ABSTRACT

Title of dissertation: AGENT MODELING IN
REPEATED STOCHASTIC GAMES

Kan Leung Cheng, Doctor of Philosophy, 2014

Dissertation directed by: Professor Dana Nau
Department of Computer Science

There are many situations in which two or more agents (e.g., human or computer decision makers) interact with each other repeatedly in settings that can be modeled as repeated games. In such situations, there is evidence that agents sometimes deviate greatly from what conventional game theory would predict. There are several reasons why this might happen, one of which is the focus of this dissertation: sometimes an agent's preferences may involve not only its own payoff (as specified in the payoff matrix), but also the payoffs of the other agent(s). In such situations, it is important to be able to understand what an agent's preferences really are, and how those preferences may affect the agent's behavior.

This dissertation studies how the notion of Social Value Orientation (SVO), a construct in social psychology to model and measure a person's social preference, can be used to improve our understanding of the behavior of computer agents. Most of the work involves the *life game*, a repeated game in which the stage game is chosen stochastically at each iteration. The work includes the following results:

- Using a combination of the SVO theory and evolutionary game theory, we studied how an agent's SVO affects its behavior in Iterated Prisoner's Dilemma (IPD). Our analysis provides a way to predict outcomes of agents playing IPD given their SVO values.

- In the life game, we developed a way to build a model of agent's SVO based on observations of its behavior. Experimental results demonstrate that the modeling technique works well.

- We performed experiments showing that the measured social preferences of computer agents have significant correlation with that of their human designers. The experimental results also show that knowing the SVO of an agent's human designer can be used to improve the performance of other agents that interact with the given agent.

- A limitation of the SVO model is that it only looks at agents' preferences in one-shot games. This is inadequate for repeated games, in which an agent's actions may depend on both its SVO and whatever predictions it makes of the other agent's behavior. We have developed an extension of the SVO construct called the *behavioral signature*, a model of how an agent's behavior over time will be affected by both its own SVO and the other agent's SVO. The experimental results show that the behavioral signature is an effective way to generalize SVO to situations where agents interact repeatedly.

AGENT MODELING IN REPEATED STOCHASTIC GAMES

by

Kan Leung Cheng

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:
Professor Dana Nau, Chair/Advisor
Professor Jennifer Golbeck
Professor Aravind Srinivasan
Professor James Reggia
Professor Ashok Agrawala

Acknowledgments

First, I have to thank my advisor, Professor Dana Nau. This work would not be possible without his mentorship, encouragement, and support. I really appreciate his teachings about how to be a good researcher and person.

I would like to thank Dr. Inon Zuckerman, Dr. Ugur Kuter, and Professor Jennifer Golbeck for their inspiring ideas, contributions, and support.

I would like to thank the rest of the committee, Professor Aravind Srinivasan, Professor James Reggia, and Professor Ashok Agrawala, for sparing their invaluable time reviewing the manuscript.

I would like to thank my fellow graduate students in LCCD, many friends of the Hong Kong student group, and many other friends I met in Maryland for their help, advice, and kindness during my years at UMD. I also want to thank the Computer Science Department and the graduate office for the education and support.

I am greatly indebted to my parents and my siblings. Without their love and support, I would not have received a good education in Hong Kong and had the opportunity of going to school in Maryland.

The work in this thesis was supported in part by ARO grant W911NF1110344, ONR grant W911NF0810144, and AFOSR grant W911NF1110344. The information in this dissertation does not necessarily reflect the position or policy of the funders, and no official endorsement should be inferred.

Table of Contents

List of Figures v

List of Abbreviations vii

1 Introduction 1
1.1 Outline of Thesis 2

2 Background and Related Work 4
2.1 The Social Value Orientation (SVO) Theory 4
2.2 The Life Game Model 5
2.3 Related Work 6

3 Social Value Orientation Theory and Evolution of Cooperation 8
3.1 Introduction 8
3.2 Evolution of Cooperation 9
3.3 Our Model 10
3.4 Analysis 12
3.5 Experiments 15
3.5.1 Prisoner's Dilemma with Constant Payoffs 15
3.5.2 Prisoner's Dilemma with Varying payoffs 16
3.6 Summary 18

4 Cognitive Strategies for The Life Game 21
4.1 Introduction 21
4.2 Strategies for the Iterated Prisoner's Dilemma 22
4.3 Strategies for the Life Game 23
4.4 Social Value Orientation Agent Models 24
4.4.1 Cooperative Model 25
4.4.2 Individualistic Model 26
4.4.3 Competitive Model 28
4.4.4 The Combined SVO Agent Modeling Strategy 28
4.5 Experiments and Results 30
4.5.1 Evaluating the SVO agent 31

4.5.2 Evaluating Trust Adaptation: to forgive or not forgive? 32
4.5.3 Evaluating the Individualistic Opponent Models 32
4.5.4 Evaluating Robustness to Number of Iterations 33
4.5.5 Analyzing the Trustworthiness of the students' agents 34
4.5.6 Evaluating the Benefit of Cooperation 35
4.6 Summary 37

5 Modeling Agents using Designers' Social Preference 38
5.1 Introduction 38
5.2 Measuring Human Social Preferences 39
5.3 Measuring Agents' Social Preferences 41
5.4 Experiments on Measuring Agents' para-SVO 42
5.4.1 Agent-human SVO Correlation 44
5.4.2 Stationary vs. Non-stationary Strategies 45
5.5 Utilizing the SVO Information 47
5.5.1 Compositing a Non-adaptive Agent 47
5.5.2 Improving an Adaptive Agent 48
5.6 Summary 52

6 Predicting Agents' Behavior by Measuring their Behavioral Signature 54
6.1 Introduction 54
6.2 Modeling Computer Agents 55
6.3 Experiments 56
6.3.1 Measuring Agents' Behavioral Signatures 57
6.3.2 Predicting Agents' Performances 58
6.4 Summary 59

7 Conclusions 62

Bibliography 64

List of Figures

2.1 Stage game for the life game. The values a, b, c, d are generated randomly as described in the text. 6
2.2 Choosing the side of the road upon which to drive. 6

3.1 The Prisoner's Dilemma game. 10
3.2 Generalized form of the Prisoner's Dilemma game, where $S < P < R < T$ and $2R > S + T$. 10
3.3 An Illustration of the social-interaction space of two players, x and y. The x and y axes show the accumulated total payoff for Players x and y, respectively. 11
3.4 An example reaction of a fair player (Player x) 12
3.5 An evolutionary simulation of IPD. The top graph shows the average theta per generation. The bottom graph shows the average payoff per generation. 16
3.6 Invasion of fair player. 17
3.7 Top graph: effect of R on average payoff. Middle graph: effect of R on average theta. Bottom graph: effect of R on the percentage of cooperative agents. 18
3.8 Top graph: effect of T on average payoff. Middle graph: effect of T on average theta. Bottom graph: effect of T on the percentage of cooperative agents. 19
3.9 Top graph: effect of P on average payoff. Middle graph: effect of P on average theta. Bottom graph: effect of P on the percentage of cooperative agents. 20

4.1 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. 23
4.2 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. 27

4.3 Procedure for a (unforgiving) SVO agent playing a game g_t at t-th iteration in a life game. 28
4.4 Average payoffs at each iteration. 34
4.5 Distribution of trustworthiness of the students' agents at each iteration. 35
4.6 Average payoffs at each iterations. 36

5.1 Format of i-th decision task of the ring measurement questionnaire . 40
5.2 A sample decision task used by the ring measurement questionnaire . 40
5.3 Procedure of measuring para-SVO of an agent x with a tester agent y after n random games. 43
5.4 SVO of 28 students. 44
5.5 Correlation of human SVO and agents' para-SVO. 45
5.6 Correlation between human SVO and agents using non-stationary strategy. 46
5.7 Performance of the simple agents playing with 21 agents written by cooperative humans. 48
5.8 Performance of the simple agents playing with 7 agents written by individualistic humans. 49
5.9 Performance of the simple and composite agents. 50
5.10 Performance of agents playing with all 28 students' agents. 51
5.11 Performance of agents playing with 21 agents written by cooperative humans. 52
5.12 Performance of agents playing with 7 agents written by individualistic humans. 53

6.1 Distribution of students' agents' SVO. 57
6.2 Average para-SVO values $\theta_{10}(x|C_\phi)$ for $\phi = -90^\circ$ to 90°, averaged over all x in the entire set of students' agents. 58
6.3 Average para-SVO values $\theta_n(x|C_\phi)$ with different tester agents C_ϕ, averaged over all x in the entire set of students' agents. 59
6.4 Correlation between predicted and actual payoffs (when student agents play in a tournament). 60
6.5 Mean square error of predicted payoffs (when student agents play in a tournament). 61

List of Abbreviations

SVO	Social Value Orientation
IPD	Iterated Prisoners Dilemma
$ALLC$	Always Cooperate
$ALLD$	Always Defect
TFT	Tit-for-Tat

Chapter 1: Introduction

In game theory, it is generally assumed that all individuals satisfy a set of mathematical assumptions known as decision-theoretic rationality — from which it follows that an individual's actions can be regarded as maximizing utility, the "currency" for everything they want. In an economic context, the rational choice theory gives rise to the concept of Homo economicus (or economic man) which suggest that the ideal man is driven by self-interested economic calculation without considering the consequences on others. It provides the central explanatory principle of many economic theories and has generated a vast academic literature. On the other hand, experiments in social and behavioral sciences show that humans rarely follow this assumption. For instance, consider the *Ultimatum Game*, in which two players interact to decide how to divide a sum of money that is given to them. The first player proposes how to divide the sum between the two players, and the second player can either accept or reject this proposal. If the second player rejects, neither player receives anything. If the second player accepts, the money is split according to the proposal. Existing experiments in this game show that the offers that are issued or accepted are closer to a "fair" division of the money ($50 for each) than the "rational" choice [1], and also involved a cultural component [2].

Indeed, it is widely accepted in social and behavioral sciences that players explicitly take into account the outcomes for other players when considering their course of action. 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 [3] in which they presented a motivational theory of choice behavior that considers both players' payoffs in game situations. This theory was later called the *Social Value Orientation* theory [4]. Most of the SVO based studies typically recognize two opposing social value orientations: a proself and prosocial orientation. A *proself orientation* is one that gives higher consideration to its own payoff, while a *prosocial orientation* gives higher regards to the payoff of the agents he is interacting with. The social orientation of a player is not an absolute value; it describes a spectrum of possible behaviors, in which one end of the spectrum denotes proself behavior and the other end denotes prosocial behavior. Note that in contrast to the diversity of the SVO theory, the conventional rationality assumption dictates that all individuals are proself, without any difference between one and another. As most social or psychological traits, the claim that SVO is a fundamental personality trait is supported by both biological and sociological findings [5]. Biological support also can be found, among others, in Van Lange's work [6] showing that the basic form of SVO is visible early in life as part of a child's temperament.

The purpose of this dissertation is to study how the notion of SVO can be used to improve our understanding of the behavior of computer agents. Most of the work involves the *life game*, a repeated game proposed by Bacharach [7] in which

the stage game is chosen stochastically at each iteration. The results of the work are described in following paragraphs.

First, we use a combination of the SVO theory and evolutionary game theory to study how an agent's SVO affects its behavior in Iterated Prisoner's Dilemma (IPD). Our analysis provides a way to predict outcomes of agents playing IPD given their SVO values. The evolutionary simulation results on IPD confirm previous findings on evolution of cooperation, and provide new insights on the evolutionary process of cooperative behavior in a society as well as on the emergence of cooperative societies.

Second, in the context of life game, we have developed a way to build a model of agent's SVO based on observations of its 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 impacts that different trust adaptation strategies have on the performance of the SVO agent.

Third, we have developed a way to quantify the social preferences of computer agents, by adapting some concepts and techniques from social psychology. We performed experiments showing that the measured social preferences of computer agents have significant correlation with that of their human designers. The experimental results also show that knowing the SVO of an agent's human designer can be used to improve the performance of other agents that interact with the given agent.

Finally, a limitation of the SVO model is that it only looks at agent's preferences in one-shot games. This is inadequate for repeated games, in which an agent's actions may depend on both its SVO and whatever predictions it makes of the other agent's behavior. To use the SVO model effectively in repeated games, it is necessary to extend the SVO model to take into account how an agent's behavior will change if it interacts repeatedly with various other kinds of agents. This dissertation includes a way to adapt and extend the SVO construct by a *behavioral signature*, a model of how an agent's behavior over time will be affected by both its own SVO and the other agent's SVO. The experimental results show that the predictions using behavioral signatures are highly correlated with the agents' actual performance in tournament settings. This shows that the behavioral signature is an effective way to generalize SVO to situations where agents interact repeatedly.

1.1 Outline of Thesis

The contents of the rest of this dissertation are as follows:

Chapter 2 presents a more detailed description of the Social Value Orientation (SVO) Theory and the Life Game. The Life Game is the game environment we used in most of the work (Chapter 4, 5, and 6) in this dissertation.

In Chapter 3, we utilize evolutionary game theory to study the evolution of cooperative societies and the behaviors of individual agents in such societies. We present a novel player model based on the SVO theory. Alongside the formal player model we provide an analysis that considers possible interactions between different

types of individuals and identifies five general steady-state behavioral patterns.

Chapter 4 presents a cognitive agent model based on the SVO theory. Our method of agent modeling can be used to learn strategies and respond to others' strategies over time, to play the game well. 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. This chapter also suggests some properties for life-game strategies to be successful in environments with such agents.

Chapter 5 presents a way to measure the social preferences of computer agents, by adapting some concepts and techniques from social psychology. We perform experiments to study the correlation between the social preferences of computer agents and that of their human designers. This chapter also describes how to use the SVO information of the human designers to improve the performance of other agents that interact with the given agent.

Chapter 6 presents a way to extend the SVO model to a *behavioral signature* that models how an agent's behavior over multiple iterations will depend on both its own SVO and the SVO of the agent with which it interacts. This chapter also provides a way to measure an agent's behavioral signature, and a way to use this behavioral signature to predicting the agent's performance.

The last chapter reviews the contributions of this dissertation and proposes directions for future work.

Chapter 2: Background and Related Work

This chapter gives a more detailed description of the Social Value Orientation (SVO) Theory and the Life Game. The Life Game is the game environment we used in most of the work (Chapter 4, 5, and 6) in this dissertation. We will also discuss related work on agent modeling.

2.1 The Social Value Orientation (SVO) Theory

There is a substantial set of evidence from the social and behavioral sciences literature showing that players explicitly take into account the outcome for the other player when considering their course of action [5]. Moreover, the choices people make depend, among other things, on stable personality differences in how they approach interdependent others. This observation can be traced back to the seminal work by Messick and McClintock [3] in which they presented a motivational theory of choice behavior that considers both players' payoffs in game situations. This theory was later denoted as the *Social Value Orientation* theory, which has since developed into a class of theorems [4].

Most of the SVO based studies typically recognize two opposing social value orientations: proself and prosocial orientations. A *proself orientation* is one that

gives higher consideration to its own payoff; while a *prosocial orientation* gives higher regards to the payoff of the agents he or she is interacting with. The social orientation of a player is not an absolute value; it describes a spectrum of possible behaviors, in which one end of the spectrum denotes proself behavior, and the other end denotes prosocial behavior. Analysis of many SVO based experiments reveal that most people are classified as cooperators (50%), followed by individualists (24%), followed by competitors (13%) [5, p. 74]. [1]

In contrast to the diversity of the SVO theory, the traditional rationality assumption dictates that all individuals are proself, without any difference among them. As most social or psychological traits, the claim that SVO is a fundamental personality trait is supported by both biological and sociological findings [5]. Biological support also can be found, among others, in Van Lange's work [6] showing that the basic form of SVO is visible early in life as part of a child's temperament. The development of the SVO from social interactions is supported by many works shown in Au and Kwong review [5]. The validity of SVO based theorems, shown both in laboratory and field studies, indicates that prosocial generally cooperate more and show greater concern for the effect of their actions on the well being of others and on the environment. For examples, McClintock and Allison [8] shows that prosocial students were more willing to contribute time to help others, and Joireman et al. [9] that shows that prosocial participants tend to take more pro-environmental and collective policies than self-interest actions.

Over the years, there have been significant advances on social dilemmas and

[1] The remaining 13% could not be classified as having a consistent SVO.

various aspects of the social value orientations since the seminal work of Messick and McClintock [3]. For example, Parks and Rumble [10] showed that different aspects of the Tit-for-Tat strategy have different effects on the cooperation rates of individuals with different SVO values. In addition, there were several other research questions that considered some relaxation of the rationality assumption in their solution, for instance, de Jong et al. [11] presented a computational model that allows for achieving fairness in multi-agent systems. Their computational model uses the *Homo Egualis* utility function that has been shown to adequately describe human behavior in several games.

2.2 The Life Game Model

Many multi-agent domains involve human and computer decision makers that are engaged in repeated collaborative or competitive activities. Examples include online auctions, financial trading, and computer gaming. Repeated games are often viewed as an appropriate model for studying these kinds of repeated interactions between agents. In a traditional, game-theoretic repeated-game model, agents repeatedly play a game called the *stage game*. Many types of games can be used as the stage game. For example, Axelrod's famous Iterated Prisoner's Dilemma (IPD) competitions showed the emergence of cooperation, even though the rational dominant equilibrium in a one-shot Prisoner's Dilemma is to defect [12]. Maynard Smith studied two-player Chicken Game with a population of Hawks and Doves [13], and Skyrms studied the evolved population when individuals were playing the Stag-hunt

game [14].

Each of these studies used a highly simplified game model in which the *same* stage game was used at every iteration. In other words, they assumed that the agents would interact with each other repeatedly in exactly the same environment. However, as pointed out by Bacharach [7, p. 100], repeatedly playing the **same** game is unlikely to be an accurate model of any individual's life. In many real-life situations, agents may interact with each other repeatedly in *different* environments.

As a 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 many 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). The rich variety of stage games also allows agents to express a larger spectrum of social preferences, resulting in an adequate playground for agents of different personalities and behaviors. We believe this makes the Life Game a better model for repeated interaction in different environments—so in this dissertation we concentrate on studying social preferences of automated agents in the Life Game of high ludic diversity.

In this dissertation, we model the *life game* as an iterated game in which each stage game is a 2x2 normal-form game that is generated randomly by choosing independent random values for the payoffs a, b, c, and d in the payoff matrix shown in Figure 2.1. The payoffs a, b, c, d are chosen from a uniform distribution over the set $[0, 9]$. At each stage, each agent knows the complete payoff matrix. After deciding on the actions, each agent will be notified of the action chosen by the other agent.

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

Figure 2.1: Stage game for the life game. The values a, b, c, d are generated randomly as described in the text.

Choosing sides game		Player 2	
		A_1=Left	A_2=Right
Player 1	A_1=Left	$(1, 1)$	$(0, 0)$
	A_2=Right	$(0, 0)$	$(1, 1)$

Figure 2.2: Choosing the side of the road upon which to drive.

