¿Por Favor? Favor Reciprocation When Agents Have Private Discounting

Christopher J. Hazard
North Carolina State University
Raleigh, NC 27695
cjhazard@ncsu.edu

Abstract

When agents can significantly increase other agents' utility at a moderate cost, the socially optimal outcome is for all agents to repeatedly provide favors to each other whenever they can. However, when agents cannot support or enforce a market system, this forms a situation similar to the repeated prisoner's dilemma because each agent can unilaterally improve its own utility by refusing to help others. We present an adaptive tit-for-tat strategy that provides a mutually beneficial equilibrium in the general cases when agents may have differing private discount factors and when favor costs and benefits are stochastic and asymmetric. This strategy allows agents to treat previously unencountered agents with caution, communicate about the trustworthiness of other agents, and evaluate past communication for deception. We discuss the details of the strategy, analytic and simulation results, and the impact of various parameterizations. We analyze one form of communication in detail and find that it causes agents to be more protective of utility.

Introduction

People and organizations routinely perform favors in a variety of settings built on norms, empathy, and trust. However, a self-interested agent acting on behalf of a person, organization, or itself, only has whatever intrinsic empathy and trust towards others with which it was designed. In many situations, the socially optimal outcome is achieved via agents exchanging resources and services.

While market-based resource allocation is often an effective tool for social optimization (Wellman 1996; Golle, Leyton-Brown, & Mironov 2001), market transactions may not be effective with self-interested agents without common currency, sufficient liquidity, means transferring resources, or effective methods to enforce a fiat currency. In such cases, agents can reciprocally perform favors instead of using markets to improve social welfare, albeit with generally less efficient outcomes. However, agents that can trust one another to reciprocate favors to form a gift economy will have a better ability to smooth out inefficient allocations over time.

Many environments lend themselves to such favor-based interactions. One widely studied example is peer-to-peer file sharing as a decentralized means of distributing data and software (Kamvar, Schlosser, & Garcia-Molina 2003; Banerjee *et al.* 2005). Complex tasks in multi-robot systems often require coordination to increase utility (Gerkey & Matarić 2002). Favor-based mechanisms are particularly useful in situations where robots are self-interested but do not have resources to enforce trade. In business-to-business settings, agents involved in procurement can decide whether to put forth extra effort to deliver goods or services exceeding the contract when the customer is in need to foster the relationship for reciprocatively beneficial behavior. Personal assistant agents may engage in similar interactions, such as transferring reservations that offer more utility to a recipient who has a history of positive reciprocations. Gift economies may also be used to augment market-based transactions to circumvent market friction, such as a burdensome taxation system, or to increase the risk in switching to another provider with an unknown reputation.

While the question of trust between agents has received considerable attention (Jøsang, Ismail, & Boyd 2007), we approach trust from a new perspective. In this paper, our main contributions are that we view trust as a matter of patience, and model trust from a strategic perspective. If an agent is patient, it will uphold its end of bargains and do favors for other agents because it knows the favor will be repaid many times over in the future if the other agents are also patient. To model an agent's patience, we use discount factors. Discount factors are a widely used method of reducing the perceived value of resources and events in the future to less than they are worth in the present. For example, an agent might prefer receiving $100 today to receiving $300 in 6 months. This mechanism accounts for the opportunity cost of not utilizing resources until the later point in time. Discounting can also be used to model uncertainty and reliability when making decisions.

Although discount factors are widely used in economics, finance, behavioral sciences, game theory, and artificial intelligence, each agent or firm is typically assumed to have a publicly known discount factor. Assuming that discount factors are public is reasonable for certain areas such as finance, but even professional economists' opinions of an appropriate discount factor can have a wide distribution (Weitzman 2001). Discount factors are influenced by the preferences of the person, agent, or organization. Though discount factors may be used strategically by opponents, many mod-

els require publicly known discount factors. Our model is unique in that it allows agents to maintain private discount factors and measure others' discount factors. We employ the commonly used exponential discounting net present value (NPV), although other decreasing discount models may be substituted. Our model also exhibits individual rationality (expected positive utility by participating in the game) and in most cases attracts agents to approximate incentive compatibility (optimal strategy is to perform honestly).

In this paper, we build a simple but applicable favor-based interaction model in which agents attempt to maximize their own utility based on their discount factor and what they expect to gain in the future. In our model, agents interact via a stochastic process, and can only choose whether to reduce their own utility in order to increase that of another. We begin by presenting related work, followed by our stochastic interaction model, and then present our adaptive reputation method, which we then use to construct our primary result: a reciprocity general strategy that works in stochastic environments. We investigate how agent communication impacts the model and evaluate the strategies via simulation. We find analytically and experimentally that agents with discount factors that allow them to retaliate the most effectively in a tit-for-tat style equilibrium achieve the highest utility. Finally, we draw some conclusions from our analysis and simulations about how communication affects agents' behavior in our model.

Related Work

Our work is similar to Sen's work (Saha, Sen, & Dutta 2003; Sen 2002) in that we build a reciprocity model on future expectations, but we allow for the discovery of private discount factors, observe possible ranges of responses rather than point values, and do not use randomization to communicate signals. Our model also resembles that of Buragohain et al. (2003) in the way we are using incentives to build trust in an environment with favors, but the primary differences are that their model has continuous interaction and does not deal with discount factors.

