

# Learning and Evaluating Human-Like NPC Behaviors in Dynamic Games

**Yu-Han Chang, Rajiv Maheswaran, Tomer Levinboim, Vasudev Rajan**

University of Southern California Information Sciences Institute

Marina del Rey, CA 90292

ychang@isi.edu, maheswar@isi.edu, tomer.levinboim@usc.edu, vasudev@usc.edu

## Abstract

We address the challenges of evaluating the fidelity of AI agents that are attempting to produce human-like behaviors in games. To create a believable and engaging game play experience, designers must ensure that their non-player characters (NPCs) behave in a human-like manner. Today, with the wide popularity of massively-multi-player online games, this goal may seem less important. However, if we can reliably produce human-like NPCs, this can open up an entirely new genre of game play. In this paper, we focus on emulating human behaviors in strategic game settings, and focus on a Social Ultimatum Game as the testbed for developing and evaluating a set of metrics for comparing various autonomous agents to human behavior collected from live experiments.

## Introduction

Development of non-player characters (NPCs) in interactive games gets at the root of the field of artificial intelligence — the creation of autonomous agents that exhibit human-like behavior. In serious games such as virtual training environments, we may need to quickly develop NPCs that emulate the behaviors of an indigenous population, which varies according to culture, socio-economic, and religious factors (Rickel 2001). In other games, realistic NPCs can lead to more engaging game-play, enabling richer development of story lines and realism of game play even in online multiplayer games. Here we focus on the need for metrics to evaluate the verisimilitude of autonomous agent behaviors relative to actual human behavior. We also describe a set of techniques for learning agent models directly from human data.

The classical Turing Test relies on human evaluation to judge the similarity of the conversation produced by the artificial agent relative to actual human conversation. In more restricted problems, such as classification, we are satisfied when a machine consistently produces the correct label (a perfect match), given a test data point. In this paper, we are concerned with domains falling somewhere in the middle, where an agent’s human-like behavior will not necessarily produce a perfect match to some predefined standards, but where we would prefer not to rely exclusively on hu-

man judgement to determine whether an agent’s outputs are “close” to real human behavior.

In particular, we are interested in multi-agent game domains where humans make sequential actions, or decisions, over time. In the emotional agents community, the degree of realism is typically evaluated by a human judge (Mao and Gratch 2004). In the machine learning and reinforcement learning community, agent “goodness” is typically evaluated relative to optimal behavior, using a metric like expected reward. This notion of optimality is ill-defined in many of the domains of interest. Optimality of one agent in a multi-agent game is dependent on the strategies employed by the other agents. Furthermore, for most games we are interested in, humans simply do not play optimally, whether it be due to bounded rationality, preference for following heuristics or narratives, altruism, or other reasons.

Luckily, human data in multi-agent domains is becoming easy to collect, given widespread access to the Internet and online interaction. Thus, we can obtain baseline collections of behavior trajectories that describe human play. The challenge is to find a way to compare collections of traces produced by autonomous agents with this existing baseline, in order to determine which agents exhibit the most realistic behavior.

In this paper, we investigate these issues in the context of the Social Ultimatum Game (SUG). SUG is a multi-agent multi-round extension of the Ultimatum Game (Henrich, Heine, and Norenzayan 2010), which has been a frequently studied game over the last three decades as a prominent example of how human behavior deviates from game-theoretic predictions that use the “rational actor” model. Data gathered from people playing SUG was used to create various classes of autonomous agents that modeled the behaviors of the individual human players. We then created traces from games with autonomous agents emulating the games that the humans played. We develop several metrics to compare the collections of traces gathered from games played by humans and games played by the autonomous agents.

From this analysis, it becomes clear that human behavior contains unique temporal patterns that are not captured by the simpler metrics. In SUG, this is revealed in the likelihood of reciprocity as a function of the history of reciprocity. The key implication is that it is critical to retain the temporal elements when developing metrics to evaluate the efficacy of

autonomous agents for replicating human strategic behavior in dynamic settings.

