Learning and Cooperation in Sequential Games

Annapurna Valluri

Adaptive Behavior 2006; 14; 195
DOI: 10.1177/105971230601400304

Adaptive **Behavior**

Learning and Cooperation in Sequential Games

Annapurna Valluri
Department of Operations and Information Management, The Wharton School of the University of Pennsylvania

The predictions of classical game theory for one-shot and finitely repeated play of many 2x2 simultaneous games do not correspond to human behavior observed in laboratory experiments. The promising results of learning models in tracking human behavior coupled with the growing electronic market and the number of e-commerce applications has resulted in an increased interest in studying the behavior of adaptive artificial agents in different economic games. We model agents with a reinforcement learning algorithm and analyze cooperative behavior in a sequential prisoner's dilemma game. Our results demonstrate the ability of artificial agents to learn cooperative behavior even in sequential games where defection is the subgame perfect Nash equilibrium. We attribute the reciprocal-like behavior to the structural flow of information, which reduces the risks of exploitation faced by the second-mover. Additionally, we analyze the impact of the second-mover's temptation payoff and payoff risks on the rate of cooperative behavior.

Keywords reinforcement learning · prisoner's dilemma · sequential games

1 Introduction and Motivation

Predictions of strategic behavior in classical game theory are based on the assumptions of perfect foresight, knowledge and complete rationality of economic actors. These assumptions collapse in real-world situations, however, and researchers in experimental economics have demonstrated, through numerous experiments encompassing a gamut of strategic interactions, how laboratory subjects very often do not behave according to the predictions of classical game theory. McKelvey and Palfrey (1992) have studied a simple centipede game where players suffer from the conflict of self-interest and mutual benefit. Although classical game theory reasons through backward induction that the first player should defect in the first round of the game, play in McKelvey and Palfrey's (1992) laboratory experiments conforms to the predicted Nash equilibrium only 5% of the time. Experimental evidence contradicting the predictions of classical game theory in sequential bargaining and iterated prisoner's dilemma games has been found by Ochs and Roth (1989) and Andreoni and Miller (1993), respectively.

Since there is strong evidence which points to players not conforming to Nash equilibrium strategies in many iterated strategic interactions, researchers in behavioral economics have started to develop learning models based on experimental play of simulated artificial agents to explain the adaptive behavior of humans. Many researchers such as Camerer and Ho (1999), Feltovich (2000), Duffy (2001), and Erev and Roth (1998) have simulated human behavior using different learn-

Correspondence to: Annapurna Valluri, Department of Operations & Information Management, University of Pennsylvania, 3730 Walnut Street, Suite 500, Jon M. Huntsman Hall, Philadelphia, PA 19104-6340.
E-mail: avalluri@wharton.upenn.edu
Tel.: +1 267 259-4527 *Fax*: +1 215 898-3664

(2006), Vol 14(3): 195–209.
[1059–7123(200612) 14:3; 195–209; 068552]

ing models in various games, and have found that the learning models track human behavior well. Most learning models can be broadly classified into belief-based models (e.g., fictitious play) and reinforcement-based models (Watkins, 1989). In belief-based models, players keep track of the history of play of the other players and form beliefs about the likely play of the other players. Players' actions are then chosen based on the expected payoffs given their beliefs about their opponent. In reinforcement-based models, on the other hand, players do not have beliefs about what the other players will do and only consider the payoffs that different strategies (actions) have yielded in the past. Actions associated from past experience with positive rewards have a higher probability of being taken by the agent. Camerer and Ho (1999) have developed a general model called "experienced-weighted attraction learning," which includes the belief-based and reinforcement-based models as special cases.

The promising results of learning models in tracking human behavior coupled with the growing electronic market and the number of e-commerce applications where such results would be useful has resulted in an increased interest in studying the behavior of adaptive artificial agents in different types of economic models/games. Fang, Kimbrough, Pace, Valluri, and Zheng (2002), for instance, address the question of how artificial agents using reinforcement-based models perform when playing an ultimatum game, whereas Wu, Kimbrough, and Zhong (2002) focus on the emergence of cooperation and trust among agents in a particular trust game. Most of the literature that has analyzed the behavior of adaptive artificial agents in discrete and small action space economic games has focused on simultaneous games. However, many real-world decision scenarios are more sequential than simultaneous in nature, meaning that different interacting economic actors could have asymmetric knowledge about the environment. Examples of some sequential decision-making situations are price setting in the market place by firms, outsourcing decisions, or bidding in an auction (either online or offline).

Prior research that has analyzed the behavior of artificial agents has done so while focusing on a particular game. Although we conduct simulations of a specific game, our focus encompasses a broader class of principal–agent problems which share certain essential features with the game we focus on. Unlike those of previous researchers, who have focused mainly on simultaneous games, however, our contributions lie in extending the research to sequential games with the focus being on principal–agent games that are fraught with the problem of moral hazard. Principal–agent relationships are a common occurrence, and the associated problems arise whenever a principal depends on an agent for production of goods, services rendered, or the completion of some other task. Play in principal–agent problems is sequential, with the second-mover (agent) having an advantage through his knowledge of the first-mover's (principal) action before he makes his decision. The second-mover's action is only indirectly observable by the first-movers through an outcome variable, where certain possible second-mover actions result in more positive outcomes for the principal than other actions. In effect, the agent can exploit his information advantage at the expense of the principal, resulting in non-cooperation. Players act in their own best interests and also face a conflict of interest, thus causing moral hazard problems that prevent the socially efficient equilibrium, which is also the cooperative solution, from being reached.

An example of a principal–agent game is one of an employer–employee relationship where the employer depends on the employee for the completion of certain tasks. While the employer faces the problem of fixing a wage rate for the employee since the efforts of the employee can't be monitored, the employee faces a disincentive to put in additional effort because he knows that his effort level can't be monitored. In this paper, we conduct simulations and analysis focusing on a specific model of one such game—the iterated sequential prisoner's dilemma (SPD)—which closely resembles the simplified version of the principal–agent game. In both cases, it is in the best interest of the principal and the agent to defect, and so cooperative outcomes are not imminent. In addition, both games are played sequentially. In the simplified version of the principal–agent game, the players have a discrete action set, the outcome variable is a one-to-one deterministic function of the agent's effort level, and so the principal can without any ambiguity ascertain the true action chosen by the agent.

The sequential iterated prisoner's dilemma (IPD) is a simple two-player multi-period game (see Figure 1). In the sequential prisoner's dilemma (SPD) game (Clark & Sefton, 2001) Player 2 chooses an action only after an action has been chosen by Player 1 and Player 1's choice of action is known to Player 2. Consequently,

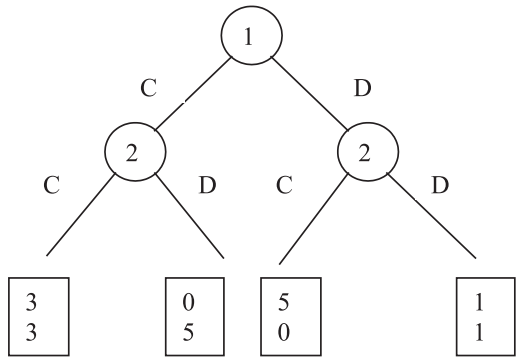

Figure 1 Sequential prisoner's dilemma.

an asymmetry is created in the information available to the two players.

