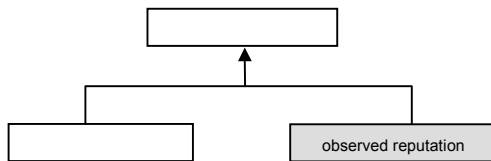
¹ *c.f.*, <http://www.reputation.com> and <http://www.openratings.com>

3.4 Direct and Indirect (individual) Reputation



One can consider individual reputation as to be derived either (1) from direct encounters or observations or (2) from inferences based on information gathered indirectly. Direct reputation refers to reputation estimates by an evaluator based on direct experiences (seen or experienced by the evaluating agent first hand). Indirect reputation refers to reputation estimates that are based on second-hand evidence (such as by word-of-mouth).

3.5 Direct Reputation



Direct experience with another agent can be further divided into (1) observations made about another agent's encounters with others, and (2) direct experience interacting with that other agent.

Observed Reputation

Reputation rating in systems such as eBay provides an example for both observed and encounter-derived reputation. These ratings are direct feedbacks from users about others whom they have interacted directly. After an encounter with a seller, a buyer can provide a rating feedback which can directly affect a seller's reputation in the system — *encounter-derived* reputation (Dewan and Hsu, 2001; Resnick and Zeckhauser, 2000b). Buyers who have not interacted with a seller need to rely on others' ratings as observations about a seller — thereby deriving *observed reputation* about the seller.

Observer based reputation plays an important role in reputation studies by evolutionary game theorists and biologists. Pollock and Dugatkin (1992) have introduced "observed tit-for-tat" (OTFT) as an evolutionarily superior strategy compared to the classic tit-for-tat strategy for the iterated Prisoner's Dilemma game. OTFT agents observe the proportion of cooperation of other agents. Based on whether a cooperation threshold is reached, an OTFT agent determines whether to cooperate or defect on an encounter with another agent. Similarly, Nowak and Sigmund (1998) use observer agents to determine agent actions in their image-score based game.

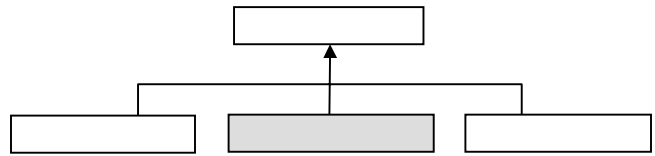
Encounter-derived Reputation

In our terminology, "observed" reputation differs from "encounter-derived" reputation in that the latter is based on actual encounters between a reputed agent and his or her evaluating agent. For example, journal impact factor as determined by citation analysis (Garfield, 1955) is an "observed" reputation based on the observed cross-citation patterns.² However, individual researchers might not agree with the impact factor based on their own readings of

² Anthropomorphically, each journal article's citation is a "rating feedback" to the cross-citation analysis observer.

individual journals.³ Each researcher revises the observed reputation based on their direct experience with each journal. Field studies by Kollock (1994) have shown that personal interactions play a more important role than indirect observations in determining whether users choose to interact with another socially.⁴

3.6 Indirect Reputations



Without direct evidence, individual reputation can be inferred based on information gathered indirectly.

Prior-derived reputation

In the simplest inference, agents bring with them prior beliefs about strangers. In human societies, each of us probably has different prior beliefs about the trustworthiness of strangers we meet. Sexual or racial discrimination might be a consequence of such prior beliefs.

For agent systems, such discriminatory priors have not yet been modeled. Mui, *et al.*, (2001)'s probabilistic model uses a uniform distribution for reputation priors. This is equivalent to an ignorance assumption about all unknown agents. Zacharia and Maes (1999)'s system give new agents the lowest possible reputation value so that there is no incentive to throw away a cyber identity when an agent's reputation falls below a starting point. Nowak and Sigmund (1998)'s agents assume neither good nor bad reputation for unknown agents.