## The Social Ultimatum Game

To ground our subsequent discussion, we begin by introducing the Social Ultimatum Game. The classical Ultimatum Game, is a two-player game where  $P_1$  proposes a split of an endowment  $e \in \mathbb{N}$  to  $P_2$  who would receive  $q \in \{0, \delta, 2\delta, \dots, e - \delta, e\}$  for  $\delta \in \mathbb{N}$ . If  $P_2$  accepts,  $P_2$  receives  $q$  and  $P_1$  receives  $e - q$ . If  $P_2$  rejects, neither player receives anything. The subgame-perfect Nash or Stackelberg equilibrium has  $P_1$  offering  $q = \delta$  (i.e., the minimum possible offer), and  $P_2$  accepting, because a “rational”  $P_2$  should accept any  $q > 0$ , and  $P_1$  knows this. Yet, humans make offers that exceed  $\delta$ , make “fair” offers of  $e/2$ , and reject offers greater than the minimum.

To represent the characteristics that people operate in societies of multiple agents and repeated interactions, we introduce the Social Ultimatum Game. Players, denoted  $\{P_1, P_2, \dots, P_N\}$ , play  $K \geq 2$  rounds, where  $N \geq 3$ . In each round  $k$ , every player  $P_m$  chooses a recipient  $R_m^k$  and makes them an offer  $q_{m,n}^k$  (where  $n = R_m^k$ ). Each recipient  $P_n$  then considers the offers they received and makes a decision  $d_{m,n}^k \in \{0, 1\}$  for each offer  $q_{m,n}^k$  to accept (1) or reject (0) it. If the offer is accepted by  $P_m$ ,  $P_m$  receives  $e - q_{m,n}^k$  and  $P_n$  receives  $q_{m,n}^k$ , where  $e$  is the endowment to be shared. If an offer is rejected by  $P_n$ , then both players receive nothing for that particular offer in round  $k$ . Thus,  $P_m$ 's reward in round  $k$  is the sum of the offers they accept (if any are made to them) and their portion of the proposal they make, if accepted:

$$r_m^k = (e - q_{m,n}^k)d_{m,n}^k + \sum_{j=1 \dots N, j \neq m} q_{j,m}^k d_{j,m}^k \quad (1)$$

The total rewards for  $P_m$  over the game is the sum of per-round winnings,  $r_m \sum_{k=1}^K r_m^k$ . A game trajectory for  $P_m$  is a time-series of proposed offers,  $O_m^k = (R_m^k, q_{m,n}^k, d_{m,n}^k)$  and received offers,  $O_{n,m}^k = (R_n^k, q_{n,m}^k, d_{n,m}^k)$ . At time  $k$ , the trajectory for  $P_m$  is  $T_m^k = (O_m^k, \{O_{n,m}^k\}_n, O_m^{k-1}, \{O_{n,m}^{k-1}\}_n, \dots, O_m^1, \{O_{n,m}^1\}_n)$ . Assuming no public information about other players' trajectories,  $T_m^k$  includes all the observable state information available to  $P_m$  at the end of round  $k$ .

A screenshot of the offer phase in our online Social Ultimatum Game is shown in Figure 1.

## Metrics

Let  $C_m$  be the collection of trajectories  $P_m$  produces by taking part in a set of Social Ultimatum Games. In other domains, these traces could represent other interactions. Our goal is to evaluate the resemblance of a set of human trace data  $C$  to other sets of traces  $\tilde{C}$ , namely those of autonomous agents. We need a metric that compares sets of multi-dimensional time series:  $d(C, \tilde{C})$ . Standard time-series metrics such as Euclidean or absolute distance, edit distance, and dynamic time warping (Mitsa 2010) are not appropriate in this type of domain.

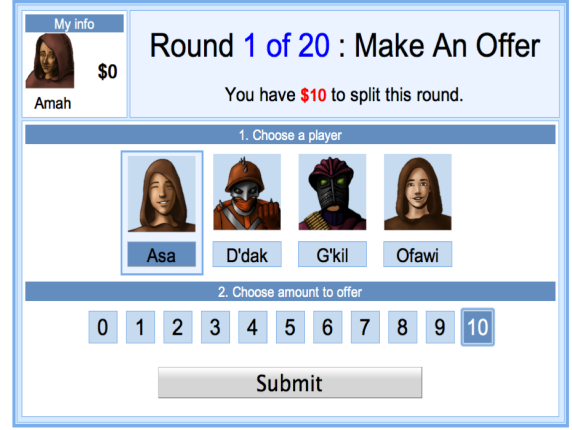


Figure 1: The Social Ultimatum Game Interface

One challenge arises because we are interested in the underlying behavior that creates the trajectories rather than superficial differences in the trajectories themselves. If we can collapse a collection of traces  $C$  to a single probability distribution  $Q$ , by aggregating over time, then we can define a *time-collapsed* metric,

