2011 IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing

The Life Game: Cognitive Strategies for Repeated Stochastic Games

Kan-Leung Cheng
Dept. of Computer Science
University of Maryland
College Park, MD, USA

Inon Zuckerman
Dept. of Industrial Engineering
and Management
Ariel University Center of Samaria
Ariel, 44837, Israel

Dana Nau
Dept. of Computer Science
University of Maryland
College Park, MD, USA

Jennifer Golbeck
College of Information Studies
University of Maryland
College Park, MD, USA

***Abstract*—Standard models in bio-evolutionary game theory involve repetitions of a single stage game (e.g., the Prisoner's Dilemma or the Stag Hunt); but it is clear that repeatedly playing the *same* stage game is not an accurate model of most individuals' lives. Rather, individuals' interactions with others correspond to many different kinds of stage games.**

In this work, we concentrate on discovering behavioral strategies that are successful for the *life game*, in which the stage game is chosen stochastically at each iteration. We present a cognitive agent model based on Social Value Orientation (SVO) theory. We provide extensive evaluations of our model's performance, both against standard agents from the game theory literature and against a large set of life-game agents written by students in two different countries. Our empirical results suggest that for life-game strategies to be successful in environments with such agents, it is important (i) to be unforgiving with respect to trust behavior and (ii) to use adaptive, fine-grained opponent models of the other agents.

***Keywords*-repeated games, non-zero-sum games, stochastic games, social value orientation**

I. INTRODUCTION

An interesting puzzle in the evolution of human societies is the dissonance between Darwin's principle of *natural selection* and cooperative actions commonly observed in human (and animal) societies. A prominent way to study this problem has been to use *repeated games*, to discover its equilibrium properties under different environmental properties, starting conditions, and reproduction mechanisms.

In the standard repeated-game model, a set of agents repeatedly play a game called the *stage game*. Many different games can be used as the stage game. For example, Axelrod's famous Iterated Prisoner's Dilemma competitions showed the emergence of cooperation, even though the rational dominant equilibrium in a one-shot Prisoner's Dilemma is to defect [3]. Maynard Smith studied two-player Chicken game with a population of Hawks and Doves [14], and Skyrms studied the evolved population when individuals were playing the Stag-hunt game [17].

Each of the above studies used a simple game model in which the same stage game was used at every iteration. However, as pointed out by Bacharach [5, p. 100], repeatedly playing the **same** game is unlikely to be an accurate model of any individual's life. As more accurate model, Bacharach proposed the *life game*, in which an individual plays a mixture of games drawn sequentially according to some stochastic process from a large set of stage games. Bacharach referred to the size and variety of this set as the game's *ludic diversity* (thus an ordinary non-stochastic repeated game has minimal ludic diversity).

Prisoner's Dilemma		Player 2	
		A_1=Cooperate	A_2=Defect
Player 1	A_1=Cooperate	$(3,3)$	$(0,5)$
	A_2=Defect	$(5,0)$	$(1,1)$

Fig. 1. Prisoner's Dilemma payoff matrix.

In this paper, we concentrate on discovering behavioral strategies that are successful in life games of high ludic diversity. These games pose difficulties when trying to describe a successful strategy. For example, well-known strategies such as the famous *tit-for-tat* strategy cannot be used verbatim, because not all iterations will have actions in which the actions correspond to "cooperate" and "defect." The complexity of the game dictates a large, complex strategy space, but our objective is to discover important general properties that characterize successful strategies.

This paper makes the following contributions. We formally describe the life game (which Bacharach only described informally), and discuss its inherent challenges. We then propose a cognitive behavioral model for agents in the life game, based upon a prominent social-preference theory called *Social Value Orientation* theory (SVO). We also refine and extensively evaluate our model using a large set of peer designed agents written by students in two different countries. Our empirical results suggest that an unforgiving strategy performs better than a tit-for-tat-like strategy. That is, in stage games where there are analogs of "cooperate" and "defect" (as in the Prisoner's Dilemma), if another agent chooses the "defect" action rather than the "cooperate" action, then we should expect them to behave similarly in future iterations, and choose our actions accordingly. The empirical work also demonstrates the importance of an adaptive, fine-grained set of opponent models in successful strategies.

978-0-7695-4578-3/11 $26.00 © 2011 IEEE
DOI

II. Background

There have been many studies on iterated games in the game theory literature. The most famous example is the *Iterated Prisoner's Dilemma (IPD)* (see Fig. 1), which is an iterated variant of the Prisoner's Dilemma that is played repeatedly an unknown number of iterations. The Prisoner's Dilemma is a widely used model for social dilemmas between two agents and has been often used to study theories of human cooperation and trust. The intriguing characteristic of the IPD is that while game theory analysis for a single iteration suggests that rational agents should "defect", cooperation often emerges when the number of iterations is unknown.

