Evolving Cooperation in Complex Behavioral Interactions through Reputation

Siang Yew Chong and Xin Yao
The Centre of Excellence for Research in
Computational Intelligence and Applications (CERCIA)
School of Computer Science
The University of Birmingham, UK.
E-mails: S.Y.Chong@cs.bham.ac.uk, X.Yao@cs.bham.ac.uk

The iterated prisoner's dilemma (IPD) has long been used to study the conditions that promote cooperative behaviors among selfish individuals. In particular, studies using the co-evolutionary learning framework have shown that cooperative behaviors can be learned through a process of adaptation of strategies based solely on direct interactions through repeated encounters in playing IPD. However, complex behavioral interactions (e.g., by humans) may involve more than just direct interactions between two individuals. In many real-world situations, it is impossible to interact with all other individuals. It is very important to study indirect interactions, e.g., in the form of estimating behaviors of future partners through their perceived reputation based on previous behaviors to other individuals. Here, we study the co-evolutionary learning of IPD with reputation where behavioral interactions for a pair of strategies depend not only on choices made in their previous moves, but also choices made to other strategies that are reflected by their reputation scores. We show that for more complex IPD interactions involving more choices where direct interaction alone is ineffective to promote cooperation, the addition of reputation helps to promote cooperation. We further show that different implementations of reputation estimation, which reflect the accuracy of reputation estimation from using memory of games from previous generation and more frequent updating of reputation scores, is for the evolution of cooperation. Finally, we also investigate the situation where strategies can misperceive their partner's reputation and show that even in the circumstances that strategies misperceive reputation, evolution to cooperation is still possible. This implies that reputation helps to increase the robustness of co-evolutionary learning of cooperative strategies.

1 Introduction

One particular interest in the study of behavioral interactions is to understand the specific conditions that can promote cooperation among selfish individuals. Here, the abstract mathematical game of IPD has long been used as the metaphor to explain why selfish individuals cooperate, whether in the context of social, economic, or biological interactions [1]. In the IPD game, two players engaged in repeated interactions are given two choices, cooperate and defect [1]. The dilemma, i.e., both players who are better off mutually cooperating than mutually defecting are vulnerable to exploitation by one of the party who defects, embodies the tension between rationality of individuals who are tempted to defect and rationality of the group where every individual is rewarded by mutual cooperation [2].

Many studies have shown that defection is not necessarily the best choice of play, and that cooperative play can be a viable strategy [2, 3]. In particular, using a co-evolutionary learning framework, it has been shown that cooperative behaviors can be learned from an initial, random population using evolutionary algorithms [4, 5, 6, 7, 8]. Unlike the classical evolutionary games approach that considers frequency dependent reproductions of predetermined strategies (e.g., "ecological approach" [2, 9]), the co-evolutionary learning approach has the added advantage in that it allows the study of how and why certain strategies can be learned through an adaptation process based on interactions by game-play.

Although earlier studies [4, 5, 6, 7, 8, 10] show that cooperation can be easily evolved for the classical *two*-choice IPD, the same cannot be said for the case when the interactions involved are more complex. For example, when there are more that two choices available for play, studies in [11, 12, 13, 14] shows that evolution of cooperation is more difficult. Their studies generally show that cooperation can still be evolved in the IPD with more choices, although the evolution of cooperation is unstable [11] and more difficult to achieve [12, 13, 14]. As such, this raises the question as to whether cooperative behaviors can still be learned in situations involving more complex interactions when the only information used to guide the adaptation process of strategy behaviors is from the outcomes of these interactions.

However, it is known that cooperation can be achieved in complex human interactions that are not limited to just direct interactions [15]. This may explain the limitation with the current co-evolutionary learning framework on IPD games that only involves direct interactions, i.e., the mechanism of direct reciprocity, i.e., repeated encounters, is less effective in promoting cooperation in more complex intractions that may involve having more choices to play. As such, we investigate one solution that uses the idea that cooperation between two players depend on prior interactions with other players, i.e., indirect reciprocity, where the mechanism for evolving cooperative behaviors is reputation [16]. Here, we incorporate reputation into the existing co-evolutionary model for IPD games. This is different from previous studies whereby strategies consider only previous moves as inputs [12, 13, 14] or reputation for decision-making only [16]. For our study, strategies consider both previous moves and reputation for input.

Following our initial study in [17], a co-evolutionary learning approach is used to show why and how the addition of reputation in the IPD game makes the evolution to cooperation more likely. Our experiments show that cooperative outcomes are obtained when strategies evolve to maintain high reputation scores, which in turn, requires highly cooperative play. One such play is the "discriminatory" play, i.e., fully cooperate with strategies that have good reputation (e.g., their reputation scores are similar or higher), otherwise fully defect against strategies with bad reputation (e.g., their reputation scores are lower).

We also investigate the impact of different implementations of reputation estimation on the evolution of cooperation. Experiments are carried out to determine how different methods for calculating reputation scores, which is based on different interpretations on how a strategy's reputation is estimated by other strategies, affect behaviors of future interactions. We focus on the impact of accurate reputation estimation on the evolutonary outcome. We show that accuracy of reputation estimation depends on how memory of games from previous generations is incorporated to calculate reputation scores, and how frequently reputation scores are updated. Increasing accuracy of reputation estimation has a significant and positive impact on the evolution of cooperation.

2 Evolution to Defection in More Complex Interactions

2.1 More Complex IPD Interactions

In the classical IPD [1], each player has two choices: cooperation and defection. The payoff a player receives depends on a payoff matrix (Fig. 1) that must satisfy the following three conditions:

1. $T > R$ and $P > S$ (Defection always pays more),
2. $R > P$ (Mutual cooperation beats mutual defection), and
3. $R > (S + T)/2$ (Alternating does not pay).

There are many possible values for T, R, P, and S that satisfy the above three conditions. Here, we use $T = 5$, $R = 4$, $P = 1$, and $S = 0$ (Fig. 2).

However, the IPD game can be easily extended with more than just two extreme choices to allow for subtle interactions involving intermediate choices [11, 12, 13, 14]. Here, we extend the the classical IPD to the more complex IPD with multiple discrete levels of cooperation following our previous work in [12, 13, 14, 17]. The n-choice IPD is based on a simple linear interpolation of the classic *two*-choice IPD using the following equation [14]:

$$p_A = 2.5 - 0.5c_A + 2c_B, \quad -1 \leq c_A, c_B \leq 1,$$

	Cooperate	Defect
Cooperate	R R	T S
Defect	S T	P P

Figure 1: The payoff matrix for the two-player IPD. The payoff listed in the lower left-hand corner is assigned to the player choosing the move.

	Cooperate	Defect
Cooperate	4 4	5 0
Defect	0 5	1 1

Figure 2: The payoff matrix for the two-player IPD used in this paper. $T = 5$, $R = 4$, $P = 1$, and $S = 0$.

where p_A is the payoff to player A, given that c_A and c_B are the cooperation levels of the choices that players A and B make, respectively.

In generating the payoff matrix for the n-choice IPD, the following conditions must be satisfied:

1. For $c_A < c'_A$ and constant c_B: $p_A(c_A, c_B) > p_A(c'_A, c_B)$,
2. For $c_A \leq c'_A$ and $c_B < c'_B$: $p_A(c_A, c_B) < p_A(c'_A, c'_B)$, and
3. For $c_A < c'_A$ and $c_B < c'_B$: $p_A(c'_A, c'_B) > \frac{1}{2}(p_A(c_A, c'_B) + p_A(c'_A, c_B))$.

One can observe that the above conditions are analogous to those for the *two*-choice IPD's. For example, the first condition ensures that defection always pays more. The second condition ensures that mutual cooperation has a higher payoff than mutual defection. The third condition ensures that alternating between cooperation and defection does not pay in comparison to just playing cooperation.