$$d(C, \tilde{C}) = KL(Q || \tilde{Q}) + KL(\tilde{Q} || Q) \quad (2)$$

where KL denotes the Kullback-Leibler divergence. The sum enforces symmetry and nonnegativity. Time-collapsed metrics for SUG include:

- **Offer Distribution.** Let  $Q^O$  be the distribution of offer values  $\{q_{m,n}^k\}$  observed over all traces and all players.
- **Target-Recipient Distribution.** Let  $Q^R$  denote the likelihood that a player will make an offer to the  $k^{th}$  most likely recipient of an offer. This likelihood is non-increasing in  $k$ . In a 5-person game, a single player may have a target-recipient distribution that looks like  $\{0.7, 0.1, 0.1, 0.1\}$  which indicates that they made offers to their most-targeted partner 7 times more often than their second-highest-targeted partner. We can produce  $Q^R$  by averaging over all games to characterize a player and further average over all players to characterize a population.
- **Rejection Probabilities.** For each offer value  $q$ , we have a Bernoulli distribution  $Q^{B_q}$  that captures the likelihood of rejection by averaging across all players, games and rounds in a collection of traces. We then define a metric:

$$d^B(C, \tilde{C}) = \sum_{q=0}^{10} KL(Q^{B_q} || \tilde{Q}^{B_q}) + KL(\tilde{Q}^{B_q} || Q^{B_q}).$$

We can also define *time-dependent* metrics that acknowledge that actions can depend on observations of previous time periods. One prominent human manifestation of this characteristic is reciprocity. We define two time-dependent metrics based on reciprocity:

- **Immediate Reciprocity** When a player receives an acceptable offer from someone, they may be more inclined

to reciprocate and propose an offer in return in the next round. We can quantify this  $p(R_m^{k+1} = n | R_n^k = m)$  across all players and games in a collection of traces. This probability defines a Bernoulli distribution  $Q^Y$  from which we can define a metric  $d^Y$  as before.

- **Reciprocity Chains** Taking the idea of reciprocity over time further, we can calculate the probability that an offer will be reciprocated, given that a chain of reciprocity has already occurred. For example, for chains of length  $c = 2$ , we  $p(R_m^{k+1} = n | R_n^k = m, R_m^{k-1} = n)$ ; for  $c = 3$ , we calculate  $p(R_m^{k+1} = n | R_n^k = m, R_m^{k-1} = n, R_n^{k-2} = m)$ . As before, these probabilities can be used to define a Bernoulli distribution  $Q^{Y_c}$  for each length  $c$ . Then, for some  $L$ , we define

$$d_L^Y(C, \tilde{C}) = \sum_{c=1}^L KL(Q^{Y_c} || \tilde{Q}^{Y_c}) + KL(\tilde{Q}^{Y_c} || Q^{Y_c}).$$

We expect that the longer a pair of players reciprocate, the higher the likelihood that they will continue doing so. The exact probabilities can be obtained from human data.

### Autonomous Agents

In this section, we describe various agent models of behavior. We first apply traditional game-theoretic analysis to the Social Ultimatum Game to derive the “optimal” behavior under rational actor assumptions. We then describe two distribution-based agents that do not model other agents but are capable of incorporating human behavior data. Finally, we describe an adaptive agent that incorporates some aspects of human behavior such as fairness and reciprocity.

#### Game-Theoretic Agents

In previous work, we analyzed the behavior of game-theoretic agents in the multi-player, multi-round Social Ultimatum Game. We showed that even though accepting all non-zero offers is not a dominant strategy, the Nash equilibrium from the one-shot, two-agent Ultimatum Game still translates to a similar equilibrium strategy in the Social Ultimatum Game.