One of the first interesting questions with respect to the IPD was the discovery and description of successful IPD strategies. These strategies and their properties were meant to help enrich the theoretical biology/evolutionary discussion on various mechanisms that complement the Darwinian process (for instance: reciprocity, kin selection, group selection). An important milestone was the completion of two publicly open IPD tournaments that were run by Robert Axelrod in the early 80s [3]. In his tournament, each strategy was paired with each other strategy for 200 iterations of a Prisoner's Dilemma, and scored on the total payoffs accumulated through the tournament. The winner of the first tournament was Anatol Rapoport's tit-for-tat strategy, which simply cooperates on the first iteration of the game, and then repeats the other agent's action from the previous iteration. Surprisingly, the same tit-for-tat strategy was also the winner in the second tournament.

Axelrod, in his post tournaments analysis, discovered that greedy strategies tended to do very poorly in the long run while cooperative strategies did better. Furthermore, by analyzing the top-scoring strategies in the tournament, Axelrod presented several properties that describe successful strategies: nice (cooperate, never be the first to defect), provocable to both retaliation and forgiveness (return defection for defection, cooperation for cooperation), non-envious (be fair with your partner), and clarity (don't try to be tricky). Since Axelrod's IPD tournaments, there has been an extensive research on finding and describing successful strategies [6], [15].

The most relevant piece of literature to our study is a recent paper [7] where the authors presented an equilibrium analysis for the emergent of cultures when playing multiple games. Nevertheless, they were not concerned with the success of individual strategies, and assumed a predefined set of 6 games with explicitly labeled actions to avoid the semantic problem.

As our main focus of this paper is the life game, in which a different stage game can be played at each iteration, there is one crucial assumption behind the above strategies: they assume that the semantic of the actions is a common knowledge to all agents. For example, in order to reciprocate or retaliate, a tit-for-tat agent needs to know the semantic of the previous action of the other agent from the perspective of the other agent. This assumption is valid for IPD, because the semantic (Cooperate and Defect) is clearly defined for all agents. However, in the life game, which is the main focus of this paper, this assumption is no longer valid. As we will see below (in Section IV), most known strategies simply cannot be generalized to the complex world of the life game, and consequently, new ones must be defined.

2x2 symmetric game		Player 2	
		A_1	A_2
Player 1	A_1	(a, a)	(b, c)
	A_2	(c, b)	(d, d)

Fig. 2. Generalized form of 2x2 symmetric games.

III. The life game model

Our *life game* model is defined as a set of iterated and symmetric 2x2 normal-form games, where the two agents are denoted as P_1 and P_2. In every iteration, a symmetric random game will be generated, and the agent's strategy space will be composed of two actions, namely A_1 and A_2. The game is a complete information game, as each agent knows the complete payoff matrix, and can compute the payoffs for each combination of actions. As a normal form games, both agents need to simultaneously choose one action from the random game. After deciding on the actions, each agent will be notified on the attained payoffs and the action chosen by the other agent. The two agents will play the games in succession, without knowing when the series of games will end. Each game is randomly and independently generated. That is, the randomized matrix in round $n - 1$ does not have any impact on the matrix generated for round n. We do not place any restrictions on the agents' memory, and they may record past matrices and the actions taken by both agents and utilize it in their strategy.

In this paper, we used a random matrix in which the payoff values were chosen uniformly from $[0, 9]$. This kind of random game is generalized enough to represent most of the interesting games found in the game theory literature. Among the well-known examples are the Prisoner's Dilemma, Chicken Game, and Stag-Hunt [16]. Fig. 2 shows a generalized form of the payoff matrix for such games, where various constraints on the payoff values can be used to define different classes of social dilemmas. Note that the semantics of the actions depend on the value of a, b, c and d. For example, if $a = 3$, $b = 0$, $c = 5$ and $d = 1$ (a Prisoner's Dilemma), then A_1 and A_2 can be considered as "Cooperate" and "Defect" respectively.

Similarly to the IPD, in the *life game* competition each agent will compete against all the agents in the population, once against each agent in a round-robin fashion. To eliminate random favorable payoff variations, we randomized the series of games, and used the same series between all agents in the population. The performance of an agent is the accumulated sum of its payoffs with each of the other agents.

IV. Strategies for the Life Game

The IPD competition was an important cornerstone for studying the evolution of cooperation and led to some interesting game strategies. However, extending the model to the life

Stag Hunt		Player 2	
		A_1=Stag	A_2=Hare
Player 1	A_1=Stag	$(2, 2)$	$(0, 1)$
	A_2=Hare	$(1, 0)$	$(1, 1)$

Fig. 3. The Stag Hunt game models two individuals go out on a hunt. If an individual hunts a stag, he must have the cooperation of his partner in order to succeed. An individual can get a hare by himself, but a hare is worth less than a stag.

game, which is a more realistic description of the interactions in a society, raises the following difficulties. First, from the semantic point of view, unlike the Prisoner's Dilemma in which actions are labeled by "cooperate" or "defect", in the life game the actions are not labeled in advance. The agents will need to define themselves the semantic of each of the actions in each round of the game. Consequently, the intentions behind the actions might be misinterpreted due to semantic differences, which also complicates the playing strategies. For example, what might look like a "cooperate" action for one agent, might be interpreted differently by another. Secondly, the semantic problems might also result in ambiguity with respect to the **intentions** behind the actions, as the agent cannot be sure whether an action is a result of an intentional strategic decision, or due to semantic differences. As such, successful strategies might require holding some form of opponent model that can be reasoned upon for issues such as mutual trust, cooperation and counter strategies.

