Modeling Reciprocal Behavior in Human Bilateral Negotiation

Ya'akov Gal
Computer Science and Artificial Intelligence Laboratory
Massachusettes Institute of Technology *
gal@csail.mit.edu

Avi Pfeffer
School of Engineering and Applied Sciences
Harvard University
avi@eecs.harvard.edu

Abstract

Reciprocity is a key determinant of human behavior and has been well documented in the psychological and behavioral economics literature. This paper shows that reciprocity has significant implications for computer agents that interact with people over time. It proposes a model for predicting people's actions in multiple bilateral rounds of interactions. The model represents reciprocity as a tradeoff between two social factors: the extent to which players reward and retaliate others' past actions (retrospective reasoning), and their estimate about the future ramifications of their actions (prospective reasoning). The model is trained and evaluated over a series of negotiation rounds that vary players' possible strategies as well as their benefit from potential strategies at each round. Results show that reasoning about reciprocal behavior significantly improves the predictive power of the model, enabling it to outperform alternative models that do not reason about reciprocity, or that play various game theoretic equilibria. These results indicate that computers that interact with people need to represent and to learn the social factors that affect people's play when they interact over time.

Introduction

A large body of evidence in the behavioral sciences has shown that people retaliate or reward each other's actions despite the absence of direct material benefit (Falk & Fischbacher 2006). Such reciprocal behavior drives people's social relationships, bringing about and maintaining cooperation, delivering punishment as well as expressing forgiveness. For example, people have been shown to pay a high cost in order to punish defectors in repeated interactions in public-good type games (Camerer 2003).

Recent technological developments have created the need for computers to interact with people in applications such as online marketplaces, military simulations, and systems for medical care (Das *et al.* 2001; Pollack 2006). While these applications differ in size and complexity, they all involve people and computer agents engaging in a series of interactions that vary the space of possible strategies and the associated rewards at each round of interaction. We refer to such interactions as *dynamic*. This paper investigates whether reasoning about reciprocity is helpful for computers that model people's behavior in dynamic interactions.

In theory, equilibrium strategies for repeated interactions can be extracted that afford reciprocal qualities (Kreps & Wilson 1982; Littman & Stone 2005). For example, strategies such as tit-for-tat allow agents to punish, forgive and reward each other in certain class of games such as the prisoners' dilemma. Other models allow agents to maximize their expected reward over time given their own beliefs about each other's decision-making process (Tesauro 2002).

However, computation of game theoretic equilibria becomes inherently difficult in dynamic settings, where players' strategies and utilities at future rounds of interaction is unknown. In addition, there is extensive evidence that shows that human behavior does not adhere to game- and decision-theoretic assumptions (Camerer 2003). Recent work has shown that it is possible to learn the social factors that affect people's play (Gal *et al.* 2004). However, this work has focused on one-shot interactions. This paper compares such models with alternative models for dynamic interaction that explicitly represents people's reciprocal reasoning.

Our theory formalizes reciprocal behavior as consisting of two factors: Players' *retrospective* benefit measures the extent to which they retaliate or reward others' actions in the past; players' *prospective* benefit measures agents' reasoning about the ramification of a potential action in the future, given that others are also reasoning about reciprocal behavior. Players' retrospective benefit depends on their beliefs about others' intentions towards them. Our model explicitly represents these beliefs in players' utility functions. This allows to learn the tradeoff people make between the retrospective and prospective benefit associated with particular actions.

We used a hierarchical model in which the higher level describes the variation between players in general and the lower level describes the variation within a specific player. This captured the fact that people may vary in the degree to which they punish or reward past behavior. We compared the performance of models that reason about reciprocal behavior with those that learn social factors in one-shot scenarios, as well as equilibrium strategies from the cooperative and behavioral game-theoretic literature. Results show that a model that learned the extent to which people reason about retrospective and prospective strategies was able pre-

*Also affiliated with the School of Engineering and Applied Sciences at Harvard University.

dict people's behavior better than the models that considered neither reciprocal factor or just one of them. These results indicate that modeling reciprocity in humans is essential for computer players that engage in dynamic interaction with people.