Moreover, while the simultaneous prisoner's dilemma has been researched thoroughly through the use of many techniques such as game theory (Poundstone, 1993), evolutionary game theory (Axelrod, 1984; Axelrod & Hamilton, 1981; Nowak, 1990; Nowak & May, 1992), behavioral economics (Andreoni & Miller, 1993; Shubik, 1970), computational economics (Ashlock, Smucker, Stanley, & Tesfatsion, 1996), and machine learning (Miller, 1996; Rubinstein, 1986; Sandholm & Crites, 1996), considerably less research has been conducted on the sequential version of the prisoner's dilemma game.

We chose to study the iterated SPD game using artificial agents with a reinforcement learning (RL) model. This was due both to the success of RL models in explaining human behavior in a variety of games (Erev & Roth, 1998) and to the large number of researchers using artificial agents with RL models. An important advantage of the RL model over game theoretic models is that it makes very few assumptions about rationality, such as knowledge of the game, its parameters and limited capacity of the agents. Moreover, it also makes fewer assumptions about the agents than other learning models. For instance, belief-based models require the agents to have prior beliefs about their opponent's strategies.

The purpose of this article is three-fold. Firstly, while we focus on the cooperative behavior of agents in the iterated SPD game, the results can be generalized to the class of simple principal–agent problems that possess the issue of moral hazard. Our simulation results indicate a high level of cooperation between players in the iterated SPD game, agreeing with experimental findings but contradicting the predictions of classical game theory; this may be because we focus on decision-making in more realistic sequential environments instead of using simultaneous scenarios. Secondly, we believe that our article provides further evidence for the applicability of adaptive learning algorithms to modeling human behaviour with artificial agents because of their ability to predict outcomes that match observed behavior. Although we leave the calibration of the simulation results with experimental data of the SPD game for future work, we provide promising results. Finally, we not only hypothesize the reasons for the observed degree of cooperation, but also shed light on factors that affect cooperative behavior in the iterated SPD game, such as bargaining power and uncertainty about the opponent's choice of action.

The article is organized as follows: The next section discusses the related literature, while Section 3 briefly introduces our methodology. We state our hypotheses in Section 4 and the results are presented in Section 5. Finally, our conclusions and future work are summarized in Section 6.

2 Literature Review

Work that is related to our article can be separated into two fields: Information and experimental economics, and computational agent-modeling, and we will examine the pertinent work in these areas separately.

2.1 Information and Experimental Economics

Markets, especially e-commerce markets, are fraught with information asymmetries. In e-commerce markets parties situated in various parts of the globe come together to transact with hardly any knowledge of the other party. As a result of information asymmetry, problems such as adverse selection and moral hazard become widespread. Moral hazard or incentive problems may arise when individuals engage in risk sharing; moreover, their actions affect the probability distribution of an outcome but cannot be contracted upon because the actions taken are private (Holmstrom, 1979). Examples of moral hazard are widespread in practice; how, for instance, does a manager measure a white-collar employee's effort in the workplace? How does a client firm (principal) measure the quality of

service and expertise that a consulting firm (agent) is providing while implementing an ERP (Enterprise Resource Planning) system?

In typical principal–agent scenarios, a principal depends on an agent for production of a good, services rendered, or for the completion of some other task. The problem arises because both players act in their own best interests and the principal is unable to monitor the agent's actions (Holmstrom, 1979). Principal contributions in this area include those of Ross (1973), Grossman & Hart (1983) and Radner (1981, 1985, 1986), among others.

In Grossman & Hart (1983), a one-period principal agent model is discussed where both the principal and the agent are self-interested players and the author demonstrates that the Nash Equilibrium is almost never the Pareto-efficient outcome. On the other hand, Radner (1981) proposed a repeated principal–agent game. In the multi-period model, there is a certain known probability with which the relationship will continue to the next round/period. He presented a specific strategy for the repeated principal–agent game by which the principal and the agent are able to sustain an efficient equilibrium, where payoffs are higher than from the inefficient Nash equilibrium in the one-shot game.

In the simplified principal–agent problem, the iterated SPD game, which we focus on in this article, classical game theory (Neumann & Morgenstern, 1944) predicts that the dominant strategy in the sequential game for both the first and second movers is defection, in both the one-shot and the finitely repeated game. The argument is made by backward induction where the game in the last period of a sequence is equivalent to a one-shot game. Laboratory experiments, however, show that real behavior does not conform to classical game theory; whether this is or is not rational is another matter. Cooperation has not only been observed in the simultaneous version but also in the sequential version of the PD game (Selten & Stoecker, 1986).

The two-player SPD game was studied by Clark and Sefton (2001) with human subjects. Their goal was to understand the primary reason for cooperation—is it conditional on first-mover cooperation, repetition, economic incentives or gender? Their examination reveals that the factor most influencing cooperation is first-mover cooperation; therefore, reciprocation is deemed an important element in games of this nature, as opposed to pure altruism. Furthermore, they experimentally observed the effects of doubling the payoffs and increasing the temptation payoff; and found that the former had no effect on the rate of cooperation, while there was a decrease in cooperation when the temptation payoff alone was doubled. Bolle and Ockenfels (1990) also studied the SPD game, both analytically and experimentally. In their experiments, however, the players played only one-shot games with the second-movers choosing actions conditional on a hypothesized choice by the first-mover. Their findings reveal that the re-valuation of the cooperative result is the cause of the degree of cooperation observed in human experiments. In other words, the utility associated with the cooperative result is more than the stated reward in the payoff structure. They discuss analytically the logic of observing greater cooperation in the sequential game than the simultaneous game, although they are unable to find a difference in the actual rates of cooperation in their experiments. However, greater cooperation was observed in the SPD game than in the simultaneous game in the laboratory experiments conducted by Oskamp (1971). In Oskamp's experiments, the first move is made by a human who plays a repeated game with a computer program that uses a fixed strategy. The fixed strategy is one of the following: Cooperate with 10% probability, cooperate with 90% probability, or tit-for-tat.

Researchers in experimental economics are concerned with finding models to explain individual human behavior by studying relatively simple games. On the other hand, the field of computational agent technology is interested in the strategic decisions taken by artificial agents when they are modeled with learning algorithms in different complex economic environments, especially electronic markets.

2.2 Computational Agent Technology

In spite of the tremendous achievements of game theory, it remains unknown how to compute equilibrium strategies for some decision-making scenarios. For instance, game theory is unable to predict which Nash equilibrium will be observed in the case of multiple equilibria for a specific game. Furthermore, it is widely reported in laboratory experiments (as discussed previously) that actual individual human behavior does not match what has been "predicted" by classical game theory. Agent-based computational economics has several substreams (Tesfatsion, 2002), one of which seeks to explore new computational models that map human behavior.