To illustrate the problem, consider two tit-for-tat-like agents (P_1 and P_2) playing in a repeated game of Stag Hunt (Fig. 3). Suppose both of them want to cooperate (hunt stag together) in the Stag Hunt game, but P_1 does not want to cooperate in the Prisoner's Dilemma (while P_2 still want to cooperate). If P_1 and P_2 play in repeated sequence of only Stag Hunt games, they will cooperate with each other forever. However, the cooperation in Stag Hunt may not emerge if we have a mix of Prisoner's Dilemma and Stag Hunt games. For example, if the first game is Prisoner's Dilemma and the second game is Stag Hunt, when P_1 defects P_2 in the first game, P_2 retaliates by "defecting" in the Stag Hunt game (i.e., hunt Hare). Therefore, P_1 will also retaliate in the next game that may lead to a chain of unnecessary retaliation.

The aforementioned difficulties as well as others bring about the need for strategies that are far more complex than the ones that came out of research on the traditional IPD. Intuitively, simple strategies such as tit-for-tat cannot be directly applied to the life game due to the labeling problem mentioned above. Our first step in developing a strategy was to look in the social and behavioral sciences literature and examine the behavioral theories that guide human behaviors in similar situations.

V. Social Value Orientation Agent Models

In the social and behavioral sciences it is widely accepted that agents explicitly take into account the outcome of the other agents' actions when considering their course of action. Moreover, the choices people make depend, among other things, on stable personality differences in the manner in which they approach interdependent others. This observation can be traced back to the seminal work by Messick and McClintock [9] in which they presented a motivational theory of choice behavior that considers both agents' payoffs in game situations. This theory was later denoted as the *Social Value Orientation* (SVO) theory, that has since developed into a class of theorems (see [8] for an excellent review).

SVO regards social values as distinct sets of motivational or strategic preferences with the weighting rule depending on the weights w_1 and w_2 of agents' payoffs:

$$\text{Utility of } P_1 = w_1 \times P_1\text{'s payoff} + w_2 \times P_2\text{'s payoff}$$

- Cooperative agent maximizes joint outcome. ($w_1 = 1, w_2 = 1$)
- Individualistic agent maximizes its own outcome. ($w_1 = 1, w_2 = 0$)
- Competitive agent maximizes its own outcome relative to other. ($w_1 = 1, w_2 = -1$)

In order to promote cooperation, both agents need to be prosocial. As mentioned in [4], "An excellent way to promote cooperation in a society is to teach people to care about the welfare of others". While various techniques to measuring the SVO values in human subjects were suggested over the years, a recent meta-analysis suggests that most people are classified as cooperators (46%), followed by individualists (38%), followed by competitors (12%) [2].

However, due to possible differences in the semantic of the games, both agents should have some way to assess mutual trust in order to deal with cases in which the different semantic interpretation were the cause of cooperation breakdown (as oppose to *intentional* breakdown). In other words, both agents need to believe that the other agent is prosocial. From the social and behavioral literature we learn that social value orientations significantly accounts for variation in trust and reciprocity. Specifically, pro-social individuals reciprocate more as the trust increases, while pro-self reciprocate less as the trust increases [10]. People with a natural inclination to cooperate are at the same time vulnerable to being exploited.

The cognitive model that will be developed in our suggested agent will be based on the above insights from the social and behavioral sciences. The following sections describe the agent model for the three most common social orientations in real world: cooperative, individualistic, and competitive.

A. Cooperative Model

A cooperative agent is one whose goal is to maximize (to some degree) the joint outcome of both agents. In the context of 2x2 symmetric games, a fully cooperative agent will choose A_1 if $a > d$ and A_2 if $a < d$. In IPD, the ALL C strategy, which always cooperates with others, can be regarded as a fully cooperative strategy. To account for the varying degrees of prosocial tendencies and cope with the aforementioned semantic problem, we need to be able to differentiate between different types of cooperative behavior. To do so we define the class of mutual-benefit games:

Definition 1 (Mutual-Benefit game): a mutual-benefit game is a 2x2 symmetric game in which there exist an unique action A_i such that the joint outcome is maximized when both agent choose A_i. Action A_i will be denoted as a cooperative action.

The varying degrees of prosocial tendencies suggest that different agents may want to restrict their cooperation to specific classes of mutual-benefit games. In general, agents with higher prosocial orientations will tend to cooperate on a larger subset of mutual-benefit games, as long as they believe that the other agent is also cooperative. We now present a possible classification to mutual-benefit games:[1]

1) $a \neq d$ and $max(a, d) > max(b, c)$
2) $a \neq d$ and $max(a, d) \geq max(b, c)$
3) $a \neq d$ and $max(a, d) \times 2 > b + c$
4) $a \neq d$ and $max(a, d) \times 2 \geq b + c$

In this classification, type i is a subset of type $i + 1$. For example, the Stag Hunt game (Fig. 3) is a member of all of the above types, while the Prisoner's Dilemma (Fig. 1) is a member of types 3 and 4 only.

In many types of symmetric games cooperation is beneficial for both agents in the long run. However, there are two major problems. First, a cooperative agent may subject to exploitation by the other agents. Second, the trustworthiness of the other agent is unknown at the beginning. Those problems will be addressed in the following trust mechanism.