The two agents will play the games in succession, without knowing when the series of games will end. We do not place any restrictions on the agents' memory, and they may record past matrices and the actions taken by both agents and use it in their strategy.

Depending on the randomly chosen values of a, b, c, and d, each stage game may or may not be an instance of a well-known social dilemma game (e.g., Prisoner's Dilemma, Chicken Game, Stag-Hunt [15], etc). Consequently, the semantics of the actions are subjective and 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." This additional layer of uncertainty might cause situation such as that when one agent considers a certain action to be a

reasonably cooperative action, it will be captured as a competitive action in the eyes of its opponent. Somes instances (e.g., Figure 2.2) of the random games may not have the notion of cooperation at all. Therefore, in addition to induce cooperation, a good agent also need to cooridinate well with other agent in some of the games.

2.3 Related Work

Previous research in agent-modeling has proposed modeling other agents by estimating agents' personalities. For example, Talman et al. [16] proposed a decision-theoretic model that explicitly represents and reasons about agents' personalities in environments in which agents are uncertain about each others' resources. Similar to our SVO-based models, their agents can identify and negotiate with those who are cooperative while avoiding those who are exploiter. Ya'akov Gal et al. [17] proposed several new decision-making models that represent, learn and adapt to various social attributes that influence people's decision-making in a group of human and computer agents.

Previous works on the Iterated Prisoner's Dilemma typically used a policy to model an agent [18–21]. A policy is a set of rules specifying the choice of C or D in every round as a function of the history of the interaction so far. For example, well-known strategies such as the famous *tit-for-tat* strategy [12] and Pavlov strategy [22] can be modeled using the policy. However, extending the policy model to the life game, which allows variations of interactions, 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. On the contrary, our SVO-based models do not require labeling actions in advance.

In most of the work (Chapter 4, 5, and 6) in this dissertation, in order to provide a richer set of strategies to perform experiments, we collected a large set of Peer Designed Agents (PDAs). Peer-Designed Agents 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 [23–25]. Lin et al. [23] provided an empirical proof that PDAs can alleviate the evaluation process of automatic negotiators, and help their designs. Efrat Manisterski et al. [24] studied how people design agents for online markets and how their design changes over time. Their results show that most human subjects modified their agents' strategic behavior; for example, they increased their means of protection against deceiving agents. Au et al. [25] used agents written by students to study enhancing agent by combining a set of interaction traces produced by other agents. In most of the work in this dissertation, we used the same evaluation methodology that we collected a set of

human-written agents and then investigate a better model and way to predict their behaviors and have better interactions with them.

Chapter 3: Social Value Orientation Theory and Evolution of Cooperation

In this Chapter, we utilize a combination of SVO theory and evolutionary game theory to study the evolution of cooperative societies and the behaviors of individual agents (i.e., players) in such societies. We present a novel player model based on the SVO theory. Alongside our formal player model we provide an analysis that considers possible interactions between different types of individuals and identifies five general steady-state behavioral patterns. We present evolutionary simulations that confirm previous findings on evolution of cooperation, and provide new insights on the evolutionary process of cooperative behavior in a society as well as on the emergence of cooperative societies.

3.1 Introduction

We consider *how cooperative societies evolve, given varying social tendencies of the individuals.* Evolution of cooperation between individuals has been studied for years, most notably starting from the seminal work of Smith [26] and Axelrod [27]. The underlying question can be summarized briefly as follows:

Why and how does cooperative behavior evolve in a Darwinian society

where the *survival of the fittest* is the prominent behavioral rule?

Existing research on this question typically utilize game theory in order to analyze highly simplified models of social dilemmas. Traditional game theory uses the *rationality assumption*, i.e., the assumption that human behavior is purely rational, self-maximizing behavior [28]. However, this assumption has received wide criticism from the behavioral science literature (e.g. [1]). For example, the *Social Value Orientation* (SVO) theory [3, 4] conjectures based on many empirical studies, that the social choices people make depend, among other things, on other people; in particular on the stable personality differences between individuals.

We describe a new formalism for studying evolution of cooperation based on the SVO theory. Our formalism captures in the player models the notion of varying and persistent social orientations exhibited in human behavior. This enables a player to reason about the relationships between the player's social orientation and how that player develops strategies when two players interact repeatedly in a game for an unknown number of times.

We present theoretical results showing how players with different social tendencies interact in a class of normal-form games. Our analysis revealed five general stable state behavioral patterns that can be explained in terms of the players' varying social orientation values.

Our experiments based on evolutionary simulations in the *Iterated Prisoner's Dilemma* (IPD) demonstrated the effects of social orientations on the evolution of cooperative behavior in individual players and on the emergence of a cooperative

society. One set of experiments showed that prosocial tendency increases with increasing reward or with decreasing temptation, thus confirming previous intuitions from [29].

In our experiments, we also found out that there are scenarios in which the definition of cooperativeness from previous studies might lead to erroneous conclusions. Previous works on the evolution of cooperation typically used the average payoff of the society as a measure of of its cooperativeness: i.e., the higher the average payoff is, the more cooperative the society is thought to be [18–20]. However, it does not happen in our model that while increasing the value of rewards or punishments results in a similar increase in the average payoff, this does not result in the emergence of the same kind of society. In particular in our experiments, increasing mutual reward typically resulted in a cooperative society, but increasing mutual punishment resulted in a divided society that includes two distinct clusters: one of highly selfish players and the other of highly cooperative players.

3.2 Evolution of Cooperation

The studies for evolution of cooperation investigate how and under what conditions individuals who pursue their own self-interest can cooperate with each other without a central authority to force them to do so. Among the prominent approaches to understanding the nature of cooperation is *evolutionary game theory*, which uses evolutionary concepts such as fitness, replication and mutation to explain the emergence of cooperation [12, 26].

Prisoner's Dilemma		Player 2	
		Cooperate (C)	Defect (D)
Player 1	Cooperate (C)	$(3, 3)$	$(0, 5)$
	Defect (D)	$(5, 0)$	$(1, 1)$

Figure 3.1: The Prisoner's Dilemma game.

Normal-form games are typically used to abstract many details at arise in situations where cooperation among individuals are to be studied. The most widely-used example of such games is the *Iterated Prisoner's Dilemma (IPD)*, an iterated variant of the prisoner's dilemma that is played repeatedly an unknown number of iterations. IPD has been the most common model for social dilemmas between two players and has been often used in order to study the evolution of cooperation.

Figure 3.1 presents the payoff matrix for Prisoner's Dilemma, where two players are both faced with a decision to either cooperate (C) or defect (D). If the game is played once, then defecting will provide a higher payoff regardless of whether the other player cooperates or defects. However, if the game is played repeatedly for an unknown number of times, cooperative behavior in an individual might emerge to increase accumulated payoffs (see [30] for an overview).

Since Axelrod's IPD tournament [12] that focused on generating winning strategies in IPD, there has been a large body of research on the various aspects of the basic IPD model: varying payoffs, number of players, and various population and structural dynamics [30].

In this Chapter, we focus on a generalized form of Prisoner's Dilemma game

2x2 symmetric game		Player 2	
		Cooperate (C)	Defect (D)
Player 1	Cooperate (C)	(R, R)	(S, T)
	Defect (D)	(T, S)	(P, P)

Figure 3.2: Generalized form of the Prisoner's Dilemma game, where $S < P < R < T$ and $2R > S + T$.

(Figure 3.2, where $S < P < R < T$ and $2R > S + T$). In the subsequent sections, we describe a new formalism based on the SVO theory, for studying evolution of cooperation. Previous works on the evolution of cooperation typically used the average payoff of the population as a measure of the cooperativeness [18–20]. In contrast, our formalism focuses on the social value orientations of the players and uses the average social values of the society as a new measure of cooperativeness.

3.3 Our Model

We formalize the *social-interaction space* of the two players in a game, namely Player x and Player y, as a two-dimensional Euclidean space, as illustrated in Figure 3.3. The x-axis represents the accumulated total payoff of Player x and the y-axis represents that of Player y.

The *social-orientation* of a player, say Player x, is a unit vector $\hat{s_x}$ such that $\hat{s_x}$'s initial point is at the origin of the social-interaction space. We represent $\hat{s_x}$ by the angle, $\theta(x)$ between $\hat{s_x}$ and the x-axis of the social-interaction space. Intuitively, the social orientation of a player is a model of its tendency to adopt a prosocial or

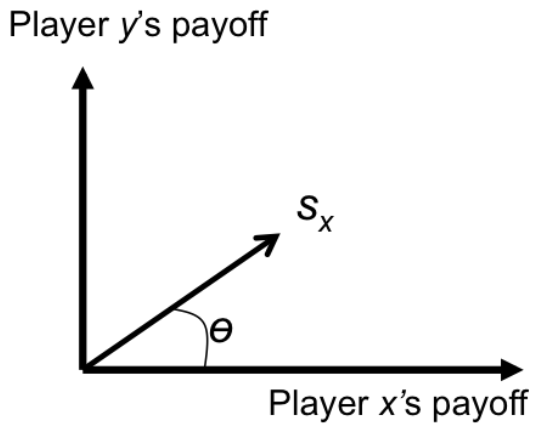

Figure 3.3: An Illustration of the social-interaction space of two players, x and y. The x and y axes show the accumulated total payoff for Players x and y, respectively.

proself behavior in a game.

For example, when $\theta_x = 0$ then Player x acts as a proself individual. If $\theta_x = \pi/4$, then this means that player is fair, i.e., it acts to balance the accumulated total payoffs of two players. When $\theta_x = \pi/2$, the player is purely prosocial, i.e., it never attempts to maximize its own payoff, but rather it tries to increase the payoff of the other player.

We define the *state* of the repeated game at any iteration point t as a vector in the social-interaction space:

$$\vec{g}_t = \langle p_x, p_y \rangle$$

where p_x is the accumulated total payoff that Player x receives from the start of the game to the point t and p_y is that of Player y. Note that both players hold the same game state that describes their accumulated total payoff, and it is the only variable state that players remember and use for deciding what to do next.

Suppose Player x, takes an action, C or D, in the game, and Player x assumes

Player y is a random player. Let $\vec{g}_t$ be the current game state. There are two possible *expected state changes* that can arise depending whether Player x takes the action C or D:

$$E[x \text{ chooses C}] = \vec{p}_C = \langle \frac{R+S}{2}, \frac{R+T}{2} \rangle$$
$$E[x \text{ chooses D}] = \vec{p}_D = \langle \frac{T+P}{2}, \frac{S+P}{2} \rangle$$

Intuitively, the state $\vec{p}_C$ is the expected change in the game state produced by the average of the payoffs that Player x could get by playing C, given Player y chooses C or D uniformly at random. This definition assumes that Player x's model of Player y is of a random player. In other words, Player x does not have any background knowledge about the other player and it cannot store and learn from the other player's actions.

The *expected successor game state*, $\vec{g}_{t+1,A}$, given that Player x chooses an action $a \in \{C, D\}$ in the current game state $\vec{g}_t$ is

$$\vec{g}_{t+1,A} = \vec{g}_t + \vec{p}_A,$$

where $\vec{p}_A$ is the expected amount of change in the state by doing an action A.

During the course of the game, each player aims to bring the game state closer to its social-orientation vector, $\hat{s_x}$. In other words, each player aims to change the game state based on its social preference. The differences between the orientations of the players create the tensions in their social interactions – hence the social dilemmas. Note that with the traditional rationality assumption, the players will try to do utility-maximization on their own payoff. In other words, the theta equals to zero and social orientation equals to $\langle 1, 0 \rangle$.

We formalize the *expected utility* of a game state $\vec{g}_{t+1,A}$ based on social orientations of players as follows. Let $\vec{g}_t$ be the current game state. The objective of each player is to minimize the in-between angle $\delta_{t,A}$ (where $A \in \{C, D\}$) between its own social-orientation vector $\hat{s}$ ($\hat{s_x}$ for Player x) and the expected utility vector $\vec{g}_{t+1,A} = \vec{g}_t + \vec{p}_A$. In other words, the utility of a game state $\vec{g}_{t+1,A}$ for Player x can be defined as $\cos \delta_{t,A}$, which can be computed as follows:

$$\cos \delta_{t,A} = \frac{\hat{s} \cdot \vec{g}_{t+1,A}}{|\vec{g}_{t+1,A}|}, \quad A \in \{C, D\}$$

Thus, at each iteration t of the game, each player will choose an action a_t such that

$$a_t = \mathsf{argmax}_{A \in \{C,D\}} \ \cos \delta_{t,A}$$

For example, consider the well-known Iterated Prisoner's Dilemma (IPD) game as depicted in Figure 3.1. Figure 3.4 shows how a fair player (i.e., $\theta = \pi/4$) interact with another player in an IPD game. For IPD, $\vec{p}_C = \langle \frac{3}{2}, 4 \rangle$, $\vec{p}_D = \langle 3, \frac{1}{2} \rangle$. At the beginning, $\vec{g} = \langle 0, 0 \rangle$, $\delta_{0,C}$ is smaller than $\delta_{0,D}$, therefore, the fair player will cooperate at the first iteration. When the other player defects at the first iteration, the game state becomes $\langle 0, 5 \rangle$, and $\delta_{1,D}$ is smaller than $\delta_{1,C}$. Therefore, the fair player defects at the second iteration. When the other player defects again, both of them get 1 and the game state becomes $\langle 1, 6 \rangle$. $\delta_{2,D}$ is smaller than $\delta_{2,C}$ again, therefore, the fair player defects until the other cooperates at some point after which the payoffs of both players becomes balanced, i.e., $\vec{g} = \langle p, p \rangle$ for some p. In other words, a fair player in IPD game behaves exactly the same as the well-known Tit-For-Tat (TFT) strategy.

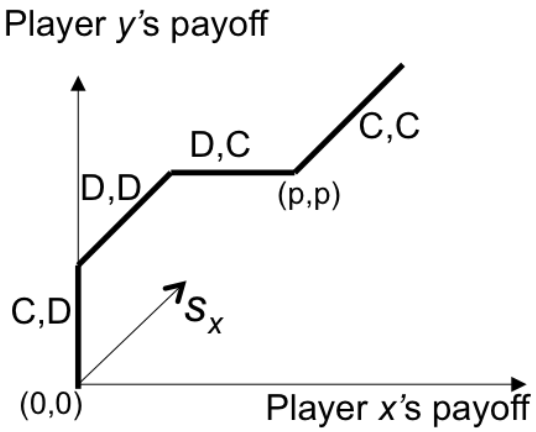

Figure 3.4: An example reaction of a fair player (Player x)

3.4 Analysis

We are interested in the dynamics of a player's behaviors (strategies), based on its social-orientation, over the course of its interaction with another player in a repeated game. This section presents an exhaustive analysis of such dynamics based on the model described earlier.

We use the definition for the social-orientation of a player in the following form. Let the two players be x and y, and their social-orientation angles be θ_x and θ_y, respectively. We define the *preference ratios* for each player as $r_x = \frac{\cos\theta_x}{\cos\theta_x + \sin\theta_x}$ and $r_y = \frac{\cos\theta_y}{\cos\theta_y + \sin\theta_y}$.

We define the following ratios for each action, C and D, in the Prisoner's Dilemma game: $r_C = \frac{R+S}{(R+S)+(R+T)}$ and $r_D = \frac{T+P}{(S+P)+(T+P)}$. Intuitively, these ratios describe the expected share of payoff that the first player will get by choosing C and D, respectively. Since $S < T$ (one of the conditions of Prisoner's Dilemma), we have $r_C < 0.5 < r_D$. In other words, this means that when a player defects, it is

expected to get a larger *share* of the total payoff. [1]

Let $\vec{g} = \langle p_x, p_y \rangle$ be the current game state. We define the current ratios for each player as $g_x = \frac{p_x}{p_x+p_y}$ and $g_y = \frac{p_y}{p_x+p_y}$. In a steady state, Player x cooperates whenever he is satisfied with his current ratio, i.e., his current ratio is greater than or equal to his preference ratio (i.e., $g_x \geq r_x$), or defects otherwise. Without loss of generality, we assume $r_x \leq r_y$. There are five possible cases in steady state:

Theorem 1. *If $r_y \geq r_x > 0.5$ (i.e., both are proself), both players always defect and get P at each game in steady state.*

Proof. When $g_x \geq r_x$, Player x cooperates while Player y defects, so g_x moves toward $\frac{S}{S+T} < r_x$. When $g_x < 1 - r_y$, Player x defects while Player y cooperates, so g_x moves toward $\frac{T}{S+T} > 1 - r_y$. Otherwise, both players defect and g_x moves toward $\frac{P}{P+P} = 0.5$. □

For example, in an IPD game, let $r_x = 0.6$ and $r_y = 0.7$. This means that Player x will always aim to get a share of 60% of the total payoff, while Player y will aim to get a share of 70% of the total payoff. Therefore, both will never be satisfied and will constantly defect to get a payoff of 1.

Theorem 2. *If $r_x \leq r_y \leq 0.5$ (i.e., both are prosocial), both players always cooperate and get R at each game in steady state.*

[1] $S < T$ is the only assumption required for the analysis. Note that this assumption is not restrictive: many well-known games satisfy this condition, including the well-known Prisoner's Dilemma, Chicken Game, and Stag-Hunt [15].

Proof. When $g_x < r_x$, Player x defects while Player y cooperates, so g_x moves toward $\frac{T}{S+T} > r_x$. When $g_x \geq 1 - r_y$, Player x cooperates while Player y defects, so g_x moves toward $\frac{S}{S+T} < 1 - r_y$. Otherwise, both players cooperate and g_x moves toward $\frac{R}{R+R} = 0.5$. □

For example, in an IPD game, let $r_x = 0.3$ and $r_y = 0.4$. This means that Player x will aim to get a share of 30% of the total payoff, while Player y will aim to get a share of 40% the total payoff. As such, they will both be easily satisfied, and therefore always cooperate and get the rewards, i.e., 3.

Theorem 3. *If $r_x < 0.5$ and $r_x + r_y = 1$, there are two cases:*

- *when $r_y > \frac{T}{S+T}$, Player x gets S while Player y gets T;*
- *otherwise, Player x gets $r_x(T + S)$, and Player y gets $(1 - r_x)(T + S)$.*

Proof. The first case above immediately follows from the fact that when $r_y > \frac{T}{S+T}$, we will have the repeated sequence of Cooperate-Defect actions in all interaction traces.

The proof for the second case is as follows. When $g_x < r_x$, Player x defects while Player y cooperates, so g_x moves toward $\frac{T}{S+T} > r_x$. When $g_x \geq r_x$, Player x cooperates while Player y defects, so g_x moves toward $\frac{S}{S+T} < r_x$. In steady state, they interact in a way that the ratio r_x (and r_y as well) is achieved, so Player x gets $r_x(T + S)$ while Player y gets $T + S - r_x(T + S)$. □

For example, in an IPD game, let $r_x = 0.4$ and $r_y = 0.6$. This means that Player x will always aim to get a share of 40% of the total payoff, while Player y

will aim to get a share of 60% of the total payoff. Therefore, they will try to grasp the share *alternatively*. In a steady state, Player x gets 2 and Player y gets 3 at each game on average.