Group-derived Reputation

Models for group can be extended to provide prior reputation estimates for agents in social groups. Tadelis (2001)'s study of the relation between firm reputation and employee reputation naturally suggests a prior estimate based on the firm that an economic agent belongs to. If the firm has good reputation, the employee can benefit with being treated as if he or she has good reputation, and vice versa. In the computer science field, both Sabater and Sierra (2001), Halberstadt and Mui (2001) have postulated different mapping between the initial individual reputation of a stranger and the group from which he or she comes from. Since the reputation of a group can be different to different agents, individual reputation derived from group reputation is necessarily personalized to the evaluating agent's perception of the group.

³ Citation analysis based impact factor has been questioned on scientific ground (Makino, *et al.*, 1998).

⁴ Our term "Encounter-derived" reputation is usually called "personalized" (Zacharia and Maes, 1999; Sabater and Sierra, 2001; Yu and Singh, 20001; Mui, *et al.*, 2001). We avoid the word "personalized" here since other notions of reputation in Figure 1 can also be described as such.

Propagated Reputation

Finally, although an agent might be a stranger to the evaluating agent, the evaluating agent can attempt to estimate the stranger's reputation based on information garnered from others in the environment. As Abdul-Rahman and Hailes (2000) have suggested, this mechanism is similar to the "word-of-mouth" propagation of information for human. Reputation information can be passed from agent to agent. Schillo, *et al.*, (2000), Mui, *et al.*, (2001) Sabater and Cieria, (2001), and Yu and Singh (2001) have all used this notion that reputation values can be transmitted from one agent to another. What differentiates these approaches is the care taken in combining the information gathered from these chains. Yu and Singh (2001) have tried to use Dempster-Shafer theory for this combination. Mui, *et al.*, (2001) have used Bayesian probability theory. The latter group has also used Chernoff Bound to propose a reliability measure for information gathered along each chain.

4. Framework for Reputation Simulations

If reputation has a utility value for the survival of an agent, we would like to design a set of experiments to test which notion of reputation provides the highest utility. We use an evolutionary version of the incomplete information game similar to that used in Kreps and Wilson (1982) and Milgrom and Roberts (1982).

4.1 Indirect Reciprocity

In the field of evolutionary game theory, several groups have applied reputation to study the "evolution of cooperation" (Pollock and Dugatkin, 1992; Nowak and Sigmund 1998). Trivers (1971) has suggested the idea of *reciprocal altruism* as an explanation for the evolution of cooperation. Altruists indirectly contribute to their fitness (for reproduction) through others who reciprocate back. Reputation can potentially help to distinguish altruists from those disguised as such, thereby preventing those in disguise from exploiting the altruists. Alexander (1987) greatly extended this idea to the notion of *indirect reciprocity*. In situations involving cooperators and defectors, *indirect reciprocity* refers to reciprocating toward cooperators indirectly through a third party. One important heuristic that has been found to pervade human societies is *reciprocity norm* for repeated interactions with the same parties (Becker, 1990; Gouldner, 1960).⁵ Therefore, a reasonable model for a human is an agent engages in reciprocal interactions.

In the following sub-section, groups of reciprocating agents are simulated against all-defecting agents. By using various notions of reputation, the reciprocating strategy can be shown to be superior from the standpoint of survivability.

4.2 Simulation Framework

For the Prisoner's Dilemma (PD) game, the action space for each agent is:

$$\text{Action} = \{ \text{cooperate}, \text{defect} \}$$

The payoff matrix for the Prisoner's Dilemma game is (where $T > R > P > S$ and $2R > T + S$. *c.f.*, Fudenberg and Tirole, 1991):

⁵ *Reciprocity norms* refer to social strategies that individuals learn which prompt them to "... react to the positive actions of others with positives responses and the negative actions of others with negative responses (Ostrom, 1998).

		agent 2	
		C	D
agent 1	C	R, R	S, T
	D	T, S	P, P

Figure 2. Payoff matrix for the prisoners' dilemma game, where C = cooperate, D = defect.