We define the trustworthiness of an agent as follow: The trustworthiness of an agent is i if and only if the agent cooperates in **all** mutual-benefit games of type i. It is easy to notice that it is riskier to cooperate in type $i + 1$ games than in type i games. Accordingly, the type number of a mutual-benefit game can be considered as the trustworthiness requirement of the game in order to cooperate with the other agent. An agent will need higher trust levels to cooperate in type $i + 1$ games, while trustworthiness of zero reflects an agent that does not cooperate at all.

Recall that according to Axelrod's analysis of the IPD competition, a "nice" strategy helps to promote cooperation. Accordingly, our trust model will assume that the other agent is trustworthy at the beginning, and will remain so as long as it cooperates in all mutual-benefit games. Specifically, with the mutual benefit games classification presented above, we initialize the trustworthiness level of the other agent to 4.

To minimize exploitation, the trustworthiness of the other agent should be decreased whenever a defect-like action is observed. Suppose the current trustworthiness of the other agent is t. Whenever the other agent defects in a mutual-benefit game of type i, we update t by $t = min(t, i-1)$. For example, if the trustworthiness of an agent is updated to 3, then our agent will cooperate only in mutual-benefit games from type 1 to 3, but not type 4. This allows the agent to maximize the amount of cooperation, while minimizing exploitation.

When an untrusted agent (with low trustworthiness) try to establish cooperation in some mutual-benefit games, one may forgive it (increase its trustworthiness) or not forgive it (trustworthiness remains unchanged). We parameterize these behaviors by a forgiving threshold, f: The trustworthiness of an agent can be restored back to t when f cooperative actions in a game of type t were observed. In IPD, a SVO agent with $f = 1$ will behave like tit-for-tat. If $f = \infty$, an untrusted agent can never be trusted again. In other words, the trustworthiness of other agent is monotonically decreasing. This replicates the grim trigger strategy in IPD, which upon defection responds with defection for the remainder of the iterated game.

Chicken game		Player 2	
		A_1=Swerve	A_2=Straight
Player 1	A_1=Swerve	$(4, 4)$	$(3, 5)$
	A_2=Straight	$(5, 3)$	$(0, 0)$

Fig. 4. The Chicken game models two drivers, both headed for a single lane bridge from opposite directions. The first to swerve away yields the bridge to the other. If neither agent swerves, the result is a potentially fatal head-on collision.

B. Individualistic Model

According to the SVO theory, an individualistic agent will try to maximize its own outcome. However, the information that an agent P_i is a self maximizing agent is insufficient to model and predict its behavior, as its behavior will depend on its belief about the strategy of the other agent P_j. For instance, its actions might be different if it assumes P_j picks its actions randomly, or tries to intentionally decrease P_i's payoff.

To cope with this problem, we suggest using two-level agent modeling. In this model, when an individualistic agent P_i is playing with another agent P_j, P_i behavior depends on the second-level model – model of P_j from P_i's perspective. With that assumption, P_i can construct a best response strategy.

These behavior models will be input to the algorithm beforehand and will depend on the underlying game. We hypothesize that a larger and more diverge set of predefined models, will allow the SVO agent to better adapt its behavior (this will be explicitly tested in Section VI). For the life game, we can suggest the following types of second-level model which represents the simplest forms of opponent reasoning in this domain: adversary, altruistic, random, and recursive. We also present the best response strategy to each of them.

We illustrate it by following example: an individualistic agent P_1 is playing the Chicken game (Fig. 4) with P_2.

- **Adversary model** - P_1 assumes that P_2 wants to minimize its outcome. Then, it reasons that (1) P_2 will choose A_2 if P_1 chooses A_1; (2) P_2 will still choose A_2 if P_1 chooses A_2. The payoffs are $(3, 5)$ and $(0, 0)$ respectively, and P_1 will choose A_1. In other words, P_1 best response is to be playing a maximin strategy.
- **Altruistic model** - P_1 assumes that P_2 is wants to maximize P_1's outcome. Then, it reasons that (1) P_2 will choose A_1 if P_1 chooses A_1; (2) P_2 will still choose A_1 if P_1 chooses A_2. The payoffs are $(4, 4)$ and $(5, 3)$ respectively, and P_1 will choose A_2. In this case P_1's best response strategy is the maximax strategy.

[1] Note that the presented classification is one possible example of coping with the semantic problem. Naturally, a finer classification might allow the agent to distinguish between finer behavioral differences.

- **Random model** - P_1 assumes P_2 is purely random with 50% chance for both A_1 and A_2. This can happen, for example, in cases where it does not have enough information. The expected payoff of choosing A_1 is $\frac{a+b}{2} = 3.5$, and of choosing A_2 is $\frac{c+d}{2} = 2.5$. P_1 will choose A_1 only if $\frac{a+b}{2} > \frac{c+d}{2}$, and choose A_2 otherwise. Therefore, P_1 will choose A_1 in the Chicken game. We will call P_1 is playing a maxi-random strategy.
- **Recursive model** - Finally, P_1 can assume that P_2 is any kind of agent described above. P_1 will first predict P_2 action using that assumption, and then choose an action to maximize its own payoff. In other words, in terms of traditional game theory, given a game, P_1's strategy is the best response to the assumed P_2's strategy. For example, P_1 can assume that P_2 is an individualistic agent with random opponent assumption (i.e., P_2 uses the maxi-random strategy). From the previous paragraph, we know that P_2 will choose A_1 in the Chicken game. Therefore, P_1 will also choose A_1 in order to maximize its own payoff. We will call P_1 is playing a maxi-maxi-random strategy.