Many others have proposed various methods of quantifying reputation and trust. Sierra and Debenham (2005) describe an information theoretic model of trust where more information yields less uncertainty in decision making. Because they use ordinal preferences instead of utility, their model works well only if all items being negotiated about are of sufficient and comparable worth. Teacy et al. (2006) use a beta distributions to model positive and negative interactions, but do not take into account the magnitude of the interaction nor strategic behavior from agents. Along similar lines of trust measurement, Yu and Singh (2002) use Dempster-Shafer theory, which represents belief and plausibility in probabilistic terms, to model trust and reputation. However, like the work of Teacy et al. and Sierra and Debenham, this model does not deal with strategic behavior nor interactions of different ranges of utility, potentially leaving agents' utility unprotected against a strategic agent.

Ramchurn et al. (2006) develop a finite-horizon negotiation mechanism based on repeated games. Because the agents are negotiating about the near future, their differing discount factors are implicitly accounted for in the negotiations, resembling the effects of our model. However, their model does not deal with agents that break promises, and thus needs exogenous enforcement.

Azoulay-Schwartz and Kraus (2004) present a favor reciprocity model of information exchange and use a punishing trigger strategy with forgivement. While their method of interaction and mechanisms resemble ours, they assume that discount factors are public and their punishment mechanism does not account for the effect on the opposing agent in relation to the cost of the punishment.

While explicitly dealing with the desiderata of incentive compatibility and individual rationality are generally considered important in game theory and auction literature, dealing with strategic behavior is more rare in the trust and reputation literature. Jurca and Faltings (2007) develop an exchange model where the client can sanction the provider if refund not given for a bad interaction. Their model achieves similar goals to ours, except that their model is built on a more complex refund-based interaction rather than simple reciprocity, and their model assumes discount factors are publicly known. While our model does not exhibit perfect incentive compatibility, we are able to leverage approximate incentive compatibility in equilibrium as an attractor to gain accuracy in modeling private discount factors.

Favor model

In our favor model, each agent $a \in A$ encounters other agents in pair-wise interactions with two roles: offering a favor and asking a favor. Each round, agents are paired with other agent in both roles. The probability that agent a_1 encounters agent a_2 in the round as offer and ask roles respectively is $r_{a_1 \to a_2}$. Similarly, the probability that a_1 encounters a_2 as ask and offer roles respectively is $r_{a_2 \to a_1}$.

When agents encounter one another, they play the game, Γ, selected from a set of possible games, $\mathfrak{G}$, as follows. The agent in the ask role knows its willingness-to-pay for a particular favor, w, and asks the other agent for the favor. While the asking agent could choose not to ask for the favor, the strategy of asking a favor always dominates not asking the favor because asking incurs no cost, reveals nothing to the opponent, and the asker will either receive a favor or gain information about the other agent. When the asking agent asks for the favor, the cost of the offering agent to perform the favor, c, is revealed to both agents. The agent in the offering role then decides whether to provide the favor, P, or to reject the request, R.

This repeated game has the obvious Nash equilibria of offering agents always playing R. However, more equilibria and interesting behaviors emerge when reputation is taken into account. An agent has no direct control over gains of its own utility, and is thus subject to the actions of other agents.

The values of c and w are drawn from the non-negative distributions of the random variables C and W respectively. While the distributions are public information, our model may be extended to have agents learn the distributions. Each agent may have a unique distribution, but we assume they all share the same distribution for clarity.

Each agent's type is comprised of its discount factor, previous observations of interactions with other agents, and information acquired from communication with other agents. As these attributes of an agent's type are private, agents must analyze other agent's actions and strategize about information revelation in order to maximize utility.

In order to deal with agents' changing preferences over time, we discount agents' histories by using a replacement process for the agents. When a replacement occurs, an agent is effectively removed and replaced with a new agent; its discount factor is redrawn from the distribution of discount factors. The agent's observations and information of other agents may be cleared when it is replaced. Agent replacements have been shown to be an effective tool for modeling how agents change over time (Mailath & Samuelson 2006). This replacement process forces agents to focus on recent observations more than old observations, and allows agents to change over time. We set agent a's replacement rate at λ_a, and use a Bernoulli process to decide when the agent is replaced. The replacements can be justified by a change in the market, agent's ownership, information, or other factors in a dynamic environment. For the model to be meaningful to real-world scenarios, the replacement rate should be sufficiently low for reputation to be significant.

Strategies With Known Discount Factors

As an agent's discount factor increases, its willingness to give favors to other agents in return for greater reward later increases. An agent with a high discount factor therefore would desire a mechanism that rewards its patience and prevents other agents with lower discount factors from taking advantage of it. Under such a mechanism, agents will reciprocate favors according to discount factors of their own and of the opposite player.

Suppose agents a_1 and a_2 are exchanging favors, where a_1 is offering all favors to a_2 that cost a_1 less than a_1's current maximal favor offering, $\overline{c}_{a_1 \to a_2}$. We can think of the value $\overline{c}_{a_1 \to a_2}$ as the minimum amount that a_1 trusts a_2 to repay back in favors. Similarly, a_2 is offering a_1 all favors that cost a_2 less than $\overline{c}_{a_2 \to a_1}$. As the only positive payoff to a_1 is controlled by a_2, and a_1 incurs cost for providing favors to a_2, a_1 has a direct incentive to reduce its costs by refusing to provide favors. Given that the cost of providing favors is known and comparable, a_1 would choose which favors to perform and control its cost by adjusting $\overline{c}_{a_1 \to a_2}$. When a_1 is playing in an offer role in game Γ, we can thus write a_1's expected total future utility of interacting with a_2 discounted by γ_{a_1} for each time step, t, given the cost of the current favor c_Γ as

$$U_{a_1} = \sum_{t=1}^{\infty} \gamma_{a_1}^t r_{a_2 \to a_1} PE(W | C < \overline{c}_{a_2 \to a_1}) - c_\Gamma - \sum_{t=1}^{\infty} \gamma_{a_1}^t r_{a_1 \to a_2} PE(C | C < \overline{c}_{a_1 \to a_2}). \quad (1)$$

We define the shorthand notation $PE(Y|X) \equiv P(X) \cdot E(Y|X)$ with $P(X)$ being the probability of event X occurring and $E(Y|X)$ is the expected value of Y given that X occurred. Equation 1 may be easily extended to a total utility by a summation over all agents.