Given the payoff equation and the three conditions above, an n-choice IPD can be formulated [12]. For example, figure 3 shows the payoff matrix for a *four*-choice IPD. Figure 3 also illustrates two important points. First, the payoffs in the four corners of an n-choice IPD payoff matrix are the same as those in the *two*-choice IPD. Second, any 2×2 sub-matrix of the $n \times n$ matrix of the n-choice IPD is itself a *two*-choice IPD.

2.2 Strategy Representation

We use a fixed-architecture feedforward neural networks in all the experiments for a more direct comparison with the study in [17]. For the experiments that consider the n-choice IPD without reputation, we use the same neural network specified in [18]. As in [17, 18], we consider deterministic, reactive, memory-one strategies (i.e., look back the one previous move only) for simplicity and also more direct comparisons. The neural network is a multilayer perceptron with four input nodes, ten hidden nodes in the only hidden layer, and one output node. The neural network then has a total of $N_w = 61$ connection

		PLAYER B			
		$+1$	$+\frac{1}{3}$	$-\frac{1}{3}$	-1
	$+1$	4	$2\frac{2}{3}$	$1\frac{1}{3}$	0
PLAYER	$+\frac{1}{3}$	$4\frac{1}{3}$	3	$1\frac{2}{3}$	$\frac{1}{3}$
A	$-\frac{1}{3}$	$4\frac{2}{3}$	$3\frac{1}{3}$	2	$\frac{2}{3}$
	-1	5	$3\frac{2}{3}$	$2\frac{1}{3}$	1

Figure 3: The payoff matrix for the two-player *four*-choice IPD used in this paper. Each element of the matrix gives the payoff for Player A [12].

weights, i.e., [(4 input weights + 1 bias) × (10 input nodes)] + [(10 connection weights + 1 bias) × (1 output node)].

For the experiments that consider the n-choice IPD with reputation, an additional input node is added. The resulting neural network is similar to that used in [17]. Altogether, there are $N_w = 71$ connection weights, i.e., [(5 input weights + 1 bias) × (10 input nodes)] + [(10 connection weights + 1 bias) × (1 output node)]. Activation function used in the nodes is a *tanh* scaled to outputs values between $+1$ and -1. The five input nodes take in the following values:

1. The neural network's previous choice, i.e., level of cooperation, in $[-1, +1]$.
2. The opponent's previous level of cooperation
3. An input of $+1$ if the opponent played a lower cooperation level compared to the neural network, and 0 otherwise.
4. An input of $+1$ if the neural network played a lower cooperation level compared to the opponent, and 0 otherwise.
5. An input of $+1$ if the opponent has an equal or higher reputation score compared to the neural network's, and -1 otherwise.

Note that during the reputation estimation stage, the fifth input node takes in a 0 value. In addition to the neural network, each strategy also stores two values used as pre-game inputs for a memory-one IPD strategy.

2.3 Co-evolutionary Learning of the IPD with More Choices

As summarized by Yao [19], real-valued weights of a neural network can be evolved using self-adaptive evolutionary algorithms, such as evolutionary programming [20], which can be implemented as a co-evolutionary procedure when fitnesses depend on interactions between members of the population [21]. The co-evolutionary model used in the experiment to investigate why more choices lead to defection outcomes in the IPD with more choices is detailed as follows:

1. Generation step, $t = 0$:
   Initialize $N/2$ parent strategies, $P_i, i = 1, 2, ..., N/2$, randomly. For each strategy i:
   (a) Weights and biases, $[w_i(j)]$, are initialized in the range of $[-1, 1]$.
   (b) Each component for the self-adaptive parameter vector, $[\sigma_i(j)]$, is intitialized to 0.05 for consistency with the initialization range of connection weights.
   (c) The two pre-game inputs are initialized randomly and can take only the n-choices value uniformly distributed in that range of $[-1, 1]$.

2. Generate $N/2$ offspring, $O_i, i = 1, 2, ..., N/2$, from $N/2$ parents. For each offspring, O_i ($[w'_i(j)]$ and $[\sigma'_i(j)]$), generated from parent, P_i ($[w_i(j)]$ and $[\sigma_i(j)]$) through a self-adaptative Gaussian mutation:

   $$\sigma'_i(j) = \sigma_i(j) * exp(\tau * N_j(0,1)), i = 1, \ldots, N/2; j = 1, \ldots, N_w,$$
   $$w'_i(j) = w_i(j) + \sigma'_i(j) * N_j(0,1), i = 1, \ldots, N/2; j = 1, \ldots, N_w,$$

   where $N_w = 61$, $\tau = (2(N_w)^{0.5})^{-0.5} = 0.2530$, and $N_j(0,1)$ is a Gaussian random variable (zero mean and standard deviation of one) resampled for every j. N_w is the total number of weights, and biases for the neural network. Each of the two pre-game inputs is mutated separately. Mutation is of the form of adding the original value where step values have a Gaussian distribution (e.g., $N_j(0,1)$).

3. Strategies compete with each other in a round-robin tournament, including itself. For N strategies in a population, every strategy competes a total of N games.

4. Select the best $N/2$ strategies based on total payoffs of all games played. Increment generation step, $t = t + 1$.

5. Step 2 to 4 are repeated until the termination criterion (i.e., a fixed number of generation) is met.

In all the experiments, the population co-evolves for 600 generations, which is sufficiently long to observe evolutionary results (e.g., persistent periods of cooperation or defection). Each experiment is repeated for 30 independent runs.

2.4 Evolution to Defection

Co-evolutionary learning of cooperative strategies require a sufficiently long duration of repeated encounters. However, such a condition may not be satisfied in complex real-world interactions. Furthermore, these interactions may involve the situation where there are more choices to play than the duration of interaction, which makes it not possible to evaluate and react to partner's behavior from all available choices. As such, these strategies are more likely to play only certain choices because other choices are not sampled (i.e., learned) during interactions in the evolutionary process.

We investigate the co-evolutionary learning of strategies involving these complex interactions. We first limit the duration of a game to only ten rounds. Then, we study two experiments. The first experiment uses the 64-choice IPD, where the number of choices is larger than the number used for game duration. The second experiment uses the *four*-choice IPD, where the number of choices is less than game duration.

Figure 4 and table 1 summarized results of the experiments. From the figure, it can be observed that for both experiments, the population did not evolve to mutual cooperation since the average payoff would be four in this case. For the *four*-choice IPD case, the population evolved to play more intermediate choices (table 1). A further increase in the number of choices to play, e.g., the 64-choice IPD case, population evolve to play defection, which can be observed from having a significantly lower average payoff (Fig. 4) and more number of runs where the final evolved population played defection (table 1).

Table 1: Comparison of results from the experiments that used the IPD with four choices and 64 choices. "$No \leq 1.5$" indicates the number of runs where the population evolved to play defection. "$1.5 < No < 3.5$" indicates the number of runs where the population evolved to play intermediate choices. "$No > 3.5\%$" indicates the number of runs where the population evolved to play cooperation.