C. Competitive Model

According to the SVO theory, a competitive agent will try to maximize (to some degree) its own outcome with respect to the other agent. In the context of 2x2 symmetric games, this amounts to maximizing the payoff differences of both agents, and will choose A_1 if $b > c$ and A_2 if $b < c$.

When we sum up the total payoffs for each agent in a tournament of a group of agents, a competitive strategy is not necessary the best one. For example, in the IPD competition, a competitive agent acts like a ALL D agent which always defects. If there are only two agents, ALL D always perform at least as good as the other agent. However, ALL D performs poorly in a group of tit-for-tat agents, because the group of tit-for-tat agent will cooperate with each other and obtains a huge amount of payoff from the cooperation [4].

D. The Combined SVO Agent Modeling Strategy

Based on the SVO agent models present above, we propose a SVO agent modeling strategy for playing with other agent in the life game. The complete procedure for our SVO agent is shown in Fig. 5. Since we assume that all agents does not have any prior knowledge about the other agent, the SVO agent does not know the social orientation of the other agent. The agent will start with some default models, and will estimate the orientation of other agent from the history of interactions.

As we mentioned before, the agent starts by assuming that the other agent is cooperative for all types of mutual-benefit games. For non-mutual-benefit games, the cooperative agent model is not applicable. For those games, the SVO agent initially assumes the other agent is random (i.e., no social orientation at all) and will use the maxi-random strategy for the first few games. After accumulating some interaction histories, the agent will learn the true trustworthiness (i.e., t in Fig. 5) and social orientation of the other agent, and will adapt and utilize it to the best of its capacity.

```
Procedure SvoAgentPlayingLifeGame
Input:
    $i$ = current iteration number
    $g_i$ = current game matrix
    $g_{i-1}$ = previous game matrix
    $B_{i-1}$ = previous opponent's action
Output:
    $A_i$ = my action for the current game $g_i$
State:
    $M$ = current set of candidate opponent models
    $t$ = current trustworthiness of the opponent
Begin procedure
    (1) update opponent's trustworthiness and models
    if $g_{i-1} \neq \emptyset$ then
        if $g_{i-1}$ is a mutual-benefit game then
            $C_{i-1} \leftarrow$ cooperative action of $g_{i-1}$
            $t_{i-1} \leftarrow$ trustworthiness requirement of $g_{i-1}$
            if $t \geq t_{i-1}$ then
                if $B_{i-1} \neq C_{i-1}$ then
                    $t \leftarrow t_{i-1} - 1$
        if $t$ is not updated above then
            increase the counters of all correct models in $M$,
                which correctly predicts $A_{i-1}$ for $g_{i-1}$
    (2) choose an action for the current game $g_i$
    if $g_i$ is a mutual-benefit game then
        $C_i \leftarrow$ cooperative action of $g_i$
        $t_i \leftarrow$ trustworthiness requirement of $g_i$
        if $t \geq t_i$ then
            return $A_i = C_i$
    if $i \leq 5$ then
        return $A_i$ = maxi-random action in $g_i$
    $m \leftarrow$ the most accurate model in $M$
    if accuracy of $m$ < 70% then
        return $A_i$ = maxi-random action in $g_i$
    $B'_i \leftarrow$ the predicted action in $g_i$ of the opponent using $m$
    return $A_i$ = the best response to $B'_i$ in $g_i$
End procedure
```

Fig. 5. Procedure for an unforgiving SVO agent playing a life game.

Similarly to humans, as long as there is some degree of cooperation, our agent will cooperate with others as much as they cooperate with it. However, when the trust model suggests that the other agent is not cooperative in some mutual-benefit games or the game itself is a non-mutual-benefit game, one should refer to a different state of mind to achieve its goal while avoiding exploitation. To better estimate whether the other agent is an individualistic agent (under the different predefined models), or a competitive one, we incorporated opponent modeling techniques. Specifically, the SVO agent will use a model-and-counter strategy, which first approximates what strategy the other agent uses and then counters that strategy. First, it creates and maintains a pool of possible individualistic or competitive models (i.e., M in Fig. 5). In this paper, we consider the following five non-cooperative opponent models described before:

1) Competitive
2) Individualistic with maximin assumption
3) Individualistic with maximax assumption

4) Individualistic with maxi-random assumption
5) Individualistic with maxi-maxi-random assumption

Each model has a counter variable for counting the number of correct predictions. If the previous action of the other agent matches the prediction of one of the model, our agent will increase the counter of that model by one. The model with the highest counter is considered as the most accurate model (i.e., m in Fig. 5). However, if the top counter is small (e.g., less than 70% of the total) when compared with the total number of counted game, our agent will assume the opponent is a random agent instead of the model with the highest count, and will use the maxi-random strategy. After knowing the most accurate model of our opponent, our agent will try to counter that strategy by maximizing its own payoff using that opponent model, i.e., it first predicts opponent's action using the opponent model, and then it chooses an action which maximizes its own payoff assuming that the other agent will choose the predicted action (i.e., B_i' in Fig. 5).