Theorem 4. *If $r_y > 1 - r_x > 0.5$, there are two cases:*

- *when $r_y > \frac{T}{S+T}$, Player x gets S while Player y gets T;*
- *otherwise, Player x gets $\bar{p}_x = \frac{SP-PT}{(P-T)-(P-S)\frac{1-r_x}{r_x}}$, and Player y gets $\bar{p}_y = \bar{p}_x \frac{1-r_x}{r_x}$.*

Proof. The proof for the first case is the same that of Theorem 3 above. The proof of the second case is as follows. When $g_x < r_x$, both players defect and get P. When $g_x \geq r_x$, Player x cooperates and gets S and Player y defects and gets T. In steady state, they will get (S, T) or (P, P) in each game in a way that r_x is achieved. Let n_{DD} be the portion of the games resulted in DD, Player x gets $\bar{p}_x$ and Player y gets $\bar{p}_y$ where $\bar{p}_x = r_x(\bar{p}_x + \bar{p}_y)$, $\bar{p}_x = Pn_{DD} + S(1 - n_{DD})$ and $\bar{p}_y = Pn_{DD} + T(1 - n_{DD})$. Solving them, we can obtain the above formula. □

For example, in an IPD game, let $r_x = 0.4$ and $r_y > 0.6$. Now we are in a situation where there is lack of resources (as $r_x + r_y > 1$) and Player y is more proself than Player x. As such, Player y will always defect, while Player x will sometimes cooperate, and sometimes defect. In a steady state, Player x will get $\frac{10}{11}$ and Player y will get $\frac{15}{11}$ at each game on average.

Theorem 5. *If $r_x < 1 - r_y < 0.5$, there are two cases:*

- *when $r_x < \frac{S}{S+T}$, Player x gets S while Player y gets T;*

- *otherwise, Player x gets $\bar{p}_x = \bar{p}_y \frac{1-r_y}{r_y}$ and Player y gets $\bar{p}_y = \frac{TR-RS}{(R-S)-(R-T)\frac{1-r_y}{r_y}}$.*

Proof. The proof for the first case is the same that of Theorem 3 above. The proof of the second case is as follows. When $g_y < r_y$, Player x cooperates and gets S and Player y defects and gets T. When $g_y \geq r_y$, both players cooperate and get R. In steady state, they will get (S, T) or (R, R) in each game in a way that r_y is achieved. Let n_{CC} be the portion of the games resulted in CC, Player x gets $\bar{p}_x$ and Player y gets $\bar{p}_y$ where $\bar{p}_y = r_y(\bar{p}_x + \bar{p}_y)$, $\bar{p}_x = Rn_{CC} + S(1 - n_{CC})$ and $\bar{p}_y = Rn_{CC} + T(1 - n_{CC})$. Solving them, we can obtain the above formula. □

For example, in an IPD game, let $r_x < 0.4$ and $r_y = 0.6$. Then, as resources are plentiful (as $r_x + r_y < 1$) and Player x is more prosocial than Player y, Player x will always cooperate, while Player y will sometimes cooperate, and sometimes defect. In a steady state, Player x will get $\frac{30}{13}$ and Player y will get $\frac{45}{13}$ at each game on average.

3.5 Experiments

We have performed several experiments in order to investigate the emergence of cooperative populations. These experiments involve evolutionary simulations on a society of players and the simulations are designed based on social orientations of individuals, as described below.

We used the replicator dynamics for evolutionary simulations [26]. We used the well-known "infinite population" setup for initializing the population as described in [18–20]. We randomly generated 10 theta values from the interval $[0, \pi/2]$ and

assumed the size of a group with a particular θ value constitutes 10% of the entire population.

In each generation, the players engaged in pairwise encounters, resulting in a payoff for each of the players that is equal to the sum of the payoffs from each individual round of the game. The expected values of the score of a player in a pairwise game in steady state are described in the previous section. After each generation, each player had a number of offspring that is proportional to its expected total payoff. Each offspring had the same social-orientation value θ as its parent. If the frequency of a group of players with a particular θ value drops below a threshold 0.001 then the group is discarded from the population.

On average, in every 100 generations, a small amount (frequency of 0.01) of new randomly generated mutant players are introduced into the population. Each simulation was performed for 10,000 generations, resulting a total of about 100 mutant strategies.

3.5.1 Prisoner's Dilemma with Constant Payoffs

Figure 3.5 shows the average population θ and average population payoff of an evolutionary process of over 10^5 generations in the Prisoner's Dilemma (PD) game. Here, the average payoff varies between P and R, which correspond to full defection and full cooperation, respectively. At the beginning of the evolutionary simulation, cooperative players in the population which have high θ values dominate the population quickly. Then, proself players (with low θ value) emerge gradually to dominate

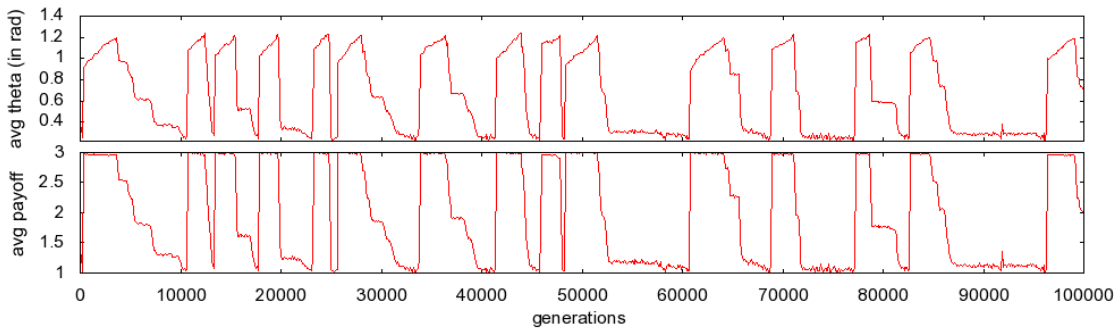

Figure 3.5: An evolutionary simulation of IPD. The top graph shows the average theta per generation. The bottom graph shows the average payoff per generation.

after about 4000 generations. At around the 10000^{th} generation, cooperative players *suddenly* regain the majority of the population. This wave-like behavior between cooperation and defection (prey-predator cycle), is a widely known phenomenon in repeated PD games, that was observed under various conditions [19, 20]. As can be seen in Figure 3.5, similar behavior also emerges in our experiments with players modeled by their social orientation values.

By examining the evolutionary traces, we found that this phenomenon is caused by mutant players introduced in the population with $\theta_{\text{mutant}} \approx \pi/4$. These fair players avoided the exploitation of proself players with $\theta_{\text{proself}} < \pi/4$, and at the same time cooperated with the other prosocial players with $\theta_{\text{prosocial}} > \pi/4$. In other words, the mutant players were using strategies similar to Tit-For-Tat.

Figure 3.6 illustrates the change of population frequencies of three types of players (selfish, altruistic, and fair) without mutation. At the beginning, altruistic players dominate the population. The theta value of an altruistic player is $\pi/2$; i.e., it will always cooperate. Therefore, it can be easily exploited and invaded by

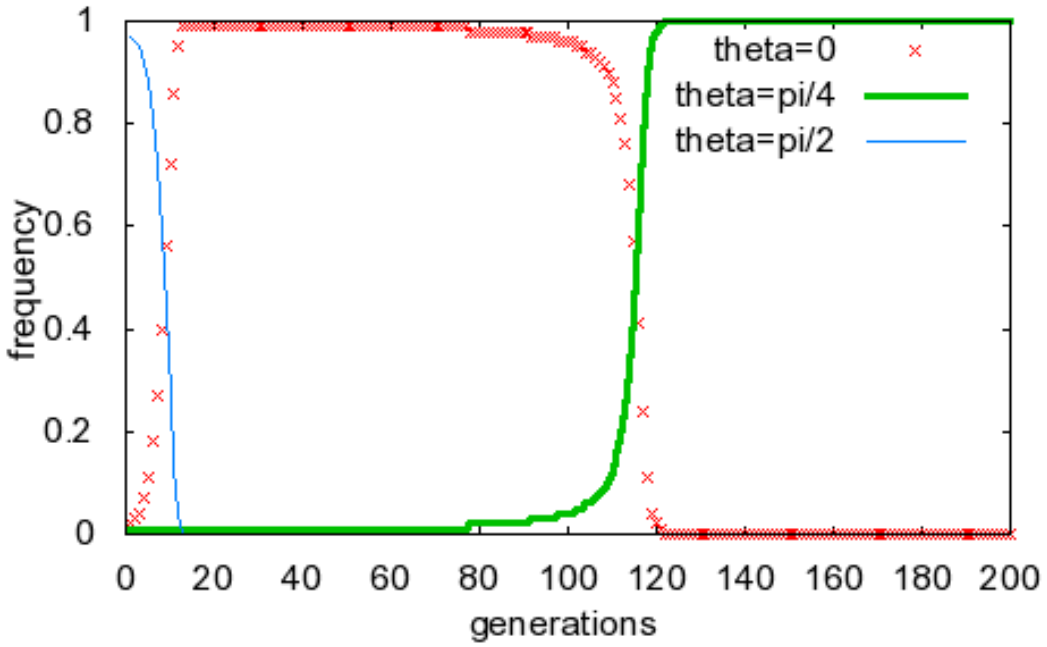

Figure 3.6: Invasion of fair player.

a selfish players ($\theta = 0$). When the altruistic players are extinct, selfish players can also be invaded by a group of fair players who will cooperate among themselves. This evolutionary pattern is similar to the one that emerges in the classical rational agent model [31].

Our results shown in Figure 3.5 also suggested that after the fair player beats the selfish player, the population enters a random drift period. Due to the random mutation, the average theta of the population increases slowly to a point at which there are many highly-cooperative players. Then, mutations introduce selfish players into the populations and their numbers grow quickly until they dominate the entire population. This pattern repeats at least until 10^7 generations. This ratifies previous findings on evolutionary cycles of cooperation and defection [32], which shows that our model based social orientations is capable of explaining those findings.

3.5.2 Prisoner's Dilemma with Varying payoffs

We also investigated the effects that different matrix values have on the result of the evolutionary process and the resulting cooperative societies. In these experiments, we varied one of the entries in the PD game matrix while keeping the others constant with their original values as well as keeping the preference relations in the PD matrix, i.e., $S < P < R < T$ and $2R > S + T$, so that the game will still be a PD game. For each matrix generated in this way, we ran 20 evolutionary simulations with 10^5 generations in each run with a total of about 1000 mutant strategies.

Figure 3.7 shows the effect of varying R on the average theta and average payoff of the population. We report the average of the data after 1000 generations because often the majority of the groups of players did not emerge before that. Increasing R provides added incentive to cooperate. Therefore, both the theta and average payoff increase with R. Note that the payoff almost reaches the maximum (i.e., R) after $R = 4.7$, i.e., it becomes always full cooperation when R is large enough. The bottom graph shows the effect of R on the percentage of *cooperative* agents which is defined by the portion of agents whose θ is greater than the $\pi/4$ (i.e., the θ of a fair player).

Figure 3.8 shows the effect of T on the average payoff and the cooperativeness of the population. These results suggest that increasing T will lead to increase in the incentive to defect. In any situation that can be modeled by a 2x2 game similar to Prisoners' Dilemma, which shows that there is a degradation in the cooperation level. Therefore, both θ and payoff decrease when T increases.

Figure 3.9 shows the effect of P on the average payoff and the cooperativeness of the population. In general, the average payoff increases when P increases. However, unlike the case for R or T, the average of θ drops sharply when P is very large compared to R. These results suggest that increasing P will lead to an increase in the average payoff, but not increase the cooperativeness of the population. In other words, using our model we are able to notice that there is no one-to-one correlation between the observed average payoff and the society's cooperativeness level. In this case, using previously suggested models one could mistakenly reason that increasing P and R has the same effect on the society, while with our new model the difference in the true cooperativeness of the society is apparent by looking at the *theta* values of its individuals.

3.6 Summary

We have described a formal model that combines game-theoretical analyses for cooperation in Iterated Prisoner's Dilemma with insights from social and behavioral sciences. Our model is not claimed to be the most accurate account of social orientations; rather, it is a simple model that takes the first step in the above direction. Unlike existing models, this formalism captures the notion of prosocial vs. proself orientations exhibited in human behavior and explicitly provides an abstract representation for how a player develops its strategies in repeated games.

We have presented theorems showing how players with different social tendencies interact. Our theorems identify five general steady-state behavioral patterns,

that can be explained in terms of the players social orientation values. We have also performed an experimental evaluation of our model using evolutionary simulations in the well-known IPD game. The results of the experiments demonstrated that our model captures the well known behavior patterns in IPD. Furthermore, it allows modeling richer behavior patterns since it does not depend on the particular game matrix.

When we varied the payoffs in the game matrix while keeping the preference relations intact in the PD game, one set of experiments showed that prosocial tendency increases when the reward (i.e., R) of the game increases or when the temptation (i.e., T) decreases. Another set of experiments identified a class of scenarios in which the evolution simulations produced a population that is not socially-oriented toward cooperation, whereas the average payoff of the population is still high. This result is contrary to the implicit assumption of all previous works that considered cooperative populations, that the high-payoff was assumed to be an indicator for cooperativeness. Our experiment showed that social orientations in a population could be a more realistic representations of the cooperativeness of the entire population.

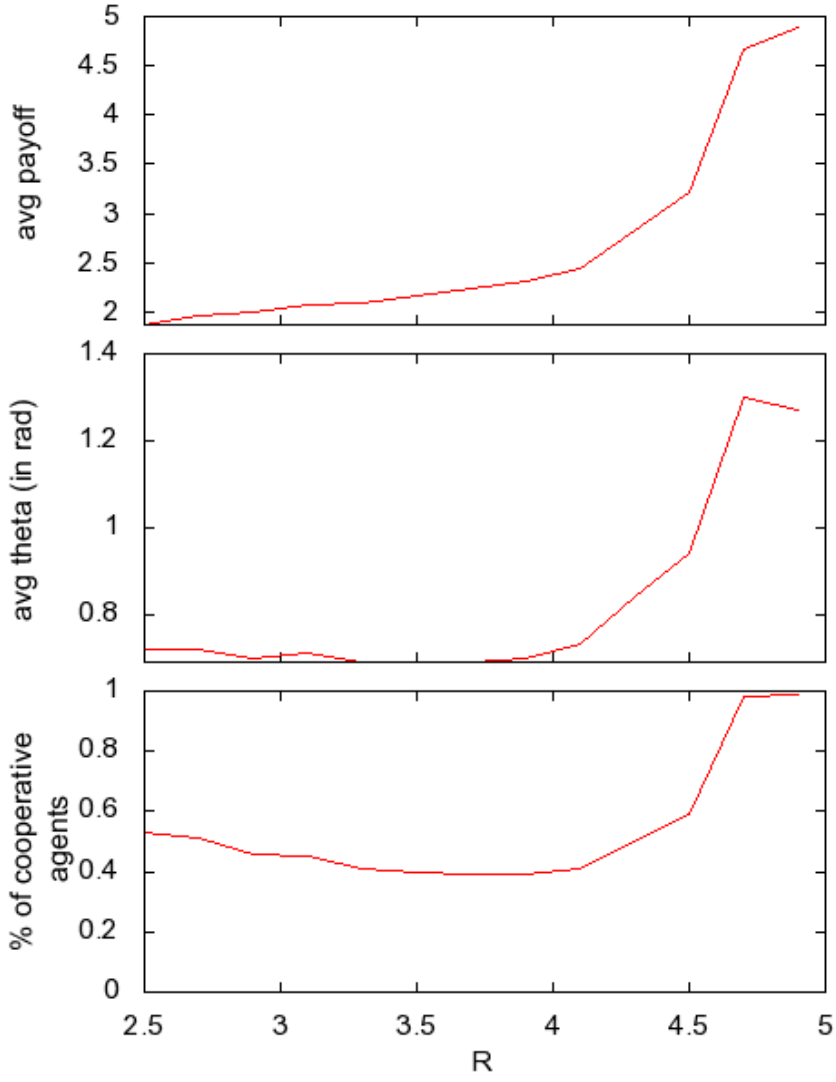

Figure 3.7: Top graph: effect of R on average payoff. Middle graph: effect of R on average theta. Bottom graph: effect of R on the percentage of cooperative agents.

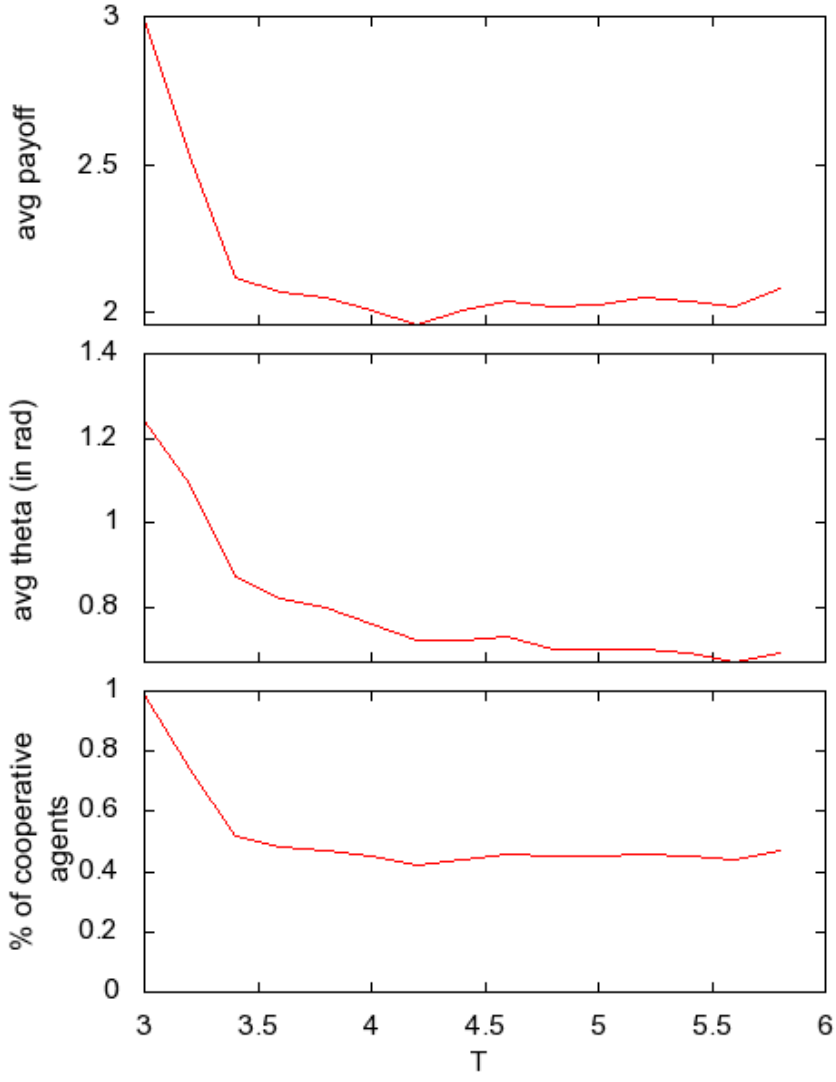

Figure 3.8: Top graph: effect of T on average payoff. Middle graph: effect of T on average theta. Bottom graph: effect of T on the percentage of cooperative agents.