Let strategies be characterized by the statistics that they produce in steady-state: the distribution of offers made by each player, where  $p_m^g(n, q)$  denotes the likelihood that  $P_m$  will give an offer of  $q$  to  $P_n$ , and the distribution of offers accepted by each player, where  $p_m^a(n, q)$  denotes the likelihood that  $P_m$  will accept an offer of  $q$  from  $P_n$ .

**Proposition 1** *In the Social Ultimatum Game, accepting all offers is not a dominant strategy.*

**Proposition 2** *In the Social Ultimatum Game, Nash equilibrium outcomes only happen when players employ strategies of the form “greedy” strategies, where*

$$p_m^g(n, q) = 0, \forall q > \delta, m, n, \quad p_m^a(n, \delta) = 1, \forall m, n, \quad (3)$$

*i.e., “greedy” strategies where players only make the minimum offers of  $\delta$ , and all players accept all minimum offers.*

Proof details are provided in (XXXXXX, 2011).

### Distribution-Based Agents

One way to create agents that satisfy a set of metrics is to use the metrics to generate the agent behavior. Using only time-collapsed metrics, one could create a distribution-based agent (DBA) as follows. Learn distributions of offer value, target recipient and rejection percentage from human data. Find the appropriate target-recipient distribution based on number of participants and assign agents to each position (i.e., most likely to least likely). In offer phases of each round, choose a target by sampling from the target-recipient distribution and an offer value by sampling from the offer distribution. For received offers, decide via Bernoulli trial based on the rejection percentage for that offer value.

The DBA has no notion of reciprocity. We also investigated a class of distribution-based reciprocal agents (DBRA) which behave like the DBA agents in all aspects other than target selection. If DBRA agents receive an offer it will decide to reciprocate based on a reciprocation percentage that is learned from human data. If multiple offers are received, the target is chosen using a relative likelihood based on the target-recipient distribution. Similarly, if it doesn’t receive any offers, it uses the target-recipient distribution. While the distribution-based agents act on the basis of data of human play, they do not have models of other agents and consequently execute an open-loop static policy. The following model introduces an adaptive model that is not based simply on fitting the metrics.

#### Adaptive Agents

In order to create adaptive agent models of human players for the Social Ultimatum Game, we need to incorporate some axioms of human behavior that may be considered “irrational”. The desiderata that we address include assumptions that people will (1) start with some notion of a fair offer, (2) adapt these notions over time at various rates based upon their interactions, (3) have models of other agents, (4) choose the best option while occasionally exploring for better deals. Each player  $P_m$  is characterized by three parameters:  $\alpha_m^0$  :  $P_m$ ’s initial acceptance threshold,  $\beta_m$  :  $P_m$ ’s reactivity and  $\gamma_m$  :  $P_m$ ’s exploration likelihood

The value of  $\alpha_m^0 \in [0, e]$  is  $P_m$ ’s initial notion of what constitutes a “fair” offer and is used to determine whether an offer to  $P_m$ , i.e.,  $q_{n,m}^k$ , is accepted or rejected. The value of  $\beta_m \in [0, 1]$  determines how quickly the player will adapt to information during the game, where zero indicates a player who will not change anything from their initial beliefs and one indicates a player who will solely use the last data point. The value of  $\gamma_m \in [0, 1]$  indicates how much a player will deviate from their “best” play in order to discover new opportunities where zero indicates a player who never deviates and one indicates a player who always does.

Each player  $P_m$  keeps a model of other players in order to determine which player to make an offer to, and how much that offer should be. The model is composed as follows:  $a_{m,n}^k$  :  $P_m$ ’s estimate of  $P_n$ ’s acceptance threshold;  $\bar{a}_{m,n}^k$  : Upper bound on  $a_{m,n}^k$ ; and  $\underline{a}_{m,n}^k$  : Lower bound on  $a_{m,n}^k$ . Thus,  $P_m$  has a collection of models for all other players  $\{[\underline{a}_{m,n}^k, a_{m,n}^k, \bar{a}_{m,n}^k]\}_n$  for each round  $k$ . The value  $a_{m,n}$  is

the  $P_m$ 's estimate about the value of  $P_n$ 's acceptance threshold, while  $\underline{a}_{m,n}^k$  and  $\bar{a}_{m,n}^k$  represent the interval of uncertainty over which the estimate could exist. For simplicity, we will assume that  $\delta = 1$ .

