GAME THEORETIC MODELING AND ANALYSIS: A CO-EVOLUTIONARY, AGENT-BASED APPROACH

QUEK HAN YANG
B.Eng (Hons., 1st Class), NUS

Summary

Game theoretic modeling and analysis is a challenging research topic that requires much attention from social scientists and researchers. The classical means of using analytical and empirical methods have presented difficulties such as mathematical intractability, limitations in the scope of study, static process of solution discovery and unrealistic assumptions. To achieve effective modeling that yields meaningful analysis and insights into game theoretic interaction, these difficulties have to be overcome together with the need to integrate realistic and dynamic elements into the learning process of individual entities during their interaction.

In view of the challenges, agent-based computational models present viable solution measures to complement existing methodologies by providing alternative insights and perspectives. To this note, co-evolutionary algorithms, by virtue of its inherent capability for solving optimization tasks via stochastic parallel searches in the absence of any explicit quality measurement of strategies makes it a suitable candidate for replicating realistic learning experiences and deriving solutions to complex game theoretic problems dynamically when conventional tools fail.

The prime motivation of this thesis is to provide a comprehensive treatment on co-evolutionary simulation modeling – simulating learning and adaptation in agent-based models by means of co-evolutionary algorithms, whose viability as a simple but complementary alternative to existing mathematical and experimental approaches is assessed in the study of repeated games. The interest in repeated interaction is due to its extensive applicability in real world situations and the added fact that cooperation is easier to sustain in a long-term relationship than a single encounter. Analysis of interaction in repeated games can provide us with interesting insights into how cooperation can be achieved and sustained.

This work is organized into two parts. The first part will attempt to verify the ability of co-evolutionary and/or hybridized approaches to discover strategies that are comparable, if not better, than solutions proposed by existing approaches. This involves developing a computer Texas Hold'em player via evolving Nash-optimal strategies that are comparable in performance to those derived by classical means. The Iterated Prisoner's Dilemma is also investigated where performance and adaptability of evolutionary, learning and memetic strategies is benchmarked against existing strategies to assess whether evolution, learning or a combination of both can entail strategies that adapt and thrive well in complex environments.

The second part of this work will concentrate on the use of co-evolutionary algorithms for modeling and simulation, from which we can analyze interesting emergent behavior and trends that will give us new insights into the complexity of collective interaction among diverse strategy types across temporal dimensions. A spatial multi-agent social network is developed to study the phenomenon of civil violence as behavior of autonomous agents is co-evolved over time. Modeling and analysis of a multi-player public goods provision game which focuses specifically on the scenario where agents interact and co-evolve under asymmetric information is also pursued. Simulated results from both contexts can be used to complement existing studies and to assess the validity of related social theories in theoretical and complex situations which often lie beyond their original scope of assumptions.

Lists of publications

The following is the list of publications that were published during the course of research that I conducted for this thesis.

Journals

1. H. Y. Quek, C. H. Woo, K. C. Tan, and A. Tay, 'Evolving nash-optimal poker strategies using evolutionary computation', *Frontiers of Computer Science in China*, vol. 3, no. 1, pp. 73-91, March 2009.

2. H. Y. Quek, K. C. Tan, C. K. Goh, and H. A. Abbass, 'Evolution and incremental learning in the Iterated Prisoner's Dilemma', *IEEE Transactions on Evolutionary Computation*, vol. 13, no. 2, pp. 303-320, April 2009.

3. H. Y. Quek, K. C. Tan, and H. A. Abbass, 'Evolutionary game theoretic approach for modeling civil violence', *IEEE Transactions on Evolutionary Computation*, vol. 13, no. 4, pp. 780-800, August 2009.

4. H. Y. Quek, K. C. Tan, and A. Tay, 'Public goods provision: An evolutionary game theoretic study under asymmetric information', *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 1, no. 2, pp. 105-120, June 2009.

Conferences

1. C. K. Goh, H. Y. Quek, E. J. Teoh, and K. C. Tan, "Evolution and incremental learning in the iterative prisoner's dilemma," in *Proceedings of the IEEE Congress on Evolutionary Computation*, Edinburgh, UK, September 2-5, vol. 3, 2005, pp. 2629-2636.

2. C. K. Goh, H. Y. Quek, K. C. Tan and H. A. Abbass, "Modeling civil violence: an evolutionary, multi-Agent, game-theoretic approach," in *Proceedings of the IEEE Congress on Evolutionary Computation*," Vancouver, Canada, July 16-21, 2006, pp. 1624 - 1631.