Interaction Scenario

Our experiments deployed a game called Colored Trails (CT) (Grosz *et al.* 2004) that explicitly manifests goals, tasks, and resources in a way that is compelling to people, yet abstracts away from a complicated underlying domain. CT is played on a 4x4 board of colored squares with a set of chips. One square on the board was designated as the goal square. Each player's icon was initially located in a random, non-goal position. Players were issued four colored chips. To move to an adjacent square required surrendering a chip in the color of that square. Players have full view of the board and each others' chips.

Our version of CT included a one-shot take-it-or-leave-it negotiation round between two agents that needed to exchange resources to achieve their goals. Players are designated one of two roles: *proposer* players could offer some subset of their chips to be exchanged with some subset of the chips of responder players; *responder* players could in turn accept or reject proposers' offers. If no offer was made, or if the offer was declined, then both players were left with their initial allocation of chips. Players' performance was determined by a scoring function that depended on the result of the negotiation. Players' roles alternated at each round (CT is not a zero sum game).

CT is a conceptually simple but strategically complex game. The number of possible exchanges at each game is exponential in the joint chip pool of both players. In our scenario, this amounted to $2^8 = 256$ possible exchanges. Varying the initial chip and board layout allows to capture a variety of dependency relationships that hold between players.

A Dynamic Model of Interaction

Each instance in our scenario includes a set of finite rounds of bilateral interaction. Each round includes a game g_k consisting of a CT board, chip allocations, and players' initial positions. The game g_k determines the set of possible exchanges $\mathbf{E}_k^g$ that the proposer can offer. The datum for a single round is a tuple $\mathbf{c}_k = (e_k^i, r_k^j)$ consisting of an observed offer $e_k^i \in \mathbf{E}_k^g$ made by proposer i and response $r_k^j \in \{\text{yes, no}\}$ made by responder j.

To capture people's diverse behavior we use the notion of types. A type captures a particular way of making a decision, and there may be several possible types. At each round, the probability of an action depends on players' types as well as on the history of interaction prior to the current round.

Let $d = (\mathbf{c}_1, \ldots, \mathbf{c}_{n_d})$ represent an instance that includes n_d rounds of interaction and let t^i, t^j denote the types for players i and j respectively. We assume each instance is generated by a unique agent pair and that agents alternate their proposer-responder roles at each round. In general, the likelihood of each round k depends on the history of former rounds. Let $\pi^i(\mathbf{c}_k)$ denote the reward at round k to i. We can write the likelihood of an instance d given t^i, t^j as

$$p(d \mid t^i, t^j) = \prod_{k=1,\text{odd}}^{n_d} \Big(p(e_k^i \mid g_k, \mathbf{c}_1, \ldots, \mathbf{c}_{k-1}, t^i, t^j) \cdot p(r_k^j \mid e_k^j, g_k, \mathbf{c}_1, \ldots, \mathbf{c}_{k-1}, t^i, t^j) \Big) \prod_{k=1,\text{even}}^{n_d} \Big(p(e_k^j \mid g_k, \mathbf{c}_1, \ldots, \mathbf{c}_{k-1}, t^i, t^j) \cdot p(r_k^i \mid g_k, e_k^j, \mathbf{c}_1, \ldots, \mathbf{c}_{k-1}, t^i, t^j) \Big) \quad (1)$$