One primary criteria of an effective economic mechanism is individual rationality, that is, an agent will expect to gain utility by participating. By applying this principle to a given pair of agents, we can find the maximum $\overline{c}$ for which each agent would be willing to provide a favor to the other while keeping the utility non-negative. We can find these values by simply setting $U_{a_1} = U_{a_2} = 0$, setting c_Γ to the corresponding $\overline{c}$ that the agent pays out in each equation, and solving to find $\overline{c}_{a_1 \to a_2}$ and $\overline{c}_{a_2 \to a_1}$.[1]

Using the method we described above to find the $\overline{c}$ values for agents does have the problem that each agent can directly increase its utility by reducing its corresponding $\overline{c}$. Our favor model is closely related to the repeated prisoner's dilemma; the Pareto efficient outcome, $\forall a, a' \in A : \overline{c}_{a \to a'} = \infty$, is not a Nash equilibrium because agents have incentive to defect from this strategy. One key feature of the repeated prisoner's dilemma is that the outcome can be Pareto efficient if agents can credibly punish others for defecting. The grim trigger strategy may easily be achieved by permanently setting $\overline{c}_{a \to a'} = 0$ when agent a' does not offer a favor. If significantly many agents use grim trigger, then agents are reluctant to ever not offer a favor, bringing about the Pareto efficient outcome. However, the grim trigger strategy is ineffective unless sufficiently many agents take on this strategy, and can be extremely pessimistic when agents and circumstances may change.

Tit-for-tat strategies are similar to grim trigger, except that the punishment is not as long lived and agents can eventually forgive others. Generally, tit-for-tat entails one agent punishing a defecting agent in a manner or magnitude similar to that of which the defector deprived the first agent of utility. If a_2 brings a_1's utility down, a_1 would like to bring a_2's utility down the same amount. Agent a_2 can decrease its costs by reducing $\overline{c}_{a_2 \to a_1}$. The rate of change of a_2's utility, U_{a_2}, with respect to a_2's rate of change of $\overline{c}_{a_2 \to a_1}$ can be written as the partial derivative $\frac{\partial U_{a_2}}{\partial \overline{c}_{a_2 \to a_1}}$. Because an agent's utility increases when its costs are reduced, this partial derivative is always negative. By changing $\overline{c}_{a_1 \to a_2}$, a_1 affects a_2's utility by the rate of $\frac{\partial U_{a_2}}{\partial \overline{c}_{a_1 \to a_2}}$.

In steady-state, agents will maintain constant values of $\overline{c}$. When an agent makes a small change to $\overline{c}$, its opponent may be able to effectively retaliate the same amount, but only if its discount factor is appropriate. We can express the equilibrium where agents' retaliations to small changes in $\overline{c}$ are equal as

$$\frac{\partial U_{a_2}}{\partial \overline{c}_{a_1 \to a_2}} = -\frac{\partial U_{a_2}}{\partial \overline{c}_{a_2 \to a_1}}. \quad (2)$$

The right hand side of the equation is negative because decreasing one's own $\overline{c}$ increases one's own utility. Further, there is usually a pair of discount factors and $\overline{c}$ values that satisfy this equality, except for extremely different values of the rates of encounter, r, or the maximum favor willing to

[1] An agent a could also have some threshold, $k_a > 0$, which it must receive in expected utility gain in order for it to participate, in which case $U_a = k_a$ would be solved instead.

be offered, $\bar{c}$. Theorem 1 in the Appendix shows that there is always a discount factor that satisfies Equation 2.

When Equation 2 holds, the two agents can equally retaliate and so neither has an incentive to deviate from its current value of $\bar{c}$. Because this, we expect the most effective cooperation would occur with agents that have the most appropriate discount factor for the given parameterization, that is, discount factors that satisfy Equation 2 using the most probable values of $\bar{c}$ because they have the most leverage to affect their opponents. We will revisit this notion when discussing the simulation results.

Modeling Reputation

We denote other agents' reputations from the vantage of agent a_1 as a set including a_1's direct observations combined with the information communicated to a_1 by other agents as $\mathcal{I}_{a_1}$. An observation, $i \in \mathcal{I}_{a_1}$, consists of the tuple $(o_i, o_i', t_i, \gamma_i^*)$, where o_i is the agent that made the observation, o_i' is the agent the observation is made about, t_i is the time of the observation, and γ_i^* is the observation range. We define an observation range as a tuple of the upper and lower bound of the discount factor, γ_i, given an observation of the action the agent took in a given game. We assume that agents' observations are accurate.