Choices	$No \leq 1.5$	$1.5 < No < 3.5$	$No \geq 3.5$
4	9	13	8
64	25	4	1

The results from the experiment is not surprising as one would expect that for the co-evolutionary learning of IPD strategies based only on short, direct interactions involving a larger number of choices to play, strategies learned to play defection rather than cooperation. This is because when strategies

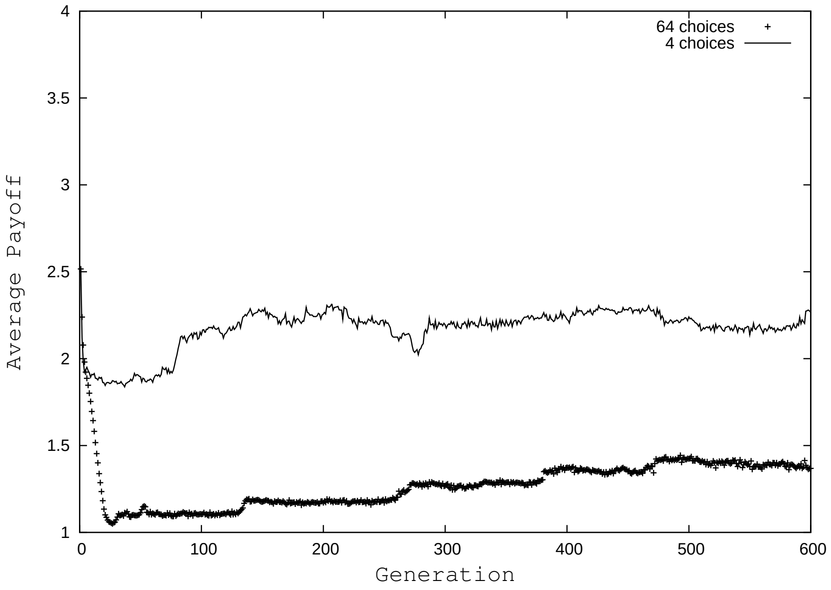

Figure 4: Comparison of the average payoff of 30 runs over 600 generations between the IPD games that consider four choices and 64 choices.

have more alternative choices to play, they are provided with more opportunity to exploit others by exploring lower cooperation levels play, leading to the situation whereby there may not be sufficient time (e.g., game duration) for strategies to engage in highly cooperative plays. As such, one would expect to observe that for the case of the 64-choice IPD, strategies are less likely to evolve to highly cooperative play compared to the case of the *four*-choice IPD because cooperative behaviors are less likely to be sampled since they do not contribute to higher payoffs when interactions are short.

3 Cooperating in the Iterated Prisoner's Dilemma through Reputation

3.1 Motivation

In the co-evolutionary learning approach, strategies can learn cooperative behaviors through interactions with other cooperative strategies, i.e., direct reciprocity allows for the evolution of cooperation. However, for more complex interactions, it was shown earlier that strategies are less able to learn cooperative behaviors, which seems to suggest a failure of the mechanism of direct reciprocity. Given that complex human interactions are not limited to direct interactions only [15], the co-evolutionary learning of IPD framework may require further extension to model complex behavioral interactions, and allows strategies to learn cooperative behaviors without relying entirely on direct interactions.

One mechanism that allows the evolution of cooperation, but does not require direct interactions (through repeated encounters), is reputation [16]. Reputation is argued and shown to be the mechanism for indirect reciprocity to occur [16, 22]. Indirect reciprocity is understood to occur when cooperation between current two individuals depends on their prior behaviors to others. With the mechanism of indirect reciprocity, an individual receives cooperation from third parties due to the individual's cooperative behaviors to others.

Nowak and Sigmund [16] studied how cooperative behaviors can be evolved through indirect reciprocity using image scoring strategies. The two-player interaction is modelled as a single-round donation game. Strategies are randomly selected to be donors and recipients. When strategies are selected to be donors, they are given two choices: cooperate (help) and defect (not help). For any donor-recipient pair, the donor pays a cost of c if it cooperates, while the recipient receives the benefit b, with $b > c$. If the donor defects, both donor and recipient receives zero payoff.

Each strategy is given a value for reputation that is known to the strategy and its opponents. There are two ways for calculating reputation: image scores [16] and standing [22]. For both methods, an individual's reputation score increases if the donor cooperates. The two methods differ in terms of how reputation score is decreased when the donor defects. For image score, the reputation score decreases whenever the donor defects [16]. For standing, the reputation score only decreases when the defection is unjustified (e.g., the recipient has a good reputation score). Otherwise, the reputation score remains unchanged [22].

A strategy's behavioral response depends on the reputation score and some additional parameters used for a threshold decision-making (i.e., the strategy representation is a set of values that includes its reputation score and parameters for thresholding). No previous moves are used. Depending on how strategies use information from reputation scores, there are three broad classes of strategies [22]. First, a strategy can consider its own score only for decision-making (e.g., "offer help when own score is less than h"). Second, a strategy can consider the recipient's score for decision-making (e.g., "offer help when recipient's score is at least k"). Third, a strategy can consider both its own and the recipient's scores (e.g., "offer help when own score is less than h and when recipient's score is at least k") [22].

Although earlier studies [15, 16, 22] focused only understanding how reputation contributes to evolution of cooperation through the mechanism of indirect reciprocity for simplicity, it is noted in [15] that complex interactions such as those found in human societies include both mechanisms of direct and indirect reciprocity, with the mechanism of indirect reciprocity occuring as a result of direct reciprocity between other individuals [15].

Here, we explore the idea of incorporating reputation in the IPD game that was traditionally modelled to study direct interactions. This study follows from our initial study that we first proposed in [17]. Our approach differs from previous studies [16, 22] on three fundamental aspects. First, in terms of the interaction, we use a different game compared to [16, 22]. Here, we consider the IPD game with multiple levels of cooperation. In [16, 22], the game is a single-round donation game with only two extreme choices. Second, our approach is different in terms of a strategy's behavioral response. We consider a strategy whose responses depend on both previous moves and reputation. In [16, 22], only reputation is used as inputs to a strategy's response. Third, we use a co-evolutionary model whereby strategies can adapt their behaviors over a range specified by the representation. Studies in [16, 22] did not consider a process of adaptation, e.g., strategies are predetermined and their replacement depends on the proportions of existing strategies in the population.

Although results from our initial study [17] indicate that evolution to cooperation is more likely with reputation, even for cases with short game durations, it was not known why and how reputation helps with evolving cooperative behaviors. Here, we will analyze the reasons behind reputation in promoting cooperation. In particular, we will study different implementation of reputation estimation and show that accuracy of estimation a strategy's reputation is important for evolution of cooperation and is related to games played in previous generation and also the frequency of updates of reputation scores. We will also study the situation where a strategy can misperceive the partner's reputation with a small probability and show that the mechanism is robust in promoting cooperation despite such an error.

3.2 How Reputation is Implemented

Following our earlier study in [17], we consider a binary reputation case, i.e., a strategy either has a good or bad reputation. Strategies decide whether their opponents's reputation by comparing their reputation scores. If the opponent has an equal or higher reputation score compared to the strategy's, then the strategy considers the opponent as having a good reputation. Otherwise, the strategy considers the opponent as having a bad reputation. By considering relative rather than absolute values for reputation, we simplify the neural network's learning task for reputation, i.e., it only needs to distinguish whether an opponent has a good or bad reputation as opposed to learning what specific values for reputation actually represent.

Since our strategy considers both previous moves and reputation for inputs, the strategy's reputation score has to be determined first. How a strategy's reputation can be estimated depends on how reputation is interpreted to reflect a strategy's possible behaviors to future opponents in light of choices it made with past opponents. Here, we consider two related issues as to how reputation should be implemented for accuracy in the estimation of an opponent's potential behavior. First, we consider memory of games played in previous generation in reflecting the strategy's behavior. Second, we consider how frequently reputation is updated in providing feedback to other strategies in the population.

The procedure for calculating reputation score is described as follows:

- Each choice is given a weight. Weights are assigned in the same way as in calculating the payoffs (e.g., for four choices, -1, $-1/3$, $+1/3$, $+1$).
- The number of times a choice is played by the strategy during interactions is recorded.
- Reputation score for each strategy is obtained by taking the sum of choices played, normalized over the total number of plays the strategy makes in all the random encounters for this stage.