**Making Offers** In each round  $k$ ,  $P_m$  may choose to make the best known offer, denoted  $\tilde{q}_m^k$ , or explore to find someone that may accept a lower offer. If there are no gains to be made from exploring, i.e., the best offer is the minimum offer ( $\tilde{q}_m^k = \delta = 1$ ), a player will not explore. However, if there are gains to be made from exploring, with probability  $\gamma_m$ ,  $P_m$  chooses a target  $P_n$  at random and offers them  $q_{m,n}^k = \tilde{q}_m^k - 1$ . With probability  $1 - \gamma_m$ ,  $P_m$  will choose to exploit. The target is chosen from the players who have the lowest value for offers they would accept, and the offer is that value:

$$q_{m,n}^k = \lceil a_{m,n}^k - \epsilon \rceil \text{ where } \epsilon \in \arg \min_{n \neq m} \lceil a_{m,n}^k \rceil \quad (4)$$

The previous equation characterizes an equivalence class of players from which  $P_m$  can choose a target agent. The  $\epsilon$  parameter is used to counter boundary effects in the threshold update, discussed below. The target agent from the equivalence class is chosen using *proportional reciprocity*, by assigning likelihoods to each agent with respect to offers made in some history window.

**Accepting Offers** For each offer  $q_{m,n}^k$ , the receiving player  $P_n$  has to make a decision  $d_{m,n}^k \in \{0, 1\}$  to accept or reject it, based on its threshold:

$$\text{If } q_{m,n}^k \geq \lceil \alpha_m^k - \epsilon \rceil, \text{ then } d_{m,n}^k = 1, \text{ else } d_{m,n}^k = 0 \quad (5)$$

**Updating Acceptance Threshold** The acceptance threshold is a characterization of what the agent considers a "fair" offer. Once an agent is embedded within a community of players, the agent may change what they consider a "fair" offer based on the received offers. We model this adaptation using a convex combination of the current threshold and the offers that are received, with adaptation parameter  $\beta_m$ . Let the set of offers that are received be defined as:  $R_m^k = \{q_{i,j}^k : j = m, q_{i,j}^k > 0\}$ . If  $|R_m^k| \geq 1$ , then  $\alpha_m^{k+1} =$

$$(1 - \beta_m)^{|R_m^k|} \alpha_m^k + \frac{(1 - ((1 - \beta_m)^{|R_m^k|})}{|R_m^k|} \sum_i q_{i,m}^k \quad (6)$$

If  $|R_m^k| = 0$ , then  $\alpha_m^{k+1} = \alpha_m^k$ . Thus, offers higher than your expectation will raise your expectation and offers lower than your expectation will lower your expectation at some rate.

**Updating Threshold Estimate Bounds** As a player makes an offer  $q_{m,n}^k$  and receives feedback on the offer  $d_{m,n}^k$ , they learn about  $P_n$ 's acceptance threshold. Using this information, we can update our bounds for our estimates of their threshold. The details can be found in an extended version of this paper.

**Updating Threshold Estimates** Once the threshold bounds are updated, we can modify our estimates of the thresholds as follows. If the player accepts the offer, we move the estimate of their threshold closer to the lower

bound and if the player rejects the offer, we move our estimate of their threshold closer to the upper bound using a convex combination of the current value and the appropriate bound as follows.

$$\begin{aligned} d_{m,n}^k = 1 &\Rightarrow \\ a_{m,n}^{k+1} &= \min\{\beta_m \underline{a}_{m,n}^{k+1} + (1 - \beta_m) a_{m,n}^k, \bar{a}_{m,n}^{k+1}\} \quad (7) \\ d_{m,n}^k = 0 &\Rightarrow \\ a_{m,n}^{k+1} &= \max\{\beta_m \bar{a}_{m,n}^{k+1} + (1 - \beta_m) a_{m,n}^k, \underline{a}_{m,n}^{k+1} + 2\epsilon\} \quad (8) \end{aligned}$$

The *min* and *max* operators ensure that we don't make unintuitive offers (such as repeating a just rejected offer), if our adaptation rate is not sufficiently high. The adaptive agent described above fulfills the properties of the desiderata prescribed to generate behavior that is more aligned with our expectations in reality.