New observations can increase the precision and accuracy of a reputation estimate or alternatively invalidate previous observations if they are conflicting. Conflicting observations typically occur because an agent has undergone replacement, but may also occur due to changes in agents' information. Given a set of observations, $\mathcal{I}$, and a new observation, i', we define the function $\mathcal{X}$, which returns the set of all observations in $\mathcal{I}$ which conflict with i', as

$$\mathcal{X}(\mathcal{I}, i') = \{i \in \mathcal{I} : o_i = o_{i'}, o_i' = o_{i'}', \gamma_i^* \cap \gamma_{i'}^* = \emptyset\}. \quad (3)$$

When agent a_1 makes a new direct observation of a_2, i', we denote the resulting relevant history of observations as $\mathcal{I}_{a_1} \oplus i'$. We define this operation of accommodating a new observation as

$$\mathcal{I}_{a_1} \oplus i' = \begin{cases} \mathcal{I}_{a_1} \cup \{i'\} & \text{iff } \mathcal{X}(\mathcal{I}_{a_1}, i') = \emptyset, \\ \{i \in \mathcal{I}_{a_1} : t_{i'} \geq \max_{j \in \mathcal{X}(\mathcal{I}_{a_1}, i')} t_j\}. \end{cases} \quad (4)$$

If agents' strategies are consistent and prevent conflicting observations, we can denote the expected number of direct observations of a_2 between replacements by a_1 as $E(|\{i \in \mathcal{I}_{a_1} : o_i = a_2\}|)$. This value is maximized just prior to replacement and 0 after the replacement. By using the equality $\sum_{t=0}^{\infty}(1-\lambda_{a_2})^t = \frac{1}{\lambda_{a_2}}$, the expression becomes

$$E(|\{i \in \mathcal{I}_{a_1} : o_i = a_2\}|) = \frac{r_{a_2 \to a_1}}{2\lambda_{a_2}}. \quad (5)$$

As the replacement rate drops to 0 and the observation history length becomes infinite, the only conflicting observations that would occur would be due to agent strategies such as misinformation or collusion. All else being equal, lower replacement rates would not decrease the expected number of *relevant observations*, that is, observations that occur at or after the last conflicting observation. Given an arbitrary additional observation, i', some replacement rate x, and a small change in replacement rate, ϵ, this can be expressed as

$$E(|\{i \in \mathcal{I}_{a_1} : o_i = a_2\} \oplus i'| \quad |\lambda_{a_2} = x) \geq E(|\{i \in \mathcal{I}_{a_1} : o_i = a_2\} \oplus i'| \quad |\lambda_{a_2} = x + \epsilon). \quad (6)$$

With a longer relevant observation history, a conflicting observation probabilistically makes more of the observation history irrelevant. If agents bias reputation toward a poor reputation when few or no relevant observations are available, then this mechanism translates into more difficulty gaining reputation than losing it. In such a system, agents would have increased value for maintaining a positive reputation because of this bias.

Agents' Utilities Under Incomplete Information

To denote the results from simultaneously solving the aforementioned constraints of individual rationality for the maximum allowable $\bar{c}$, $U_{a_1} = U_{a_2} = 0$, we introduce the function $\gamma^*_{\text{offer}} : \mathfrak{G} \times \{P, R\} \to ([0,1) \times [0,1))$. This function returns an observation of the discount factor for the offering agent based on the action performed by the offering agent and the parameters on the game, $\Gamma \in \mathfrak{G}$, where $\mathfrak{G}$ is the set of all possible games that the agent could play. Each game consists of the favor cost c_Γ and the agents involved. The range returned for γ_{offer} will be one of $[\gamma, 1)$ or $[0, \gamma]$, depending on whether the offering agent played P or R respectively. We will refer to this range as γ_i^* for the outcome of game Γ_i. Further, we introduce the function $c^* : [0,1) \times [0,1) \to \mathbb{R}$, which takes in discount factors (or estimations thereof) of two agents, γ_{a_1} and γ_{a_2}, and returns the maximum value of a favor the first agent would offer to a_2, $\bar{c}_{a_1 \to a_2}$, for the parameters given. We use the shorthand notation $c^*_{a_1 \to a_2}$ to indicate $c^*(E(\gamma|\{i \in \mathcal{I} : o_i' = a_1, o_i = a_2\}, E(\gamma|\{i \in \mathcal{I} : o_i' = a_2, o_i = a_1\}))$.

Both functions γ^*_{offer} and c^* are used as inputs to Equation 1 for the corresponding unknown values. A system of equations of total utility may be used to include all agents with a separate equation for each agent within the connected components of the communication graph. These connected components include the agents of interest, all other agents that those agents trust sufficiently to accept communicated observations, all agents trusted by those agents to accept an observation, etc., so all agents may be included. These equations are evaluated from an individual agent's perspective, so the values are accurate only to the accuracy of the agent's information. While these systems of equations can be difficult or impossible to solve in closed form, numerical methods such as multivariate secant or bisection are effective.

An agent's utility in the current game for a given action is the sum of the value obtained by the action plus the expected future value, V, of all future games given the agent's reputation after having performed the action. In order to find this future valuation of an agent's reputation, an agent a_1 must evaluate other agents' discount factors from the observations, $\mathcal{I}_{a_1}$, it has made. When a_1 is playing in the offer role, given that a_2 has chosen to ask the favor by playing P, a_1's expected discounted utility can be expressed relative to its discount factor for the current game, Γ, as

$$U_{a_1}(s) = V_{a_1}(\mathcal{I}_{a_1} \oplus (a_1, a_2, t, \gamma^*_{\text{offer}}(\Gamma, s))) - \delta_{s,P} \cdot c_\Gamma, \quad (7)$$

where the result of function V yields the future value of a_1's reputation given that a_2 will observe a_1 playing $s \in \{P, R\}$ in a game with the value of c_Γ. If a_1 plays P, then its utility will be reduced by c_Γ.[2]