Another sub-stream aims to capture the behavior of artificial agents modeled with adaptive capabilities from the former sub-stream in order to predict the dynamics of human decision-making and the off-equilibrium paths taken in complex economic environments. See, for example, the seminal work by Axelrod and Hamilton (1981), where they studied the iterated prisoner's dilemma game.

Several researchers in the multi-agent learning literature have studied the adaptive behavior of artificial agents when interacting with other agents in unfamiliar and complex environments. Some researchers have modeled agents with RL algorithms, while others have used models such as belief-based algorithms (Camerer & Ho, 1999) fictitious play, and cumulative best response (Shoham & Tennenholtz, 1993), among others. Suguwara and Lesser (1993) developed their own heuristic learning algorithm to assist agents in learning to coordinate their strategies, whereas Littman (2001) presented an algorithm based on Q-learning called Friend-or-Foe Q-learning, which has strong convergence properties and enables agents to learn the optimal strategies in many scenarios. In Littman (1994), agents play a zero-sum game and try to find optimal mixed strategies to play against their opponents. Sandholm and Crites (1996) model agents with Q-learning, a form of RL, to study cooperative behavior in an iterated prisoner's dilemma game, while Ashlock et al. (1996) use a model related to RL for partner selection in the iterated prisoner's dilemma game. Tesauro (1992) demonstrates the ability of agents using temporal difference learning to successfully adapt and find strategies even in more complicated games such as backgammon with zero built-in knowledge. Sutton and Barto (1998), and Kaelbling, Littman, and Moore (1996) provide a detailed survey of the literature on artificial agents modeled with different RL algorithms in various games.

More recently, and very relevant to our work here, is a series of trust games (stag hunt game—Fang et al., 2002; ultimatum game—Zhong, Kimbrough, & Wu, 2002; mad mex trust game—Wu et al., 2002) that aimed to systematically investigate whether, and how, social trust or cooperation can emerge as a systematic property where communities of self-interested agents interact with each other in the internet environment. These papers focus on the question of whether learning by agents in games can result in cooperation and whether learning agents can outperform fixed-strategy agents. Their focus on the benefits of learning prevents them from shedding light on the structure of the games themselves. We not only corroborate experimental findings by demonstrating the ability of artificial agents to learn to cooperate, but also study different factors of the game that are conducive to promoting cooperation among players.

In contrast to the study of 2x2 trust and coordination games using adaptive agents, Meidinger and Terracol (2002) studied a more realistic game situation—a sequential two-player investment game. In this game, the first player decides whether or not to invest, and the second player decides how much return to give the investor and how much of a cut to take for himself. The authors model players with a reciprocation strategy using RL and evaluate the predictions of the model against collected data. Findings indicate that the RL strategy is able to capture the trend of the observed frequency of returns by the second mover, although it tends to initially underestimate the observed frequency of investment by the first mover.

An important theoretical commonality in these papers is agent learning, where history, experience and memory matter. Among various classes of learning algorithms used previously, of particular interest is the class of learning algorithms called *reinforcement learning*, which itself is an active research frontier in artificial intelligence. The key aspect of RL algorithms is that they reinforce good actions and weaken bad actions. The next section provides a brief overview of the particular RL algorithm we use.

3 Methodology

In our experiments, we model artificial adaptive agents as reinforcement learners, specifically, Q-learners (Watkins, 1989). This algorithm is one of the most widely pursued learning algorithms. Its attractiveness lies in its lack of any need for a model of the environment and its applicability to online learning. Q-learning rewards actions that turn out to be positive and punishes those that yield negative results. The value of taking a particular action, a_i, by a player while in state s, at time t, is represented by $Q(s_t, a_i^t)$. The value function representing the value of taking a particular action in state s is updated by:

$$Q(s_t, a_i^t) = Q(s_t, a_i^t) + \alpha[r_t + \gamma \max_{a_i} Q(s_{t+1}, a_i) - Q(s_t, a_i^t)]$$

where γ is the discount factor, r_t represents the payoff or reward obtained by taking action a_i^t while in state s and moving to another state at time, $t + 1$. The learning rate, α, captures the "recency effect," which weights recent rewards more heavily than past ones, and is an important feature in a dynamic environment. We use the most popular exploration rate called Softmax. According to this policy, the probability of choosing a particular action a_i^t, $p_t(a_i)$, at time t is:

$$p_t(a_i) = \frac{e^{Q(s_t, a_i)/\tau}}{\sum_{i=1}^{n} e^{Q(s_t, a_i)/\tau}} \quad \text{where } \tau \geq 0$$

and τ is a positive parameter called temperature. When the temperature value is high, all actions have equal probability of being chosen; however, with low values of τ, the more highly valued actions are favored and, finally, a zero value corresponds to no exploration but only exploitation of gathered knowledge.

The agents have a memory of one, which means that they remember only the most recent state they were in. In other words, assuming that the first letter represents the action the agent took in the previous round, and the second letter represents his opponent's action, then an agent with memory one and only two possible actions {C, D}, will have four possible states {CC, CD, DC, DD}.

The program for running the simulations, including the modeling of agents, is written in C++, in object-oriented code. The learning and discount rates are 0.2 and 0.95, respectively. These parameters are taken from the pioneering work by Sandholm and Crites (1996) on multi-agent learning, in which they study the iterated prisoner's dilemma game. All results are averaged over 100 games and each game is played for 300,000 episodes. The exploration rate is decreased exponentially at a rate of 0.999 and starts off with a value of 5, unless otherwise stated.

4 Experimental Design and Hypotheses

According to classical game theory, defection is the dominant strategy in the one-shot prisoner's dilemma and the finitely repeated PD game (for both the simultaneous and sequential games), where the argument holds through the use of backward induction. Contrary to the predictions of classical game theory, but based on the observations of cooperative behavior in experimental settings, we hypothesize the following:

Hypothesis 1 Contrary to the predictions of classical game theory, artificial agents (modeled with an RL algorithm) playing the iterated SPD game will not only learn to cooperate but also cooperate more than in the simultaneous version of the game.[1]

Bolle and Ockenfels (1990) discuss the logic for observing more cooperation in the sequential game than in the simultaneous game, but failed to observe it in experiments. In their experiments the game is one-shot, where the second-movers are unaware of the actual choices made by their opponents and make choices based on hypothesized choices made by their first-mover opponents. However, if the game is played repeatedly, especially against the same opponent, then the action of the first-mover observed by the second-mover can be interpreted as a direct consequence of the second-mover's action in the previous period. Unsurprisingly, therefore, more cooperation was observed in the sequential game in the experiments conducted by Oskamp (1971) than in the one-shot game played by the subjects of Bolle and Ockenfels.