Participants in an encounter are chosen randomly from the population. After the first participant is selected, a second participant is randomly selected. At the end of a generation (where a certain number of dyadic encounters between agents have occurred), an agent begets progeny in the next generation proportional to that agent's total fitness. The total population size is fixed, so any increase in the number of one type of agent is balanced by a decrease in the numbers of other types of agents.

4.3 Simulation Parameters

For each of the simulation experiments, 50 agents with strategy TFT and 50 agents with strategy ALLD are mixed into a shared environment. A total of 30 generations are simulated per experimental run (during which no new agents are introduced into the system). The payoff values (*c.f.*, Figure 2) are: $T = 5$, $R = 3$, $P = 1$, $S = 0$.

4.4 Agent Strategies

We studied agent strategies in which the decision for an encounter with an agent is based on the last interaction with that agent. Each strategy is characterized by five probabilities for cooperation: an initial probability and four probabilities for each of the possible outcomes of the last encounter. We extended these strategies by adding a reputation threshold that determines how an agent will act. Example agent strategies for this game are:

- *Cooperate (C)*: always cooperates.
- *Defect (D)*: always defects.
- *Tit-for-tat (TFT)*: initially cooperates, and then does what the other agent did in the last round.
- *Reputation tit-for-tat (RTFT)*: initially cooperates depending on the reputation of the other agent, and then does whatever the other agent did in the last round

The reputation referred to for RTFT agents is determined using one of several reputation notions as described below. If the reputation of the target agent is less than a minimum reputation threshold, then the RTFT agent defects, otherwise it cooperates.

Strategies	I	T	R	P	S
Cooperate (C)	1	1	1	1	1
Defect (D)	0	0	0	0	0
Tit-for-tat (TFT)	1	1	1	0	0
Reputation Tit-for-tat (RTFT)	*	1	1	0	0

Figure 3. Probabilities of cooperation for different strategies. The column labeled I gives the initial probability for cooperation, while those labeled T , R , P , and S give the probabilities for cooperation given that the outcome (payoff) of the previous encounter was temptation, reward, punishment, or sucker. The initial probability for RTFT (*) depends on opponent's reputation and the reputation threshold used.

4.5 Goal of Simulation

In our simulations, we studied the conditions under which TFT agents are evolutionarily stable when they use different notions of reputation to judge agents with whom they interact. Specifically, we examined the “number of encounters per generation” (EPG) threshold for reputation-enhanced TFT (RTFT) to become the evolutionarily stable strategy (ESS, *c.f.* Maynard Smith, 1982). Reputation should aid agents more when more information about other agents’ behavior is available. When no agents have met each other before, there is no information to calculate any reputation. As more encounters per generation occur, the more chances each RTFT agent has to learn the real reputation of the opponent agents. Note that each agent does not know the strategy of the other agents. Agents can only observe the behavior of other agents. Therefore, it is not true that once an agent is observed acting *defect*, it is therefore an AllD agent.

4.6 Notions of Reputation Experimented

Encounter-derived individual reputation r_e is simulated by having each TFT agent remember encounters it has with every agent it has met before. Encounter-derived individual reputation is then the ratio of number of cooperation directly encountered over total number of encounters with a specific opponent. Such an RTFT agent defects if $r_e < r_c$ where r_c represents a critical threshold point of defection, which can be variable across agents. In our simulation, $r_c = 0.5$ for all agents.

Observed individual reputation is simulated in a similar way as encounter-derived reputation with the addition of observers. The setup mirrors observer-based image collection by Nowak and Sigmund (1998). Each agent a_i designates 10 random agents in the environment as being observed. All encounters by these 10 observed agents are recorded by a_i . The reputation of agent a_j in the eyes of a_i is r_{ij} which is the ratio of number of cooperation observed by a_i among its 10 observed agents’ encounters over the number of defection. Such an RTFT agent a_i defects an opponent a_j if $r_{ij} < r_c$ where r_c is also set at 0.5 in the actual simulations.