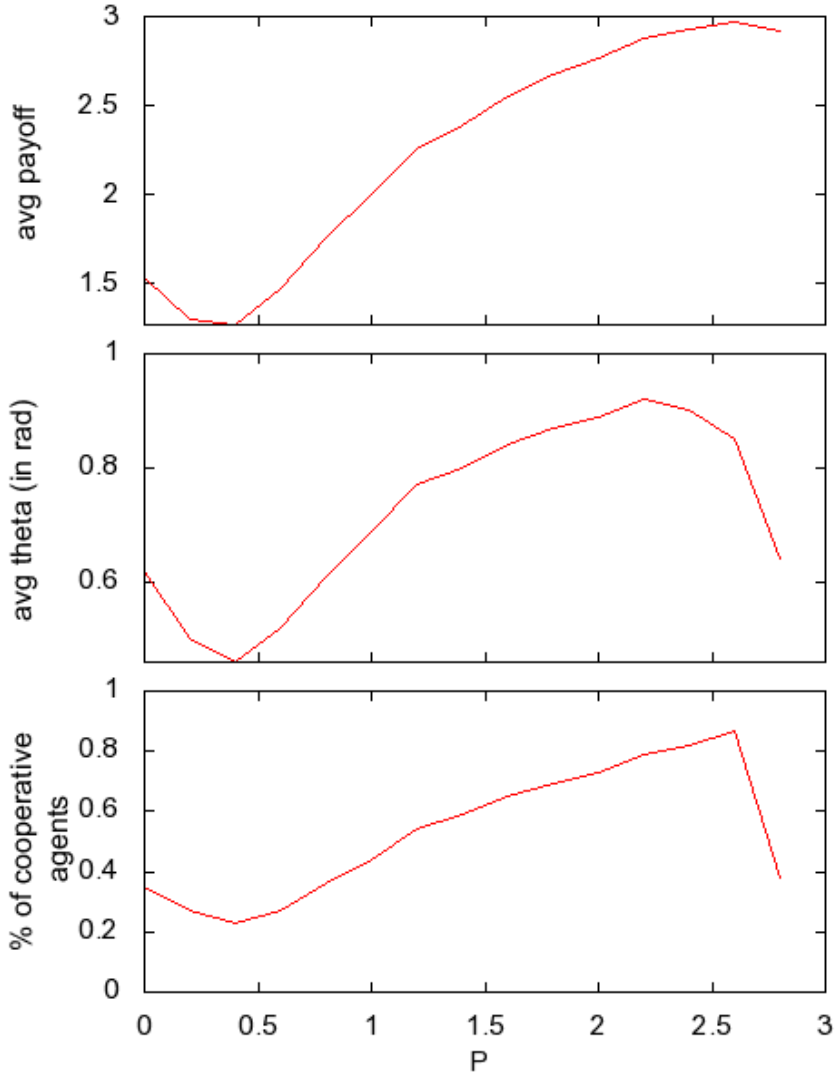

Figure 3.9: Top graph: effect of P on average payoff. Middle graph: effect of P on average theta. Bottom graph: effect of P on the percentage of cooperative agents.

Chapter 4: Cognitive Strategies for The Life Game

Standard repeated-game model 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 chapter, 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.

4.1 Introduction

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 [12]. Maynard Smith studied two-player Chicken game with a population of Hawks and Doves [13], and Skyrms studied the evolved population when individuals were playing the Stag-hunt game [14].

Each of the above studies used a simple game model in which the same stage game was used at every iteration. However, as discussed in Section 2.2 repeatedly playing the **same** game is unlikely to be an accurate model of any individual's life. As a 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 many stage games. We formally described the Life Game in Section 2.2. In this chapter, we concentrate on discovering behavioral strategies that are successful in life games. 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.

The most relevant piece of literature to our study is a recent paper [33] 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.

This chapter makes the following contributions. We 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.

4.2 Strategies for the Iterated Prisoner's Dilemma

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 Figure 3.1), 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 [12]. 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 [22, 34].

As our main focus of this chapter 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 knowl-

edge 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 chapter, this assumption is no longer valid. As we will see below (in Section 4.3), most known strategies simply cannot be generalized to the complex world of the life game, and consequently, new ones must be defined.

4.3 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 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.

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

Figure 4.1: 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.

To illustrate the problem, consider two tit-for-tat-like agents (x and y) playing in a repeated game of Stag Hunt (Figure 4.1). Suppose both of them want to cooperate (hunt stag together) in the Stag Hunt game, but x does not want to cooperate in the Prisoner's Dilemma (while y still want to cooperate). If x and y 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 x defects y in the first game, y retaliates by "defecting" in the Stag Hunt game (i.e., hunt Hare). Therefore, x 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.

4.4 Social Value Orientation Agent Models

According to the SVO theory (described in Section 2.1), the choices people make depend, among other things, on stable personality differences in the manner in which they approach interdependent others. 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:

$$\text{Utility} = w_1 \times \text{my payoff} + w_2 \times \text{other's payoff}$$

- Altruistic agent maximizes other agent's outcome.

  $(w_1 = 0, w_2 = 1)$

- 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.

  $(w_1 = 1, w_2 = -1)$

- Adverse agent minimizes other agent's outcome.

  $(w_1 = 0, w_2 = -1)$

In order to promote cooperation, both agents need to be prosocial. As mentioned in [27], "An excellent way to promote cooperation in a society is to teach people to care about the welfare of others". 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, prosocial individuals reciprocate more as the trust increases, while proself reciprocate less as the trust increases [35]. 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.

4.4.1 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 $max(a, d) \geq max(b, c)$

[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.

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 τ is a subset of type $\tau + 1$. For example, the Stag Hunt game (Figure 4.1) is a member of all of the above types, while the Prisoner's Dilemma (Figure 3.1) 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 λ if and only if the agent cooperates in **all** mutual-benefit games of type τ. It is easy to notice that it is riskier to cooperate in type $\tau + 1$ games than in type τ 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 $\tau + 1$ 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 λ. Whenever the other agent defects in a mutual-benefit game of type τ, we update λ by $\lambda = min(\lambda, \tau - 1)$. 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 λ when f cooperative actions in a game of type λ 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.

4.4.2 Individualistic Model

According to the SVO theory, an individualistic agent will try to maximize its own outcome. However, the information that an agent x 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 y. For instance, its actions might be different if it assumes y picks its actions randomly, or tries to intentionally decrease x's payoff.

To cope with this problem, we suggest using two-level agent modeling. In this model, when an individualistic agent x is playing with another agent y, x behavior depends on the second-level model – model of y from x's perspective. With that assumption, x 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 4.5). 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.

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

Figure 4.2: 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.

We illustrate it by following example: an individualistic agent x is playing the

Chicken game (Figure 4.2) with y.

- **Adversary model** - x assumes that y wants to minimize its outcome. Then, it reasons that (1) y will choose A_2 if x chooses A_1; (2) y will still choose A_2 if x chooses A_2. The payoffs are $(3, 5)$ and $(0, 0)$ respectively, and x will choose A_1. In other words, x best response is to be playing a maximin strategy.

- **Altruistic model** - x assumes that y is wants to maximize x's outcome. Then, it reasons that (1) y will choose A_1 if x chooses A_1; (2) y will still choose A_1 if x chooses A_2. The payoffs are $(4, 4)$ and $(5, 3)$ respectively, and x will choose A_2. In this case x's best response strategy is the maximax strategy.

- **Random model** - x assumes y 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 $\frac{c+d}{2} = 2.5$. x will choose A_1 only if $\frac{a+b}{2} > \frac{c+d}{2}$, and choose A_2 otherwise. Therefore, x will choose A_1 in the Chicken game. We will call x is playing a maxi-random strategy.

- **Recursive model** - Finally, x can assume that y is any kind of agent described above. x will first predict y 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, x's strategy is the best response to the assumed y's strategy. For example, x can assume that y is an individualistic agent with random opponent assumption (i.e., y uses the maxi-random strategy). From

the previous paragraph, we know that y will choose A_1 in the Chicken game. Therefore, x will also choose A_1 in order to maximize its own payoff. We will call x is playing a maxi-maxi-random strategy.

4.4.3 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 [27].

4.4.4 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 Figure 4.3. Since we assume that all agents does not have any prior knowledge about the other agent, the SVO agent does not know the

Procedure SvoAgentPlayingLifeGame

Input and Notation:

g_t, g_{t-1} = current, and previous game matrix

A = previous opponent's action

M = current set of candidate opponent models

λ = current trustworthiness of the opponent

$\tau(g)$ = trustworthiness requirement of g

$C(g)$ = cooperative action of g if g is a mutual-benefit game, $\emptyset$ otherwise

Output: An action for the current game g_t

Begin procedure

(1) Update opponent's trustworthiness and models (when $A \neq \emptyset$ and $g_{t-1} \neq \emptyset$).

If $C(g_{t-1}) \neq \emptyset$ and $\lambda \geq \tau(g_{t-1})$ and $A \neq C(g_{t-1})$, then $\lambda \leftarrow \tau(g_{t-1}) - 1$

Increase the counters of all models (in M) which correctly predict A for g_{t-1}.

(2) Choose and return an action for the current game g_t.

If $C(g_t) \neq \emptyset$ and $\lambda \geq \tau(g_t)$, then return $C(g_t)$.

Else

$m \leftarrow$ the most accurate model in M

If $i \leq 5$ or accuracy of $m < 70\%$, then return maxi-random action of g_t.

Else return the best response to m's prediction in the game g_t.

End procedure

Figure 4.3: Procedure for a (unforgiving) SVO agent playing a game g_t at t-th iteration in a life game.

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., λ in Figure 4.3) and social orientation of the other agent, and will adapt and utilize it to the best of its capacity.

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 Figure 4.3). In this chapter, 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 Figure 4.3). 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., A in Figure 4.3).

4.5 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 students' agents / 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 [23–25]. Lin et al. provided an empirical proof that PDAs can alleviate the evaluation process of automatic negotiators, and facilitate their designs [23].

To obtain a large collection of students' agents, 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 Figure 2.1. 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).

4.5.1 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 students' agents + 5 standard strategies), the total number of participant in each run of the competition

Table 4.1: Average payoffs and rankings of the SVO agent, standard agents and top three students' agents

Agent	Rank and (Avg Payoff)
SVO agent	1 (5.836)
The best students' agent	2 (5.831)
The 2nd best students' agent	3 (5.792)
The 3rd best students' agent	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)

is $m = 54$. The tournament is similar to Axelrod's IPD tournaments [12] and the 2005 IPD tournament [36]. Each participant played against every participant including itself (thus a tournament among m agents consists of m^2 iterated games). The number of iterations in one life game was set to $n = 200$. 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 $100 \times n \times m$ payoffs, where n is the number of iterations in each life game and m is the number of participating agents. Hence, each average payoff is computed from averaging the payoffs of 1080000 games.

Table 4.1 shows average payoffs and rankings of the SVO agent, standard agents and the top three students' agents. 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 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 students' agents 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

Table 4.2: 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)

the SVO algorithm which is relatively short and simple, their algorithms are much longer and complicated.

4.5.2 Evaluating Trust Adaptation: to forgive or not forgive?

As mentioned in Section 4.4.1, 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 4.2, the average payoffs of all of the forgiving agents

Table 4.3: 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)

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 students' agents (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 students' agents constituted a biased sample. We observed that if a students' agent 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.

4.5.3 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 4.3 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.

4.5.4 Evaluating Robustness to Number of Iterations

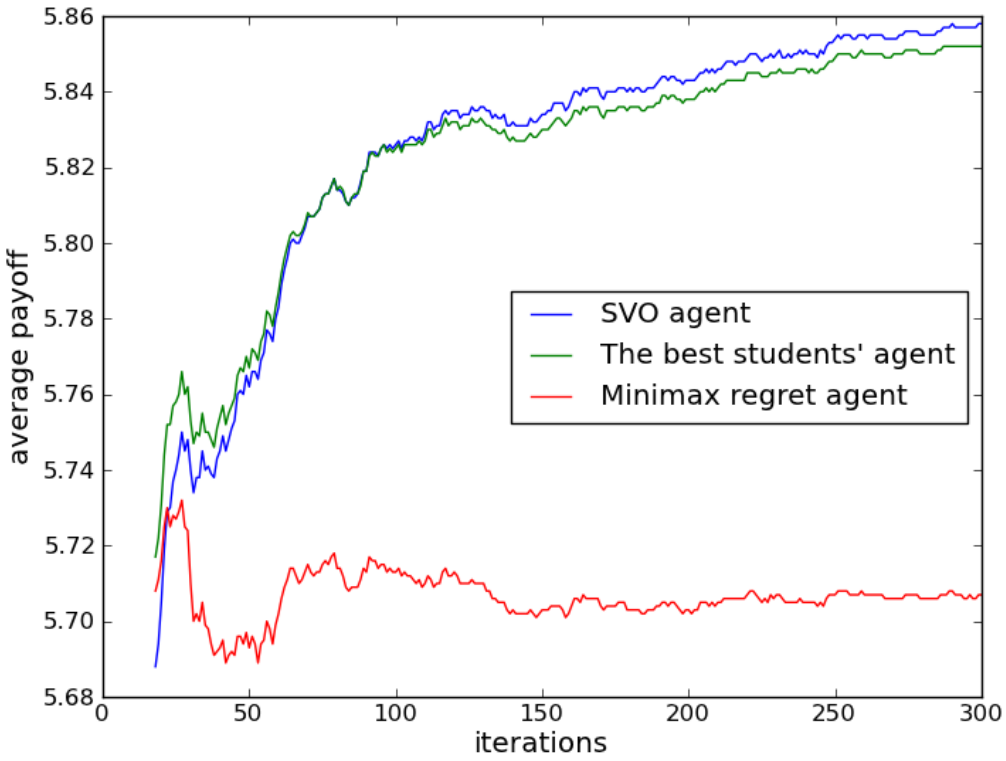

Figure 4.4: Average payoffs at each iteration.

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. Figure 4.4 shows the trends of the average payoffs of the SVO agent, the best students' agent and the best standard agent (i.e., the minimax regret agent). With an increasing number of iterations, both SVO agent and the best students' agent 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 students' agent are

doing opponent modeling. With an increase in the 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 students' 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.

4.5.5 Analyzing the Trustworthiness of the students' agents

With this analysis we seek to explore the trustworthiness of the students' agents that were written by students, and investigate the significance of each mutual-benefit game type. We did as follows: each of the students' agents played against our SVO agent. While in game, at each iteration, the distribution of all five types of cooperative agents was recorded. Figure 4.5 shows the portion of the five types of cooperation classification at each iteration. At the beginning, as we mentioned in Section 4.4.4, the SVO agent assumed that all the other agents are trustworthy, so

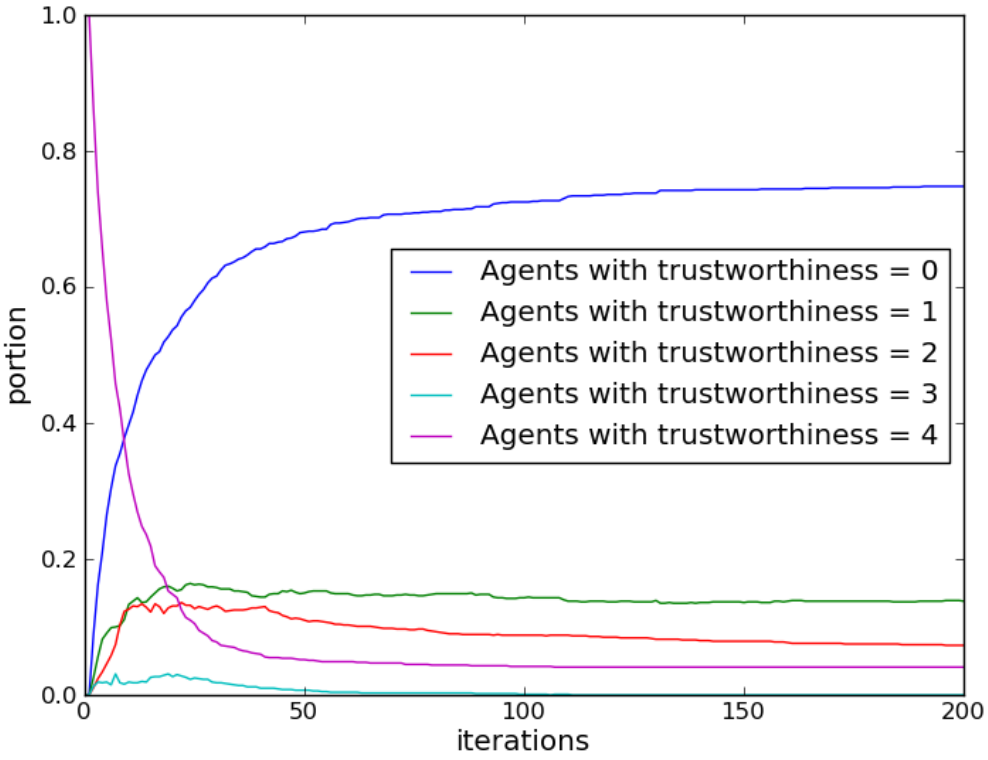

Figure 4.5: Distribution of trustworthiness of the students' agents at each iteration.

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.

4.5.6 Evaluating the Benefit of Cooperation

Figure 4.6: Average payoffs at each iterations.

According to Axelrod's analysis cooperation is a crucial aspect of successful strategies in the IPD tournament. While this hypothesis seems highly intuitive and had many justifications from theoretical biology, in this experiment we went examined this hypothesis explicitly.

To study the amount of the benefit of cooperation, we implemented a paranoid variant of the SVO agent, which do not trust other agent at any time. The paranoid SVO agent assumes that all of the other agents are non-cooperative (type 0), so it will not try to cooperate with other at all. In other words, we eliminated the "nice"

property of the SVO agent so it does not trust anyone at any time.

We test our original SVO agent, the paranoid SVO agent, and the best students' agent with the test set agents. The average payoffs of all three agents are recorded at each iteration. Figures 4.6 shows the trends of their average payoffs. With the increasing number of iterations, all agents obtained more payoffs in general and level off after 200th iteration. It is because all of the three agents are doing opponent modeling. When there is enough history of interaction, the modeling will be more accurate, and so they can better coordinate with their opponent to get more payoffs. The payoff of the original SVO agent is low at the beginning of the repeated games, because it will try to be nice to all other agents at the beginning. 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. Its performance catches up quickly and outperforms others after 100th iteration, because it will stop cooperating with the defectors and keep cooperating with the cooperators. The other two agents do not fully cooperate with others, so they cannot get the benefit from mutual cooperation. Therefore, they cannot obtain higher payoff in long run.