We hypothesize the reason for increased cooperation in the sequential version of the iterated prisoner's dilemma game as compared with the simultaneous game to be as follows:

Hypothesis 2 The informational flow in the iterated SPD game, where the second-mover has knowledge of the first-mover's action before choosing his own, results in the second-mover learning the optimal state-action function faster than in the simultaneous game.

Our third hypothesis states that the cooperative behavior of artificial agents is related to the payoff structure; specifically, the risks of defection faced by the second-mover. We define the risk faced by the second-mover in the iterated SPD game along the lines used by Yang, Weimann, and Mitropoulos (2001).[2] The risk faced by the second-mover in choosing to defect decreases if the payoff received by the second-mover when both players defect increases (see Table 2a for the payoffs). Note, however, that the bargaining power or risk is different from the temptation payoff, where temptation is defined as the incentive to defect when the opponent cooperates. The risks faced by the

second-mover, on the other hand, depend on the payoff he receives in the sub-game perfect Nash equilibrium solution and the socially optimal (cooperative) solution. Therefore, as the second-mover's risk associated with defection increases relative to the first-mover, the first-mover's bargaining position to induce cooperation is affected.

Hypothesis 3 When adaptive artificial agents play the iterated SPD game, cooperative behavior and the first-mover's relative bargaining position are affected by the payoff risks of defection faced by the second-mover.

In the extant literature, there are several papers that discuss the first-mover advantage or bargaining power in sequential games. Yang et al. (2001) studied four classical games, namely rent-seeking, best-shot, trust, and the ultimatum game, by transforming the games from the continuous space to simplified *two-action sequential* games. The goal of the authors was to study bargaining games where there is a conflict of interest between the first and the second-movers. The subgame perfect Nash equilibrium in all these games remains the same. While the first-movers prefer the subgame perfect Nash equilibrium, the second-movers in these sequential games prefer them not to select these actions. The preferred action of the first-movers results in a dramatically decreased payoff for the second-mover as compared with the non-preferred action of the first-mover. The first-movers upon choosing the preferred action can in some of the games face punishment by the second-movers; on the other hand, upon choosing the non-preferred action, the first-movers face the risk of being exploited by the second-movers. In the event that the first-movers choose the non-preferred action, the second-movers have the choice of rewarding the first-movers and sharing the payoffs equally with them. The experimental findings illustrate that first-movers differ in terms of their choice away from the subgame perfect equilibrium action depending on the bargaining power of the second-mover in the various games.

We further qualify the statement of the first-mover advantage by noting that in the iterated SPD game with adaptive learning agents, the payoffs received by the first-mover depend on the payoff risks of defection faced by the second-mover. In other words, as the second-mover's payoff risk associated with defection decreases, he has less incentive to cooperate. As a result, not only is there a decrease in the convergence to the cooperative outcome in the iterated SPD game, but the first-mover's relative bargaining position is weaker, which results in a decrease in the payoffs received by the first-mover.

Related to the bargaining power is the relationship of the temptation payoff to the observed rate of cooperation in the iterated SPD game. We therefore hypothesize the following:

Hypothesis 4 There is a non-linear relationship between the risk faced by the second-mover (cooperating incentive) and the observed cooperation rate when adaptive agents play the iterated SPD game.

Clark and Sefton (2001) demonstrate that if the temptation payoff is doubled while keeping everything else constant, the observed rate of cooperation will decrease. We take this analysis one step further and hypothesize that there is a non-linear reduction in the rate of cooperation when the temptation payoff is increased.

Not all information in the real world is perfect and, therefore, we also wish to analyze the impact that imperfect information regarding an opponent's action has on the rate of cooperation in the iterated SPD game.

Hypothesis 5 Uncertainty in the second-mover's action choice decreases cooperation more than the uncertainty in the first-mover's action choice when adaptive agents play the iterated SPD game.

Uncertainty, or noise, is defined as the chance of an error in the information regarding an opponent's action choice. In other words, noise of 5% means that the opponent's action choice is perceived incorrectly 5% of the time; hence, the cooperative action of an agent is perceived by its opponent as defection 5% of the time and defection is viewed by the opponent as a cooperative move 5% of the time. Huck and Miller (2000) find that experimental subjects playing sequential games can tolerate noise up to 10%. Second-movers observing the first-movers' actions with some noise converge to the subgame perfect Nash equilibrium when the noise level is below 10%, and converge to the simultaneous Nash equilibrium when noise levels increase. Similarly, we are also interested in analyzing the effects of noise on the level of cooperation

in the iterated SPD game. We compare the results of the second-mover observing noisy signals of the first-mover's action to the results of noisy signals being observed by the first-mover of the second-mover's action. We hypothesize that the noisy signals of the second-mover's action cause a greater reduction in the rates of cooperation. The reason for lower cooperation is that the noisy signal of the second-mover's action increases the uncertainty faced by the first-mover which, in turn, amplifies the risks of exploitation faced by the first-mover.

5 Simulation Results

Bolle and Ockenfels (1990) broadly classified players in the prisoner's dilemma games as "moral players" and "egoistic players" based on prior research on the motivations and preferences of experimental subjects playing different versions of the prisoner's dilemma game. Both types of players are averse to cooperating in the simultaneous game due to the incomplete information about their opponent's type. However, second-movers in the sequential game possess greater knowledge about their opponent. Therefore, the authors argue that first-movers who are moral players have a greater incentive to cooperate in the sequential game because of the greater probability of their opponent reciprocating a cooperative move if he is also a moral player. Our simulation findings corroborate Bolle and Ockenfels' (1990) logic of increased cooperation in the sequential game. Table 1 shows the payoff matrix used for the prisoner's dilemma game in our simulations.

Observation 1 Contrary to the predictions of classical game theory, artificial agents playing the iterated SPD game not only learn to cooperate but also cooperate more than in the simultaneous version of the game.

Table 1 Payoff for the prisoner's dilemma game with two actions C and D, where C represents cooperation and D represents defection

	C	D
C	0.3, 0.3	0.0, ***0.5***
D	***0.5***, 0.0	***0.1***, ***0.1***

In spite of defection being the initially dominant strategy for the agents we find that the agents learn to cooperate over time in the sequential game, as shown in Table1a. The cells in Table 1a show the percentage of game rounds in which the players find themselves in each of the four possible states, and demonstrates the ability of artificial agents equipped with Q-learning to adapt and learn to cooperate with a high probability in a sequential setting. This corroborates the findings of Clark and Sefton (2001), and also contradicts the predictions of classical game theory which assumes perfect rationality and predicts defection to be the subgame perfect Nash Equilibrium strategy for both players.