Group-derived reputation is simulated by grouping all agents with the same strategy into a group. The **group reputation** is calculated as the ratio of number of cooperation performed by members of a group over total number of encounters with a given agent. Reputation derived from group depends on individual experience and is therefore not the same for all agents. When an RTFT agent meets an unknown agent, it uses the group reputation as the prior estimate for this unknown’s reputation r_g . Such an RTFT agent defects an unknown opponent if $r_g < r_c$ where r_c is also set at 0.5 in the actual simulations. After the first encounter with an unknown agent, all subsequent decisions are based on encounter-derived individual reputation as discussed above.

Propagated reputation is simulated by having each RTFT agent recursively ask agents whom it has encountered before for their reputation estimate of an unknown agent. Propagation is checked by a MAX_TRAVERSAL limit. All gathered results are tallied using a Bayesian algorithm as described in Mui, *et al.* (2001) to calculate the propagated reputation r_p for an unknown opponent agent. If the calculated reputation $r_p < r_c$, the RTFT agent defects on the unknown opponent. Again, r_c is also set at 0.5 in the actual simulations. After the first encounter with an unknown agent, all subsequent decisions are based on encounter-derived individual reputation as discussed above.

Our hypothesis is that reputation should lower the threshold of EPG necessary for TFT agents to dominate over AllD. By making TFT agents use separate notions of reputation, we would like to compare how effective each reputation notion allows the TFT agent to discriminate between AllD and other TFT agents.

5. Experimental Results

Figure 4 shows the evolution of TFT population size in a simulation starting with 50 AllD and 50 TFT agents. (No additional reputation measure is used by TFT agents except the 1 slot memory for the TFT for every one of its opponents.) The legend of Figure 4 (and all other graphs in this section) indicates the number of encounters per generation (EPG). As the chance for repeated encounter is enhanced with increases in EPG, the TFT strategy dominates over AllD when EPG is greater than approximately 12000.

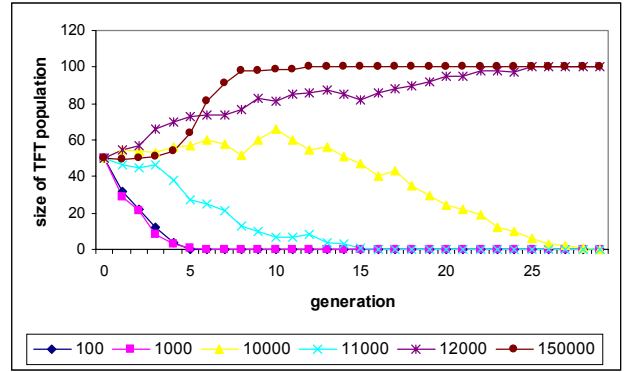


Figure 4. Base case when no reputation is used for TFT agents.

The same experiment as that shown in Figure 4 is performed for each of the 5 notions of reputation as discussed in the last section. The EPG thresholds for RTFT strategies to dominate over AllD are summarized in Figure 5.

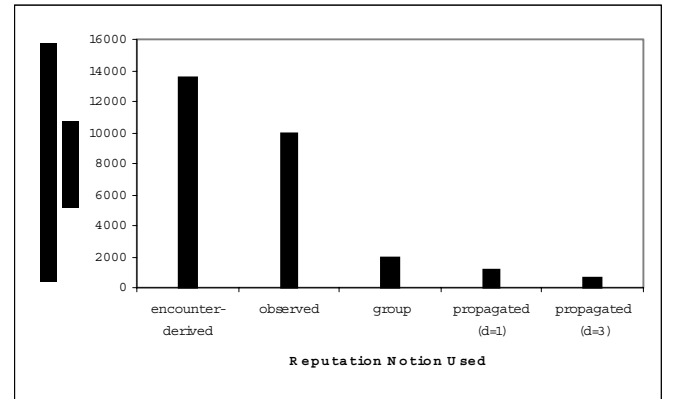


Figure 5. Threshold number of encounters per generation (EPG) for RTFT agents to become evolutionarily stable over AllD agents. The 5 notions of reputation used are shown by the horizontal axis labels.