A formal way to describe how reputation score is calculated for each strategy is to consider the distribution of choices the strategy plays that contributes to the strategy's reputation. If $f(x)$ gives the frequency distribution of the n choices, $x = \{x_1, x_2, ..., x_n\}$, played by a strategy, and that the value of x specifies the weight for each choice, then the reputation score, R, is given by:

$$R = \sum_{i=1}^{n} x_i * f(x_i). \tag{1}$$

The most important difference between our implementation and that of [17] is how reputation score is calculated. In [17], a strategy's reputation score is calculated based on the payoffs received. Here, a strategy's reputation score is calculated based on the choices that a strategy play. We believe this approach to be more accurate compared to the approach used in [17] because it better reflects the actual cooperativeness of strategies. In [17], the use of average payoffs as an indication of cooperativeness can be misleading, e.g., higher average payoff can be a result of mutual cooperation and not just one strategy exploiting the other.

As in other studies [], we also investigate whether the mechanism provided by reputation is robust to errors in promoting cooperation. In particular, we focus on error in perception, i.e., there is a small probability that a strategy misperceives the opponent's reputation. For example, if the opponent has a good reputation (based on the comparison of reputation scores with the strategy), there is a probability of 0.01 that the strategy perceives the opponent to have a bad reputation.

3.3 The Impact of Reputation on Evolution of Cooperation

We first examine a two-stage procedure that we first proposed in [17]. In the first stage, which we named the reputation estimation stage, a strategy's reputation score is calculated based on choices the strategy played in some random n-choices IPD games in the current generation. After obtaining reputation scores, strategies compete in the second stage of the n-choice IPD with reputation.

It should be noted that the approach here (and also [17]) may not be realistic due to two assumptions made on how reputation is estimated. First, it is assumed that there is no memory of games played in previous generations, i.e., a strategy's reputation score is always calculated anew every generation, and choices it made in the past has no impact on its current reputation. Second, it is assumed that a strategy reputation is static for the duration of its interactions with other strategies in the population until the next round of update, which occurs in the next generation. However, the approach does simplify analyzing the evolutionary dynamics of the game by separating the procedure for estimating reputation from the strategy interactions that involved both previous moves and reputation scores.

3.3.1 Co-evolutionary Model for the IPD with More Choices and Reputation

In the case of experiments whereby strategies also consider reputation in addition to previous moves, we used the following model:

1. Generation step, $t = 0$:
   Initialize $N/2$ parent strategies, $P_i, i = 1, 2, ..., N/2$, randomly. For each strategy i:
   (a) Weights and biases, $[w_i(j)]$, are initialized in the range of $[-1, 1]$.
   (b) Each component for the self-adaptive parameter vector, $[\sigma_i(j)]$, is intitialized to 0.05 for consistency with the initialization range of connection weights.
   (c) The two pre-game inputs are initialized randomly and can take only the n-choices value uniformly distributed in that range of $[-1, 1]$.

2. Generate $N/2$ offspring, $O_i, i = 1, 2, ..., N/2$, from $N/2$ parents. For each offspring, O_i ($[w'_i(j)]$ and $[\sigma'_i(j)]$), generated from parent, P_i ($[w_i(j)]$ and $[\sigma_i(j)]$) through self-adaptation:

$$\sigma'_i(j) = \sigma_i(j) * exp(\tau * N_j(0, 1)), i = 1, \ldots, N/2; j = 1, \ldots, N_w,$$
$$w'_i(j) = w_i(j) + \sigma'_i(j) * N_j(0, 1), i = 1, \ldots, N/2; j = 1, \ldots, N_w,$$

   where $N_w = 71$, $\tau = (2(N_w)^{0.5})^{-0.5} = 0.2436$, and $N_j(0, 1)$ is a Gaussian random variable (zero mean and standard deviation of one) resampled for every j. N_w is the total number of weights, and biases for the neural network. Each of the two pre-game inputs is mutated separately. Mutation is of the form of adding the original value where step values have a Gaussian distribution (e.g., $N_j(0, 1)$).

3. Reputation score, $R_{current}$, for each strategy in the population (parent and offspring), is initialized to 0. In addition, each strategy's frequency distribution of choices played, $f(x), x = \{x_1, x_2, ..., x_n\}$, is initialized to 0 for all x.

4. Strategies compete with each other in a two-stage game:
   (a) Stage 1 (Reputation Estimation):
      i. For every unique $(N^2 - N)/2$ pairwise interactions (games) between N strategies ($N/2$ offspring + $N/2$ parents), there is a fixed probability, p_r, a pair is sampled (without replacement) for this stage.
      ii. Interactions are in the form of n-choice IPD game.
      iii. For each strategy, reputation score is calculated as follows, after finish competing with all, randomly selected opponents:

$$R_{current} \quad = \quad \frac{\text{Sum of all choices made}}{\text{Total number of moves made}} \quad = \quad \sum_{i=1}^{n} x_i * f(x_i), \tag{2}$$

      which is obtained from equation 1.
      iv. Each strategy is also given IPD payoffs from the interactions.
   (b) Stage 2 (n-choice IPD with reputation):
      i. For the remaining pairs not sampled for Stage 1, interactions are in the form of n-choice IPD with reputation.
      ii. Each strategy is given IPD payoffs from the games.

5. Select the best $N/2$ strategies based on combined payoffs from Stage 1 and 2. Increment generation step, $t = t + 1$.

6. Step 2 to 5 are repeated until the termination criterion (i.e., a fixed number of generation) is met.

Comparing our co-evolutionary implementation with the earlier implementation used in [17], there are several noticeable differences. In terms of how reputation score of a strategy is calculated, our approach uses choices that a strategy plays, while in [17], reputation score is calculated based on payoffs received from playing a fixed number of random games. In addition, our approach also differs in how random games are selected. Here, we use a fixed probability to determine whether two strategies will participate in the reputation estimation stage. In [17], each strategy plays a fixed number of random games, with the implementation requiring random shuffling of strategies' positions in the population. In terms of co-evolutionary models, we evolved neural network weights directly, while in [17], neural networks were evolved using genetic algorithms (e.g., variation operators operate on neural networks coded as binary strings of 0s and 1s).

It should be noted that selection of strategies are based on combined payoffs of games from both stages (reputation estimation and n-choice IPD with reputation) of the two-stage procedure described earlier. In general, the probability for selecting random games for the reputation estimation stage, p_r, should be small and less than 0.5. This procedure is consistent with our motivation of explicitly modelling both mechanisms of direct and indirect reciprocity in the interaction, and how the mechanism of indirect reciprocity occurs as a result of direct reciprocity between other individuals [15]. We will discuss the implication of combining payoffs of games from the two stages later.

3.3.2 Reputation Leads to Cooperative Play

For more direct comparisons between experiments with and without reputation, we use the same 64-choice IPD with ten rounds of game duration. For the experiments with reputation, we investigated different p_r values (e.g., 0.05, 0.15, and 0.25). Figure 5 and table 2 summarizes the results of the experiments. In general, results show that when when reputation was included, evolution to cooperation was more likely to occur. Comparison of average payoff of 30 runs between the experiments shows that the higher average payoffs for all experiments with reputation are statistically significantly different compared to the experiment without reputation. Given that when reputation was included, the population evolved to cooperation in more runs (table 2), results suggest that reputation has a positive impact in promoting the learning of cooperative behaviors.