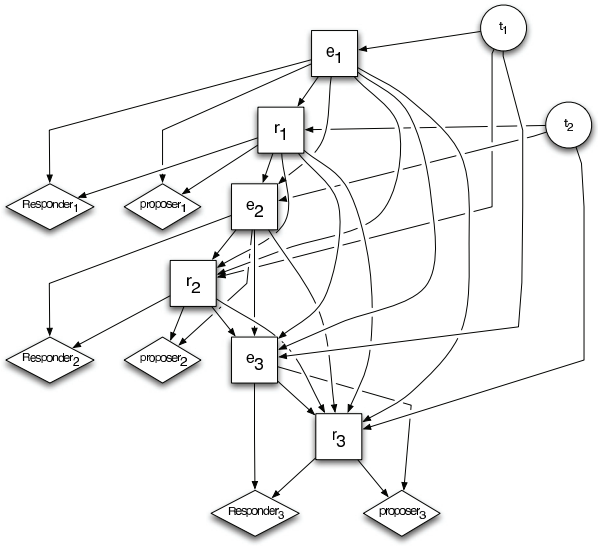

Figure 1: Three rounds of interaction

We illustrate this interaction using Multi-Agent Influence diagrams (Koller & Milch 2001). A MAID is a directed acyclic graph containing three kinds of nodes: chance nodes denoted by ellipses, decision nodes denoted by rectangles, and utility nodes denoted by diamonds. Each chance node has an associated conditional probability distribution (CPD). A utility node has an associated deterministic function from values of its parents to the real numbers. The parents of a decision node represent information that is known to the decision maker at the time of making the decision, and are called informational parents. Each decision and utility node is associated with a particular agent. A MAID for three rounds of interaction is shown in Figure 1. [1] In general, there is a high degree of dependency in the network. Each action depends on the type that generated the action as well as on the entire history of past play.

We attempt to reduce the dependency of the model by making agents' actions depend on an unobserved *state*. The state encapsulates all the information for an agent that is salient for making its decision. Thus, the state information should include the types for both agents, the current round and a function that summarizes the past history of

[1] For simplicity, we have not included nodes representing games in this diagram.

play. We selected a function that determined players' beliefs about each other's intentions towards them in the game, suggested by Rabin (1993). For each agent, we define a scalar called *merit*, denoted $\mathbf{m}_k = (m^i, m^j)$ that is given an initial value of zero at the onset of an instance and is updated by the agents given the observations at each round. A positive merit value for agent i implies that agent j believes that i has positive intentions towards j, and conversely for a negative merit value. This belief depends on the relative difference between the benefit from the proposed offer to agent j and an action deemed "fair" to agent j. The state at round k is thus a tuple $\mathbf{s}_k = (\mathbf{m}, \mathbf{t})$ consisting of agents' merits and types. Given an initial state $\mathbf{s}_1$ and a game g_k we can rewrite Equation 1 as

$$p(d \mid \mathbf{s}_1) = \begin{array}{ll} \prod_{k=1,\text{odd}}^{n_d} & p(e_k^i \mid g_k, \mathbf{s}_k) \cdot p(r_k^j \mid e_k^j, g_k, \mathbf{s}_k) \\ \prod_{k=1,\text{even}}^{n_d} & p(e_k^j \mid g_k, \mathbf{s}_k) \cdot p(r_k^i \mid e_k^j, g_k, \mathbf{s}_k) \end{array} \tag{2}$$

We chose the "fair" action at each round to correspond with the Nash bargaining strategy, a common solution concept of cooperative game theory that maximizes the product of agents' benefits from any potential exchange. This concept satisfies several attractive social criteria, such as Pareto optimality, beneficial to agents (no player can get less from not bargaining), and symmetric with respect to agents' changing roles. If for a given agent the difference in benefit to the other between its action and the fair action is positive then the agent has behaved kindly towards the other, and nastily if the difference is negative. Merit is thus an aggregate measure of players' kindness at each round k. We assume that agents have common knowledge of each others' merits and update them in the same way as follows:

$$m_{k+1}^i(\mathbf{c}_k) = m_k + \frac{\pi_k^j(\mathbf{c}_k) - \pi_k^j(nb)}{\pi_k^j(max) - \pi_k^j(min)} \tag{3}$$

where max and min represent the exchanges that incur maximum and minimum benefit to player j in round k. If i is a proposer the quantity $\pi^j(\mathbf{c}_k)$ refers to the associated benefit from exchange e_k^i to responder j; if i is a responder the quantity $\pi^j(\mathbf{c}_k)$ refers to the benefit for proposer j from exchange e_k^j if r_k^i is "yes", and zero otherwise. Each round of interaction depends only on the state that generated it. This state-space representations allows us to represent an instance compactly, as shown in Figure 2.