3. H. Y. Quek, and C. K. Goh, "Adaptation of Iterated Prisoner's Dilemma strategies by evolution and learning," in *Proceedings of the IEEE Symposium Series on Computational Intelligence*, Computational Intelligence and Games, Honolulu, Hawaii, USA, April 1-5, 2007, pp. 40-47.

4. C. S. Ong, H. Y. Quek, K. C. Tan, and A. Tay, "Discovering Chinese Chess strategies through co-evolutionary approaches," in *Proceedings of the IEEE Symposium Series on Computational Intelligence*, Computational Intelligence and Games, Honolulu, Hawaii, USA, April 1-5, 2007, pp. 360-367.

5. H. Y. Quek, and A. Tay, "An evolutionary, game theoretic approach to the modeling, simulation and analysis of public goods provisioning under asymmetric information," in *Proceedings of the IEEE Congress on Evolutionary Computation*, Singapore, September 25-28, 2007, pp. 4735-4742.

6. H. Y. Quek, and K. C. Tan, "A discrete particle swarm optimization approach for the global airline crew scheduling problem," in *Proceedings of the International Conference on Soft Computing and Intelligent Systems and International Symposium on Advanced Intelligent Systems*, Nagoya University, Nagoya, Japan, September 17-21, 2008.

Book Chapters

1. H. Y. Quek, H. H. Chan, and K. C. Tan, "Evolving computer Chinese Chess using guided learning," in *Biologically-Inspired Optimisation Methods: Parallel Algorithms, Systems and Applications*, Studies in Computational Intelligence, Vol. 210, A. Lewis, S. Mostaghim, and M. Randall, Eds. Berlin / Heidelberg, Springer, 2009, pp. 325-354.

Acknowledgements

The course of completing my doctoral dissertation has been a fulfilling journey of intellectual curiosity, personal accomplishment and purposeful reflections. It has taught me much about the multi-faceted geometry of life - one that encompasses much uncertainty, asymmetry, intricate inter-dependencies and new perspectives of understanding and making sense of our existence. To this end, I would like to convey my heartfelt thanks to many people who have made this journey possible.

First and foremost, I would like to thank my thesis supervisor, Assoc. Prof Tan Kay Chen for giving me the opportunity to pursue this multi-disciplinary area of research. His guidance, understanding and kind words of encouragement and advice have always served as a strong motivational force which kept me on track throughout my candidature. I would also like to thank my co-supervisor Assoc. Prof. Arthur Tay for his relentless support and belief in me; Prof. H. A. Abbass for providing much assistance and suggestions that helped improve my research work, Assoc. Prof. Vivian Ng for nurturing me under the ECE outreach program, also to Ms Chua for all the fruitful discussions about human relations and everyone else who had kindly contributed ideas towards the completion of this thesis.

I am grateful to a bunch of happy folks in the Control and Simulation Lab for making my four years' stay fun and enjoyable: Chi Keong aka Zhang Lao for all his timely advice, Dasheng for sharing his research experiences, Eu Jin for his profound discussions, Brian and Chun Yew for their fair share of jokes, Chiam for playing big brother, Chin Hiong for his great tips; Chen Jia and Vui Ann for their jovial presence which spice up the entire lab atmosphere; not forgetting Sara and Hengwei for giving their utmost technical and logistical support from time to time.

I would also like to extend my gratitude to members of the outreach team: Li Hong, Teck Wee, Swee Chiang, Mo Chao, Yen Kheng, Siew Hong, Kai Tat, Yit Sung, Marsita and Elyn, for making my stay a fun, educational and enriching one; to my personal friends for their encouragement through my ups and downs; to my travel buddies for the wonderful backpacking experiences together, and to all my volunteering compatriots for accompanying me on the beautiful journey of giving and sharing the joy that goes beyond spoken words.