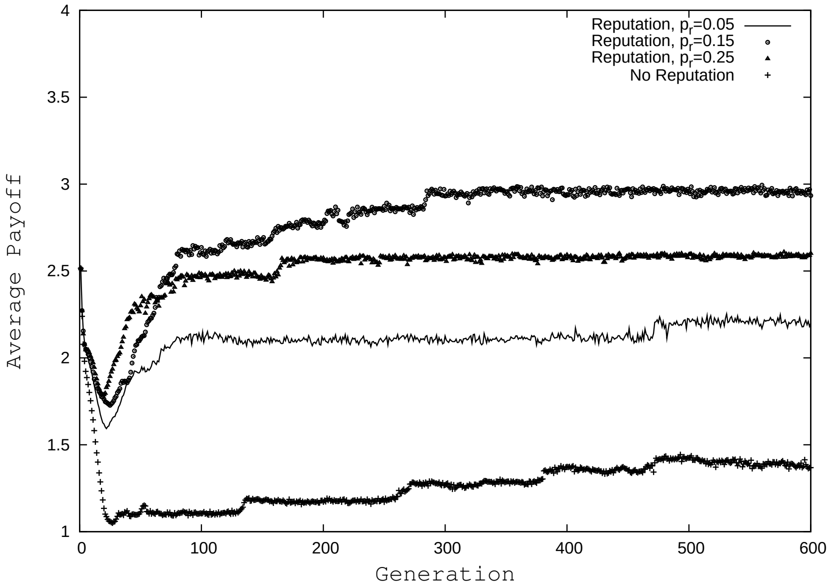

Figure 5: Comparison of the average payoff of 30 runs over 600 generations between the 64-choice IPD games with and without reputation.

3.3.3 How and Why Reputation Helps to Promote Cooperation

We answer the first question of how reputation helps in promoting more complex IPD interactions by first analyzing behavioral responses of the final best evolved strategy. In particular, in runs where population evolved to play mutual cooperation, we observe strategies discriminate opponents according to their

Table 2: Comparison of results of the experiments that consider the 64-choice IPD game with and without reputation. "$No \leq 1.5$" indicates the number of runs where the population evolved to play defection. "$1.5 < No < 3.5$" indicates the number of runs where the population evolved to play intermediate choices. "$No > 3.5\%$" indicates the number of runs where the population evolved to play cooperation.

Reputation	$No \leq 1.5$	$1.5 < No < 3.5$	$No \geq 3.5$
Not included	25	4	1
Included, $p_r = 0.05$	16	6	8
Included, $p_r = 0.15$	10	0	20
Included, $p_r = 0.25$	14	0	16

reputation in the second stage of 64-choice IPD with reputation. That is, strategies play "all cooperate" if the opponents have good reputation (Fig. 6), but play "all defect" if the opponents have bad reputation (Fig. 7).

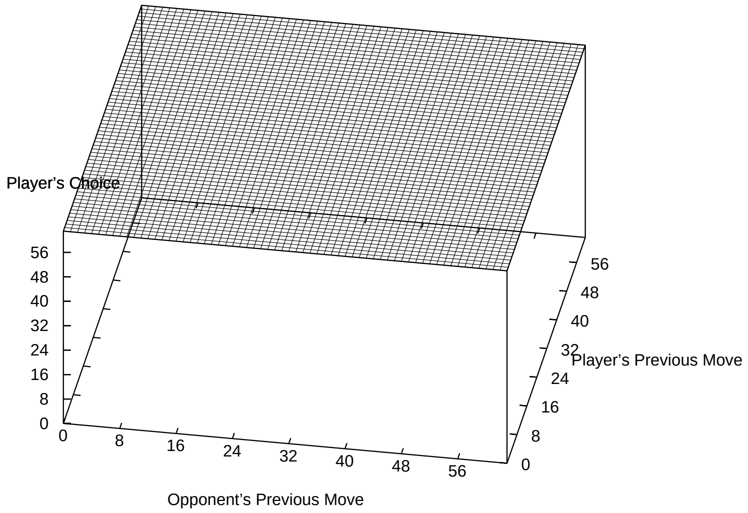

Figure 6: Best strategy's behavioral response during 64-choice IPD with reputation stage at the end of an evolutionary run that resulted in highly cooperative play. 0 indicates full defection, while 63 indicates full cooperation. This strategy started with a move of 63 (full cooperation) and only responds with the same move, effectively playing "all cooperate"

These results indicate that strategies evolved to consider reputation only even though both reputation and previous moves are taken as input. Closer inspection on the neural networks shows that the evolution of "discriminatory" behavior was achieved when the fifth input for reputation is weighted much higher compared to the other four inputs for previous moves. We also note that such behaviors were achieved when we allowed for higher selective pressure on reputation inputs by setting p_r to a value that is less than 0.5, thus making sure that there are more games whereby a strategy use reputation (i.e., second stage) that contribute to its fitness compared to games without considering reputation (i.e., first stage). As such, there is a feedback to evolution to place higher selective pressure on traits associated with reputation.

In addition, we also observe that in runs where population evolved to play mutual cooperation, the population is dominated by these "discriminatory" strategies. A "discriminatory" strategy is successful because it is able to engage in full cooperation play throughout the entire game with similar strategies, thus obtaining high payoffs. Here, reputation acts as a sort of signalling mechanism for strategies to engage in mutual cooperation. High reputation scores reflect a strategy's willingness to engage in highly cooperative play. Lower reputation scores are indicative of a strategy's unwillingness to engage in highly cooperative play. As such, by comparing reputation scores, a "discriminatory" strategy knows

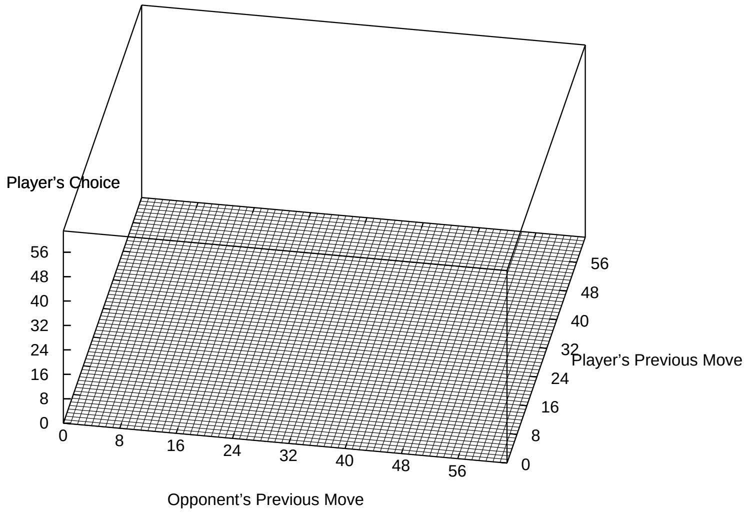

Figure 7: Best strategy's behavioral response during 64-choice IPD with reputation stage at the end of an evolutionary run that resulted in highly cooperative play. 0 indicates full defection, while 63 indicates full cooperation. This strategy started with a move of 0 (full defection) and only responds with the same move, effectively playing "all defect"

beforehand, which strategy to trust by playing all cooperate, and which strategy it should not trust by uncompromisingly playing all defect.

However, for reputation to promote cooperation, we will need to address the issue of whether it pays for strategies to maintain good reputation, which will then allow us to answer the second question of why reputation helps with cooperation. This is important because the additional input consideration for reputation introduces a subtle and complex relationship between responding to the current interaction (gaining fitness from the current game) and responding for future interactions (gaining probable fitness from other members of the population by maintaining favorable reputation scores).

We observed that in all our runs where the population evolved to play mutual cooperation, the best strategy of the population has high reputation scores of +1, which means that the strategy will always be viewed as having a good reputation since the strategy's reputation score is equal to or higher than their opponents'. In these runs, we observe that the rise of the average population payoff, which will indicate mutual cooperation play, coincides with the rise of reputation score of the best evolved strategy (Fig. 8 plots the observation for a typical run). We further observe that the best strategy evolved plays nice and continues to be highly cooperative (Fig. 9).

These results suggest that evolved strategies cooperate during the initial reputation estimation stage to obtain high reputation scores that will later help them elicit mutual cooperative play from similar strategies with comparable reputation scores, thus allowing promoting cooperation play in the population. However, these strategies are not naive. They will defect against other strategies with lower reputation scores because they are likely not to engage in highly cooperative play. In this way, reputation helps with cooperation because it discourages strategies from playing other alternative choices that lead to lower cooperation levels, especially in the situation of games with more choices and shorter durations.