Each observation i loses potency as elapsed time increases since the observation was made, with loss rate based on the replacement rate of the agent observed. The uniform distribution satisfies the principal of maximum entropy given the maximum and minimum value of an observation. We can find the value of the probability density function (PDF) of an agent's discount factor γ, given a single observation i and current time T, discounted by the replacement rate as

$$f_i(T,\gamma) = \begin{cases} \frac{1-(1-\lambda^{T-t_i})(1-(\sup\gamma_i^* - \inf\gamma_i^*))}{\sup\gamma_i^* - \inf\gamma_i^*} & \text{if } \gamma \in \gamma_i^*, \\ 1-\lambda^{T-t_i} & \text{if } \gamma \notin \gamma_i^*. \end{cases} \tag{8}$$

We can then use Bayesian inference to combine the PDFs of the relevant observations to find what a given agent will expect another agent to believe of its discount factor, $E(\gamma|T,\mathcal{I})$, as

$$E(\gamma|T,\mathcal{I}) = \int_0^1 x \frac{\prod_{i\in\mathcal{I}} f_i(T,x)}{\int_0^1 \prod_{i\in\mathcal{I}} f_i(T,y)dy} dx. \tag{9}$$

To find the total future utility for a given reputation, an agent needs to determine its expected gain from encounters with every agent. By taking the expected discount factor via Equation 9 for each situation, combining relevant observations, finding the corresponding maximum favor value c^*, and using the results in the manner of Equation 1, we find

$$V_{a_1}(\mathcal{I}) = \frac{\gamma_{a_1}}{1-\gamma_{a_1}} \sum_{a\in A} \big(r_{a\to a_1} PE(W|C < c^*_{a\to a_1}) - r_{a_1\to a} PE(C|C < c^*_{a_1\to a}) \big). \tag{10}$$

In addition to using observation and communication to evaluate others' discount factors, agents can also strategize how to influence others' perception of their own discount factor. Agents would prefer to have other agents overestimate their own discount factor, as then an agent would be able to take advantage of other agents that are willing to give up short-term gains for larger long-term gains. Similarly, agents do not want other agents to underestimate their own discount factor, because then they will be missing gains for which they would be willing to reciprocate favors.

Despite the incentive to convince others of an artificially high reputation, our model is approximately incentive compatible in the steady state because the cost to convince another agent of a better reputation is more than the expected future gain. Incentive compatibility is important because without it, agents cannot accurately deduce discount factors from other agents that strategically give larger favors than they should. While our model does not ensure incentive compatibility, it generally ensures that the region of incentive compatible state space is an attractor. When agents are not in a region of the state space that is incentive compatible, agents are incentivized to correct others' beliefs.

[2]The Kronecker delta, $\delta_{i,j}$, yields 1 if i = j, 0 otherwise.

The three exceptions where agents are not approximately incentive compatible are as follows. First, while agents prefer opponents to overestimate their discount factors, spending utility to inflate reputation costs more than an agent will receive from the inflated reputation. However, if an agent already has a stronger reputation than its discount factor, then the agent is incentivized to use the reputation and play R for games below $\overline{c}$ while obtaining favors from the other agent. Second, if an agent's reputation is much lower than its actual discount factor, then the agent may not always offer small favors for small values of $\overline{c}$ because they may not sufficiently increase reputation to be worthwhile. While not incentive compatible, these two cases correct other agents' beliefs.

The third exception to incentive compatibility is if agents' discount factors are high and the value received from a favor is disproportionately large relative to the cost of offering a favor. In this case, an agent may know that its opponent's high discount factor will prevent a decrease in reputation from dramatically decreasing utility because its opponent's expectations of the future far outweighing the utility lost. In these cases, agents with the highest discount factors may not necessarily end up with the highest utility.

Communicating Reputation Information

When communicating information about other agents' reputations, agents have a variety of ways to divulge information. An agent could send another agent its entire list of observations for a given agent, or alternatively just its estimate of the other agent's discount factor. While supplying more detailed information can be more helpful to the recipient, more degrees of freedom of this information make it more complex for the recipient to evaluate whether the information is accurate and truthful, particularly if other agents are colluding. Further, because agents can gain utility when others to overestimate their discount factors, a self-interested agent may be reluctant to divulge extra information that might reduce another agent's belief of its discount factor.

Our communication model offers similar effects to that of the model proposed by Procaccia et al. (2007), although their method uses randomization to communicate reputation instead of ranges of discount factors. Agents maintain observations and communications which are used in aggregation to evaluate each other and give future recommendations.

We focus on simple yet plausible forms of communication of the following forms. Suppose agent a_1 asks a_3 a question about the trustworthiness of a_2 to reciprocate favors in the future. Agent a_3 can answer yes, no, or refuse to answer. Similarly, a_1 can choose to use or ignore a_3's advice and may solicit advice from other agents.

Agent a_1 could ask of a_3 whether a_3 itself would provide a favor to a_2 for the current game, and whether a_3 would recommend that a_1 provide a favor to a_2. These two questions can yield different answers, because a_1's and a_3's discount factors and reputations may be different. For the former question, a_3 obviously has more information about itself, and can provide a more accurate answer in that regard, but a_1 might not have much information about a_3. However, if a_1 does not have much information about a_3, then a_1 should not ask a_3 because a_1 cannot effectively evaluate a_3's an-

swer. For the latter question, a_3 has less information about a_1, and could be punished for giving a bad recommendation only for having inaccurate information about a_1.