In order to investigate the reasons for the high rate of cooperation in the iterated SPD game, we analyzed the percentages of states visited as play progresses in the game. Our findings indicate that over time, as the

Table 1a The proportion of time the two players converge to each of the four possible states when they play the sequential prisoner's dilemma game. The table illustrates that the game converges a greater proportion of time to the cooperative state (CC) in the sequential version of the prisoner's dilemma game than in the simultaneous game (Table 1b)

State	Player 1 – First-mover	Player 2 – Second-mover
CC	80.0000	83.0000
CD	3.0000	0.0000
DC	3.0000	0.0000
DD	14.0000	17.0000

Table 1b The proportion of time the two players converge to each of the four possible states when they play the simultaneous prisoner's dilemma game with two actions

State	Player 1	Player 2
CC	31.5000	31.5000
CD	2.0000	2.5000
DC	2.5000	2.0000
DD	64.0000	64.0000

Table 1c The mean payoffs for the entire game and over the last 100 episodes received by the two players in the simultaneous and sequential versions of the Prisoner's Dilemma Game, with standard deviations shown in parentheses. The table illustrates that both players receive higher mean payoffs both overall and in the last 100 episodes of the game in the sequential prisoner's dilemma game than in the simultaneous game

	Overall mean Player 1	Last 100 episodes	Overall mean Player 2	Last 100 episodes
Sequential	0.267681 (0.068274)	0.269 (0.069551)	0.267638 (0.068329)	0.269 (0.069551)
Simultaneous	0.17136 (0.094898)	0.171 (0.096708)	0.168904 (0.092163)	0.1685 (0.09392)

agents adapt to one another's moves and learn the payoff function, there is a decrease in the occurrences of the state CD for the first-mover in the sequential game. This indicates that either the first-mover cooperates less often upon adapting, or that the second-mover adapts to defect less on seeing a cooperative move by the first-mover. The former argument can be discarded, because of the increase in occurrences of the state CC for the second-mover, implying that either one or both of the agents have increased their rate of cooperation. Consequently, we argue that the second-mover learns the high payoffs associated with cooperation and therefore learns to defect less often on seeing a cooperative move by the first-mover.

A comparison of Tables 1a and 1b shows that there is more cooperation in the sequential version than in the simultaneous version of the iterated prisoner's dilemma game; specifically, the proportion of time Player 1 finds himself in the cooperative (CC) state is close to 80% in the sequential version but only about 31.5% in the simultaneous version. As a result, the mean payoffs (see Table 1c) received by the two players in the sequential version are (0.268, 0.268), which are much larger than in the simultaneous version, (0.171, 0.169). This leads us to our next observation.

Observation 2 The informational flow in the iterated SPD game, where the second-mover has knowledge of the first-mover's action before choosing his own, results in the second-mover learning the optimal state-action function faster than in the simultaneous game.

Theories of observed cooperation in sequential games attribute cooperation to reciprocal behavior, whether positive or negative. Reciprocal behavior is defined as "…an action that would not otherwise be taken [in a given situation] and if it is undertaken in response to the action of another" (Cox & Deck, 2005). Sufficient evidence suggests that cooperation on the part of players is not motivated by pure altruism but by positive reciprocity. For instance, McCabe and Smith (2000) observe reciprocity in a trust game while Guth (1995) observes reciprocal fairness in the ultimatum game.

Since the players in our experiments are artificial agents and we do not explicitly model reciprocal behavior, we cannot attribute increased cooperation to reciprocal behavior. Instead, we hypothesize that the reason for increased cooperation in the iterated SPD game is the second-mover's knowledge of the first-mover's action, which eliminates the uncertainty and associated risk of a cooperative move. In the simultaneous game, Player 2 (who is the second-mover in the sequential game) determines the action to take in any given time period based on the simultaneous actions taken by himself and his opponent in the previous period. In the sequential game, however, Player 2 (second-mover) determines his action based on his previous action and Player 1's (first-mover) action in the current period, where the first-mover's action can be interpreted as a direct effect of the second-mover's action in the previous period. Consequently, if indeed Player 1's actions are based on the past behavior of Player 2, it is more difficult for Player 2 to decipher this relationship between his actions and those of Player 1 in a simultaneous game. However, in the sequential game, by having knowledge of the action taken by Player 1 (first-mover) in the current period, Player 2 (second-mover) is more easily able to discern

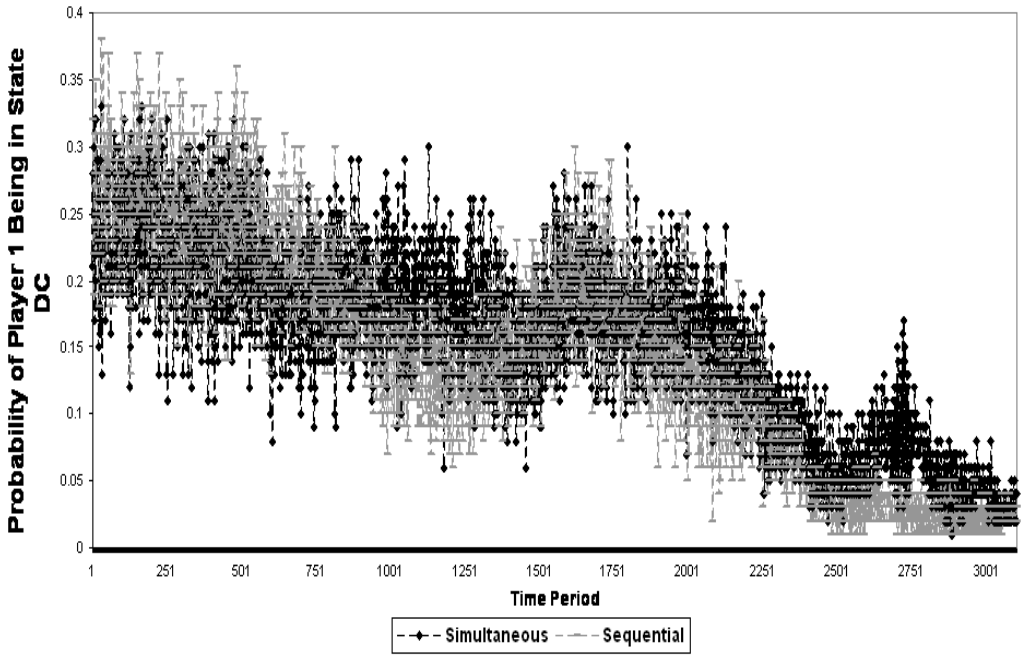

Figure 2 Comparison of the probability of Player 1 (first-mover in the sequential game) being in state DC in the iterated SPD game and the iterated simultaneous game. The figure shows that Player 2 learns to defect faster upon seeing defection by Player 1 in the sequential game than in the simultaneous game.