Credit Assignment

In our model, the total accumulated reward for some action at state s_k in round k depends on two factors: the immediate benefit of the action at the game g_k and the ramification of the action over future rounds given that agents update their models of each other at each round. We denote $FR_P^i(e_k^i \mid g_k, \mathbf{s}_k)$ to be the future ramification for proposer i for taking action e_k^i at round k at state $\mathbf{s}_k$. We can now define the total reward ER_P^i for the proposer to be

$$ER_P^i(e_k^i \mid g_k, \mathbf{s}_k) = \sum_{r_k^j} p(r_k^j \mid e_k^i, g_k, \mathbf{s}_k) \cdot \left(\pi^i(\mathbf{c}_k) + FR_P^i(e_k^i \mid g_k, \mathbf{s}_k)\right) \tag{4}$$

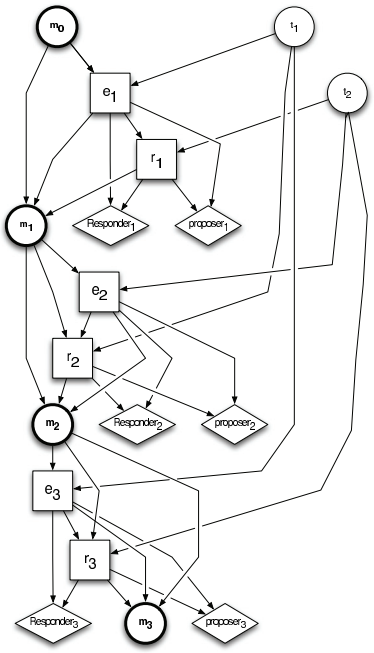

Figure 2: Three rounds of interaction with state representation

We define the total accumulated reward ER_R^i for the responder to be

$$ER_R^i(r_k^i \mid e_k^j, g_k, \mathbf{s}_k) = \pi^i(\mathbf{c}_k) + FR_R^i(r_k^i \mid e_k^j, g_k, \mathbf{s}_k) \tag{5}$$

The future ramification for proposer i at round k depends on the total accumulated reward for i at round $k+1$ in which i will be playing the role of a responder, and j will be playing the role of a proposer. The games played and states reached in future rounds are unknown, but we can sum over the likelihood of each possible game and future state and define the future ramification as

$$\begin{aligned} FR_p^i(e_k^i \mid g_k, \mathbf{s}_k) = & \int p(\mathbf{s}_{k+1} \mid e_k^i, g_k, \mathbf{s}_k) \cdot \\ & \Big(\sum_{g_{k+1}} \sum_{e_{k+1}^j} p(e_{k+1}^j \mid g_{k+1}, \mathbf{s}_{k+1}) \\ & \sum_{r_{k+1}^j} p(r_{k+1}^i \mid e_{k+1}^j, g_{k+1}, \mathbf{s}_{k+1}) \cdot \\ & ER_R^i(r_{k+1}^i \mid e_{k+1}^j, g_{k+1}, \mathbf{s}_{k+1}) d\mathbf{s}_{k+1} \Big) \end{aligned} \tag{6}$$

In a similar fashion, the term $FR_R^i(r_k^i \mid e_k^j, g_k, \mathbf{s}_k)$ represents the total accumulated reward for the responder at round $k+1$, in which the responder and proposer roles are reversed:

$$\begin{aligned} FR_R^i(r_k^i \mid e_k^j, g_k, \mathbf{s}_k) = & \int p(\mathbf{s}_{k+1} \mid \mathbf{c}_k, g_k, \mathbf{s}_k) \cdot \\ & \Big(\sum_{g_{k+1}} \sum_{e_{k+1}^i} p(e_{k+1}^i \mid g_{k+1}, \mathbf{s}_{k+1}) \cdot \\ & ER_P^i(e_{k+1}^i \mid g_{k+1}, \mathbf{s}_{k+1}) d\mathbf{s}_{k+1} \Big) \end{aligned} \tag{7}$$

Representing Social Factors

We define the following features $\mathbf{x} = \{x_1, \ldots, x_4\}$ representing social factors for agents i, j in round k, that depend on the current state $\mathbf{s}_k$:

Individual Benefit This feature represents the immediate benefit to agent i from the round $(x_1 = \pi^i(\mathbf{c}_k))$.