Because a_1 knows what information it has about a_3 better than a_1 knows what information a_3 has on a_1, asking the question of whether a_3 would provide a favor to a_2 will provide a_1 with the most reliable information. In choosing this question, a_1 is incentivized to ask advice from agents it knows the most about, but may also choose to ask advice from other agents for the purpose of learning about them. Agent a_1 is further incentivized to ask advice from agents with similar discount factors, because their answers would be similar. However, if a_3's discount factor is different than a_1's and a_1 has some knowledge about a_3's discount factor, then a_1 can still use a_3's advice to learn more about a_2.

When a_1 asks a_3 whether a_3 would offer a_2 the favor in the current game, a_1 would like to make an observation about a_3 in addition to the observation of a_2. Agent a_1 can combine a_3's advice with a_2's action and utilize future observations of a_2 to more accurately reevaluate the observation made about a_3's answer. If a_3 refuses to answer when asked, a_1 can assume a_3 is not confident in its information about a_2, and cannot make an observation. If, on the other hand, a_3 has sufficient information, a_1 will judge a_3 based on the recommendation either positively or negatively.

Now we examine the implications of how a_3's recommendation will affect a_1's perception of a_3. Suppose a_3 answered it would play P in the queried game. In this case, a_1 has a P observation of a_2 for this recommendation, where a_1 will compute the observation of a_2 as if it were from a_3's perspective using its belief of a_3's discount factor. Later, a_1 may reevaluate this observation when deciding to offer a favor to a_3 by using a_1's current knowledge of a_2's discount factor and the parameters to the game in which a_3 had given its recommendation. Agent a_1 will believe that a_3's recommendation is accurate if a_1 believes that a_2's discount factor is within the bounds of a_1's observation of a_3's recommendation. Similarly, if a_3 recommends R to a_1 and a_1 later finds out that a_2 had had a low discount factor, then a_3's reputation will be increased by the P observation in a_1's hypothetical game between a_2 and a_3.

Suppose a_3 answered it would play R with a_2 and a_1 finds out later that a_2 had had a high discount factor. Agent a_1's interpretation of the observation of a_3's recommendation would be that a_3 has a discount factor below that required to play P with a_1, making the observation an upper bound on a_3's discount factor.

Finally, the most complex case is when a_3 answers it would play P with a_2, but a_1 later finds out that a_2 had had a low discount factor. If a_1 finds that a_2's discount factor is low and does not return favors, then a_1 can reason that a_3 gave the answer P so that a_1 would increase its expected value of a_3's discount factor. However, the answer would indicate that a_3's discount factor is so low that it was not concerned with a_1 other than extracting a favor that it would not have to repay. Because a_3's discount factor cannot be measured with respect to this false information for the hypothetical game, and because this false answer indicates that a_3's discount factor is arbitrarily low, a_1 would be prudent to throw away its previous observations of a_3 to reset its expected discount factor to a low value.

Note that these hypothetic observations from recommendations should be observed as when the recommendation was given, not at the time of computation. If a_1 believes a_2 underwent replacement since the observation of a_3's recommendation, but a_3 has not undergone replacement, a_1 should only use its observation history about a_2 from before a_2's replacement when computing a_3's expected discount factor.

The number of observations from communication can scale up with the number of agents as much as $|A|^3$ because each agent can communicate with all others about all others. However, in scale-free networks and other network topologies found in real-world applications, relevant communication usually scales under that bound. An agent can use various techniques to determine which agents to ask advice. A simple technique is for an agent to ask other agents that it has interacted with at least a minimum number of times; we use this in our simulations. For agents with very low replacement rates, more complex evaluations may be used, such as asking agents with similar discount factors, or agents whose answers are believed to offer the maximum amount of information entropy about a given agent.

When an agent is reevaluating the impact of observations from recommendations on other observations from recommendations, the order of evaluation will have some impact on the end results. For example, if a_1 accepts conflicting recommendations from a_3 and a_4 about a_2, one of a_3 and a_4 may be incorrectly punished depending on the order of evaluation. However, by reevaluating these observations using the agent's best knowledge at every step, this impact is minimized. Because an agent does not know which ordering is best since agents need not be truthful, reasonable solutions include evaluating observations in chronological order or minimizing conflicting observations. We use the chronological ordering in our simulations.

Simulation Results

We conducted simulations of our model to assess its behavior on groups of agents. For the simulations without communication, we used 32 agents and ran each experiment with 100 rounds. For simulations with communication, we used 16 agents and 50 rounds (due to the more significant simulation time). Random numbers were only used to set up the games and agents, so using the same seed with different algorithms provided the same set of games. While we examined the behaviors with larger and randomized samples, we used the same seed for randomizing the games across variations to remove noise in the graphs depicted in this section. The replacement rate was chosen uniformly to be .02 to reflect the mean expected agent life of 50 rounds.

We searched the set of parameterizations and used $C \sim U(0, 200)$ and various values for W because it offered a good range of behavior with respect to the interaction between agents with full knowledge of each others' discount factors. We also experimented with exponential distributions of C with mean of 100. From what we have observed with different parameterizations and distributions, the behaviors discussed in this section are typical.

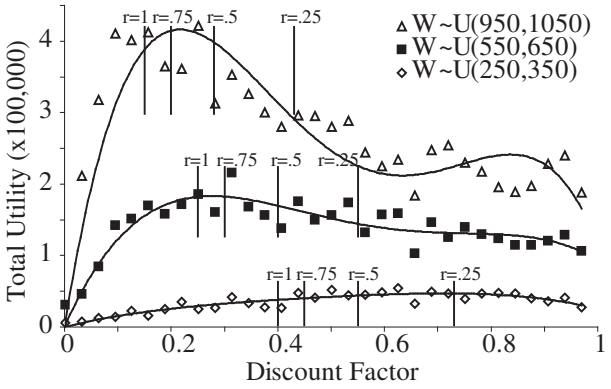

Figure 1: Typical simulation results of 32 agents each with a uniform distribution of encounter rates.