4.6 Summary

In this chapter we have described several new challenges posed by the life game poses, 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 impacts that different trust adaptation strategies have on the performance of the SVO agent.

Chapter 5: Modeling Agents using Designers' Social Preference

According to the SVO theory (discussed in Section 2.1), there is evidence that in interactions with others, humans have preferences that depend partly on a stable, measurable personality trait called *Social Value Orientation (SVO)*. Thus, if a human writes an agent to act as the human's delegate in a multi-agent environment, one might expect the agent's behavior to be influenced by the human designer's SVO. In this chapter, we present experimental evidence that show that the social preferences of computerized agents correlate positively with the social preferences of their human designers. We also show that the human designers' social preferences can provide useful predictions of their agents' behaviors.

5.1 Introduction

Human social preferences have been shown to play an important role in many areas of decision-making; e.g., interaction in labor markets [37], bilateral or small-group bargaining [38, 39], social welfare considerations [40]. The social preferences depend partly on a stable, measurable personality trait called a human's *Social Value Orientation (SVO)* (Section 2.1). In multi-agent systems, agents are often written by humans to serve as their delegates when interacting with other agents. Thus, one

might expect an agent's behavior to be influenced by the SVO of its human designer. The purpose of this chapter is to explore the correlation between the social preferences of human designers and theirs computer agents.

There are many methods to gauge human social preferences or even personalities, as it is well studied in social psychology. Several measurement methods for quantifying variations in SVO across individuals have been developed [6, 41, 42]. In addition to taking a psychology test that may be impractical in some situations, personality of humans can also be estimated by social media, where users present themselves to the world, revealing personal details and insights into their lives. For example, Golbeck et al. [43,44] presented methods by which a user's personality can be accurately predicted through the publicly available information on their Facebook or Twitter profile. Trust and distrust between users in social network can also be accurately computed by an inference algorithm [45].

To obtain the social preferences of computer agents, we needed to have some measurement methods. Quantifying the social preferences of computer agents presents several challenges. The tests that were designed for humans cannot be easily converted to automated agents. For instance, an agent that was constructed to play in a simple repeated game cannot provide answers to questions out of the game contexts, such as "what will you do in this situation?" More precisely, human SVO is usually measured by one-shot games by giving the following instruction: "You have been randomly paired with another person whom we refer to simply as other. You will never knowingly meet or communicate with this other, nor will (s)he ever knowingly meet or communicate with you." In this chaper, we propose to use ideas

and techniques derived from SVO theory to measure computer agents' social preference. To the best of our knowledge, this is the first attempt to quantify the social preference of computer agents using a theory from social psychology, and we will explore the challenges and provide some ideas to address that challenge.

We collected a set of students' agents playing life game and conducted psychological SVO based evaluation to get the corresponding SVO value of the human who constructed the agent. We estimated the social preference of computer agents by the proposed methods, and studied the correlation. The results show that the SVO of human designer is highly correlated with the social preference of the corresponding agent.

We also show the value of having the SVO of the designer of other agents by presenting an application of using that information on agent modeling. Specifically, we take the life game automated agent described in Chapter 4 and improve it by using the knowledge of the SVO of the human delegator of the agent whom it is playing against. The improvement mainly comes from a more accurate initial agent model that can help the agent avoiding exploitation by some selfish agents.

To sum up, our main goal was to explore the correlation between the Social Value Orientation of computer agents, and the human who designed them. As such our main results can be summarized as follows:

- We explore, discuss, and provide a solution to the question of how SVO tests that were designed for humans can be used to evaluate agents' social preferences (Section 5.3).

- We show that in our example domain (the life game) there is a high positive correlation between the social preferences of agents and their human designers (Section 5.4.1).

- We exemplify how the SVO information of the designer of computer agents can be used to improve the performance of some other agents playing against those agents (Section 5.5).

5.2 Measuring Human Social Preferences

There are many measurement methods proposed by social psychologist for measuring human SVO [6,41,42,46]. To measure SVO of a person x, x is usually asked a series of questions in which he needs to select between certain distributions of resources, some amount to himself/herself, and some amount to be allocated to some other randomly determined person y. The examiner will ask x to imagine that the points involved with the decisions have value to you: specifically, the more of them you accumulate the better. Similarly, x needs to imagine that the other person y feels about his/her own points the same way. It is told that x and y will remain mutually anonymous during and after the decision is made, and there is nothing y can do to affect x in any way. In other words, it is an one-shot game. Hence the choice made by x is not a strategic decision, but rather this is a one-shot individual decision under certainty. Nonetheless this choice has a social dimension, as x's action will affect y's behavior and x is aware of this potential effect.

For example, one well-known technique for measuring SVO used in social psy-

Choose between:

A: $p_{\text{self},i}(A)$ for me, and $p_{\text{other},i}(A)$ for other

B: $p_{\text{self},i}(B)$ for me, and $p_{\text{other},i}(B)$ for other.

Figure 5.1: Format of i-th decision task of the ring measurement questionnaire

Choose between:

A: 26 for me, and 97 for other

B: 38 for me, and 92 for other.

Figure 5.2: A sample decision task used by the ring measurement questionnaire

chology is the Ring measure [41]. Typically, the ring measure involves a series of 24 decision tasks between two options. Figure 5.1 shows the format of each decision task. The participants are told to be randomly paired with another person whom the question refers to as "other." In the decision task, the participants will be making choices by circling the letter "A" or "B" on a response sheet. The participants' choices will produce points/money for themselves and the other. The options involve combinations of own outcome and other outcome. A sample decision task used by the ring measurement questionnaire is shown in Figure 5.2.

Adding up the chosen amounts separately for the self and for the other player provides an estimation of the weights assigned by the participant to own and other's payoffs. These weights are used to estimate the SVO angle (θ) of the participant by the formula below:

$$\text{SVO angle} = \theta = \arctan\left(\frac{\sum p_{\text{other},i}(r_i)}{\sum p_{\text{self},i}(r_i)}\right), \text{ where } r_i = \text{i-th response} \tag{5.1}$$

All angles between 112.5° and 67.5° were classified as altruistic; those between

67.5° and 22.5° were classified as cooperative; those between 22.5° and 337.5° as individualistic, and angles between 337.5° and 292.5° as competitive. [1] [41]

Since the total number of points a participant receives on each decision problem is determined by the combination the choices of both participants, the participants are in fact playing the following symmetric game for the i-th decision problem:

		Player 2	
		A	B
Player 1	A	$p_{\text{self},i}(A) + p_{\text{other},i}(A)$	$p_{\text{self},i}(A) + p_{\text{other},i}(B)$
	B	$p_{\text{self},i}(B) + p_{\text{other},i}(A)$	$p_{\text{self},i}(B) + p_{\text{other},i}(B)$

For example, the sample decision task mentioned above can be written as the following symmetric game:

		Player 2	
		A	B
Player 1	A	123	118
	B	135	130

There are several other techniques for measuring social preferences, such as the decomposed game measure, the triple dominance measure, and the slider measure. In a decomposed game, participants choose between three options that offer points to the self and another person. The most commonly used measure of SVO is the

[1] The boundary between cooperative and individualistic is $\frac{45^\circ+0^\circ}{2} = 22.5^\circ$. Other boundary angles can be derived similarly.

9-item triple-dominance measure. Typically, participants are classified as one of three orientations (cooperators, individualists, or competitors) if they make 6 out of 9 choices consistent with the orientation. Like the ring measure, the slider measure can help us estimate the SVO angles of participants; and it has been reported that the slider measure has better test-retest reliability [47]; but as described in the next section, Ring measure is the only one that can be adapted for use in a repeated-game setting.

5.3 Measuring Agents' Social Preferences

In order to model the behavior of an agent, we would like to have a precise quantitative measurement (like SVO angle) on computer agents. For example, a Maximin agent maximizes its worst-case payoff, so its SVO angle always equals to $0°$. A Minimax agent minimizes other agent's best-case payoff, so its SVO angle always equals to $-90°$. Except for Ring method, all the other measurement methods cannot be transformed into 2x2 games, so they cannot be used to measure social preferences of computer agents playing 2x2 games (e.g., life game). Therefore, we use a modified version of Ring method to measure social preference of agents.

Although the choice questions in Ring measurement can be presented as 2x2 normal form games, most of the payoff values of the game matrices are not valid for the life game model we used. In the life game model we used, the payoff values must be in the range $[0, 9]$. To apply Ring measurement on automated agents, we modified the game matrices to G_{Ring} by downsampling, scaling, and translating, so

that all payoff values will sit within $[0, 9]$.

Another problem is that the agents were designed under the assumption that they may have repeated interactions with the other agents. It is likely that the social preference of an agent varies with the current number of iterations and the behavior of the other agent. For example, an aggressive partner might trigger an aggressive behavior even from an initially cooperative agent. This is so because the decision may involve many factors like behavior of other agent, social preference and competence of the human designer. However, one-shot games are used in all of the SVO measurement methods for human, and the participants are told that they will remain mutually anonymous during and after their decisions are made. This non-repetitive interaction assumption is not valid in most multi-agent environments. In the environment we used, the repeated interaction is modeled by a repeated game with unknown number of iterations.

In repeated games, an agent's social preference can be influenced not only by the agent's own SVO, but also by how the agent reacts to the other agent's SVO. For example, let x be an agent whose SVO is $45°$ (i.e., it prefers equal payoffs for both agents) and y be a memoryless agent whose SVO is $0°$ (i.e., y cares only about maximizing its own payoff in the current iteration). If x and y interact repeatedly, then after repeated observations of y's behavior, x might decide that the best way to equalize both agents' cumulative payoffs might be for x to try to maximize its own payoff at each iteration. Consequently, if we perform a Ring measurement of x after it has had many interactions with y, x's "apparent" SVO value may be closer to $0°$ than $45°$. We will call this x's *para-SVO* against y.

The para-SVO, $\theta_n(x|y)$, of agent x at the n-th iteration with tester agent y is measured by applying the modified Ring measurement on the agent at the $(n+1)$-th iteration after it interacted with the tester agent y for n iterations. In this chapter, we use a random agent as the tester agent y. [2] The parameter n is introduced, because we would like to measure social preference, which might change during the interactions, at a specified iteration. Our measurement algorithm uses the behavioral data of the agent at the last iteration, therefore the para-SVO represents just the *latest* social preference of the agent after n games. Figure 5.3 shows the complete procedure for measuring $\theta_n(x|y)$ using the modified game matrices G_{Ring}. It will get the responses from the testee agent at the last game which is one of the games in G_{Ring}, and then calculate the para-SVO using Formula (5.1). We verified the validity of the measurement by applying it on some simple agents with known para-SVO angles (e.g., para-SVO angles of maximin, minimax, and a prosocial agent are $0°$, $-90°$, and $45°$ respectively).

5.4 Experiments on Measuring Agents' para-SVO

In this section, we will present some results on the relationship between social preferences of agents and that of their designers. We collected a set of peer-designed agents (PDAs) by asking students in several advanced-level AI and Game theory classes to contribute agents. 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

[2] In the next chapter, we futher extend the notion of para-SVO by having a special set of tester agents.

Procedure MeasureParaSvo

Input:

n = number of random games (before measurement)

r = number of runs

G_{Ring} = set of games of modified Ring measurement

Output:

para-SVO of x with a tester agent y after n random games

Begin procedure

$p_x \leftarrow 0$ /* p_x = total payoff of agents using x's strategy */

$p_y \leftarrow 0$ /* p_y = total payoff of agents using y's strategy */

Repeat for r times:

For each game g in G_{Ring}:

create new agents x' and y' which use the same

strategies of x and y respectively

pair up x' and y' for a repeated game with n

random games and then g as the last game

$p_x \leftarrow p_x +$ (last gain of x' due to x''s last action)

$p_y \leftarrow p_y +$ (last gain of y' due to x''s last action)

End For

End Repeat

Return $\arctan(\frac{p_y}{p_x})$

End procedure

Figure 5.3: Procedure of measuring para-SVO of an agent x with a tester agent y after n random games.

a round-robin fashion). We also asked the students to participate in an online SVO measure [48].

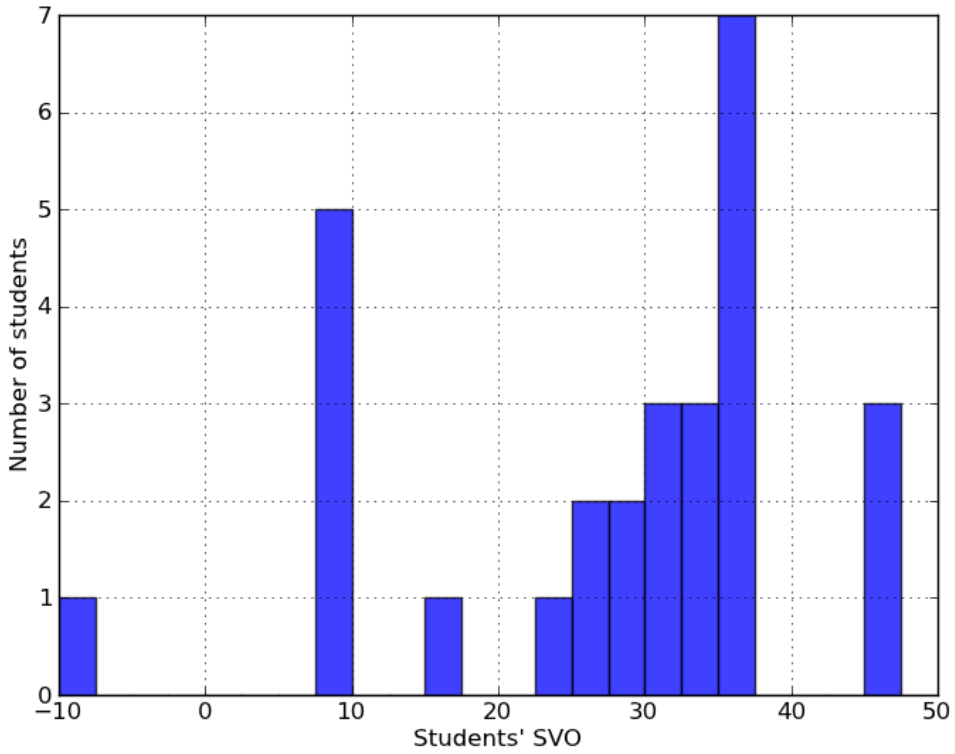

Figure 5.4: SVO of 28 students.

We collected 28 agents with SVO of the corresponding designer. Figure 5.4 shows the SVO distribution of all 28 students. 21 of them are cooperative ($67.5° >$ SVO angle $> 22.5°$ [41]), and 7 of them are individualistic ($22.5° >$ SVO angle $> -22.5°$).[3]

[3]It skews toward cooperative orientation, possibly because we collected the data from students voluntarily responding to our survey.

5.4.1 Agent-human SVO Correlation

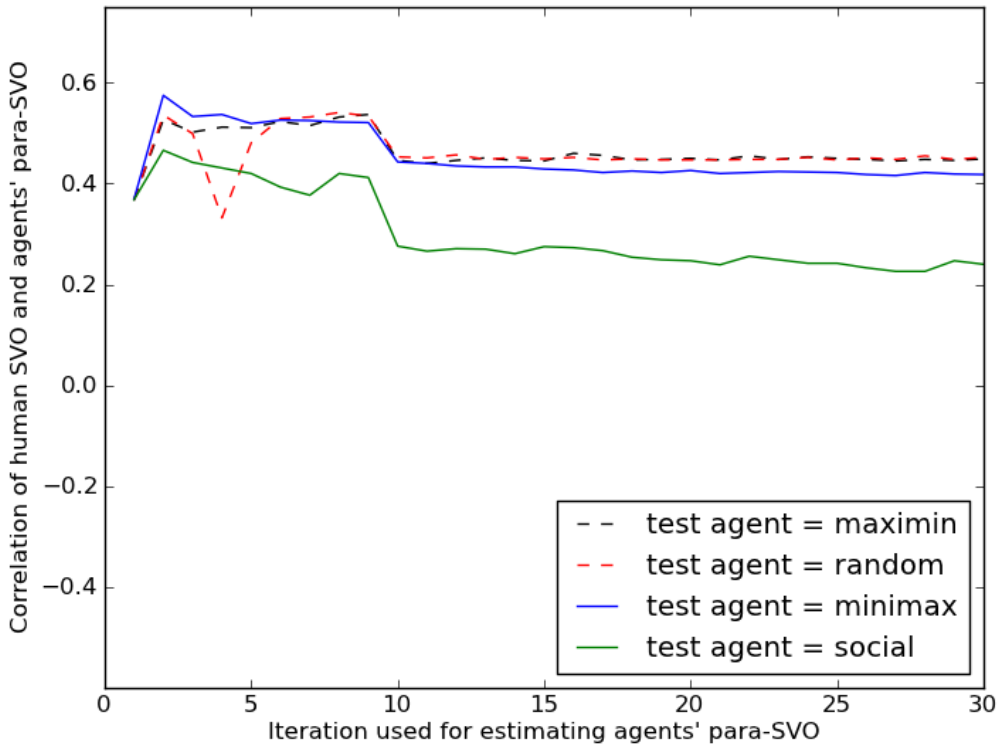

Figure 5.5: Correlation of human SVO and agents' para-SVO.

We used the modified ring method presented in Section 5.3 to measure the para-SVO of all computer agents using different tester agents, and then calculated the Pearson correlation of the agents' social preferences and human SVO. Figure 5.5 shows the correlation of human SVO and agents' para-SVO measured by the modified ring method using different tester agents.

The x-axis of Figure 5.5 is the number of iterations (n) used for measuring the agent SVO. The correlation at the first iteration is about 0.4, and then rises to about 0.55, which is considered high in behavioral sciences [49], for several iter-

ations. After reading the source codes of the agents, we found that some students wrote some codes for the first iteration only. In other words, they hard coded some initial behaviors that may be different from other iterations. Figure 5.5 shows that the correlation between agents' para-SVO and human SVO rises to a peak at second iteration, and then decreases and level off for the rest of the repeated game. From examining the code, we found that many of the agents try to build a model of the other agent in the game, based on the history of interactions. As the game progresses, such an agent's behavior will come to depend partly on the social preference of the designer, and partly on the agent's predictions of the other agent's behavior. We surmise that this effect is responsible for the correlations shown in Figure 5.5.