Other's Benefit This feature represents the immediate benefit to the other agent j from the round $(x_1 = \pi^j(\mathbf{c}_k))$.

Retrospective Benefit This feature represents the benefit to an agent from rewarding and punishing past actions of the other agent. We define this to be the product of the other's benefit and its merit $(x_3 = \pi^j(\mathbf{c}_k) \cdot m^j)$. Players' retrospective benefit increase when rewarding others that have positive merit, and when punishing others that have negative merit.

Prospective Benefit This feature represents the benefit the agent expects to receive in the future as a result of taking its current action. In our terminology this is the future ramification for the agent ($x_4 = FR^i_p$ for a proposer or $x_4 = FR^i_R$ for a responder). Players' prospective benefit depends on the consequences of their action in future rounds, given that players update their beliefs about each others' types and intentions.

Using and Learning the Model

A state $\mathbf{s}_k$ includes a type t^i that generates weights $\mathbf{w}^i$ associated with the features $\mathbf{x}$. The social utility for i at round $\mathbf{c}_k$ is a weighted sum of the feature values given the state $\mathbf{s}_k$.

$$u_i(\mathbf{c}_k \mid \mathbf{s}_k) = \mathbf{w}^i \cdot \mathbf{x}$$

Note that the social utility depends on the joint actions of both proposer and responder agents as well as their beliefs over each other's merits.

To compute the probability $p(e^i_k \mid g_k, \mathbf{s}_k)$ of an exchange and the probability $p(r^i_k \mid e^j_k, g_k, \mathbf{s}_k)$ of a response we use a soft-max function. The probability of any action depends on the utility of the action compared to other actions.

$$p(e^i_k \mid g_k, \mathbf{s}_k) = \frac{e^{u^i((e^i,\text{yes})|\mathbf{s}_k)}}{\sum_{e'_k \in \mathbf{E}^g_k} e^{u^i((e'_k,\text{yes})|\mathbf{s}_k)}} \quad (8)$$

$$p(r^i_k \mid e^j_k, g_k, \mathbf{s}_k) = \frac{e^{u^i((e^j_k,r^i_k)|\mathbf{s}_k)}}{\sum_{r'_k \in \{\text{yes.no}\}} e^{u^i((e^j_k,r'_k)|\mathbf{s}_k)}} \quad (9)$$

This function captures certain aspects of human behavior: Its stochasticity makes the likelihood of actions associated with a high social utility to greater, while still allowing players to deviate from this principle; the likelihood of choosing an exchange that incurs a high social utility will increase if there are few other similar exchanges that incur high utility, and will decrease if there are many other similar exchanges.

We now wish to learn a mixture model consisting of a distribution over people's types, and the weight values for each type. To this end, we adapt an algorithm proposed by Gal & Pfeffer (2006) in which gradient descent is used to update the weight after each observation for each type. The likelihood that type t^i generated an instance d can be computed using Bayes rule.

$$p(t^i \mid d) \cong p(d \mid t^i) \cdot p(t^i) = p(t^i) \int p(\mathbf{s}_1) p(d \mid \mathbf{s}_1) d\mathbf{s}_1 \quad (10)$$

where $p(d \mid \mathbf{s}_1)$ is given in Equation 2.