At this point, we note that our results of strategies evolving to consider only reputation corresponds to the results, and also justify the motivation, of previous studies [16, 22] that modelled behavioral responses based solely on reputation for input and decision-making. As mentioned by the same authors in [15], the mechanism of direct reciprocity was removed from their experiments both for simplicity and also to focus on how the mechanism of indirect reciprocity can result with cooperation. However, they also pointed out that indirect reciprocity is a result of direct reciprocity that occurs between other individuals. Here, we made this process that connects these two mechanisms explicit: reputation scores of strategies are estimated from choices that they made in a small number of interactions that do not

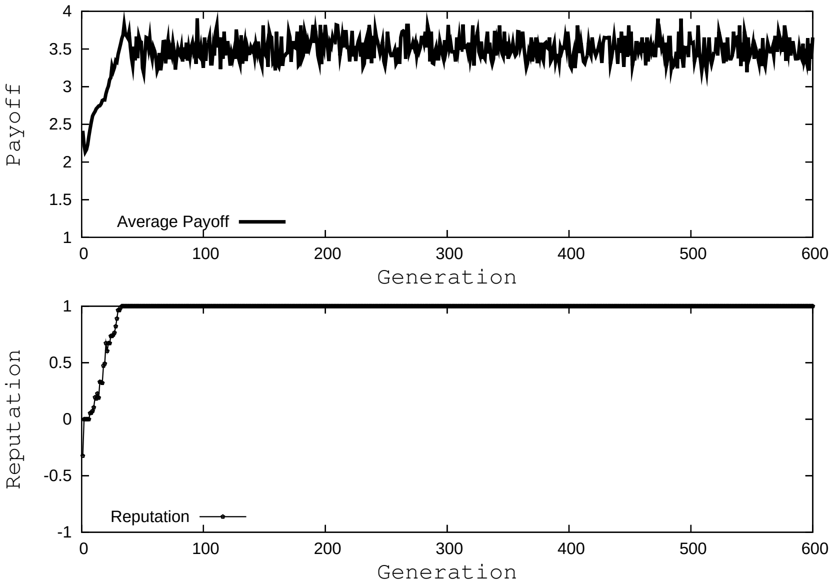

Figure 8: Evolutionary dynamics for a sample run that evolved to cooperative play. The top graph plots the average population payoff across generations. The lower graph plots the reputation score of the best evolved strategy of the population across generations.

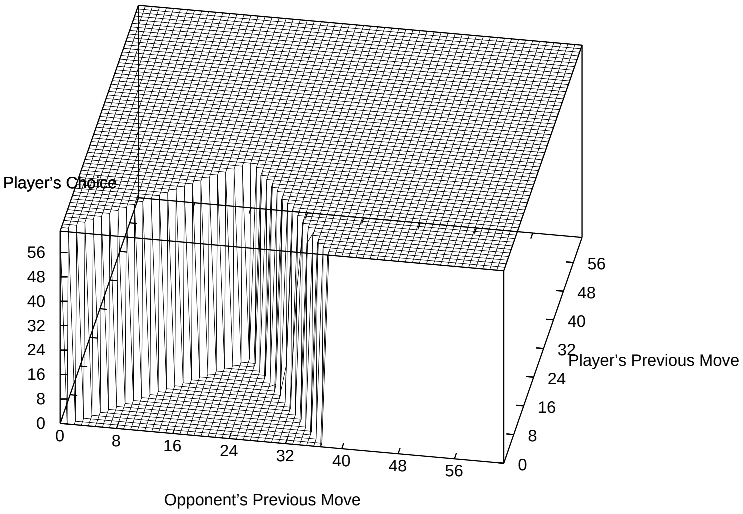

Figure 9: Best strategy's behavioral response during reputation estimation stage at the end of an evolutionary run that resulted in highly cooperative play. 0 indicates full defection, while 63 indicates full cooperation. This strategy started with a move of 63. Note that the triangular ridge of full defection responses is a result of evolution of earlier strategies that played intermediate choices. At this point, the strategy has effectively evolve to play "all cooperate" since it starts with full cooperation, and continues to cooperate regardless of the opponent's response. The triangular regions are never accessed during a behavioral exchange.

involve reputation in the population, which are then used in interactions that involve reputation with other remaining strategies. From these two mechanisms, the co-evolutionary process produces strategies with cooperative behaviors.

3.4 The Impact of Different Reputation Estimations

Although earlier we showed that reputation can help the evolution of cooperation in the more complex IPD interactions that involve more choices and short game durations, the two-stage procedure that was studied might not be realistic. First, the two-stage procedure does not allow a near simultaneous operation of the mechanisms of direct and indirect reciprocity since reputation scores must be determined first. Second, since the reputation scores are estimated anew every generation, there is no memory of games from previous generations in the calculation, and as such can affect the accuracy of the estimation. For example, a strategy that obtains a high reputation score because of interactions with cooperative opponents can obtain a low reputation score in the following generation because it defects against random opponents that play low cooperation levels only. Furthermore,

We address these major limitations by considering an approach with a single stage procedure whereby in every generation, all strategies in the population compete in the n-choice IPD with reputation. This allows near simultaneous operation of both the mechanism of direct and indirect reciprocity, which is more realistic. We address the other limitation that involves incorporating memory of games from previous generations by having each strategy storing its frequency distribution of choices played, $f(x)$, and updating it for every game played. Surviving strategies also carry forward their distributions to the next generation. Although this increases the memory requirement, reputation scores can be accurately determined from choices played by applying equation 1.

More importantly, with the new procedure, we can study how different interpretations on estimating a strategy's reputation impact on the evolutionary outcome. In particular, we focus on the issue of the accuracy of reputation estimation, and how it is related to the procedure of incorporating memory of games from previous generations to calculate reputation scores and how frequently reputation scores are updated.

We note here that the procedure for selection and variation of strategies remain the same. As for behavioral responses of the representation, strategies now consider, in addition to previous moves, the case of whether their opponents have good or bad reputations. That is, the fifth input of the neural network now takes either $+1$ or -1. It does not take in 0 anymore, which was originally used in the reputation estimation stage of the two-stage procedure.

3.4.1 Accurate and Frequent Update for Reputation Estimation Leads to further Evolution of Cooperation

The co-evolutionary learning model for the alternative reputation estimation differs from the earlier model involving a two-stage game. In particular, each strategy now also stores its frequency distribution of choices played, $f(x), x = \{x_1, x_2, ..., x_n\}$, which is initialized to 0 for all x. The offspring inherits the parent's frequency distribution of choices played, $f(x)$. Reputation score for each strategy is also calculated using equation 1 at the start of every generation, which is unchanged once it is calculated. At the start of the evolutionary process, due to the initialization of strategies' $f(x)$s, all strategies are arbitrarily assigned random reputation scores from a uniform distribution in the range of $[-1, 1]$. We refer to this experiment as *RepA*1. We also conduct additional experiment where reputation scores are updated more frequently (after every game played), which we refer as *RepA*2.

Figure 10 and table 3 summarizes the experimental results. In particular, we compare the results for *RepA*1 and *RepA*2 experiments with the original two-stage approach *Rep* (we consider the experiment with $p_r = 0.05$ since at this setting, the co-evolutionary learning involves the highest number of IPD interactions with reputation, which would allow a more direct comparison) and also the experiment without reputation *NoRep*. We first found that in terms of averge payoff of 30 runs (Fig. 10), both results obtained through *RepA*1 and *RepA*2 are statistically significant compared to *NoRep*. As further shown by results in table 3, the higher average payoffs obtained through *RepA*1 and *RepA*2 were due to more runs where the population evolved to play cooperation.