6. Discussion and Conclusion

Based on the encounters per generation (EPG) threshold, in order for RTFT agents to dominate over ALLD agents, the following utility order is derived for the different notions of reputation in our simulations (where $a > b$ indicates a is preferred over b)

$$r_{p3} > r_{p1} > r_g > r_o > r_e$$

An initial glance at Figure 5 might be surprising until one realizes that in the iterated PD game that is simulated, the reciprocating agents use TFT strategy. Encounter-derived individual reputation does not “kick-in” to warn an RTFT agent against an ALLD agent until TFT agents have already cooperated once with an ALLD agent. Therefore, the notion of direct encounter-derived reputation is not useful for this TFT-ALLD game since repeated encounters between any two agents is not numerous. This is not to say that such notion of reputation is not useful. As mentioned earlier in the paper, several existing systems have used this notion of reputation and have derived useful results.

Based on the size of drop in the number of encounters per generation (EPG) threshold, propagated reputation seems to provide a significant utility to TFT agents against ALLD agents. Whether the order of strength among the different notions of reputation holds in other types of game can only be speculated at present. Our immediate future work is to extend the results shown in this paper to other types of games.

This paper has proposed a typology for different notions of reputation that have been studied by various researchers and implemented in real world systems. The typology serves a useful function in unifying the diverse literature on reputation. Based on this typology, this paper has studied the relative strengths of different notions of reputation in a set of evolutionary games. Whereas these notions of reputation could only be compared qualitatively before, our simulation framework has enabled us to compare them quantitatively.

ACKNOWLEDGEMENTS

We would like to thank the freedom and support that Professor Peter Szolovits has given us in his laboratory for this work. This work is partially supported by fellowship support from the NIH/NLM.