VI. Experiments and Results

In this section our goal is to evaluate the performance of our SVO agent and investigate the properties of successful strategies in the life game. As such, we implemented an automated SVO based agent and in order to evaluate its performance we implemented the following agents that represent well-known strategies in the game theory literature:

1) Nash agent – chooses pure Nash equilibrium strategy if it's unique; else plays mixed Nash equilibrium strategy.
2) Maximin agent – maximizes its min. possible payoff.
3) Minimax agent – minimizes other agent's max. payoff.
4) Minimax-regret agent – minimizes its worst-case regret (difference between actual payoff and the payoff had a different action been chosen).
5) Random agent – probability 1/2 of either action.

To the best of our knowledge, the above standard strategies represent the best available strategies from the literature of repeated games, which are applicable to the life game. As discussed earlier, other strategies such as the successful tit-for-tat cannot be generalized and used in the life game.

Due to the novelty of the life game, and in order to provide a richer set of strategies to evaluate the SVO agent, we collected a large set of Peer Designed Agents (PDAs). PDAs have been recently used with great success in AI to evolve and evaluate state-of-the-art cognitive agents for various tasks such as negotiation and collaboration [12], [13], [1]. Lin et al. provided an empirical proof that PDAs can alleviate the evaluation process of automatic negotiators, and facilitate their designs [12].

To obtain a large collection of PDAs, we asked students in several advanced-level AI and Game theory classes to contribute agents. To attain a richer set of agents, we used two different universities in two different countries: University of Maryland in the USA, and Bar-Ilan University in Israel. The students were told that their agent would compete against all the agents of the other students in the class (once against each agent in a round-robin fashion). The instructions stated that at each iteration, they will be given a symmetric game with a random payoff matrix of the form shown in Fig. 2. Following Axelrod's methodology, we did not tell the students the exact number of iterations in each life game. The total agent's payoff will be the accumulated sum of payoffs with each of the other agents. For motivational purposes, the project grade was positively correlated with their agents overall ranking based on their total payoffs in the competition. Overall, we collected 48 agents (24 from the USA and 24 from Israel).

TABLE I
Average payoffs and rankings of the SVO agent, standard agents and top three PDAs

Agent	Rank and (Avg Payoff)
SVO agent	1 (5.836)
The best PDA	2 (5.831)
The 2nd best PDA	3 (5.792)
The 3rd best PDA	4 (5.789)
Minimax regret agent	6 (5.695)
Maximin agent	35 (5.453)
Nash agent	43 (5.271)
Random agent	52 (4.351)
Minimax agent	54 (3.954)

A. Evaluating the SVO agent

The first experiment was meant to assess the competence of the suggested SVO based agent. The version that was used in this experiment was with $f = \infty$ (unforgiving trust method), in which following a (perceived) defection and a consequent lost of trust level, it cannot be recovered.

We ran tournaments with the unforgiving SVO agent and all the other agents in the test set. Since the test set is composed of 53 agents (48 PDAs + 5 standard strategies), the total number of participant in each run of the competition is 54. The tournament is similar to Axelrod's IPD tournaments [3] and the 2005 IPD tournament [11]. Each participant played against every participant including itself (thus a tournament among n agents consists of n^2 iterated games). The number of iterations in one life game was set to 200. In each experiment, we calculated the average payoff per game for each agent. Since the values in the payoff matrix are chosen uniformly from $[0, 9]$, the expected average payoff of a random agent who played with another random agent is 4.5. In order to have a fair comparison, we used the same sequence of random games for each of the pairs in the experiment. We repeated the experiment 100 times using different random seeds, so each average payoff is an average of $100 \times k \times n$ payoffs, where k is the number of iterations in each life game and n is the number of participating agents. Hence, each average payoff is computed from averaging the payoffs of 540000 games.

Table I shows average payoffs and rankings of the SVO agent, standard agents and the top three PDAs. The SVO agent has the highest average payoff, and so it ranked number one. Because the standard agents are not adaptive, and cannot learn from the history of interactions, their performances are bad in general, except the minimax regret agent. The minimax regret agent performed well in the tournament, unexpectedly. One

TABLE II
EVALUATING TRUST ADAPTATION – RESULTS

Agent	Rank and (Avg Payoff) in Each Tournament
Unforgiving SVO agent ($f = \infty$)	1 (5.836)
Forgiving SVO agent ($f = 1$)	6 (5.689)
Forgiving SVO agent ($f = 2$)	5 (5.722)
Forgiving SVO agent ($f = 3$)	5 (5.744)
Forgiving SVO agent ($f = 4$)	5 (5.757)

TABLE III
EVALUATING THE INDIVIDUALISTIC OPPONENT MODELS – RESULTS

Agent	Rank and (Avg Payoff) in Each Tournament
SVO agent	1 (5.836)
Maxi-maxi-rand-only agent	2 (5.800)
Maxi-rand-only agent	5 (5.721)
Maximin-only agent	5 (5.700)
Maximax-only agent	9 (5.681)