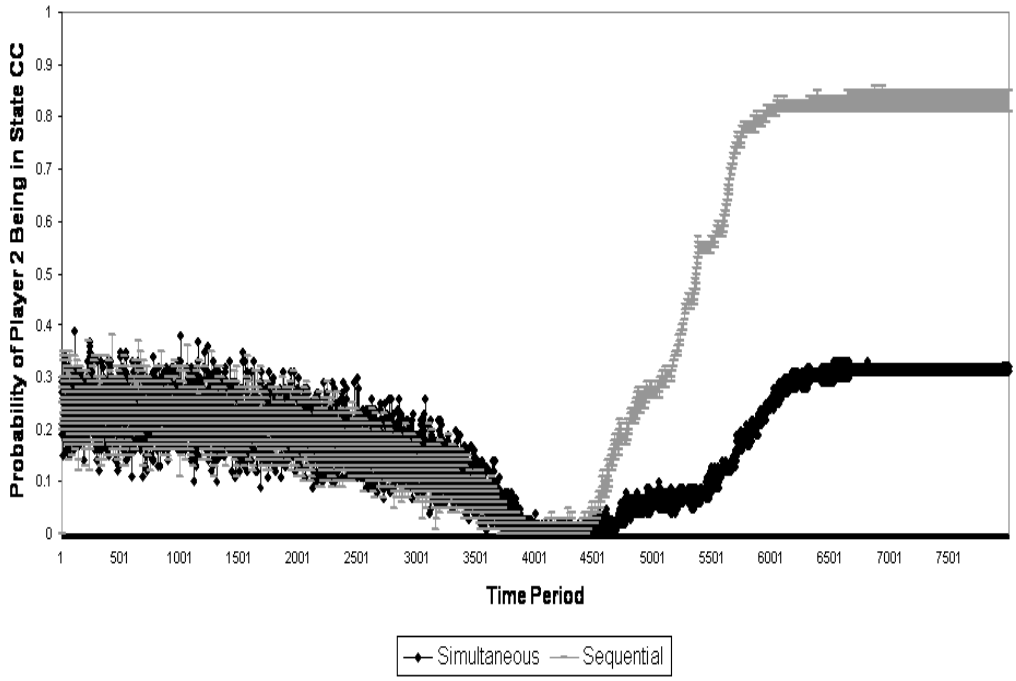

Figure 3 Comparison of the probability of Player 1 (first-mover in the sequential game) being in state CC in the iterated SPD game and the iterated simultaneous game. The figure shows that Player 2 learns the benefits of cooperation and learns to cooperate faster upon seeing a cooperative move by Player 1 in the sequential game than in the simultaneous game.

the consequences of his actions. In other words, the players (artificial agents) are able to discern the benefits of cooperation in the sequential game even though neither agent has knowledge of the payoff structure.

In order to determine the increase in the rate of cooperation in the iterated SPD game, we compared the percentages of visits to each state as play progresses in the sequential and simultaneous games. Our findings indicate Player 2 (second-mover) learns to defect when Player 1 (first-mover) defects sooner in the sequential game than in the simultaneous game. In the simultaneous game, Player 2 faces an uncertainty in the action that Player 1 is going to take and so might end up cooperating when Player 1 defects. However, in the sequential game uncertainty regarding Player 1's action is eliminated by the information received by Player 2, who is thus able to more quickly learn the low payoff associated with cooperating when Player 1 defects. Additionally, since Player 2 learns the disadvantages of cooperating upon seeing Player 1 defect, Player 1 learns that he will not be able to exploit Player 2. Figure 2 compares the probability of Player 1 finding himself in State DC (the condition where Player 1 defects and Player 2 cooperates) in the sequential game and in the simultaneous game.[3] As the figure shows, with time the artificial agents adapt and in the sequential game Player 2 learns that cooperating upon seeing defection by Player 1 results in low payoffs for Player 2. Although the probability of Player 1 finding himself in State DC in both games decreases, the trend is for Player 2 to learn faster in the sequential game. Furthermore, the faster learning in the sequential game also applies to Player 2 learning the benefits of cooperating when Player 1 cooperates. In Figure 3 we compare the probability of Player 2 finding himself in State CC, the socially efficient state, in the sequential and simultaneous games. Due to the elimination of the uncertainty about Player 1's action in the sequential game, Player 2 is able to more quickly discern the benefits of cooperating upon seeing a cooperative move by Player 1.

We have thus illustrated how the informational flow of Player 1's action to Player 2 in the sequential game decreases the risk faced by Player 2 in cooperating. The reduction in risk, in turn, allows Player 2 in the sequential game to adapt and learn the higher payoffs associated with cooperating. Thus, we observe a higher rate of cooperation in the iterated sequential prisoner's dilemma game than in the simultaneous version of the game.

We will now study the impact of different factors on the observed rates of cooperation when adaptive artificial agents play the iterated SPD game. The first factor that we examine is the risk of defection faced by the second-mover.

Observation 3 When adaptive artificial agents play the iterated SPD game, cooperative behavior and the first-mover's relative bargaining position are affected by the payoff risks of defection faced by the second-mover.

We demonstrate below the significance of the bargaining power possessed by the second-movers in promoting convergence to the cooperative outcome. The basic structure of the prisoner's dilemma game is maintained and we modify only the payoff received by the second-mover when both players defect (see Table 2). Note that the increase in the payoff received by the second-mover for defecting when the first-mover defects results in a decrease in the payoff risk associated with defection for the second-mover. In addition, the difference between the payoff received by the second-mover when both players cooperate and that when both players defect decreases as well. Therefore,

Table 2 Payoff table for the modified prisoner's dilemma game where Player 2 (the second-mover) faces a lower payoff risk when he defects than in the original prisoner's dilemma game (Table 1)

	C	D
C	0.3, 0.3	0.0, ***0.5***
D	***0.5***, 0.0	***0.1***, ***0.2***

Table 2a The proportion of time the two players converge for each of the four possible states when they play the modified sequential prisoner's dilemma game with the payoff structure given in Table 2

State	Player 1 – First-mover	Player 2 – Second-mover
CC	67.000	68.000
CD	1.000	0.000
DC	1.000	0.000
DD	31.000	32.000

Table 2b The mean payoffs for the entire game and the last 100 episodes received by the two players in the modified prisoner's dilemma game (Table 2) when the risk of defection faced by Player 2 decreases, with standard deviations in parentheses

Overall mean Player 1	Last 100 episodes	Overall mean Player 2	Last 100 episodes
0.235978 (0.090972)	0.237 (0.092556)	0.267688 (0.045512)	0.268 (0.046341)

the relative bargaining position of the first-mover to induce cooperative behavior decreases and, since the second-mover has less incentive to cooperate, the overall rate of cooperation decreases. Table 2a shows the proportion of time periods for which the two players are in each of the four different states once adaptation is completed and their behavior has converged for the game with the payoff structure outlined in Table 2. The results indicate a dramatic decrease in the observed rate of cooperation in the modified iterated SPD game versus the classical iterated SPD game (Table 1). Cooperation under the original payoff structure (Table 1) is close to 80%, whereas under the modified payoff structure the rate of cooperation decreases to 67%. In addition, as shown in Table 2b the mean payoffs received by Player 1 in the original payoff structure and the modified structure are 0.268 and 0.238, respectively.

Keeping with the theme of the bargaining power, we will now analyze the relationship between the temptation payoff and the rate of cooperation. We experimented with the temptation payoff for the second-mover and found an inverse non-linear relationship between rates of cooperation and temptation levels (see Tables 3, 3a, 4 and 4a).