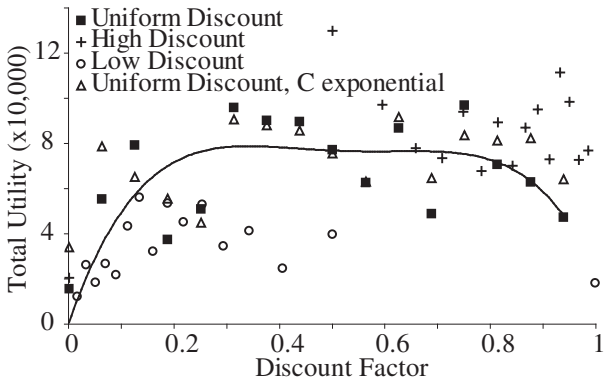

Figure 2: Various simulation results of 16 agents for 50 time steps each with a uniform distribution of encounter rates.

We ran the experiments with two topologies for agents to encounter each other. The first is a uniform topology, such that each $r_{a_1 \to a_2}$ was chosen from a uniformly distributed random number from 0 to 1. The second topology is that of a scale-free network. To construct this network, we begin with two agents having corresponding encounter rates set to 1, and add agents individually, randomizing the new agent's rate of encounter with every other agent proportional to the sum of the other agent's rate of encounters relative to the sum of all encounter rates. In this scale-free network, agents commonly encounter a small set of agents and occasionally encounter an outside agent.

To determine the effect of the distribution of agents' discount factors, we used four different distributions: uniform, all the same, 4^{th} root of uniform distribution, and 1 minus the 4^{th} root of uniform distribution. The uniform distribution allowed us to see the effects in a population of all agents, whereas the second test made sure that the behavior did not dramatically change when the agents all had the same discount factor. The last two distributions were to bias toward high or low discount factors respectively, but with one agent that had an opposite discount factor.

Figure 1 shows a small but indicative subset of our results of agents' performance given different distributions of W using a uniform topology and uniform distribution of C. Each point represents an agent's final utility, and the trend lines are depicted by the best fit quartic polynomial. Simulations with higher expected values of W obviously have higher final utilities, but the interesting feature is how the group of agents with discount factors that yield the highest final utility change with respect to W.

The vertical lines near each simulation set represent the lowest discount factor that satisfies Equation 2 with the lowest value of c solved for the symmetric rates of encounter of $r \in \{1, .75, .5, .25\}$. That means that each of these lines represents the start of a region where two opposing agents can equally balance off their retaliations to each other, forming an even tit-for-tat strategy. Because we are setting $r = r_{a_1 \to a_2} = r_{a_2 \to a_1}$ in these derivations, the results are approximations to the actual interactions. The approximation is due to agents having higher value interacting with agents from which they are likely to frequently have the chance to receive favors. However, it is clear that these analytical results give a decent approximation as to which discount factors will receive the highest payoffs.

Figure 2 depicts variations with $W \sim U(550, 650)$ for 16 agents with 50 time steps. The points labeled Uniform Discount and the corresponding trend line are the same parameterization as in Figure 1 as a reference point. The two 4^{th} root distributions (labeled High Discount and Low Discount) show results when agents have high or low discount factors. Groups of agents with high discount factors outperform the uniform distribution and groups with low discount factors underperform the uniform distribution, but that is not necessarily the case within a given distribution. The trends from the distributions matched up regardless of the topology or communication; throughout all our data, the trend is that agents with discount factors that allow them to equally retaliate achieving higher payoffs, which may not necessarily be the highest discount factors. The results from the scale-free distribution (not depicted) were similar to that of the uniform distribution with the only notable differences being an increased variance and mean in payoffs. The results from using communication (not depicted) were that agents with low discount factors tended to have 10-30% lower payoffs. Agents with high discount factors were less affected, although some attained higher and lower utilities than without communication.

Conclusions

Our favor model offers a mechanism for self-interested agents to achieve cooperation when agents can only decrease their own utility to increase others' utility. While it does not necessarily achieve the maximum possible social utility, it maximizes an agent's utilities under its own private discount factor while ensuring that agents can expect to not lose utility by helping others. Using adaptive discount factor modeling allows analysis to bridge the gap between reputation and rational strategy. This modeling also allows agents to use discount factors in other contexts besides favors. For

example, if agents were performing market transactions or playing other repeated games with one another, our favor model can supplement such interaction systems.

Agents learn which agents have high discount factors and exploit the reciprocity. Agents also have the ability to avoid loss by refusing favors to agents with low discount factors or inconsistent strategies. Our strategy converges to a steady-state equilibrium and is locally optimal with respect to the agents' discount factors. For these reasons, our model approximately meets the criteria described by Vu et al. (2006) for effective learning algorithms in multi-agent systems.

Strategic models of trust such as the one we present are required in open agent communities if the strategies are to be evolutionary stable, that is, resilient to invasion by undesirable strategies. While the model we present does not encompass all possible favor scenarios, it provides a foundation from which to build. Choosing which agent to ask a favor (similar to the multi-armed bandit problem), and using and learning joint probability distributions between C and W are extensions we leave to future work.