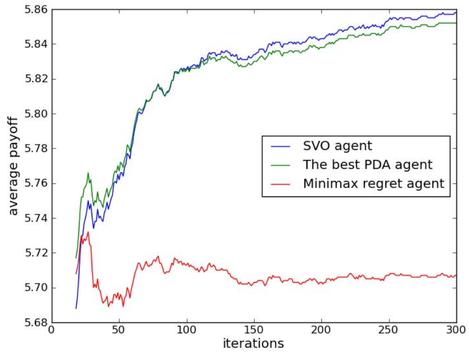

Fig. 6. Average payoffs at each iteration.

possible reason is that it does not have any assumption on its opponent, and focus on minimizing its own possible regret.

The performances of the top PDAs are very close to our SVO agent. In our post-experiment analysis, we found that most of them are doing some sort of opponent modeling by counting (i.e., similar to our counting method), but none of them are modeling the other agent using trust or SVO. Moreover, in contrast to the SVO algorithm which is relatively short and simple, their algorithms are much longer and complicated.

B. Evaluating Trust Adaptation: to forgive or not forgive?

As mentioned in Section V-A, our agent trust adaptation approach can be set using the f parameter. We would like to study if a forgiving approach is a better-suited approach in repeated stochastic 2x2 symmetric games. As such, we varied the f parameter from the unforgiving approach ($f = \infty$) to SVO agents with different forgiveness thresholds, and ran four additional tournaments for each forgiving SVO agent ($f = 1, 2, 3, 4$). The methodology to evaluate an agent P was to run a tournament with P and all the other agents in the test set. In other words, for each SVO agent P, we reran the previous tournament with the original SVO agent replaced by P.

As we can see in Table II, the average payoffs of all of the forgiving agents are lower than that of the unforgiving agent. This result is interesting as it may contradict to some extent the "forgiving" property of successful strategies in IPD as described by Axelrod. On the other hand, there is a possible confounding factor in our experiments. In particular, we have some preliminary results suggesting that the PDAs (against which we tested our agents) behaved in ways that were correlated with some of the personality characteristics of the students who wrote those agents. As those students were primarily young males, it is possible that the PDAs constituted a biased sample. We observed that if a PDA defects on another agent at the beginning of the life game, it is very likely that it will defect again later. Therefore, the risk and cost of a forgiving approach is high, which probably explains the decrease in performance.

C. Evaluating the Individualistic Opponent Models

One of our hypotheses during the model's construction was that a larger set of opponent models would provide a more refined playground to differentiate and classify different models, which in turn will allow the agent to provide better responses to their strategies. To investigate the significance of each component of individualistic model, we implemented four simplified versions of the SVO agent, where each contained a single, predefined opponent model:

1) Maximin-only agent – uses the maximin model for individualistic agent modeling.
2) Maximax-only agent – uses the maximax model for individualistic agent modeling.
3) Maxi-rand-only agent – uses the maxi-random model for individualistic agent modeling.
4) Maxi-maxi-rand-only agent – uses the maxi-maxi-random model for individualistic agent modeling.

We tested the above four agents by running four additional tournaments for each of them. Table III shows the average payoffs and rankings of the four agents in each of the tournament, as well as the complete SVO agent. We can see that the average payoffs of all of the four simplified agents are less than that of the complete SVO agent. These results ratify our hypothesis that a single individualistic opponent model is not refined enough for successful opponent modeling.

D. Evaluating Robustness to Number of Iterations

To investigate the performance of the SVO agent at different number of iterations, we recorded the average payoffs the agent accumulated at different iteration in the tournament. Fig. 6 shows the trends of the average payoffs of the SVO agent, the best PDA and the best standard agent (i.e., the minimax regret agent). With an increasing number of iterations, both SVO agent and the best PDA obtained higher payoffs and level off after 200^{th} iteration, while the payoff of the minimax regret agent remains the same most of the time. The impact is probably due to the fact that both SVO agent and the best PDA are doing opponent modeling. With an increase in the

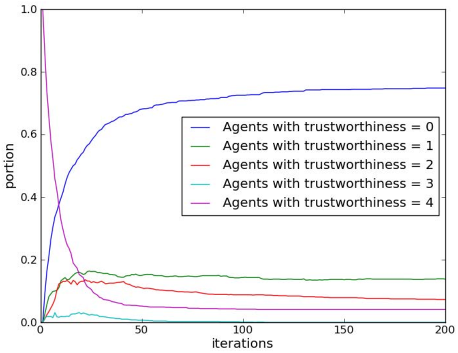

Fig. 7. Distribution of trustworthiness of the PDAs at each iteration.

number of interactions, the modeling will be more accurate, and so they can better coordinate with their opponents to get higher payoffs. On the contrary, the minimax regret agent does not change its strategy, so its performance remains unchanged most of the time. The payoff of the SVO agent is low at the beginning of the life games, because it begins by applying the "nice" strategy towards all other agents. If the other agent is non-cooperative, the SVO agent may be exploited for the first few mutual-benefit games, and lose some payoffs at the beginning. However, its performance catches up quickly and outperforms others after the 100^{th} iteration, because it will stop cooperating with the defectors and keep cooperating with the cooperators. The best PDAs agents do not have trust modeling and cannot fully cooperate with others, so it cannot get the full benefit from mutual cooperation. Therefore, the SVO can obtain higher payoff in long run, while the other agents cannot.