7. REFERENCES

- [1] A. Abdul-Rahman, S. Hailes (2000) “Supporting Trust in Virtual Communities,” 33rd *Hawaii International Conference on System Sciences*.
- [2] R. D. Alexander (1987) *The Biology of Moral System*, New York: Aldine de Gruyter.
- [3] J. Andreoni, J. H. Miller (1992) “Rational Cooperation in the Finitely Repeated Prisoner’s Dilemma: Experimental Evidence,” *The Economic Journal*, 103 (418), pp. 570-585.
- [4] R. Axelrod (1984) *The Evolution of Cooperation*. New York: Basic Books.
- [5] H. Baumgartner, R. Pieters (2000) “The Influence of Marketing Journals: a Citation Analysis of the Discipline and its Sub-Areas,” *Center for Economic Research Paper No. 2000-123*. <http://citeseer.nj.nec.com/baumgartner00influence.html>
- [6] L. C. Becker, (1990) *Reciprocity*. Chicago: University of Chicago Press.
- [7] P. Bajari, A. Hortacsu (1999) “Winner’s Curse, Reserve Prices and Endogenous entry: Empirical Insights from eBay Auctions,” *Stanford Institute for Economic Policy Research Policy paper No. 99-23*.
- [8] R. Boyd, P. Richerson (1988) “The Evolution of Reciprocity in Sizeable Groups,” *Journal of Theoretical Biology*, 132, pp. 337-356.
- [9] C. Castelfranchi, R. Conte, M. Paolucci (1998) “Normative Reputation and the Costs of Compliance,” *J. Artificial Societies and Social Simulations*, 1(3).
- [10] C. Dellarocas (2000) “Immunizing Online Reputation Reporting Systems Against Unfair Ratings and Discriminatory Behavior,” *Proc. 2nd ACM Conference on Electronic Commerce*.
- [11] S. Dewan, V. Hsu (2001) “Trust in Electronic Markets: Price Discovery in Generalist Versus Specialty Online Auctions,” <http://databases.si.umich.edu/reputations/bib/papers/Dewan&Hsu.doc>.
- [12] R. Dingleline, M. J. Freedman, D. Molnar (2001) “Free Haven,” *Peer-to-Peer: Harnessing the Power of Disruptive Technologies*, O’Reilly.
- [13] L. C. Freeman (1979) “Centrality in Social Networks: I. Conceptual Clarification,” *Social Networks*, 1, pp. 215-239.
- [14] J. Friedman (1971) “A Non-cooperative Equilibrium for Supergames,” *Review of Economic Studies*, 38, pp. 1-12.
- [15] D. Fudenberg, E. Maskin (1986) “The Folk Theorem in Repeated Games with Discounting and Incomplete Information,” *Econometrica*, 54, pp. 533-554.
- [16] D. Fudenberg, J. Tirole (1991) *Game Theory*, Cambridge, Massachusetts: MIT Press.
- [17] E. Garfield (1955) “Citation Indexes for Science,” *Science*, 122, pp. 108-111.
- [18] A. W. Gouldner (1960) “The Norm of Reciprocity: A Preliminary Statement,” *American Sociological Review*, 25, pp. 161-78.
- [19] M. Granovetter (1985) “Economic Action and Social Structure: the Problem of Embeddedness,” *American Journal of Sociology*, 91, pp. 481-510.
- [20] A. Halberstadt, L. Mui (2001) “Group and Reputation Modeling in Multi-Agent Systems,” *Proc. Goddard/JPL Workshop on Radical Agents Concepts*, NASA Goddard Space Flight Center.
- [21] J. Harsanyi (1967) “Games with Incomplete Information Played by Bayesian Players,” *Management Review*, 14, pp. 159-182, 320-334, 486-502.
- [22] D. E. Houser and J. Wooders (2001) “Reputation in Internet Auctions: Theory and Evidence from eBay,” working paper: http://w3.arizona.edu/~econ/working_papers/Internet_Auctions.pdf.
- [23] M. Kandori (1992) “Social Norms and Community Enforcement,” *The Review of Economic Studies*, 59 (1), pp. 63-80.