Last but not least, I wish to express my sincere appreciation to my family – brothers, sisters, nephews and nieces for their love and support which have always been a constant source of strength for me; but most importantly my parents for making so much sacrifice to raise me up painstakingly, educating me, showering me with unconditional love and always tolerating my random eccentricities and irrationality with enduring patience and care. To them, I dedicate this thesis…

"The best and most beautiful things in the world cannot be seen or even touched but must be felt within the heart."
                                                                                                    ~ Helen Keller

"If it's true that we are here to help others, then what exactly are the others here for?"
                                                                                                    ~ George Carlin

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"In terms of the game theory, we might say the universe is so constituted as to maximize play. The best games are not those in which all goes smoothly and steadily toward a certain conclusion, but those in which the outcome is always in doubt. Similarly, the geometry of life is designed to keep us at the point of maximum tension between certainty and uncertainty, order and chaos…"

~ George B. Leonard

"No man is an island"   *~ John Donne, Meditation XVII*

Game theory [1] is the study of strategic behavior and interaction among two or more decision making entities - typically referred to as *players*, in interdependent situations where the outcomes of interaction are not determined unilaterally by any one player but collectively by the combination of choices of all players. In such contexts, all players involved in the strategic interaction – coined a *game*, decide their course of action based on a set of rules e.g. *strategy* and are generally concerned only with the maximization of their own individual well-bring or *payoff*. However, as each is fully aware that his actions can and will affect one another's success, and literally takes this fact into account during the process of decision making, it becomes complex but interesting at the same time to analyze how players would prefer to act in different scenarios, and the corresponding nature of outcomes which arises eventually amidst the interaction.

By virtue of its nature, game theory - a branch of applied mathematical discipline that spans socio-economic origins; constitutes a powerful framework to which we can study multi-person decision problems [2] in many real life contexts. Its assemblage of associated ideas and theorems provides a rational basis to model

1

and replicate complex, inter-weaving relationships which subsist very much in the day-to-day interaction between social entities. More often than not, game theoretic analysis can shed light and provide us with a potential channel to gain fruitful insights into the behavioral complexities and interconnections which characterize real world interaction at numerous levels of contact – between genes, animals, individuals, groups, firms, stakeholders or even nation states. Such understanding will be of concern and importance to social scientists, policy makers, economists, biologists, psychologists and cognitive researchers, perhaps even laymen as well.

1.1 Essential elements of game theory

In game theory, there are several essential elements that are common ingredients to all situations of strategic interaction. These include basic terminologies like *player*, *strategy*, *payoff*, *game*, as well as the important concepts of *dominance* and *Nash Equilibrium* (NE) [3]. Defined below, these fundamental aspects are crucial and constitute the crux of game theoretic modeling and analysis.

Definitions of core terminologies and concepts

Player:

A single, indivisible, decision making entity that is participating in the strategic interaction, has a nontrivial set of strategies (more than one) and selects among possible strategies based on payoffs.

Strategy:

A complete plan that defines the moves or actions which a player should execute for every possible scenario of interaction in a given game, regardless of whether a

scenario does arise. For example, a strategy for checkers would define a player's move at every possible position which is attainable during the course of the game. The set of all strategies that is available to a player is called its strategy space. In the game theoretic context, a player is typically driven to find an optimal strategy in the huge space of possible strategies in order to maximize its well-being in the associated environment of interaction.

Payoff*:*

A numerical figure that quantifies the utility or level of satisfaction e.g. profit, welfare etc, which a player derives from the outcome of a strategic interaction. It reflects the motivations and represents the usual means of measuring success for a player's strategy within the game. In most games, the payoff to any player in every situation is expressed in the form of a payoff matrix or function that maps an input strategy profile (specification of strategies for every player) to an output payoff profile (denoting payoff values for every player).

Game*:*

A strategic interaction among mutually aware players (usually rational and seeks payoff maximization), where the decision of one impacts the payoffs of others and vice versa. A game can be completely specified and described by its players e.g. their types (which include the information known and used by each player for the basis of decision making, and how each player values the possible outcomes or utilities that result from making choices in strategic interaction), each player's strategies, resulting payoffs awarded for each outcome (denoting a particular combination of choices made by all players) and the order in which players make their moves (in the case of sequential game).

Dominance*:*

The concept establishes the relationship between strategies such that one is better than another for a player regardless of the profile of actions which other players may choose to play. In this context, a strategy is dominant if it is always better than other strategies e.g. earns a larger payoff. Similarly, a strategy is dominated if it is always better to play some other strategy e.g. earns a smaller payoff.

Nash Equilibrium (NE)*:*

A set of strategies, one for each player, such that no player has the incentive to unilaterally change his action. This occurs when a change in strategy by any one player would lead to a lower corresponding payoff for that player, given that all others do not change the strategies that they have currently adopted for use. The concept is typically used as an avenue to analyze and possibly predict the outcome of strategic interaction among several decision makers but does not necessarily imply a situation with best cumulative payoff for all the players involved.