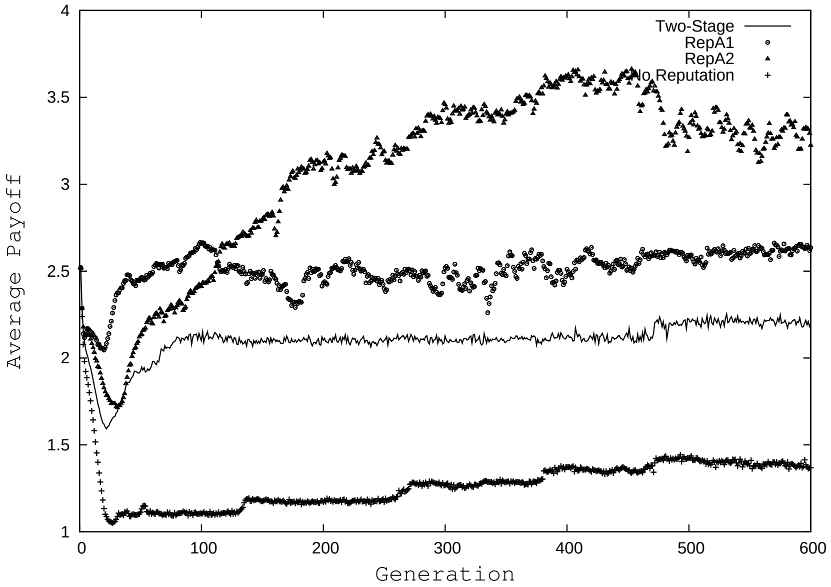

Figure 10: Comparison of the average payoff of 30 runs over 600 generations between the 64-choice IPD games with reputation for various reputation schemes and without reputation.

Table 3: Results of experiments that use alternative reputation estimation based on frequency distribution of choices played, $f(x)$, that is accumulated for surviving strategies throughout evolution. Unlike the original two-stage procedure, *Rep*, the alternative approaches, *RepA*1 and *RepA*2 evolved strategies to play the game of IPD with more choices and reputation directly. "$No \leq 1.5$" indicates the number of runs where the population evolved to play defection. "$1.5 < No < 3.5$" indicates the number of runs where the population evolved to play intermediate choices. "$No > 3.5\%$" indicates the number of runs where the population evolved to play cooperation.

Experiment	$No \leq 1.5$	$1.5 < No < 3.5$	$No \geq 3.5$
NoRep	25	4	1
Rep	16	6	8
*RepA*1	12	3	15
*RepA*2	5	5	20

When comparing between results obtained from different implementations of reputation estimation, we found that the average payoff of 30 runs obtained through *RepA*1 is not statistically significant compared to that of *Rep*. This is despite results showing that with *RepA*1, a higher average payoff (Fig. 10) from having more runs where the population evolved to play cooperation (table 3) was obtained. However, the average payoff of 30 runs obtained through *RepA*2 is statistically significant compared to that of *Rep*. These observations suggest the importance of accumulating memory of games from previous generation to calculate reputation scores and having them updated after every game, and how subsequent increases to the accuracy of reputation estimation lead to further evolution of cooperation.

In terms of how strategy behaviors were evolved, experiments for both *RepA*1 and *RepA*2 did not result with strategies evolving to purely "discriminatory" play based on reputation alone. This result can be explained by considering one fundamental difference between the two-stage procedure used in the earlier section and the alternative, single stage procedure for reputation estimation used here. With the two-stage procedure, "discriminatory" play is observed due to higher selective pressure on reputation input than on previous moves' inputs. With the single stage procedure, because all games are played using previous moves and reputation scores simultaneously, selective pressure did not favor one type of input over the other. We did not observe markedly higher connection weights for the input corresponding to reputation (the same observation is also made for the two-stage procedure when strategies played more games for the second stage that involve reputation).

3.5 Does Reputation Still Help to Promote Cooperation Despite Misperception

Here, we study the impact of error in perception for the co-evolutionary learning of IPD with reputation. Leimar and Hammerstein [22] had earlier studied two possible errors in evolution of strategies based on reputation only, i.e., errors in implementation (mistakes that occur during a strategy's prior choice that affect the calculation of reputation scores, and hence, its reputation) and in perception (a strategy misperceive the opponent's reputation, e.g., instead of good reputation, it is perceived as bad reputation). For our investigation, we focus only on errors in perception for simplicity since for the IPD with reputation interaction that we studied, only reputation is affected by the error.

In particular, we investigate for all the three implementations of reputation (*Rep*, *RepA*1, and *RepA*2) that are effected by error in perception. We consider the situation where there is a small probability 0.01 that a strategy might misperceive its opponent's reputation. In general, we observe that with a small possibility of strategies misperceiving opponents' reputation, evolution to cooperation is still possible. For example, figures 11, 12, and 13 show that although the introduction of the error results with a slightly lower average payoff compared to the error-free experiments, comparison between error and error-free experiments for each corresponding implementation of reputation estimation show that the difference in results are not statistically significant. However, comparison between experiments with error and the experiment without reputation, the difference in results are statistically significant. These results suggest that evolved cooperative strategies are robust even when errors in perception can occur.

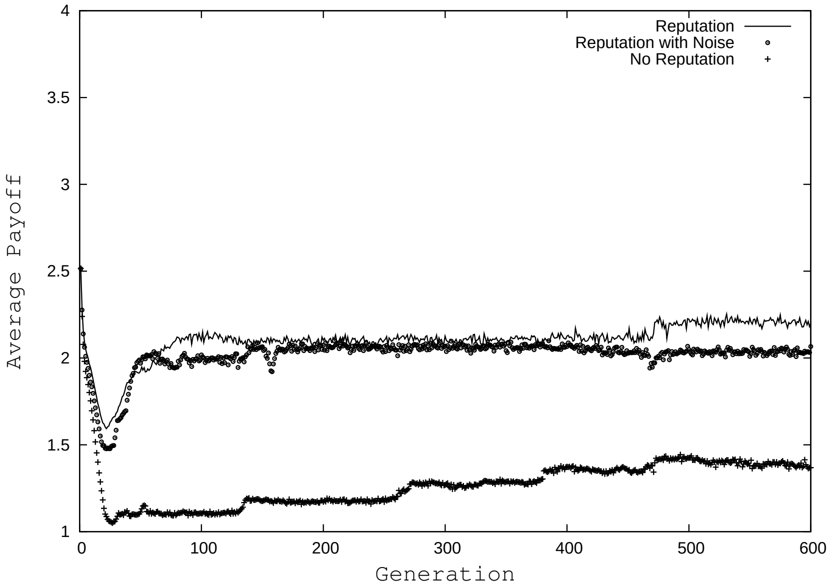

Figure 11: Comparison of the average payoff of 30 runs over 600 generations between the 64-choice IPD games for *Rep* with and without error in perception of reputation.

4 Conclusion

We have extended the co-evolutionary learning framework for IPD games with reputation. This extension is motivated from the view that complex human interactions are not limited to only direct interactions but also indirect interactions. We have shown that reputation helps to promote evolution of cooperation even for situations of more complex behavioral interactions where direct interations alone are not sufficient for the learning of cooperative behaviors. Reputation helps in the evolution of cooperation of more complex IPD interactions by providing a mechanism to estimate behaviors of future partners prior to interactions. Through reputation, cooperative strategies are able to elicit mutual cooperation play right from the start. With reputation, there is incentive for strategies to cooperate to obtain good reputation, which will make it more likely for strategies to engage in mutual cooperation with future partners.