We make the degree to which a training example contributes to learning the weights in a network be proportional to the probability that the model actually generated the data. To this end, we assign a learning rate $\alpha_i = \alpha \cdot p(t^i \mid d)$ that depends on the likelihood that type t^i generated d, where α is a constant less than 1. Taking the derivative of the error function with respect to the weights, we get the following update rule at each round k of instance d for agent i (proposer or responder) and action $a \in \{e^i_k, r^i_k\}$:

$$\mathbf{w}^i = \mathbf{w} + \alpha_i \cdot (err^i_k)^2 \cdot (1 - err^i_k) \cdot (\mathbf{x}^{a,*}_k - \mathbf{x}^a)$$

Here, the term $(\mathbf{x}^{a,*}_k - \mathbf{x}^a)$ denotes the difference between the features associated with the observed action a^*_k in the training set and any possible action a^i_k. For the proposer, when we expanding the error function according to Equation 8 we get that for any potential offer $e^i_k \in \mathbf{E}^g_k$ the error function is equal to

$$err^i_k = \frac{1}{1 + e^{-(u^i(\mathbf{c}^*_k|\mathbf{s}_k) - u^i(\mathbf{c}^*_k|\mathbf{s}_k))}}$$

The weight update for the proposer is similar.

We proceed to update the weights for each instance in training set for every agent i and for every possible instantiation of the state $\mathbf{s}_1$. In our formalism, this corresponds to summing over the possible types t^i. Players' types are unobserved during training, but after each epoch computing the new parameters is done using a variant of the EM algorithm. We compute for each instance $d \in D$ $E(N_t|D) = \sum_{d \in D} p(t^i \mid d, \mathbf{s}_1) + p(t^j \mid d, \mathbf{s}_1)$ and normalize.

To determine the social utility, it is necessary to compute the future ramification factors of Equation 7 and 6. There are difficulties associated with this computation. First, it requires combinatorial search in the depth of the number of rounds. Second, computing the update belief state $\mathbf{s}_{k+1}$ requires to sum over all possible values of s_k.

However, there is structure in our domain that reduces this computation. First, players' types do not change between rounds so they do not depend on the state. Second, agents' merits are common knowledge, and can be computed at round $k+1$ using Equation 3. The only unobserved element in $\mathbf{s}_{k+1}$ is the type for the other player, and we can therefore rewrite Equation 5 to be

$$FR^i_R(r^i_k \mid g_k, \mathbf{s}_k) = \sum_{t^j} p(t^j) \Big(\sum_{g_{k+1}} \sum_{e^i_{k+1}} p(e^i_{k+1} \mid g_{k+1}, \mathbf{s}_{k+1}) \cdot ER^i_P(e^i_{k+1} \mid g_{k+1}, \mathbf{s}_{k+1}) \Big)$$

where $\mathbf{s}_{k+1} = (t^i, t^j, \mathbf{m}_{k+1})$ is the updated state at round $k+1$ and $\mathbf{m}_{k+1}$ represents the updated merit values. Sim-

ilarly, we can compute the future ramification for the proposer as follows:

$$FR^i_P(e^i_k \mid g_k, \mathbf{s}_k) = \sum_{t^j} p(t^j)\Big(\sum_{g_{k+1}} \sum_{e^j_{k+1}} p(e^j_{k+1} \mid g_{k+1}, \mathbf{s}_{k+1}) \cdot \sum_{r^i_{k+1}} p(r^i_{k+1} \mid e^j_{k+1}, g_{k+1}, \mathbf{s}_{k+1}) \cdot ER^i_R(r^i_{k+1} \mid e^j_{k+1}, g_{k+1}, \mathbf{s}_{k+1})\Big)$$

These computations still requires to sum over all possible games and all possible actions, for a horizon that equals the number of rounds. The number of possible games is enormous. To simplify the computation, we sample a small number of games. Different games are sampled for each instance and each round, but the same sampled games are used to compare different actions, to make the comparison as fair as possible.

The number of actions at each game is not so large, typically about fifty, but since we need to compute future ramifications many times for learning this is also a bottleneck, so we sample a few actions. However, to get the probability of an action, according to Equation 8 we need to know the utility of all actions, not just the ones we sample. To get around this, we explicitly compute the future ramification of a small number of sampled actions. We then fit a linear function where the independent variable is the merit change for taking a particular action, and the dependent variable is the future ramification. We use this function to estimate the future ramification of actions that are not sampled, and so compute their utility. The probability of all actions can then be computed.