1.2   Types of games

Games generally capture intrinsic aspects of complex, real world problems while being simple enough to enable extensive in-depth analysis. They can be broadly classified into a variety of basic types, depending on differences in the inherent nature of information structure, mode of game play and the interaction outcome. Some common distinctions in each category are listed and described as follow.

1.2.1 Information structure

- ***Perfect versus Imperfect***

A game is said to have *perfect information* if all players know all the moves that have taken place thus far. Examples include Chess, Tic tat toe, Go etc. In contrast,

a game of *imperfect information* is one in which some information of the game is not revealed to all players e.g. in card games like Poker, Blackjack etc, where each player's cards are hidden from other players.

- ***Complete versus Incomplete***

Complete information is used to describe a game in which players have access to knowledge e.g. payoffs and available strategies, of all players; while *incomplete information* denotes otherwise. Though similar, *complete* and *perfect information* are not identical. The prior refers to a state of knowledge about the game structure and objective functions of players, while not necessarily implying knowledge of actions in the game e.g. one may have complete information in the context of the Prisoner's Dilemma (PD) [4], but yet still subjected to the bounds of imperfect information, since one does not know the action of the other player.

- ***Symmetric versus Asymmetric***

Though not widely considered, there is a crucial need to define this category of distinction between games. *Symmetric information* games refer to those in which players subscribe to the same type of information and subjected to identical set of available strategies for the basis of decision making. In contrast, players subscribe to different types of information and strategy sets for the a*symmetric* case. The latter can arise due to differences in beliefs (which cause fundamental differences in the inherent strategy structures) or the degree of accessibility to information (some players might have access to more or different information as compared to others) for different players. A popular example pertains to the market for lemons [5] where information asymmetry exists between buyers and sellers.

1.2.2 Mode of game play

- *Simultaneous versus sequential*

Simultaneous games are those where players execute their moves concurrently, or if they do not, the players who move later are unaware of the actions that are made by players who move earlier. On the opposite note, *sequential* games are those where some players will choose their actions before others and players who move later can use knowledge about earlier actions as a basis to make their decisions.

- *One-shot versus repeated*

One shot games are those in which players only participate in one single round of interaction with each other. For games played in the *repeated* manner, players interact over a series of rounds which can be either *finitely* or *infinitely* repeating, depending on the time horizon of consideration. Unlike *one-shot* games, *repeated* games capture the idea that a player will have to take into account the impact of his current actions on the future actions of other players.

- *Two player versus multi-player*

Games where interaction always takes place in a pair-wise manner between any two entities are called *two-player* games. *Multi-player* games are those in which the mode of interaction is between *N players* where *N > 2*. In some sense, *two player* games can be considered a special case of *multi-player* games where *N = 2*.

1.2.3 Interaction outcome

- *Zero sum versus non-zero sum*

In *zero sum* games, total benefit to all players for any combination of strategies always adds up to zero. This is equivalent to implying that available resources can

neither increase nor decrease such that one can benefit only at an equal expense of others. Poker, Chess and Go exemplify such games because one wins exactly the amount the opponents lose. In *non-zero sum* games, however, some outcomes can have net results that are greater or less than zero. As such, one's gain does not necessarily correspond to a loss of another. Examples of such nature include the IPD, Battle of the Sexes etc.

- *Cooperative versus non-cooperative*

A game is *cooperative* if players are able to make enforceable contracts and form binding commitments through the presence of an external party e.g. legal system. In *non-cooperative* games, players are unable to enforce contracts beyond those specifically modeled in the game and the act of cooperation must be self-enforcing. Epitomizing the nature of many real world problems, *non-cooperative* games are generally concerned with situations with some conflict of interests among players in the game but for which there is no natural incentives for anyone to cooperate. As such, using relevant concepts in *non-cooperative* game theory to analyze the decisions which players make and the collective outcomes of their interaction can help enhance the understanding and resolution of conflicts and rivalry.

- *Transitive versus Intransitive*

A *transitive* game is one in which the relations between *A* and *B; B* and *C* directly implies the relation between *A* and *C* e.g. *(A > B) and (B > C)* → *(A > C)*. In the context of game theory, *A, B,* and *C* denote three distinct strategies employed in the course of game play and the inherent relation for any strategy pair denotes the order of dominance between the relevant component strategies. For *intransitive* games, however, the above relations are not always preserved.