References

Azoulay-Schwartz, R., and Kraus, S. 2004. Stable repeated strategies for information exchange between two autonomous agents. *Artificial Intelligence* 154(1-2):43–93.

Banerjee, D.; Saha, S.; Dasgupta, P.; and Sen, S. 2005. Reciprocal resource sharing in P2P environments. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, 853–859.

Buragohain, C.; Agrawal, D.; and Suri, S. 2003. A game theoretic framework for incentives in p2p systems. In *Proceedings of the Third International Conference on Peer-to-Peer Computing*, 48– 56.

Gerkey, B. P., and Matarić, M. J. 2002. Sold!: Auction methods for multirobot coordination. *IEEE Transactions On Robotics And Automation* 18:758–768.

Golle, P.; Leyton-Brown, K.; and Mironov, I. 2001. Incentives for sharing in peer-to-peer networks. In *Proceedings of the ACM Conference on Electronic Commerce*, 264–267.

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

Jurca, R., and Faltings, B. 2007. Obtaining reliable feedback for sanctioning reputation mechanisms. *Journal of Artificial Intelligence Research* 29:391–419.

Kamvar, S. D.; Schlosser, M. T.; and Garcia-Molina, H. 2003. The eigentrust algorithm for reputation management in P2P networks. In *Proceedings of the 12th international conference on World Wide Web*, 640 – 651.

Mailath, G. J., and Samuelson, L. 2006. *Repeated Games and Reputations: Long-Run Relationships*. New York: Oxford University Press. chapter 18, 568–579.

Procaccia, A. D.; Bachrach, Y.; and Rosenschein, J. S. 2007. Gossip-based aggregation of trust in decentralized reputation systems. In *The Twentieth International Joint Conference on Artificial Intelligence (IJCAI)*, 1470–1475.

Ramchurn, S. D.; Sierra, C.; Godo, L.; and Jennings, N. R. 2006. Negotiating using rewards. In *AAMAS '06: Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems*, 400–407. New York, NY, USA: ACM.

Saha, S.; Sen, S.; and Dutta, P. S. 2003. Helping based on future expectations. In *Proceedings of the Second Intenational Joint Conference on Autonomous Agents and Multiagent Systems*, 289–296.

Sen, S. 2002. Believing others: Pros and cons. *Artificial Intelligence* 142(2):179–203.

Sierra, C., and Debenham, J. 2005. An information-based model for trust. In *AAMAS '05: Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, 497–504. New York, NY, USA: ACM.

Teacy, W. T.; Patel, J.; Jennings, N. R.; and Luck, M. 2006. Travos: Trust and reputation in the context of inaccurate information sources. *Autonomous Agents and Multi-Agent Systems* 12(2):183–198.

Vu, T.; Powers, R.; and Shoham, Y. 2006. Learning against multiple opponents. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multi-Agent Systems*, 752–759.

Weitzman, M. L. 2001. Gamma discounting. *The American Economic Review* 91(1):260–271.

Wellman, M. P. 1996. Market-oriented programming: Some early lessons. In Clearwater, S., ed., *Market-Based Control: A Paradigm for Distributed Resource Allocation*. River Edge, NJ: World Scientific. chapter 4, 74 – 95.

Yu, B., and Singh, M. P. 2002. Distributed reputation management for electronic commerce. *Computational Intelligence* 18(4):535–549.

Supporting Theorem

Theorem 1. *Given two agents a_1 and a_2 that have equal encounter rates, $r_{a_1 \to a_2} = r_{a_2 \to a_1}$, and equal maximum favor limits, $\overline{c}_{a_1 \to a_2} = \overline{c}_{a_2 \to a_1}$, there exists a discount factors for each agent that gives each agent the ability to change their opponent's utility at the same rate as their opponent.*

Proof. Given the probability density function (PDF) of C, $f_C(\cdot)$, $\frac{\partial U_{a_2}}{\partial \overline{c}_{a_1 \to a_2}} = \frac{\gamma_2}{1-\gamma_2} r_{a_1 \to a_2} E(W) f_C(\overline{c}_{a_1 \to a_2})$ and $\frac{\partial U_{a_2}}{\partial \overline{c}_{a_2 \to a_1}} = -1 - \frac{\gamma_2}{1-\gamma_2} r_{a_2 \to a_1} \overline{c}_{a_2 \to a_1} \cdot f_C(\overline{c}_{a_2 \to a_1})$. $\lim_{\gamma_{a_2} \to 0} \left\{ \frac{\partial U_{a_2}}{\partial \overline{c}_{a_1 \to a_2}} < -\frac{\partial U_{a_2}}{\partial \overline{c}_{a_2 \to a_1}} \right\}$ because all terms become 0 except for the 1, leaving $0 < 1$. $\lim_{\gamma_{a_2} \to 1} \left\{ \frac{\partial U_{a_2}}{\partial \overline{c}_{a_1 \to a_2}} > -\frac{\partial U_{a_2}}{\partial \overline{c}_{a_2 \to a_1}} \right\}$ because the constant 1 becomes irrelevant and the remaining values can be divided off leaving $E(W) > \overline{c}$, which was assumed in the problem for individual rationality. Because $\frac{1-\gamma_{a_2}}{\gamma_{a_2}}$ is continuous and differentiable over the interval of $[0, 1)$, by the intermediate value theorem, there exists a γ_{a_2} that satisfies the equality $\frac{\partial U_{a_2}}{\partial \overline{c}_{a_1 \to a_2}} = -\frac{\partial U_{a_2}}{\partial \overline{c}_{a_2 \to a_1}}$. The same derivation holds for the opposite agent. □