In addition, we use dynamic programming to compute future ramification. This cannot be done directly as Equations 4 and 5 are written because merit is a continuous variable. So we discretize the merit into a small number of bins, keeping the number of states small. To simplify things even further, we make the assumption that the other agent's merit does not change in future rounds; only changes to an agent's own merit are taken into account. Changes to the other player's merit are a second-order effect — they do not affect how the other player responds to your actions, and have a small effect on how your future actions are perceived. With these optimizations, it takes about four seconds to compute future ramifications for a horizon of eight.

Results and Discussion

We performed human subject trials to collect data of people playing our dynamic CT scenario. There were ten subjects that generated 54 instances of dynamic interactions. Each instance consisted of between four and ten rounds of CT games in which the players alternated proposer-responder roles at each round. CT games were sampled from a distribution that varied the number of possible negotiation strategies, their associated benefit and the dependency relationships between players, i.e. who needed whom to get to the goal. A total of 287 games were played.

An initial analysis of the data revealed that players engaged in reciprocal behavior when they interacted with each other. For example, a significant correlation ($r^2 = 0.371$) was found between the benefit to the responder from one round to the next. This shows that if the proposer was nice to the responder in one round, the responder (who became a proposer in the next round) rewarded the gesture by making a nice offer. Similar correlations were apparent across multiple rounds, suggesting that the effects of reciprocal reasoning on people are long lasting.

We learned several computer models of human players. The model named *no-learning* used weights learned from data collected from single-shot games. Including this model allowed us to test to what degree behavior from single-shot games carries over to the dynamic setting. The model named *no-recip* learned only the weights for the individual benefit and other's benefit, without accounting for reciprocal behavior. This model has been shown to capture people's play in one-shot scenarios (Gal *et al.* 2004). The model named *recip* learned all four features, the two non-time-dependent features and the features representing both retrospective and prospective benefit. The models named *retro* and *prosp* learned three features: the two non-time-dependent features, and either the retrospective benefit or the prospective benefit, respectively. We also compared to two game-theoretic concepts: the Nash Bargaining solution (J.Nash 1950), and the Fairness Equilibrium (Rabin 1993), which attempts to capture player's psychological benefit from rewarding or punishing others who they think will be nice or nasty to them. The following table shows the fit-to-data results, all of which are significant within the 95% confidence interval.

Model	Likelihood	class (Proposer)	class (Responder)
no-learning	27.17	0.42	0.68
no-recip	26.5	0.46	0.66
recip	24.4	0.50	0.66
Nash B.	—	0.42	0.52
Fair Eq.	—	0.38	0.42
retro	24.3	0.42	0.72
prosp	24.3	0.51	0.7

The first column shows the negative log likelihood of the data. Since the Nash bargaining solution and the fairness equilibrium do not define probability distributions, no likelihood is shown for them. The likelihoods are fairly similar to each other because the probability distribution is relatively flat, because there are a large number of offers available with similar utility. According to this metric, we see that all the models that learn reciprocal behavior do better than the *no-recip* model which does not. Furthermore, the *no-learning* model does worst, indicating that we cannot simply reuse the weights learned for a single-shot game. However there is little difference between the three-feature models and the four-feature model (the four-feature model actually doing marginally worse). So there is evidence that people do employ reciprocal reasoning, and that it is useful to model this computationally, but according to this metric this can be captured just as well with a single reciprocal feature, and either retrospective or prospective benefit will do just as well.

Looking more closely at the results provides more information. The next column, labeled class(Proposer), indicates the fraction of times that the actual offer made by the proposer was in the top 25% of offers predicted by the model. For the computer models this meant the top 25% in order of probability; for the Nash Bargaining solution this meant the top 25% in order of product of benefits for the proposer and responder; and for the Fairness Equilibrium this meant the top 25% in order of total benefit including psychological benefit. The results for this metric show that *recip* and *prosp* perform best, while *retro* performs as badly as *no-learning* and the game-theoretic models. This provides evidence that it is prospective benefit that is particularly important. The final column shows a similar metric for the responder decision; in this case it is the fraction of time that the model correctly predicted acceptance or rejection. We see here that the two game-theoretic equilibria did significantly worse than the other models, showing that we cannot simply borrow game-theoretic notions and expect them to work well. There was little differentiation between the other models, with the two three-parameter models doing slightly better. Taking all the metrics into account, we conclude that there is no evidence that people simultaneously reason prospectively and retrospectively, and of the two temporal features prospective reasoning are more important.