5.4.2 Stationary vs. Non-stationary Strategies

The guess of the previous subsection motivated us to classify and investigate the agents based on the complexity of their strategies. We divided the agents' strategies into two groups, stationary and non-stationary, according to the variance of their para-SVO:

1. For agents using stationary strategy, given a tester agent, their para-SVOs remain the same all the time, because their choices at each iteration depend only on the payoff matrix of current game. They usually have shorter and simpler codes. For example, some students' agents use a simple competitive strategy that choose action A_1 when $b > c$, and choose action A_2 otherwise. We have about 12 agents using stationary strategy among those students'

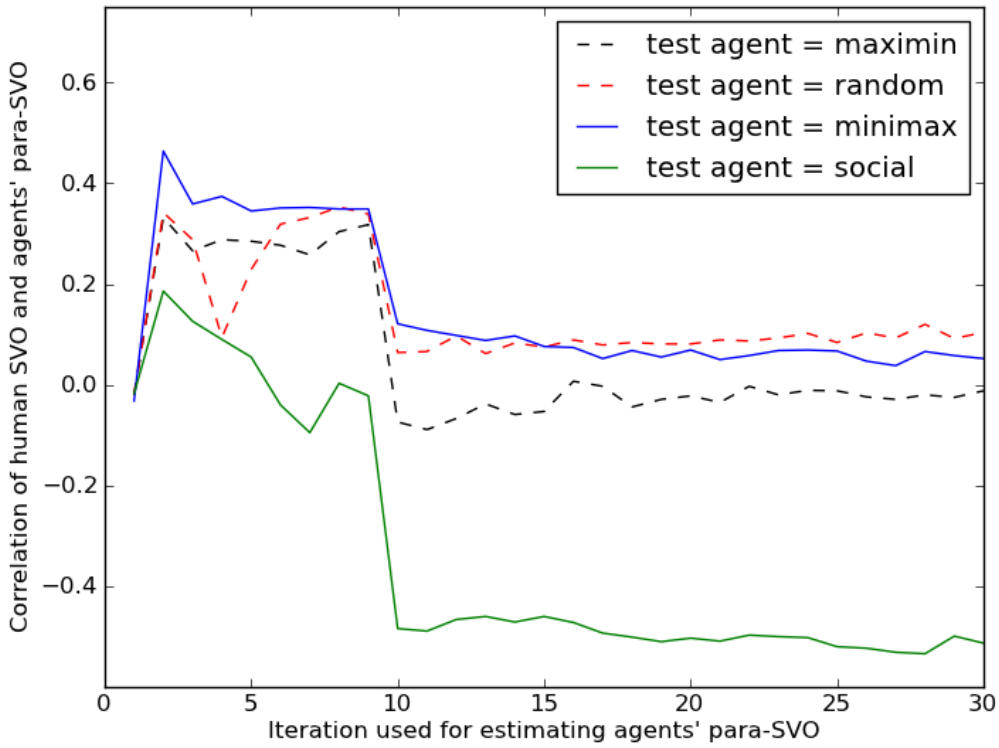

Figure 5.6: Correlation between human SVO and agents using non-stationary strategy.

agents, and the correlation is approximately 0.6.

2. For agents using non-stationary strategy, given a tester agent, their para-SVOs change as the game progresses, because their choices at each iteration depend on the previous history of the interactions. They may build predictive models of some kind and make some strategic decisions based on the model. For example, some students' agents estimate the probability of the other agent choosing some kind of action in different situations, and then respond accordingly.

Figure 5.6 shows the correlation of human SVO and para-SVO of agents using non-stationary strategy. Comparing with the correlation for agents using stationary strategy ($\approx$ 0.6 for all iteration), the correlation for agents using non-stationary strategy is lower (ranging from -0.4 to 0.4). This is consistent with our previous guess that para-SVOs of agents using non-stationary strategies correlate less with the social preferences of their designers.

5.5 Utilizing the SVO Information

There are many possible applications of utilizing the SVO information. Having the SVO information, we can predict the behaviors of the other agents in various situations, and our agents can interact with them in some better ways. For example, our agent can avoid possible exploitation by agent that is probably competitive. On the other hand, if we know that the other agent is possibly cooperative according to the SVO information, our agent can possibly increase mutual benefits by working closely with the other agent. If the other agent is neither competitive nor cooperative, our agent can start with some safe actions and learn a more accurate model of the other agent from interaction. Using that kind of strategy, we can develop a collaborative agent to enhance safety and productivity by utilizing the SVO information.

In this section, we present two examples that use the SVO information. First, we show how we can use the SVO information to composite two simple and non-adaptive agents to form a better non-adaptive agent. Second, we present how we can improve an adaptive agent [50] by using the SVO information.

5.5.1 Compositing a Non-adaptive Agent

In this subsection, we present a way to use the data of other agents' designer to combine two simple agents to form a better agent. The two agents we used are social agent and maximin agent. Both of them are non-adaptive that they do not apply any agent-modeling technique during the game. Social agent always chooses an action that maximizes the sum of payoff of itself and other, so its SVO angle is 45° and it performs better if the other agent is also cooperative. Maximin agent always chooses an action that maximizes its own minimum possible payoff, so its SVO angle is 0° and it can avoid being exploited by other non-cooperative (individualistic or competitive) agents.

It would be better if we can combine the advantage of both agents in following way: if the other agent is cooperative, our agent will act like the social agent to gain the benefit of mutual cooperation; if the other agent is non-cooperative, our agent will act like the maximin agent to avoid being exploited by them. However, as we do not know the exact social preference of other agents before interacting with them, we propose to approximate the social preference of the other agents by the SVO of their designer. In other words, if the SVO of other agent's designer is in the cooperative range ($\geq 22.5°$), our agent will act like a social agent; otherwise ($< 22.5°$), our agent will act like a maximin agent.

We implemented the simple agents and the proposed composite agent described above, and compared their performance in tournaments (10000 runs) with the 28 students' agents. Figure 5.7 and 5.8 shows the average payoffs when the

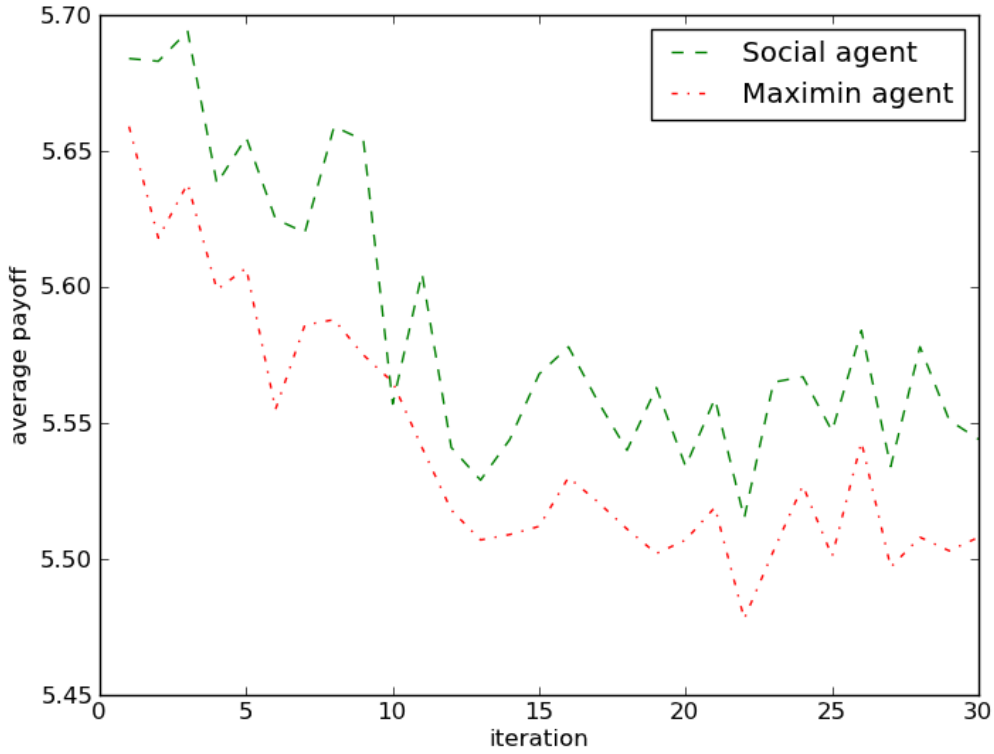

Figure 5.7: Performance of the simple agents playing with 21 agents written by cooperative humans.

simple agents play with 21 agents written by cooperative human and 7 agents written by individualistic human. Social agent performs better with agents written by cooperative human, while maximin agent performs better with agents written by individualistic human. This is so because human SVO is a good approximation of the social preference of the agents. Figure 5.9 shows the average payoffs when the three agents playing with all 28 students' agents. It shows that the average payoff of the composite agent is (almost, except one point) always higher than both simple agents, because it has the strengths of both agents in different situations. The re-

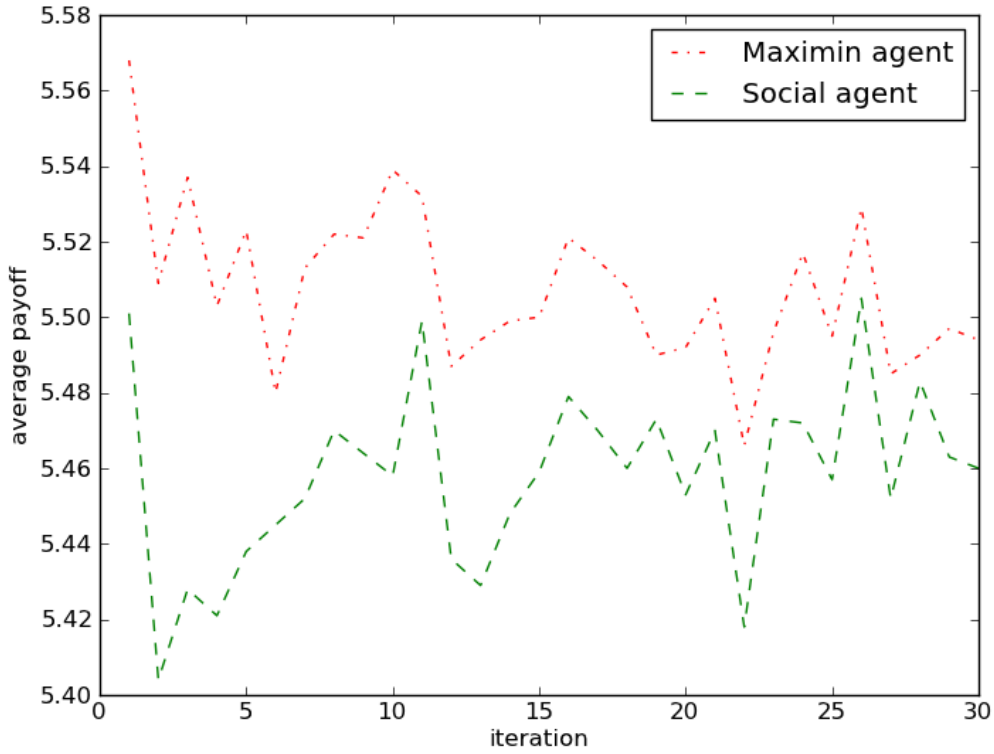

Figure 5.8: Performance of the simple agents playing with 7 agents written by individualistic humans.

sult also shows that the performances of the simple and composite agents drop when there is more iterations. It is because some students' agents apply agent-modeling technique that they can easily exploit agents using stationary strategies (including the simple and composite agents) after they have enough interaction data. As we shown in Section 5.4.2, if the other agents use non-stationary strategy, designers' SVO is not enough to model them later in the game. We still need agent-modeling techniques during interaction to being exploited by those agents.

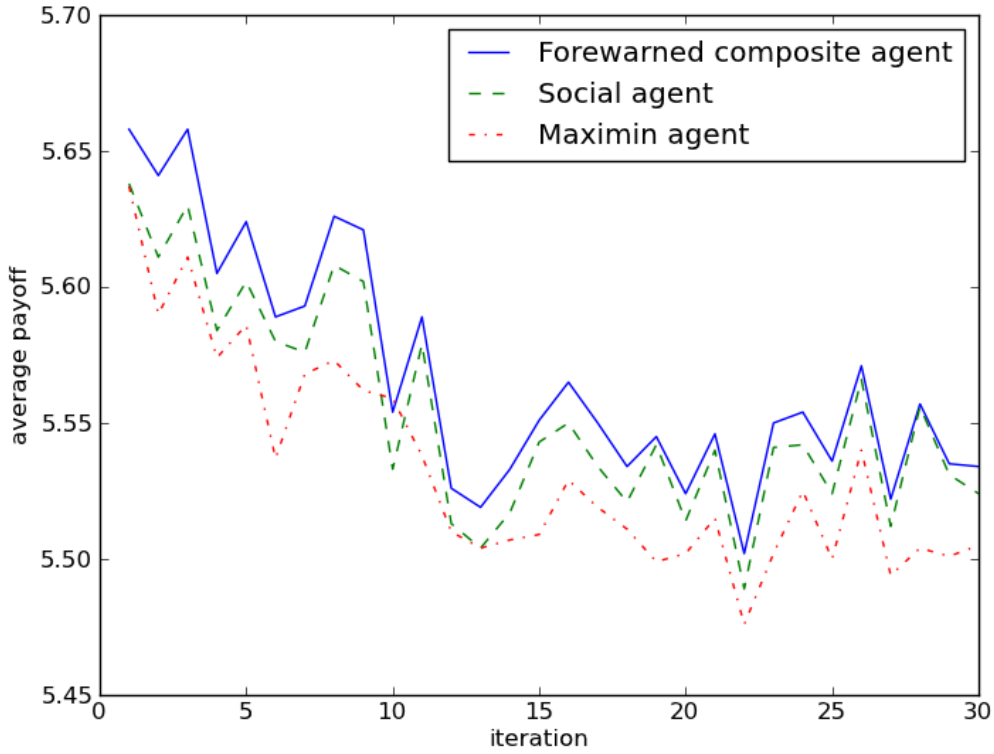

Figure 5.9: Performance of the simple and composite agents.

5.5.2 Improving an Adaptive Agent

In this subsection, we show a way to use the data of other agents' designer to improve an adaptive agent that apply agent-modeling techniques. We use the adaptive life game agent described in Chapter 4. [50] The agent is an automated agent for the *life game* which performs agent-modeling using a cognitive agent model based on the Social Value Orientation (SVO) theory. In this subsection, we exemplify a way to modify that agent with the newly discovered SVO correlation results, by providing it the SVO data of other agent's designer.

Since the original adaptive agent does not have any prior knowledge about

the other agent, the agent does not know the social preference of the other agent. Therefore, the agent will start with some default models, and will estimate the orientation of other agent from the history of interactions. More precisely, the agent starts by assuming that the other agent is fully cooperative. After accumulating some interaction histories, the agent will learn the true social orientation of the other agent, and will adapt and use it to the best of its capacity (for example, if the other agent is cooperative, the agent will also be cooperative). To minimize exploitation, the estimated trustworthiness of the other agent is decreased whenever a defect-like action is observed. There are five types of trustworthiness: type 0 (fully non-cooperative), 1, 2, 3, and 4 (fully cooperative). Agents with higher trustworthiness will tend to cooperate on a larger subset of games.

Although it can prevent future exploitation by decreasing the estimated trustworthiness of the other agent whenever a defect-like action is observed, it cannot prevent the initial exploitation of non-cooperative agents. Avoiding initial exploitation is importance, especially when the expected number of iteration is small. We propose to use the SVO information of the designer of other agent to minimize the exploitation by selfish agent. If the designer of other agent has higher SVO angle, the higher initial estimated trustworthiness of the agent. More precisely, instead of initializing the estimated trustworthiness (λ_x) of other agent (x) to fully cooperative (4), we initialize it according to the agent designer's SVO (θ_x) using following formula:

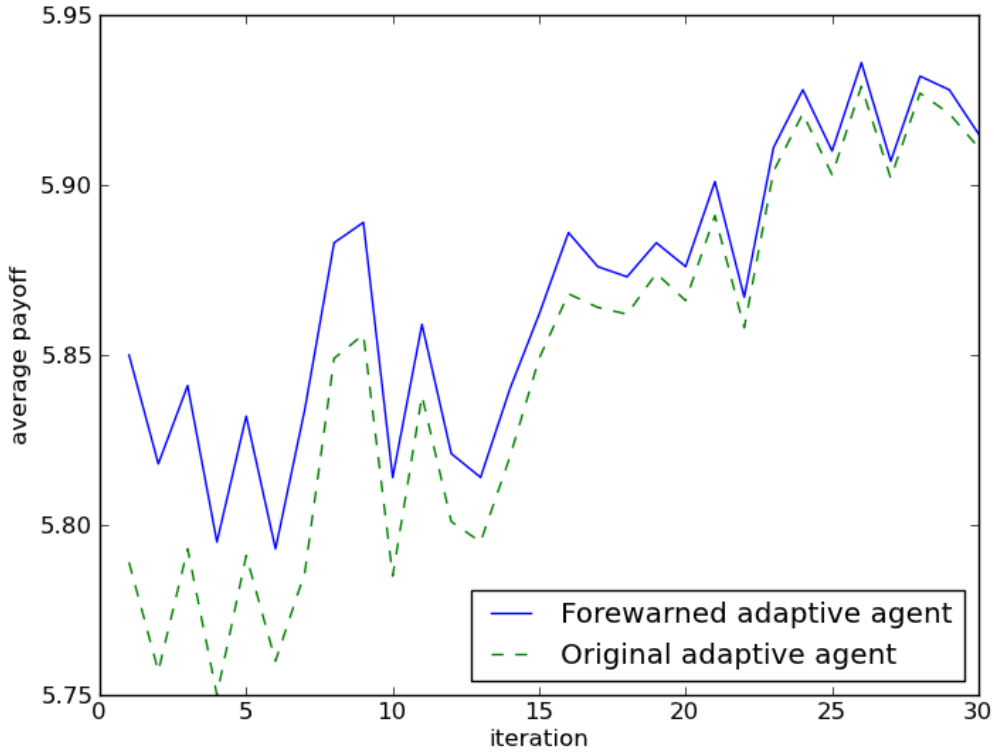

Figure 5.10: Performance of agents playing with all 28 students' agents.

$$\lambda_x = \max(\min(\left\lfloor \frac{\theta_x}{10} \right\rfloor, 4), 0)$$

In other words, we have a higher λ_x value for an agent x written by a designer having higher SVO angle θ_x.