Observation 4 There is a non-linear relationship between the risk faced by the second-mover (cooperating incentive) and the observed cooperation rate when adaptive agents play the iterated SPD game.

When we increase the temptation of the second-mover from the base level of 0.5 to 0.7, we observe a decrease in the proportion of time periods in which he finds himself in the cooperative state. The decrease is about 12% (Table 3a). However, when the temptation is increased to 0.9, the decrease in the proportion of cooperative state is approximately 28% (Table 4a). While the increase in the temptation payoff is only doubled, the decrease in the percentage of the cooperative state is more than double.

Table 3 Payoff table for the modified prisoner's dilemma game where Player 2 (the second-mover) has less incentive to cooperate than in the original prisoner's dilemma game (Table 1) because of the greater temptation payoff associated with defecting when Player 1 (the first-mover) cooperates

	C	D
C	0.3, 0.3	0.0, ***0.7***
D	***0.5***, 0.0	***0.1***, ***0.1***

Table 3a The proportion of time the two players converge for each of the four possible states when they play the sequential prisoner's dilemma game with the payoff structure shown in Table 3

State	Player 1 – First-mover	Player 2 – Second-mover
CC	68.0000	72.5000
CD	4.5000	0.0000
DC	4.5000	0.0000
DD	23.0000	27.5000

Table 4 Payoff table of the modified prisoner's dilemma game where Player 2 (the second-mover) has less incentive to cooperate than in the original prisoner's dilemma game (Table 1) because of the greater temptation payoff associated with defecting when Player 1 (the first-mover) cooperates

	C	D
C	0.3, 0.3	0.0, ***0.9***
D	***0.5***, 0.0	***0.1***, ***0.1***

We next tested the significance of perfect information in promoting cooperation. We conducted two sets of simulations with a noise level of 10%; this

Table 4a The proportion of time the two players converge for each of the four possible states when they play the sequential prisoner's dilemma game with the payoff structure given in Table 4

State	Player 1 – First-mover	Player 2 – Second-mover
CC	38.0000	55.5000
CD	17.5000	0.0000
DC	17.5000	0.0000
DD	27.0000	44.5000

level of noise was chosen based on the tolerance level of subjects observed in the experiments of Huck and Miller (2000). In the first set of simulations, the action taken by the first-mover is known to the second-mover with 10% uncertainty, while in the second set the action taken by the second-mover is known to the first-mover at the end of each period with 10% uncertainty.

Observation 5 Uncertainty in the second-mover's action choice decreases cooperation more than the same level of uncertainty in the first-mover's action choice when adaptive agents play the iterated SPD game.

Table 5 contains the means and standard deviations of the payoffs received by the two players under the two types of noise. We find that uncertainty in the knowledge to the first-mover of the action taken by the second-mover decreases the payoffs for the two agents much more than uncertainty in the knowledge about the action choice taken by the first-mover transmitted to the second-mover. Our reasoning is that the noisy signal of the second-mover further increases the uncertainty faced by the first-mover, which increases the risks of exploitation faced by him and also hinders his adaptive learning process.

We conclude with a summary of our work and provide directions for future extensions in the next section.

6 Conclusions and Future Work

In this article we modeled artificial agents with reinforcement learning algorithms so that they can learn to strategically interact. Although our focus was on the broad class of principal–agent games, we conducted the simulations and analysis while focusing on a specific game; namely, the iterated sequential prisoner's dilemma game. The iterated SPD game closely resembles the simplified version of the principal–agent game. Play in principal–agent problems is also sequential with the second-mover (agent) having an advantage, since he has knowledge of the first-mover's (principal) action before he makes his decision. Players act in their best interests and also face a conflict of interest, thus causing moral hazard problems that prevent the socially efficient equilibrium, which is the cooperative solution, from being reached. Principal–agent relationships are a common occurrence; and the associated problems arise whenever a principal depends on an agent for production of goods, services rendered, or for the completion of some other tasks.

To summarize, the research described in this article has demonstrated the ability of artificial agents to learn to cooperate over time in the iterated SPD game, where classical game theory predicts defection as the dominant strategy. Our simulation results agree with experimental findings that observe high rates of cooperation in the iterated SPD game among human sub-

Table 5 The mean payoffs for the entire game and over the last 100 episodes received by the two players under the two types of noise in the sequential version of the prisoner's dilemma game (Table 1), with the standard deviations in parentheses

	Overall mean Player 1	Last 100 episodes	Overall mean Player 2	Last 100 episodes
First-mover action affected by noise	0.25249 (0.045035)	0.251 (0.049633)	0.252716 (0.050103)	0.25135 (0.054975)
Second-mover action affected by noise	0.198074 (0.054136)	0.19935 (0.056506)	0.19088 (0.094344)	0.192 (0.097646)

jects. Researchers studying the behavior of human subjects playing the iterated SPD game in laboratory experiments have attributed reciprocal behavior to issues of fairness and equity. We argue that the reason for observing greater levels of cooperation or reciprocal behavior in the sequential versus the simultaneous iterated prisoner's dilemma game is due to the additional knowledge received by the second-mover. This additional knowledge eliminates the second-mover's uncertainty about the first-mover's action, which decreases the risk of exploitation and is conducive to agents learning the socially optimal solution.

We also investigated the impact that different factors of the structure of the game have on the observed rates of cooperation. For instance, we find that uncertainty in the action taken by the second-mover has a greater impact on cooperation than the uncertainty in the flow of information from the first-mover to the second-mover about the former's action choice.

In terms of future work, it would be interesting to compare our results with those achieved using different models of learning such as other reinforcement learning algorithms or belief-based systems. It would also be valuable to study the behavior of artificial agents under more complicated models of principal–agent problems. Additionally, we could study the impact of other factors in the structure of the game on observed rates of cooperation. For instance, we could introduce uncertainty in the payoffs by adding a random error factor every period to the payoffs and study the effectiveness of learning in such an environment. We could also introduce costs to the second-mover for observing the action of the first-mover and thereby study the value of the first-mover's action to the second-mover under different game scenarios. Finally, we would also like to corroborate our hypotheses with experimental work.

Notes

1 The hypotheses and corresponding simulation results assume that the artificial agents are modeled using an RL algorithm, even where this is not explicitly stated.

2 The 2x2 games analyzed by Yang et al. are different from ours. However, we have defined risk in our article along the lines of their definition while making the corresponding modifications.

3 Note that the probabilities presented in Figures 2 and 3 are the average of 100 games with different random seeds.

References

Andreoni, J., & Miller, J. H. (1993). Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence. *The Economic Journal, 103*, 570–585.

Ashlock, D., Smucker, M. D., Stanley, E. A., & Tesfatsion, L. (1996). Preferential partner selection in an evolutionary study of prisoner's dilemma. *BioSystems*, *37*(1–2), 99–125.

Axelrod, R. (1984). *The evolution of cooperation*. New York: Basic Books.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science, 211*, 1390–1396.