E. Analyzing the Trustworthiness of the PDAs

With this analysis we seek to explore the trustworthiness of the PDAs that were written by students, and investigate the significance of each mutual-benefit game type. We did as follows: each of the PDAs played against our SVO agent. While in game, at each iteration, the distribution of all five types of cooperative agents was recorded. Fig. 7 shows the portion of the five types of cooperation classification at each iteration. At the beginning, as we mentioned in Section V-D, the SVO agent assumed that all the other agents are trustworthy, so 100% of them are of the type 4 (most trustworthy). However, the population of type 4 cooperation drops quickly as most of the agents start defecting in mutual-benefit games. Therefore, the population of the others, less trustworthy agents, increases quickly. The fastest population growth is of type 0 agents, which are not cooperative at all. After 100^{th} iteration, the distribution starts to stabilize. At the end, the whole population consists of 74.8% type 0 agent, 13.8% type 1 agent, 7.3% type 2 agent, 0% type 3 agent, and 4.1% type 4 agent. This also shows that our classification of cooperative agent is effective. For example, without that classification, we would expect our agent to fail to cooperate with those 25.2% cooperative agents.

VII. Conclusions and Future Work

In this paper we have described the life game, a stochastic repeated game for studying social interactions. The life game poses several new challenges, for example, strategies that work well in conventional iterated games cannot be used directly.

In order to develop a successful cognitive strategy for the life game, we utilized SVO theory, a motivational theory for human choice behavior. Our method of agent modeling can be used to learn strategies and respond to others' strategies over time, to play the game well. Our experiments demonstrated that our SVO based agent outperformed both standard repeated games strategies and a large set of peer designed agents. Furthermore, our experimental work illustrates the importance of adaptive and fine-grained opponent modeling, as well as the impact that different trust adaptation strategies have on the performance of the SVO agent.

In the future, we intend to investigate other versions of the life game (e.g., a noisy version). We also plan to incorporate the notion of nondeterministic trust models, as well as to extend our algorithm for continuous SVO estimation and modeling. Another interesting direction would be to account for power relations as a way to describe different types of intentional biases (e.g., cultural, personal or relational).

References

[1] T.-C. Au, S. Kraus, and D. Nau. Synthesis of strategies from interaction traces. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems - Volume 2*, AAMAS '08, pages 855–862, 2008.

[2] W. Au and J. Kwong. Measurements and effects of social-value orientation in social dilemmas. *Contemporary psychological research on social dilemmas*, pages 71–98, 2004.

[3] R. Axelrod. *The Evolution of Cooperation*. Basic Books, 1984.

[4] R. Axelrod and W. Hamilton. The evolution of cooperation. *Science*, 211(4489):1390, 1981.

[5] M. Bacharach, N. Gold, and R. Sugden. *Beyond individual choice: teams and frames in game theory*. Princeton Univ Pr, 2006.

[6] B. Beafuils, J.-P. Delahaye, P. Mathieu, C. G. Langton, and T. Shimohara. *Our Meeting With Gradual: A Good Strategy For The Iterated Prisoners Dilemma*, pages 202–209. MIT Press, 1996.

[7] J. Bednar and S. Page. Can game (s) theory explain culture? *Rationality and Society*, 19(1):65, 2007.

[8] S. Bogaert, C. Boone, Declerck, and C. Declerck. Social value orientation and cooperation in social dilemmas: A review and conceptual model. *Brit. Jour. Social Psych.*, 47(3):453–480, Sept. 2008.

[9] C. G. M. David M. Messick. Motivational bases of choice in experimental games. *Experimental Social Psychology*, 1(4):1–25, 1968.

[10] K. Kanagaretnam, S. Mestelman, K. Nainar, and M. Shehata. The impact of social value orientation and risk attitudes on trust and reciprocity. *Journal of Economic Psychology*, 30(3):368–380, 2009.

[11] G. Kendall, X. Yao, and S. Chong. *The iterated prisoners' dilemma: 20 years on*. World Scientific Pub Co Inc, 2007.

[12] R. Lin, S. Kraus, Y. Oshrat, and Y. K. Gal. Facilitating the evaluation of automated negotiators using peer designed agents. In *AAAI*, 2010.

[13] E. Manisterski, R. Lin, and S. Kraus. Understanding how people design trading agents over time. In *AAMAS (3)*, pages 1593–1596, 2008.

[14] J. Maynard-Smith. Evolution and the theory of games, 1982.

[15] Nowak and Sigmund. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoners dilemma game. *Nature*, 364(6432):56–58, 1993.

[16] A. Rapoport. *Two-Person Game Theory. The Essential Ideas*. The University of Michigan Press, Ann Arbor, 1966.

[17] B. Skyrms. *The stag hunt and the evolution of social structure*. Cambridge Univ Pr, 2004.