To evaluate the performance improvement of the agent with the help of the human SVO data, we implemented a forewarned adaptive agent described above, and evaluated its performance in tournaments (10000 runs) with students agents. Figure 5.10 shows the average payoffs the agents at different iteration when they play with all 28 students' agents. Figure 5.11 and 5.12 shows the average payoffs

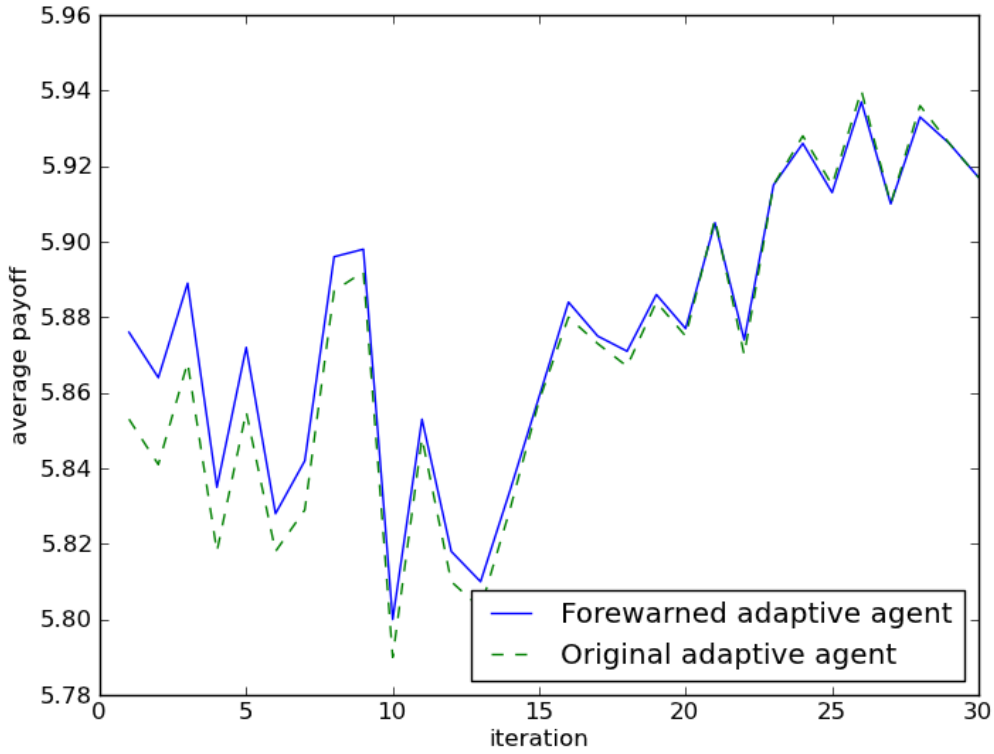

Figure 5.11: Performance of agents playing with 21 agents written by cooperative humans.

when they play with 21 agents written by cooperative human and 7 agents written by individualistic human.

The payoff of the original agent is very low at the beginning, because it begins by assuming the other agent is fully cooperative and so applying the "nice" strategy towards all other agents. If the other agent is non-cooperative, the original agent may be exploited for the first few games, and lose some payoffs at the beginning. On the other hand, the forewarned adaptive agent has higher payoff at the beginning, because it prevents some of the exploitation by having a more accurate initial model.

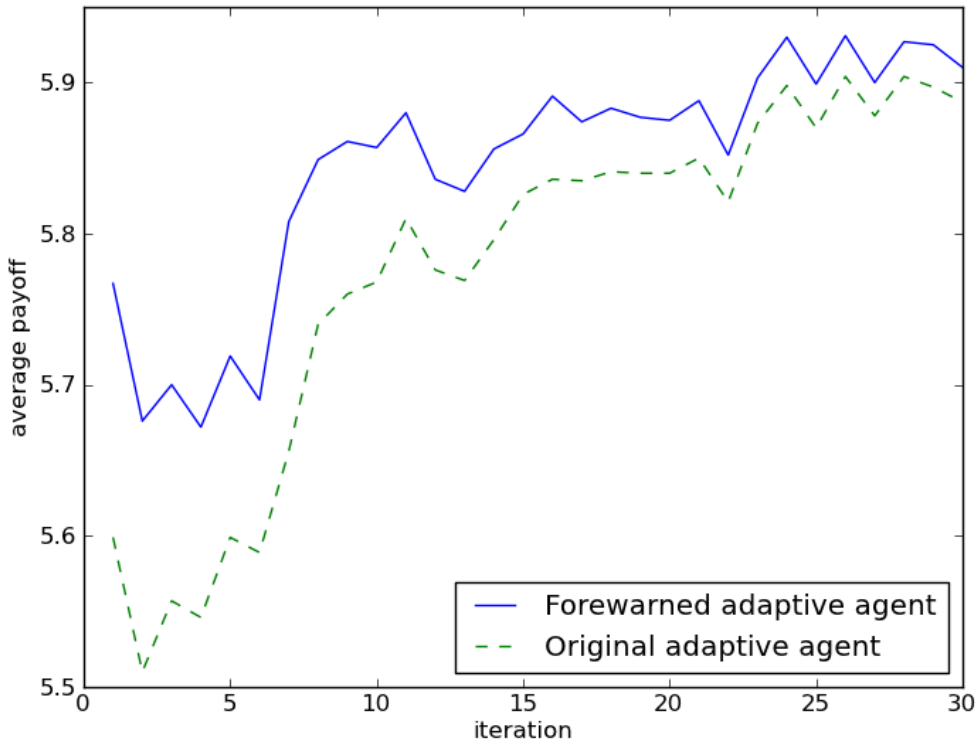

Figure 5.12: Performance of agents playing with 7 agents written by individualistic humans.

That is also the reason why the payoff difference is very large in Figure 5.12, but small in Figure 5.11. It also shows that the advantage of forewarned adaptive agent mainly comes from avoiding being exploited by agents written by individualistic human. The Human SVO data can help the agent to estimate the trustworthiness of other agents, rather than learning through interaction.

Comparing Figure 5.11 and 5.12, the average payoff obtained is higher when the adaptive agents are playing with students' agents written by cooperative humans, because those agents also tend to be more cooperative that give us benefit

of mutual cooperation. This shows that maintaining cooperation with cooperative agents is very important.

With an increasing number of iterations, both agents' performances improve and converge. It is probably because both agents are doing agent modeling. With more interaction data, the modeling will be more accurate, and so they can better predict other agents' action to get higher payoffs. For example, even though the original agent always starts with being nice, when it knows more about the other agents, it will stop cooperating with the defectors and keep cooperating with the cooperators.

In summary, there are at least three main factors for a good life game agent: (1) apply agent-modeling techniques during interaction; (2) start with a more accurate model; (3) maintain mutual cooperation with other agents if possible.

5.6 Summary

We have developed a way to measure the social preferences of computer agents, by adapting some concepts and techniques from social psychology. In our study of agents that were designed to play a repeated stochastic game (the life game), we have found a strong correlation between the agents' social preferences and the social preferences of their human designers. We have shown that this correlation can be used to make useful predictions of what choices an agent will make over the course of a game, and have shown that these predictions can be used to improve the performance of other agents that interact with the given agent.

Chapter 6: Predicting Agents' Behavior by Measuring their Behavioral Signature

The original SVO model was limited to one-shot games, and assumed that each individual's behavioral preferences remain constant over time—an assumption that is inadequate for repeated-game settings, where an agent's future behavior may depend not only on its SVO but also on its observations of the other agents' behavior. We extend the SVO model to take this into account. Our experimental evaluation, on several dozen agents that were written by students in classroom projects, show that our extended model works quite well.

6.1 Introduction

Many multi-agent domains involve human and computer decision makers that are engaged in repeated collaborative or competitive activities. Examples include online auctions, financial trading, and computer gaming. Repeated games are often viewed as an appropriate model for studying these kinds of repeated interactions between agents. Compared to one-shot games, repeated games are much more complex as they allow agent to adapt their behavior between the rounds. The relevant literature contains many demonstrations of how an agent's behavior can change as it develops

a better understanding of the other agents' behavior. [21, 51–54]

In order to model the behavior of an agent and predict its performance, we adapt and extend a construct, Social Value Orientation (SVO), from social psychology [4]. SVO theory assumes that in interpersonal interactions, an individual's choices depend not only on his/her own payoffs but also on his/her preferences for the *other* individual's payoff, and that these preferences remain stable over time. SVO theory provides a way to measure these preferences, and experimental validations of these measurements on human subjects.

If a human writes an agent to act as the human's delegate in a multi-agent environment, one might expect the computerized agent to have social preferences as well. Knowing an agent's social preference would make it possible to make informed guesses about the agent's future actions.

A critical limitation of the SVO model is that it only looks at agent's preferences in one-shot games. This is inadequate for repeated games, in which an agent's actions may depend on both its SVO and its model of the other agent's behavior. To use the SVO model effectively in repeated games, it is necessary to extend the SVO model to take into account how an agent's behavior will change if it interacts repeatedly with various other kinds of agents.

Our contributions in this chapter are as follows. First, we extend the SVO model by developing a *behavioral signature*, a model of how an agent's behavior over time will be affected by both its own SVO and the other agent's SVO. Second, we provide a way to measure an agent's behavioral signature, and methods for using behavioral signatures to predict agents' performance. Third, we present

experimental results using agents that students wrote to compete in repeated-game tournaments. The experimental results show that our predictions are highly correlated with the agents' actual performance in tournament settings. This shows that our proposed model is an effective way to generalize SVO to situations where agents interact repeatedly.

6.2 Modeling Computer Agents

In repeated games, an agent's social preference can be influenced not only by the agent's own SVO, but also by how the agent reacts to the other agent's SVO. For example, let x be an agent whose SVO is 45° (i.e., it prefers equal payoffs for both agents) and y be a memoryless agent whose SVO is 0° (i.e., y cares only about maximizing its own payoff in the current iteration). If x and y interact repeatedly, then after repeated observations of y's behavior, x might decide that the best way to equalize both agents' cumulative payoffs might be for x to try to maximize its own payoff at each iteration. Consequently, if we perform a Ring measurement of x after it has had many interactions with y, x's "apparent" SVO value may be closer to 0° than 45°. We called this x's *para-SVO* against y in Chapter 5. In this chapter, we will define a *behavioral signature* for x to be a vector $\sigma_n(x)$ that includes x's SVO and a collection of para-SVO values for x against several different "constant-SVO agents":

$$\Theta_n(x) = (\theta_0(x), \theta_n(x|C_{-90}), \theta_n(x|C_{-80}), \theta_n(x|C_{-70}), ..., \theta_n(x|C_{90})),$$

where $\theta_0(x)$ is x's SVO, and $\theta_n(x|C_\phi)$ is x's para-SVO at the n-th iteration when x plays with the agent C_ϕ defined below.

Each agent C_ϕ is a memoryless agent whose SVO is ϕ degrees. C_ϕ always tries to maximize the quantity $p_{\text{self}} \cos\phi + p_{\text{other}} \sin\phi$, where p_{self} is it's expected payoff and p_{other} is other agent's expected payoff if C_ϕ plays against an agent that chooses each action with equal probability.[1] For example, if $\phi = 0$ and the game matrix is the one shown in Figure 2.1, the Constant-SVO agent will choose A_1 if $a+b > c+d$, otherwise it will choose A_2.

The para-SVO, $\theta_n(x|C_\phi)$, of agent x at the n-th iteration with tester agent C_ϕ is measured by applying the modified Ring measurement on the agent at the $(n+1)$-th iteration after it interacted with the tester agent C_ϕ for n iterations. Figure 5.3 shows the complete procedure for measuring $\theta_n(x|C_\phi)$ using the modified game matrices G_{Ring} (tester agent $y = C_\phi$).

If we know the behavioral signatures of two agents x and y, we can estimate the cumulative payoff when x and y play with each other. We will study and evaluate two methods, $E_0(x, y)$ and $E_n(x, y)$, for estimating x's average payoff when it plays with y for N iterations (where $N > n$). Both methods use a E_C function to approximate the payoff. $E_C[\phi_1, \phi_2]$ is the payoff of Constant-SVO agent C_{ϕ_1} when it plays with another Constant-SVO agent C_{ϕ_2} for N total number of iterations. Note that $E_C[\phi_1, \phi_2]$ can be computed quickly because the Constant-SVO agents are very simple.

[1]The "equal probability" assumption is needed to calculate the expect payoff for each action. It can be shown that this assumption is compatible with the para-SVO measurement.

E_0 estimation:

$$E_0(x, y) = E_C[\theta_0(x), \theta_0(y)]$$

E_n estimation:

$$E_n(x, y) = E_C[\theta_n(x|C_\beta), \theta_n(y|C_\alpha)],$$

where $\alpha = \theta_0(x)$ rounded off to the nearest tens digit, and $\beta = \theta_0(y)$ rounded off to the nearest tens digit.

The first method uses (initial) SVO values of x and y as the input to E_C. The second method uses a more sophisticated input that involves the behavioral signatures of both agents. Note that $E_0(x, y)$ is a degenerated case of $E_n(x, y)$ when $n = 0$, because all elements in the behavioral signature equal to SVO of the agent when $n = 0$.

6.3 Experiments

We have evaluated our model experimentally on a large collection of agents that were written by students in several advanced-level AI and Game Theory classes. In each case, the students wrote their agents to compete in a round-robin tournament among all the agents in their class. To attain a richer set of agents, the classes were held at two different universities in two different countries: one in the USA, and one in Israel.

Our experimental studies involved measuring the agents' behavioral signatures, playing round-robin tournaments among the entire set of agents, and comparing the agents' performance with the predictions made by our model. To eliminate random

favorable payoff variations, we randomized the series of games, and used the same series between all agents in the population. The instructions stated that at each iteration, they will be given a symmetric game with a random payoff matrix of the form shown in Figure 2.1. 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 71 agents (47 from the USA and 24 from Israel).

6.3.1 Measuring Agents' Behavioral Signatures

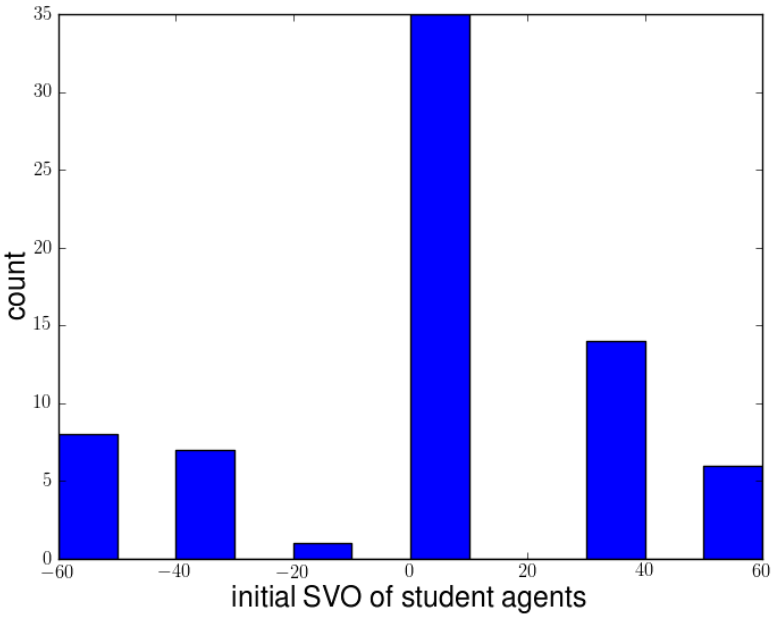

Figure 6.1: Distribution of students' agents' SVO.

We use the para-SVO measurement procedure (shown in Figure 5.3) to find

the behavioral signatures of all students' agents. Figure 6.1 shows the distribution of (initial) SVO of students' agents.[2] While most of them are individualistic (to different degrees), there were some who had competitive and cooperative orientations.

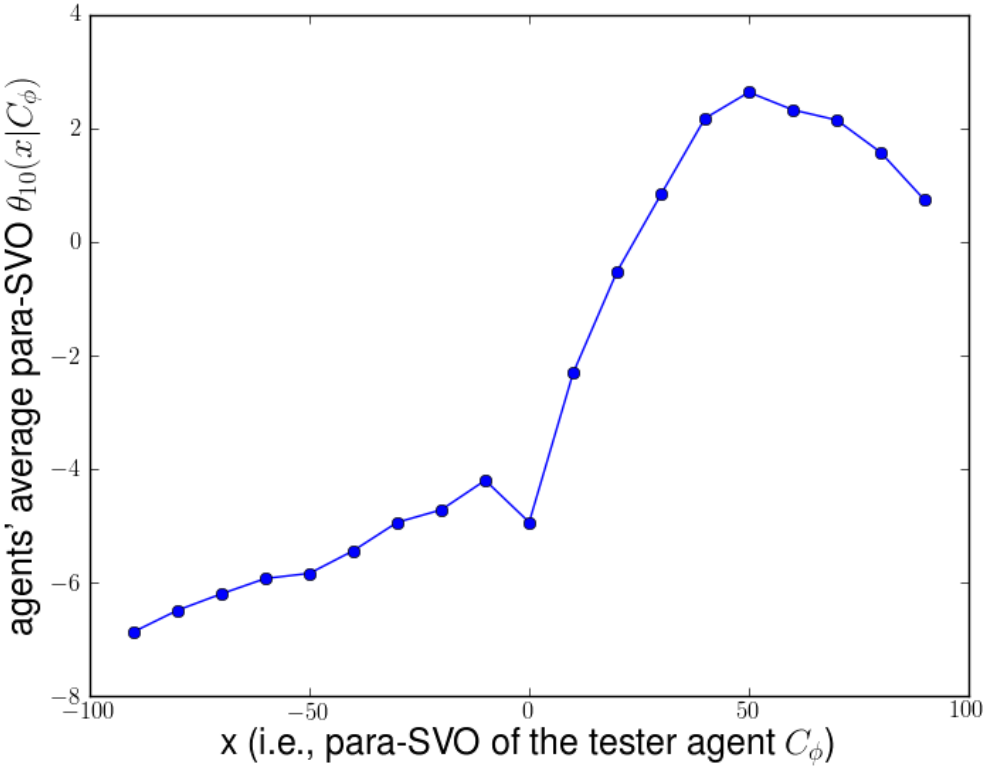

Figure 6.2: Average para-SVO values $\theta_{10}(x|C_\phi)$ for $\phi = -90°$ to $90°$, averaged over all x in the entire set of students' agents.

Figure 6.2 shows the average, over all of the students' agents, of the para-SVO value $\theta_{10}(x|C_\phi)$. Recall that $\theta_{10}(x|C_\phi)$ is agent x's para-SVO value at 10th iteration against a memoryless agent C_ϕ whose SVO is ϕ degrees. Notice that the average para-SVO of the students' agents increases with the para-SVO of the tester agents,

[2]SVO of x is measured by testing x with one-shot games, i.e., it is equals to the initial para-SVO of x.

because it is beneficial to be more cooperative if the other agent is more cooperative. The magnitude of change of the average is not large, because para-SVO values of about 45 (out of 71) agents remain constant across different tester agents.