## Experiments

Data was collected from human subjects recruited from undergraduates and staff at the Univ. of (XXXXX). In each round, every player is given the opportunity to propose a \$10 split with another player of their choosing. Games ranged from 20 to 50 rounds. A conversion rate of 10 ultimatum dollars to 25 U.S. cents was used to pay participants, i.e., \$5 per 20 rounds per player in an egalitarian social-welfare maximizing game. The subjects participated in organized game sessions and a typical subject played three to five games in one session. Between three and seven players participated in each game. During each session, the players interacted with each other exclusively through the game's interface on provided iPads, shown in Figure 1. We have collected data from 27 human subject games thus far. In this paper, we focus on the seven 5-person games in the dataset. By restricting our attention to five-player games, we avoid biases that may be introduced if we attempted to normalize the data from the other games to reflect a five-person composition. Analysis on the games of other sizes yields similar results.

To create the Distribution-Based Agent and Distribution-Based Reciprocal Agent to the collected data, we calculated the appropriate distributions (offer value, rejection percentage by value, targeted-recipient), by counting and averaging over all games and all players. For the Adaptive Agents, we analyzed the traces of each game, and estimated game-specific  $\alpha, \beta, \gamma$  parameters of each of the participating players, as follows. For each player  $P_m$  in the game,

- $\alpha_m$  : This is set as the player's first offer in this game.
- $\beta_m$  : When a player decreases his offer to a specific player from  $q_1$  to  $q_2$  after  $K$  steps (not necessarily consecutive), we find and store the best  $\beta$  value such that  $K$  applications of  $\beta q_2 + (1 - \beta) q_1$  yields a result less than  $\frac{q_1 + q_2}{2}$  (so that the next offer should be closer to  $q_2$  than it is to  $q_1$ ). We then take  $\beta_m$  to be this stored  $\beta$  value.
- $\gamma_m$  : This is the likelihood that a player's offer is less than the minimum known accepted offer, where the minimum accepted offer at a given round  $k$  is the minimum offer known to be accepted by any player at time  $k - 1$ .

Having estimated the population parameters of each game, we then use them as input to create an autonomous agent for each player, and simulate each game ten times to produce ten traces. Within each of these games, each of the five players uses the parameters corresponding to one of the five original human players.

## Evaluation

These experiments and simulations result in a collection of game traces for each of the five types of agent discussed: Human, Adaptive, DBA, DBRA, and Game-theoretic (GT). Table 1 shows the similarity between the collection of human traces and each of the four collections of autonomous agent traces, according to the metrics discussed earlier.

	Adaptive	DBA	DBRA	GT
$d^O$	0.57	0.008	0.008	33.26
$d^R$	0.21	0.0005	0.01	0.19
$d^B$	11.74	0.008	0.11	32.83
$d^{Y_s}$	4.22	16.34	20.10	97.02

Table 1: Similarity to human play, based on various metrics.

Note that we do not normalize the metrics to fall within the same range because the relative importance of the metrics is unclear. Thus, the similarity between behaviors should be viewed independently for each metric.

The DBA and DBRA agents score very well on the three metrics based on offer value, rejection percentage, and target-recipient. We fully expect this result as both these agents generate their behavior by sampling from these distributions. It is also clear that the GT agent performs very differently from the human data, based on most of the metrics. Naturally, the Adaptive Agent scores worse than the distribution-based agents on the temporally-independent distribution metrics  $d^O$ ,  $d^R$ , and  $d^B$ , since it does not explicitly optimize for these metrics, but its behavior is still relatively close to human behavior. On the temporally-dependent reciprocation-chain metric  $d^{Y_s}$ , the Adaptive Agent scores much better in similarity to the human traces.

To get a more intuitive sense of the differences in the trace data, we also display the actual distributions that underlie the metrics in Figure 2 which shows the distributions of offer amounts for each of the agent types, the probability of rejection given each offer amount, the distribution of offer recipients, ordered from most likely to least likely, and the probability that an offer will be reciprocated, given that a chain of  $c$  offers have been made between the players in the past  $c = 1, 2, \dots, 8$  time periods.