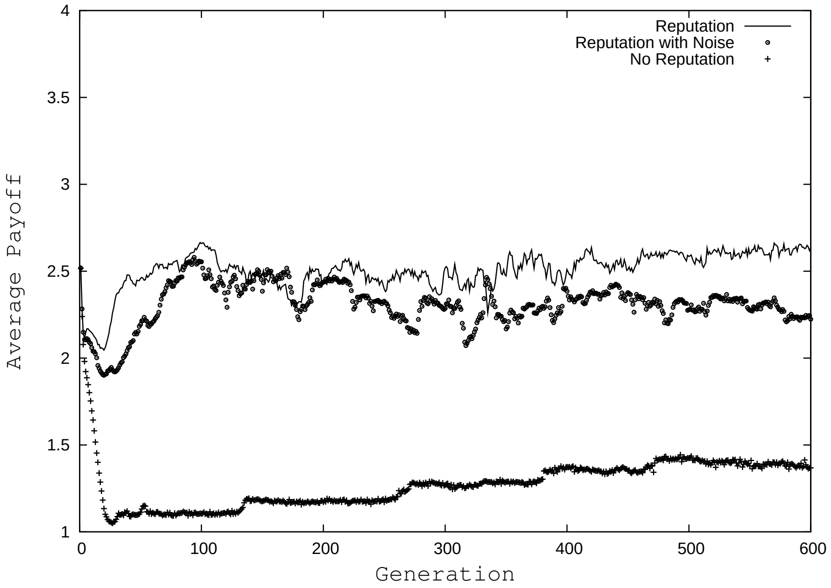

Figure 12: Comparison of the average payoff of 30 runs over 600 generations between the 64-choice IPD games for *RepA*1 with and without error in perception of reputation.

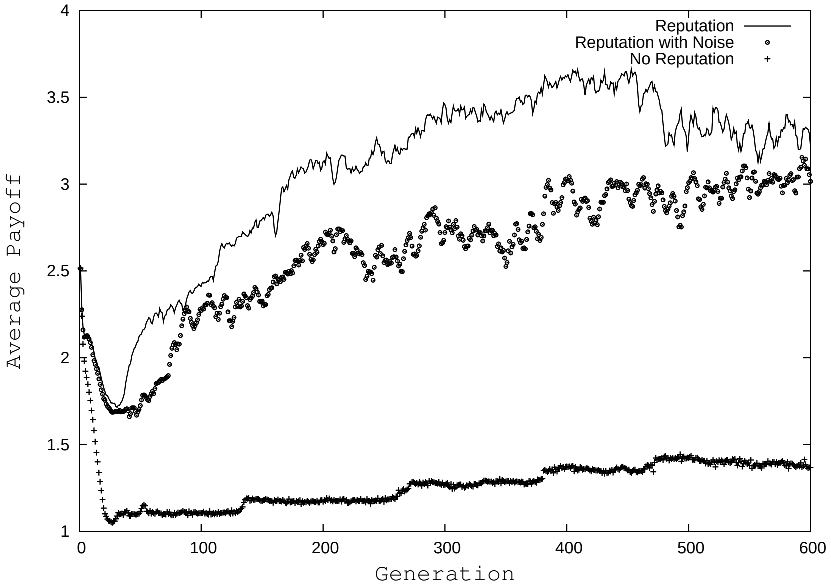

Figure 13: Comparison of the average payoff of 30 runs over 600 generations between the 64-choice IPD games for *RepA*2 with and without error in perception of reputation.

A strategy's reputation can be estimated in many different ways. We investigated the impact of different reputation implementations on the evolution of cooperation, focusing on one major issue: the accuracy of reputation estimation. Experiments show that incorporating memory of games from previous generations using actual choices played from games leads to a higher accuracy in reputation estimation that has a positive impact on the evolution to cooperation. Further evolution to cooperation is obtained when one considers an almost instantaneous feedback of a strategy's behavior to others in the population by updating reputation score for every game played.

We also studied the impact of error in perception on the evolution of strategy behavior. We have shown that even with the possibility that strategies misperceiving opponents' reputation, evolution of cooperation is still possible. That is, cooperative strategies learned through co-evolutionary learning are robust to such errors.

References

[1] R. Axelrod. *The Evolution of Cooperation.* Basic Books, New York, 1984.

[2] R. Axelrod. More effective choice in the prisoner's dilemma. *The Journal of Conflict Resolution*, 24(3):379–403, September 1980.

[3] R. Axelrod. Effective choice in the prisoner's dilemma. *The Journal of Conflict Resolution*, 24(1):3–25, March 1980.

[4] R. Axelrod. The evolution of strategies in the iterated prisoner's dilemma. In L. D. Davis, editor, *Genetic Algorithms and Simulated Annealing*, chapter 3, pages 32–41. Morgan Kaufmann, New York, 1987.

[5] D. B. Fogel. The evolution of intelligent decision making in gaming. *Cybernetics and Systems: An International Journal*, 22:223–236, 1991.

[6] D. B. Fogel. Evolving behaviors in the iterated prisoner's dilemma. *Evolutionary Computation*, 1(1):77–97, 1993.

[7] D. B. Fogel. On the relationship between the duration of an encouter and the evolution of cooperation in the iterated prisoner's dilemma. *Evolutionary Computation*, 3(3):349–363, 1996.

[8] B. A. Julstrom. Effects of contest length and noise on reciprocal altruism, cooperation, and payoffs in the iterated prisoner's dilemma. In *Proc. 7th International Conf. on Genetic Algorithms (ICGA'97)*, pages 386–392, San Francisco, CA, 1997. Morgan Kauffman.

[9] R. Axelrod and W. D. Hamilton. The evolution of cooperation. *Science*, 211:1390–1396, 1981.

[10] P. Darwen and X. Yao. On evolving robust strategies for iterated prisoner's dilemma. In *Progress in Evolutionary Computation*, volume 956 of *Lecture Notes in Artificial Intelligence*, pages 276–292, 1995.

[11] P. G. Harrald and D. B. Fogel. Evolving continuous behaviors in the iterated prisoner's dilemma. *BioSystems: Special Issue on the Prisoner's Dilemma*, 37:135–145, 1996.

[12] P. Darwen and X. Yao. Does extra genetic diversity maintain escalation in a co-evolutionary arms race. *International Journal of Knowledge-Based Intelligent Engineering Systems*, 4(3):191–200, July 2000.

[13] P. Darwen and X. Yao. Why more choices cause less cooperation in iterated prisoner's dilemma. In *Proc. 2001 Congress on Evolutionary Computation (CEC'01)*, pages 987–994, Piscataway, NJ, 2001. IEEE Press.

[14] P. Darwen and X. Yao. Co-evolution in iterated prisoner's dilemma with intermediate levels of cooperation: Application to missile defense. *International Journal of Computational Intelligence and Applications*, 2(1):83–107, 2002.

[15] M. A. Nowak and K. Sigmund. The dynamics of indirect reciprocity. *Journal of Theoretical Biology*, 194:561–574, 1998.

[16] M. A. Nowak and K. Sigmund. Evolution of indirect reciprocity by image scoring. *Nature*, 393:573–577, 1998.

[17] X. Yao and P. Darwen. How important is your reputation in a multi-agent environment. In *Proc. 1999 Conf. on Systems, Man, and Cybernetics (SMC'99)*, pages 575–580, Piscataway, NJ, 1999. IEEE Press.

[18] S. Y. Chong and X. Yao. Behavioral diversity, choices, and noise in the iterated prisoner's dilemma. *IEEE Transactions on Evolutionary Computation*, 9(6):540–551, 2005.

[19] X. Yao. Evolving artificial neural networks. *Proc. IEEE*, 87(9):1423–1447, September 1999.

[20] X. Yao, Y. Liu, and G. Lin. Evolutionary programming made faster. *IEEE Transactions on Evolutionary Computation*, 3(2):82–102, 1999.

[21] K. Chellapilla and D. B. Fogel. Evolution, neural networks, games, and intelligence. *Proc. IEEE*, 87(9):1471–1496, September 1999.

[22] O. Leimar and P. Hammerstein. Evolution of cooperation through indirect reciprocity. *Proceedings of the Royal Society of London, Series B*, 268:745–753, 2001.