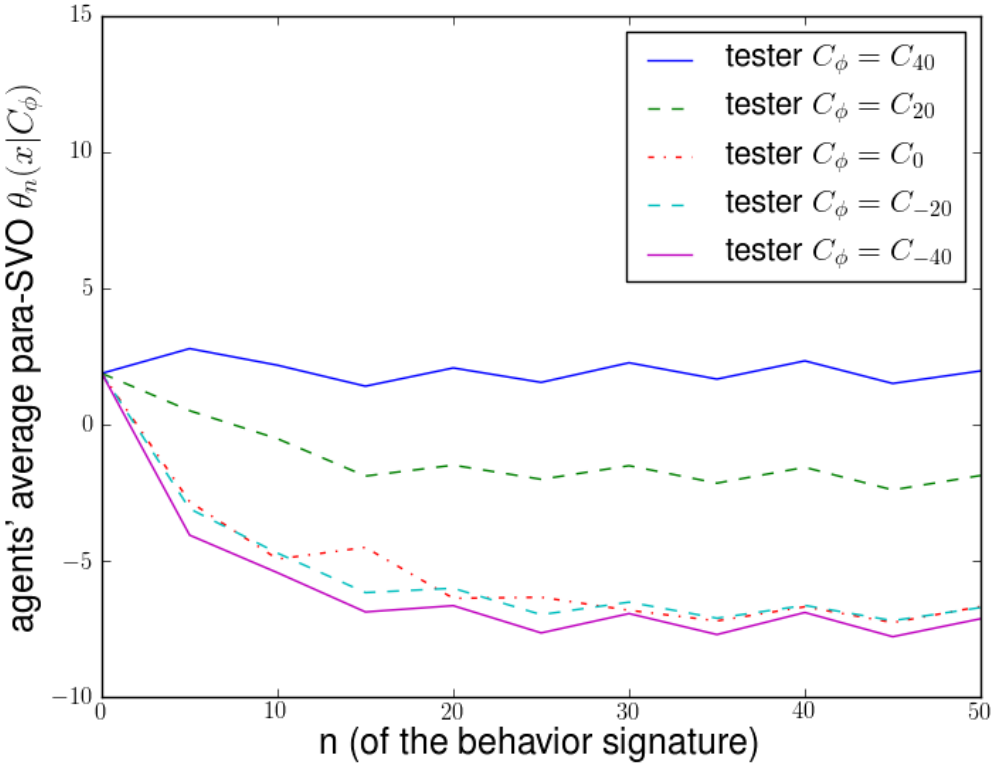

Figure 6.3: Average para-SVO values $\theta_n(x|C_\phi)$ with different tester agents C_ϕ, averaged over all x in the entire set of students' agents.

Figure 6.3 shows the average para-SVO of students' agents when the tester agents are Constant-SVO agents with SVO $= -40°, -20°, 0°, 20°$, or $40°$. Again, the results show that the average para-SVO of the students' agents increases with the para-SVO of the tester agents. Moreover, when n increases, most of the averages decrease, and all of them level off after about 20 iterations. From examining the

code, we found that many of the agents try to build a model of the other agent in the game, based on the history of interactions, and the model tend to be stabilized after some number of iterations. All of the above results show that the apparent social preferences of agents change with the behaviors of other agents, because the action of an agent is usually determined by both its SVO and its prediction of the opponent action.

6.3.2 Predicting Agents' Performances

Our next goal was to evaluate the accuracy of our prediction algorithms. In the following experiments, the total number of iterations (N) is 100, and the number of runs is also 100. We predicted the average payoff of all possible games of any two students' agents (including playing with itself, i.e., 71×71 data points for each run), using the method mentioned in Section 6.2.

Figure 6.4 and Figure 6.5 show the correlation and mean square error between predicted payoffs and actual payoffs. Regardless the value of n, the predicted payoffs have high correlation with the actual payoffs. Their mean square errors are low, comparing with the average payoff ≈ 5.5. When $n = 0$, the accuracy of E_n is good (mean square error = 0.284). As n increases, the accuracy of E_n also increases until $n = 20$, at which point it levels off (similar to Figure 6.3).

When $n = 0$, E_n degenerates to E_0 which only considers the (initial) SVO value of the agents. When $n > 0$, E_n takes the agents' adaptive behaviors into account by considering their behavioral signatures. The better performance of E_n shows that

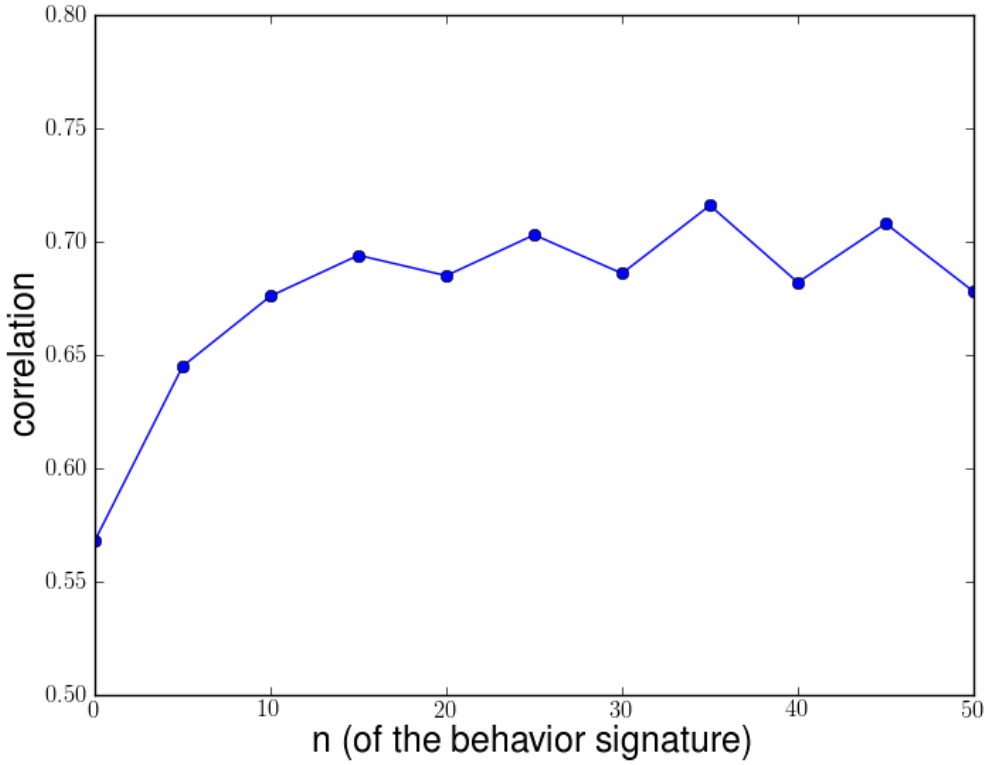

Figure 6.4: Correlation between predicted and actual payoffs (when student agents play in a tournament).

our extended SVO model works better in repeated games than the original SVO model.

6.4 Summary

We have extended the SVO model from social psychology, to provide a *behavioral signature* that models how an agent's behavior over multiple iterations will depend on both its own SVO and the SVO of the agent with which it interacts. We have provided a way to measure an agent's behavioral signature, and a way to use this

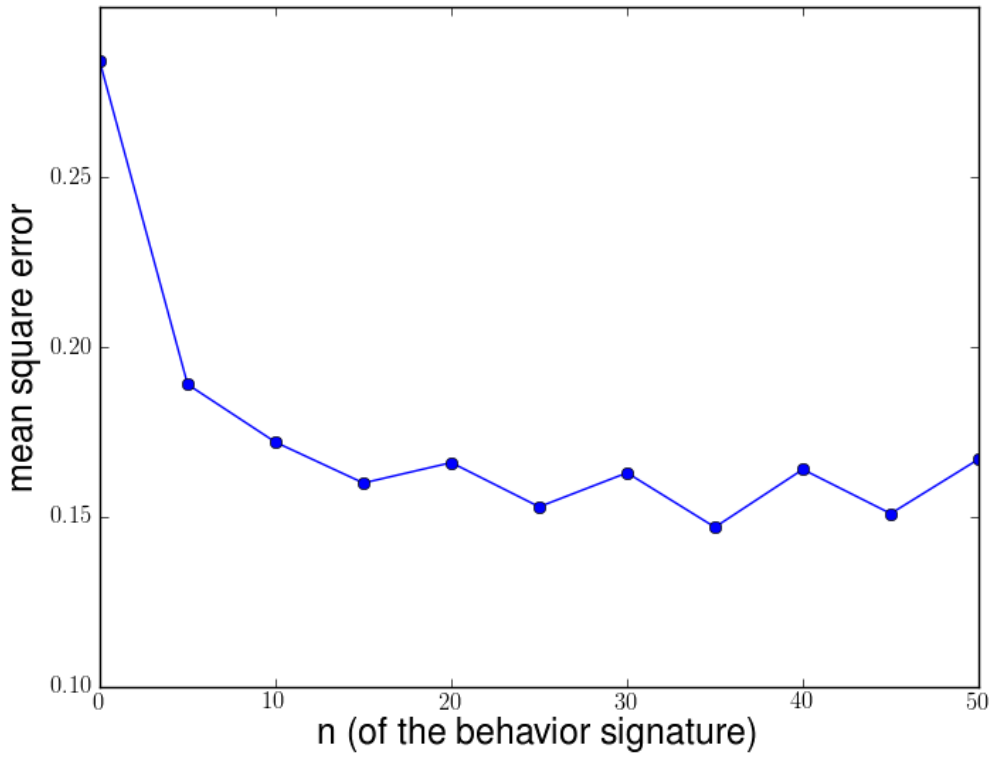

Figure 6.5: Mean square error of predicted payoffs (when student agents play in a tournament).

behavioral signature to predicting the agent's performance. In our study of agents that were designed to play a repeated stochastic game (the Life Game) in classroom tournaments, the predictions made by our model were highly correlated with the agents' actual performance.

Chapter 7: Conclusions

Human social preferences — i.e., human preferences for the outcomes of their interactions with others — have been shown to play an important role in many areas of decision-making. As agents are developed that exhibit more autonomy and take an increasing role in interacting with other human and agents, it is becoming important to understand the social preferences of agents as well as humans. This dissertation presents a step in this direction, by studying how the notion of SVO can be used to improve our understanding of to the behavior of computer agents. This chapter summarizes the contributions in this dissertation and proposes new directions for future work.

The four main contributions of this thesis are:

1. Chapter 3 described a formal model that combines game-theoretical analyses for cooperation in Iterated Prisoner's Dilemma with insights from social and behavioral sciences. Our model is not claimed to be the most accurate account of social orientations; rather, it is a simple model that takes the first step in the above direction. Unlike existing models, this formalism captures the notion of prosocial vs. proself orientations exhibited in human behavior and explicitly provides an abstract representation for how a player develops its strategies

in repeated games. We have presented theorems showing how players with different social tendencies interact. Our theorems identify five general steady-state behavioral patterns, that can be explained in terms of the players social orientation values. We have also performed an experimental evaluation of our model using evolutionary simulations in the well-known IPD game. The results of the experiments demonstrated that our model captures the well known behavior patterns in IPD. Furthermore, it allows modeling richer behavior patterns since it does not depend on the particular game matrix.

2. Chapter 4 described a successful cognitive strategy for the life game by utilizing SVO theory. 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 impacts that different trust adaptation strategies have on the performance of the SVO agent.

3. In Chapter 5, we have developed a way to measure the social preferences of computer agents, by adapting some concepts and techniques from social psychology. In our study of agents that were designed to play the life game, we have found a strong correlation between the agents' social preferences and the social preferences of their human designers. We have shown that this correlation can be used to make useful predictions of what choices an agent

will make over the course of a game, and have shown that these predictions can be used to improve the performance of other agents that interact with the given agent.

4. In Chapter 6, we have extended the SVO model, to provide a *behavioral signature* that models how an agent's behavior over multiple iterations will depend on both its own SVO and the SVO of the agent with which it interacts. We have provided a way to measure an agent's behavioral signature, and a way to use this behavioral signature to predicting the agent's performance. In our study of agents that were designed to play the life game in classroom tournaments, the predictions made by our model were highly correlated with the agents' actual performance.

A limitation to most of the above works is that they were restricted to the life game. However, we believe there is a strong potential for extending the results to other contexts. One topic of future work would be to generalize our model and analysis to other kinds of repeated games. Another limitation of our study is that the algorithm for measuring an agent's behavioral signature requires a collection of interaction trace between the agent and a specific group of agents in some special sequences of games. An interesting direction would be to estimate the behavioral signatures of agents using a set of interaction traces produced by different pairs of agents in arbitrary two-player repeated games. Such extensions may provide both an improved understanding of agent behavior, and ways to improve the effectiveness of agents in their interactions with others.

Bibliography

[1] Daniel Kahneman, Jack Knetsch, and Richard H. Thaler. Fairness and the assumptions of economics. *Journal of Business*, 59(4):S285–300, 1986.

[2] Hessel Oosterbeek, Randolph Sloof, and Gijs van de Kuilen. Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental Economics*, 7(2):171–188, 2004.

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

[4] 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, September 2008.

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

[6] P.A.M. Van Lange, E. De Bruin, W. Otten, and J.A. Joireman. Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. *Journal of personality and social psychology*, 73(4):733, 1997.

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

[8] Charles G. McClintock and Scott T. Allison. Social value orientation and helping behavior. *Journal of Applied Social Psychology*, 19(4):353 – 362, 1989.

[9] Joireman, Lasane, Bennett, Richards, and Solaimani. Integrating social value orientation and the consideration of future consequences within the extended norm activation model of proenvironmental behavior. *British Journal of Social Psychology*, 40:133–155, 2001.

[10] Craig D. Parks and Ann C. Rumble. Elements of reciprocity and social value orientation. *Personality and Social Psychology Bulletin*, 27(10):1301–1309, 2001.

[11] Steven de Jong, Karl Tuyls, and Katja Verbeeck. Artificial agents learning human fairness. In *AAMAS*, pages 863–870, 2008.

[12] Robert Axelrod. *The Evolution of Cooperation*. Basic Books, 1984.

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

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

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

[16] Shavit Talman, Meirav Hadad, Ya'akov Gal, and Sarit Kraus. Adapting to agents' personalities in negotiation. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, AAMAS '05, pages 383–389, New York, NY, USA, 2005. ACM.

[17] Ya'akov Gal, Barbara Grosz, Sarit Kraus, Avi Pfeffer, and Stuart Shieber. Agent decision-making in open mixed networks. *Artificial Intelligence*, 174(18):1460–1480, 2010.

[18] M.A. Nowak and K. Sigmund. Tit for tat in heterogeneous populations. *Nature*, 355(6357):250–253, 1992.

[19] Martin Nowak and Karl Sigmund. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature*, 364(6432):56–58, July 1993.

[20] Anders Eriksson and Kristian Lindgren. Evolution of strategies in repeated stochastic games with full information of the payoff matrix. In *GECCO*, pages 853—859, 2001.

[21] T.-C. Au and D. Nau. Is it accidental or intentional? a symbolic approach to the noisy iterated prisoners dilemma. *The iterated prisoners' dilemma: 20 years on*, 4:231, 2007.

[22] 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.

[23] Raz Lin, Sarit Kraus, Yinon Oshrat, and Ya'akov (Kobi) Gal. Facilitating the evaluation of automated negotiators using peer designed agents. In *AAAI*, 2010.

[24] Efrat Manisterski, Raz Lin, and Sarit Kraus. Understanding how people design trading agents over time. In *AAMAS (3)*, pages 1593–1596, 2008.

[25] T.-C. Au, Sarit Kraus, and Dana 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, Richland, SC, 2008.

[26] John Maynard Smith. *Evolution and the Theory of Games*. Cambridge university press, 1982.

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

[28] John Von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

[29] M.A. Nowak. Five rules for the evolution of cooperation. *Science*, 314(5805):1560, 2006.

[30] Robert Hoffmann. Twenty years on: The evolution of cooperation revisited. *J. Artificial Societies and Social Simulation*, 3(2), 2000.

[31] Jack Hirshleifer and Juan Carlos Martinez Coll. What strategies can support the evolutionary emergence of cooperation? *Journal of Conflict Resolution*, 32(2):367–398, June 1988.

[32] L.A. Imhof, D. Fudenberg, and M.A. Nowak. Evolutionary cycles of cooperation and defection. *Proceedings of the National Academy of Sciences*, 102(31):10797, 2005.

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

[34] Bruon Beafuils, Jean-Paul Delahaye, Philippe Mathieu, Christopher G Langton, and Taksunori Shimohara. *Our Meeting With Gradual: A Good Strategy For The Iterated Prisoners Dilemma*, pages 202–209. MIT Press, 1996.

[35] 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.

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

[37] Ernst Fehr, Georg Kirchsteiger, and Arno Riedl. Does fairness prevent market clearing? an experimental investigation. *The Quarterly Journal of Economics*, 108(2):437–459, 1993.

[38] Ernst Fehr and Klaus M Schmidt. A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, 114(3):817–868, 1999.

[39] Gary E Bolton and Axel Ockenfels. Erc: A theory of equity, reciprocity, and competition. *American economic review*, pages 166–193, 2000.

[40] Gary Charness and Matthew Rabin. Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, 117(3):817–869, 2002.

[41] W.B.G. Liebrand and C.G. McClintock. The ring measure of social values: A computerized procedure for assessing individual differences in information processing and social value orientation. *European journal of personality*, 2(3):217–230, 1988.

[42] Ryan Murphy, Kurt Ackermann, and Michel Handgraaf. Measuring social value orientation. *Available at SSRN 1804189*, 2011.

[43] Jennifer Golbeck, Cristina Robles, Michon Edmondson, and Karen Turner. Predicting personality from twitter. In *Proceedings of the 3rd IEEE International Conference on Social Computing*, pages 149–156, Boston, Massachusetts, USA, 2011.

[44] Jennifer Golbeck, Cristina Robles, and Karen Turner. Predicting personality with social media. In *Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, pages 253–262, New York, NY, USA, 2011. ACM.

[45] Thomas Dubois, Jennifer Golbeck, and Aravind Srinivasan. Predicting trust and distrust in social networks. In *Proceedings of the 3rd IEEE International Conference on Social Computing*, Boston, Massachusetts, 2011.

[46] R.O. Murphy and K.A. Ackermann. A review of measurement methods for social preferences. *Judgment and Decision Making*, 2012.

[47] R.O. Murphy, K.A. Ackermann, and M.J.J. Handgraaf. Measuring social value orientation. *Judgment and Decision Making*, 6(8):771–781, 2011.

[48] R.O. Murphy. Svo slider measure. `http://vlab.ethz.ch/svo/SVO_Slider/SVO_Slider.html`, 2013.

[49] Jack Cohen. *Statistical power analysis for the behavioral sciencies*. Routledge, 1988.

[50] Kan-Leung Cheng, Inon Zuckerman, Dana Nau, and Jennifer Golbeck. The life game: Cognitive strategies for repeated stochastic games. In *IEEE International Conference on Social Computing*. IEEE, 2011.

[51] D. Egnor. Iocaine powder explained. *ICGA Journal*, 23(1):33–35, 2000.

[52] Colin F Camerer, Teck-Hua Ho, and Juin-Kuan Chong. Sophisticated experience-weighted attraction learning and strategic teaching in repeated games. *Journal of Economic Theory*, 104(1):137–188, 2002.

[53] Robert L Slonim. Competing against experienced and inexperienced players. *Experimental Economics*, 8(1):55–75, 2005.

[54] Ananish Chaudhuri, Sara Graziano, and Pushkar Maitra. Social learning and norms in a public goods experiment with inter-generational advice. *The Review of Economic Studies*, 73(2):357–380, 2006.