While the Adaptive Agent may not have been the most human-like agent according to the other three metrics, the form of its distributions still reasonably resembled the distributions produced by human play. However, on the time-dependent reciprocation-based metric, it is very clear that the Adaptive Agent is the only one that exhibits behavior that is similar to human play. This temporal dependence is crucial to creating agent behavior that emulates human behavior.

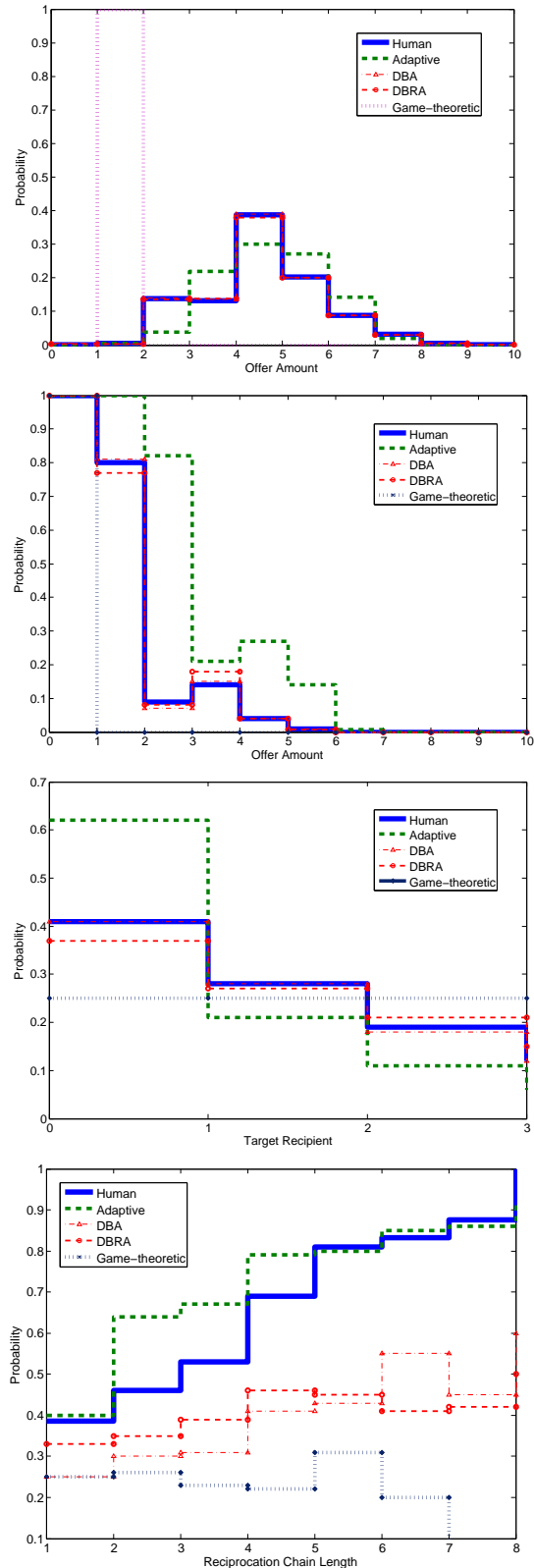


Figure 2: (Top to bottom) Distribution of offer amounts; Rejection probabilities given offer amounts; Target recipient distribution; Reciprocation probabilities given a chain of reciprocation of length  $c = 1, 2, \dots, 8$ .

Finally, note that the aim of the paper is to develop this set of metrics and discuss the importance of time-dependence in constructing and choosing metrics. Thus, we do not perform validation of the learned agents, since they are meant as illustrations to show the pros and cons of the metrics rather than as algorithmic results themselves.

## Related Work

Our choice to investigate the Ultimatum Game was motivated by its long history in the field and the fact that it is a leading example of where game-theoretic reasoning fails to predict consistent human behaviors (Oosterbeek, Sloof, and van de Kuilen 2004; Henrich, Heine, and Norenzayan 2010). Economists and sociologists have proposed many variants and contexts of the Ultimatum Game that seek to address the divergence between the “rational” Nash equilibrium strategy and observed human behavior, for example, examining the game when played in different cultures, with members of different communities, where individuals are replaced by groups, where the players are autistic, and when one of the players is a computer. Interestingly, isolated non-industrialized cultures, people who have studied economics, groups, autists, and playing with a computer all tend to lead to less cooperative behavior (Oosterbeek, Sloof, and van de Kuilen 2004; Mascha van’t Wout and Aleman 2006; Hill and Sally 2002; Carnevale 1997; Frank, Gilovich, and Regan 1993). Learning human game data is a promising approach for quickly learning realistic models of behavior. In the paper, we have demonstrated this approach in SUG, and proposed metrics that evaluate the similarity between autonomous agents’ game traces and human game traces.