Bolle, F., & Ockenfels, P. (1990). Prisoner's dilemma as a game with incomplete information. *Journal of Economic Psychology, 11*, 69–84.

Camerer, C., & Ho, T. (1999). Experience-weighted attraction learning in normal form games. *Econometrica, 67*(4), 827–874.

Clark, K., & Sefton, M. (2001). The sequential prisoner's dilemma: Evidence on reciprocation. *The Economic Journal, 111*, 51–68.

Cox, J. C., & Deck, C. A. (2005). On the nature of reciprocal motives. *Economic Inquiry, 43*(3), 623–635.

Duffy, J. (2001). Learning to speculate: Experiments with artificial and real agents. *Journal of Economic Dynamics and Control, 25*(3–4), 295–319.

Erev, I., & Roth, A. (1998). Predicting how people play games: Reinforcement learning in games with unique strategy equilibrium. *American Economic Review, 88*, 848–881.

Fang, C., Kimbrough, S. O., Pace, S., Valluri, A., & Zheng, Z. (2002). On adaptive emergence of trust behavior in the game of stag hunt. *Journal of Group Decision and Negotiation, 11*(6), 449–467.

Feltovich, N. (2000). Reinforcement-based vs. beliefs-based learning in experimental asymmetric-information games. *Econometrica, 68*, 605–641.

Grossman, S. J., & Hart, O. D. (1983). An analysis of principal–agent problem. *Econometrica, 51*, 7–45.

Guth, W. (1995). On ultimatum bargaining experiments—a personal review. *Journal of Economic Behavior and Organization, 27*, 329–344.

Holmstrom, B. (1979). Moral hazard and observability. *The Bell Journal of Economics, 10*(1), 74–91.

Huck, S., & Miller, W. (2000). Perfect versus imperfect observability—an experimental test of Bagwell's result. *Games and Economic Behavior, 31*, 174–190.

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research, 4*, 237–285.

Littman, M. L. (1994). Markov games as a framework for multi-agent reinforcement learning. *Proceedings of the Eleventh International Conference on Machine Learning* (pp. 157–163). San Francisco: Morgan Kaufmann.

Littman, M. L. (2001). Friend-or-foe Q-learning in general-sum games. *Proceedings of the Eighteenth International Conference on Machine Learning* (pp. 322–328). San Francisco: Morgan Kaufmann.

McCabe, K., & Smith, V. (2000). A two-person trust game played by naïve and sophisticated subjects. *Proceedings of the National Academy of Sciences, 97*(7), 3777–3781.

McKelvey, R., & Palfrey, T. (1992). An experimental study of the centipede game. *Econometrica, 60*(4), 803–836.

Meidinger, C., & Terracol, A. (2002). Reciprocation and reinforcement learning model in the investment game. Paper presented at the *19th Conference on Journees de Micro-economic Appliquee*.

Miller, J. H. (1996). The coevolution of automata in the repeated prisoner's dilemma. *Journal of Economic Behavior and Organization, 29*(1), 87–112.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economics behavior.* Princeton, NJ: Princeton University Press.

Nowak, M. A. (1990). Stochastic Strategies in the Prisoner's Dilemma. *Theoretical Population Biology, 38*, 93–112

Nowak, M. A., & May, R. M. (1992). Evolutionary games and spatial chaos. *Nature, 359*, 826–829.

Ochs, J., & Roth, A. E. (1989). An experimental study of sequential bargaining. *American Economic Review, 79*, 355–384.

Oskamp, S. (1971). Effects of programmed strategies on cooperation in the prisoner's dilemma and other mixed-motive games. *Journal of Conflict Resolution, 15*, 225–229.

Poundstone, W. (1993). *Prisoner's dilemma: John von Neumann, game theory, and the puzzle of the bomb.* Oxford, UK: Oxford University Press.

Radner, R. (1981). Monitoring cooperative agreements in a repeated principal–agent relationship. *Econometrica, 49*, 1127–1148.

Radner, R. (1985). Repeated principal–agent games with discounting. *Econometrica, 53*(5), 1173–1198.

Radner, R. (1986). Repeated partnership games with imperfect monitoring and no discounting. *The Review of Economic Studies, 53*(1), 43–57.

Ross, S. (1973). The economic theory of agency: The principal's problem. *American Economic Review, 63*, 134–139.

Rubinstein, A. (1986). Finite automata in the repeated prisoner's dilemma. *Journal of Economic Theory, 39*, 83–96.

Sandholm, T. W., & Crites, R. H. (1996) Multi-agent reinforcement learning in the iterated prisoner's dilemma, *BioSystems, 37*(1–2), 47–166.

Selten, R., & Stoecker, R. (1986). End behavior in sequences of finite prisoners' dilemma super-games. *Journal of Economic Behavior and Organization, 7*, 47–70.

Shoham, Y., & Tennenholtz, M. (1993). Co-learning and the evolution of coordinated multi-agent activity.

Shubik, M. (1970). Game theory, behavior, and the paradox of the prisoner's dilemma: Three solutions. *Journal of Conflict Resolution, 14*, 181–194.

Suguwara, T., & Lesser, V. (1993). On-line learning of coordination plans. *Computer Science Technical Report* 93–27, Amherst, MA: University of Massachusetts.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Tesauro, G. J. (1992). Practical issues in temporal difference learning. *Machine Learning, 8*, 257–277.

Tesfatsion, L. (2002). Agent-based computational economics: Growing economies from the bottom up. *Artificial Life, 9*(1), 55–82.

Watkins, C. J. C. H. (1989). *Learning from delayed rewards*. PhD Thesis, Cambridge University, UK.

Wu, D. J., Kimbrough, S., & Zhong, F. (2002). Artificial agents play the mad mex trust game: A computational approach. In R. H. Sprague, Jr. (Ed.), *Proceedings of the Thirty-fifth Annual Hawaii International Conference on System Sciences (HICSS-35, CD ROM)*, Los Alamitos, CA: IEEE Computer Society Press (pp. 335–347).

Yang, C., Weimann, J., & Mitropoulos, A. (2001). Game structure and bargaining power in sequential mini-games: An experiment.

Zhong, F., Kimbrough, S. O., & Wu, D. J. (2002). Cooperative agent systems: Artificial agents play the ultimatum game. *Journal of Group Decision and Negotiation, 11*(6), 433–447.

About the Author

Annapurna Valluri will receive her Ph.D. in operations & information management from The Wharton School of Business in August 2006. Her thesis focused on studying the impact of information technology on the outsourcing decisions of firms. Valluri's research has involved employing computational methodologies to analyze economic/business decisions, where artificial agents modeled with learning capabilities determine their optimal or efficient strategic decisions through repeated interactions. She obtained her Master's degree in managerial science and applied economics from The Wharton School of Business and a dual degree in computer science and finance from the University of Maryland, College Park. *Address*: Suite 500, Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104, USA. *E-mail*: avalluri@wharton.upenn.edu