- [24] M. Kandori (2002) "Introduction to Repeated Games with Private Monitoring," *Journal of Economic Theory*, 102, pp. 1-15.
- [25] L. Katz (1953) "A New Status Index Derived from Sociometric Analysis," *Psychometrika*, 18, pp. 39-43.
- [26] P. Kollock (1994) "The Emergence of Exchange Structures: An Experimental Study of Uncertainty, Commitment, and Trust," *American Journal of Sociology*, 100(2), pp. 313-345.
- [27] D. Krackhardt, M. Lundberg, L. O'Rourke (1993) "KrackPlot: A Picture's Worth a Thousand Words," *Connections*, 16, pp. 37-47.
- [28] D. M. Kreps, R. Wilson (1982) "Reputation and Imperfect Information," *Journal of Economic Theory*, 27, pp. 253-279.
- [29] D. Lucking-Reiley, D. Bryan, N. Prasa, D. Reeves (1999) "Pennies from eBay: The Determinants of Price in Online Auctions," <http://eller.arizona.edu/~reiley/papers/PenniesFromEBay.pdf>
- [30] J. Makino, Y. Fujigaki, and Y. Imai (1997) "Productivity of Research Groups – Relation between Citation Analysis and Reputation within Research Community," *Japan Journal of Science, Technology and Society*, 7, pp. 85-100.
- [31] P. V. Marsden, N. Lin (eds.) *Social Structure and Network Analysis*, Newbury Park, CA: Sage.
- [32] J. Maynard Smith (1982) *Evolution and the Theory of Games*, Cambridge: Cambridge University Press.
- [33] P. R. Milgrom, J. Roberts (1982) "Predation, Reputation and Entry Deterrence," *Journal of Economic Theory*, 27, pp. 280-312.
- [34] L. Mui, M. Mohtashemi, C. Ang, P. Szolovits, A. Halberstadt, A. (2001) "Ratings in Distributed Systems: A Bayesian Approach," *11th Workshop on Information Technologies and Systems (WITS)*, New Orleans.
- [35] L. Mui, M. Mohtashemi, A. Halberstadt (2002) "A Computational Model for Trust and Reputation," *35th Hawaii International Conference on System Sciences*.
- [36] M. A. Nowak, and K. Sigmund (1998) "Evolution of Indirect Reciprocity by Image Scoring," *Nature*, 393, pp. 573-577.
- [37] M. A. Nowak, and K. Sigmund (2000) "Cooperation versus Competition," *Financial Analyst Journal*, July/August, pp. 13-22.
- [38] M. Okuno-Fujiwara, A. Postlewaite (1995) "Social Norms and Random Matching Games," *Games and Economic Behavior*, 9, pp. 79-109.
- [39] E. Ostrom (1998) "A Behavioral Approach to the Rational-Choice Theory of Collective Action," *American Political Science Review*, 92(1), pp. 1-22.
- [40] G. B. Pollock, L. A. Dugatkin (1992) "Reciprocity and the Evolution of Reputation," *Journal of Theoretical Biology*, 159, pp. 25-37.
- [41] W. Raub, J. Weesie (1990) "Reputation and Efficiency in Social Interactions: An Example of Network Effects," *American Journal of Sociology*, 96(3), PP. 626-654.
- [42] P. Resnick, K. Kuwabara, R. Zeckhauser, E. Friedman (2000a) "Reputation Systems," *Communications of the ACM*, 43(12), pp. 45-48.
- [43] P. Resnick, R. Zeckhauser (2000b) "Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System," *NBER Workshop on Empirical Studies of Electronic Commerce Paper*.
- [44] J. Rouchier, M. O'Connor, F. Bousquet (2001) "The Creation of a Reputation in an Artificial Society Organized by a Gift System," *Journal of Artificial Societies and Social Simulations*, 4(2).
- [45] J. Sabater, C. Sierra (2001) "REGRET: A reputation Model for Gregarious Societies," *4th Workshop on Deception, Fraud and Trust in Agent Societies*.
- [46] M. Schillo, P. Funk, M. Rovatsos (2000) "Using Trust for Detecting Deceitful Agents in Artificial Societies," *Applied Artificial Intelligence, Special Issue on Trust, Deception and Fraud in Agent Societies*.
- [47] R. Selten (1978) "The Chain Store Paradox," *Theory and Decision*, 9, pp. 127-159.
- [48] S. Tadelis (1999) "What's in a Name? Reputation as a Tradeable Asset," *American Economic Review*, 89(3), pp. 548-563.
- [49] S. Tadelis (2000) "Firm Reputation with Hidden Information," *Stanford Economics Working Paper*.
- [50] J. Tirole (1996) "A Theory of Collective Reputation (with Applications to the Persistence of Corruption and to Firm Quality)," *The Review of Economic Studies*, 63(1), pp. 1-22.
- [51] R. L. Trivers, (1971) "The Evolution of Reciprocal Altruism," *Quarterly Review of Biology*, 46, pp. 35-57.
- [52] US Department of Justice (2001) *Press Release*. <http://www.usdoj.gov/criminal/cybercrime/ebayplea.htm>
- [53] S. Wasserman, K. Faust (1994) *Social Network Analysis: Methods and Applications*. Cambridge University Press.
- [54] B. Yu, M. P. Singh (2001) "Towards a Probabilistic Model of Distributed Reputation Management," *4th Workshop on Deception, Fraud and Trust in Agent Societies*, Montreal, Canada.
- [55] G. Zacharia, P. Maes (1999) "Collaborative Reputation Mechanisms in Electronic Marketplaces," *Proc. 32nd Hawaii International Conf on System Sciences*.
- [56] P. R. Zimmerman (1995) *The Official PGP User's Guide*, Cambridge, Massachusetts: MIT Press.