Recently, there has also been other work attempting to model human behavior in multi-agent scenarios, primarily in social network and other domains modeled by graphical relationship structures (Judd, Kearns, and Vorobeychik 2010). In contrast, our work focuses on multi-agent situations where motivated agents make sequential decisions, thus requiring models that include some consideration of utilities and their interplay with psychological effects. Our Adaptive Agent is a simple model, with parameters that are fit to the collected data, that demonstrates this approach.

Finally, a critical aspect of this line of work must include the development of appropriate metrics for evaluating the verisimilitude of the autonomous agent behaviors to human behavior. While there is a long literature on time-series metrics (Mitsa 2010), in this paper, we show that these metrics do not capture the temporal causality patterns that are key to evaluating human behaviors, and thus are insufficient to evaluate agent behaviors when used alone.

## Conclusion

Our goal is to develop approaches to create autonomous agents that replicate human behavior in multi-player games. We begin with a simple abstract game, the Social Ultimatum Game, which captures many aspects of the sequential decision-making in such games. To create and evaluate these agents, one needs appropriate metrics to characterize the deviations from the source behavior. The challenge is that a

single source behavior in dynamic environments produces not a single decision but instead multiple traces where each trace is a sequence of decisions. Thus, the challenge is to find a way to compare collections of traces.

We developed time-collapsed and time-dependent metrics to evaluate such collections. We showed that agents built on time-collapsed metrics can miss key characteristics of human play, in particular an accurate model of temporal reciprocity. While our adaptive agent was able to perform closer to this metric, the key is the identification of time-dependent metrics as a key factor in evaluating emulation agents. This also has implications on the type of agent model necessary to have as a substrate upon which one can learn from human data.

Going forward, we will consider more complex games and potential corresponding models. We will require both general, parameterized models that can be learned from data, as well as more formal methods for constructing appropriate temporal metrics to automatically evaluate the realism of the learned behaviors.

**Acknowledgements.** This material is based upon work supported by the AFOSR Award No. FA9550-10-1-0569.

## References

- Carnevale, C. R. P. J. 1997. Group choice in ultimatum bargaining. *Organizational Behavior and Human Decision Processes* 72(2):256–279.
- Frank, R. H.; Gilovich, T.; and Regan, D. T. 1993. Does studying economics inhibit cooperation? *The Journal of Economic Perspectives* 7(2):159–171.
- Henrich, J.; Heine, S. J.; and Norenzayan, A. 2010. The weirdest people in the world? *Behavioral and Brain Sciences* 33(2-3):61–83.
- Hill, E., and Sally, D. 2002. Dilemmas and bargains: Theory of mind, cooperation and fairness. Working paper, University College, London.
- Judd, S.; Kearns, M.; and Vorobeychik, Y. 2010. Behavioral dynamics and influence in networked coloring and consensus. In *Proceedings of the National Academy of Science*.
- Mao, W., and Gratch, J. 2004. Social judgment in multiagent interactions. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 1*, AAMAS '04, 210–217. Washington, DC, USA: IEEE Computer Society.
- Mascha van’t Wout, René S. Kahn, A. G. S., and Aleman, A. 2006. Affective state and decision-making in the ultimatum game. *Experimental Brain Research* 169(4):564–568.
- Mitsa, T. 2010. *Temporal Data Mining*. CRC Press.
- Oosterbeek, H.; Sloof, R.; and van de Kuilen, G. 2004. Differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental Economics* 7:171–188.
- Rickel, J. 2001. Intelligent virtual agents for education and training: Opportunities and challenges. In de Antonio, A.; Aylett, R.; and Ballin, D., eds., *Intelligent Virtual Agents*, volume 2190 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg. 15–22.