Conclusion and Future Work

In this paper we have studied the computational modeling of people's behavior in dynamic bilateral negotiation scenarios. This initial investigation has shown that reasoning about reciprocal behavior improves the predictive power of computers that model people. One cannot simply borrow an approach that worked for one-shot games and expect it to work well for dynamic interactions. Also, standard equilibrium notions from game theory do not work well. We presented a theory of reciprocity that distinguishes between the immediate benefit of agents' actions in the present, the benefit agents gain from rewarding or punishing past behavior, and the effects of the actions on their future well-being, and described the relationship between these factors formally. We provided an algorithm to learn the relative contribution of these factors on people's play in a domain that includes uncertainty over future interactions with a large number of possible strategies.

In future, we plan to conduct further experiment to see whether the finding that prospective reasoning is more affective than prospective reasoning generalizes to environments that vary the negotiation protocol, game size and other factors. Also, we will design computer agents that use these learned models of reciprocal behavior. We envision two kinds of agents. Emulators will utilize the learned models to mimic people's play, either by choosing the option that has the highest probability according to the model, or else randomizing according to the distribution specified by the model. Maximizers will attempt to maximize their expected utility given their model of how the other agent will play. These players will need to tailor their model of how people behave in general to how individual, new people behave in different circumstances, and update their beliefs about the other player based on the observed behavior of the other player. We plan to test these computer players in experiments in which they participate in negotiations with people.

Acknowledgments

This work was supported by AFOSR under contract FA9550-05-1-0321 Development and dissemination of the Colored Trails formalism is supported in part by the National Science Foundation under Grant No. CNS-0453923.

References

Camerer, C. 2003. *Behavioral Game Theory. Experiments in Strategic Interaction*. Princeton University Press. chapter 2.

Das, R.; Hanson, J. E.; Kephart, J. O.; and Tesauro, G. 2001. Agent-human interactions in the continuous double auction. In Nebel (2001).

Falk, A., and Fischbacher, U. 2006. A theory of reciprocity. *Games and Economic Behvior* 54(2):293–315.

Gal, Y., and Pfeffer, A. 2006. Predicting people's bidding behavior in negotiation. In Stone, P., and Weiss, G., eds., *Proc. 5th International Joint Conference on Multi-agent Systems (AAMAS'06)*.

Gal, Y.; Pfeffer, A.; Marzo, F.; and Grosz, B. 2004. Learning social preferences in games. In *Proc. 19th National Conference on Artificial Intelligence (AAAI'04)*.

Grosz, B.; Kraus, S.; Talman, S.; and Stossel, B. 2004. The influence of social dependencies on decision-making. Initial investigations with a new game. In *Proc. 3rd International Joint Conference on Multi-agent Systems (AAMAS'04)*.

J.Nash. 1950. The bargaining problem. *Econometrica* 18:155–162.

Koller, D., and Milch, B. 2001. Multi-agent influence diagrams for representing and solving games. In Nebel (2001).

Kreps, D., and Wilson, R. 1982. Reputation and imperfect information. *Journal of Economic Theory* 27:253–279.

Littman, M., and Stone, P. 2005. A polynomial-time Nash equilibrium algorithm for repeated games. *Decision Support Systems*. repeated games, nash bargaining.

Nebel, B., ed. 2001. *Proc. 17th International Joint Conference on Artificial Intelligence (IJCAI'01)*.

Pollack, M. 2006. Intelligent technology for an aging population: The use of AI to assist elders with cognitive impairment. *AI Magazine* 26(9).

Rabin, M. 1993. Incorporating fairness into game theory and economics. *American Economic Review* 83:1281–1302.

Tesauro, G. 2002. Efficient search techniques for multi-attribute bilateral negotiation strategies. In *Third International Symposium on Electronic Commerce*.