1.3   Scope of analysis

Depending on the area of interest and concern, game theoretic interaction can be analyzed from a number of different perspectives such as strategy, outcomes of interaction, mechanism of game play, as presented in the following subsections. Apart from seeing and evaluating each viewpoint separately, varied perspectives can complement one another to give us a holistic picture into the richness of complex interaction among multiple intelligent entities, which is otherwise quite difficult to observe and make sense of in the actual real world context.

1.3.1 Strategy

From the strategy perspective, analysis looks at game theoretic interaction through the lens of an individual player. It is concerned with action plans that lead to the maximization of one's expected payoff, which is closely tied to the approach of maximizing the expected value of numerical utility function for an individual in *decision theory* [6]. The only difference, as opposed to decision theory, is that the analysis is essentially framed in the context of a multi-person decision theory – one which is concerned with the study of rational utility maximization behavior of each entity given that others are maximizing their utilities concurrently as well.

Using this perspective of study, we can verify the existence of optimum strategies and in turn decipher their inherent nature if they do exist. As far as the individual is concerned, the dominance relationship between different strategies can also be examined to give us a better understanding of the traits that constitute a good strategy. This can then provide an explanation as to why good strategies have an edge over inferior ones, which allows us to draw possible insights into how rational, self-interested players will tend to behave and act under different

circumstances. Such information is pertinent and can certainly serve as a useful guide for decision making in the likely event that the notion of optimal strategies might not even exist in numerous complex situations.

1.3.2 Outcomes of interaction

As opposed to the micro perspective of analyzing strategies which are adopted by individuals, the second perspective takes a macro view at the outcomes of game theoretic interaction. Instead of seeing things from the position of a single player in the game and concerning ourselves with one's payoff maximization behavior, the nature of collective outcomes and overall payoffs from scenarios of interaction that involve a relatively large number of individuals e.g. stock markets, auctions, public goods provision etc, are of primarily interest here.

By virtue of the complex interconnectedness that exists between players' actions and collective outcomes of interaction in the game theoretic context, it is insufficient for us to understand the entire picture of strategic interaction by analyzing solely from the individualistic strategy perspective. In numerous contexts, the mapping which couples actions and outcomes is always never straightforward - the maximization of individual payoffs using individualistically optimal strategies is typically not equivalent and does not necessarily translate to the maximization of group/overall payoffs. As such, the wider perspective of examining interaction outcomes can actually complement analysis from the prior perspective and help us, in particular policy makers, to gain a fuller and better understanding of the consequences of interaction. In the process, we also seek to identify and study interesting emergent behavior and trends amidst the collective interaction of different player strategies over time.

1.3.3 Mechanism of game play

The third perspective of analysis involves the design of the underlying rules and mechanism of game play so as to achieve the desired objectives for game theoretic interaction. Instead of adhering to just a fixed set of rules, mechanism design [7] differs from the two prior modes of analysis in that it asks about the consequences of different types of rules. It is not concerned merely about the collective outcome of interaction for a particular scenario but those arising from different mechanisms of game play. It questions generic factors which affect the outcomes and analyzes how the consequences of interaction can be improved if they are undesirable – depending on the objectives that policy makers have in mind, an outcome, though in NE, might not necessarily be deemed desirable to achieve in nature. Examples of mechanism design can encompass compensation and wage agreements which effectively spread risk for the firm while maintaining incentives for the employees, optimal auctions that maximize revenue and allocate resources efficiently etc.

1.4   Development and applications of game theory

Contrary to its theoretical foundations as a mere tool for economic analysis, the theory of games has seen extensive development since its fundamental and formal conception by Von Neumann and Morgenstern [8]. Distinguished Nobel laureates in game theory have since been honored for their contributions in pioneering the analysis of equilibria in non-cooperative games [3], [9], devising the economic theory of incentives under asymmetric information [5], [10], [11], enhancing our understanding of conflict and cooperation through game theoretic analysis [12], [13] and laying the foundations of mechanism design theory [14] - [16].

In line with the advances in theoretical concepts, the applications of game theory have spanned cross-disciplinary boundaries. This budding trend derives a vital need for researchers to negotiate multiple fields of expertise. Social scientists and computer scientists, for instance, have successfully applied relevant concepts to study the possibility of attaining and sustaining cooperation in both the classical and extended variants of the IPD [17] – [23]. Military strategists have turned to game theory to study conflicts of interest that are resolved through "war games" [24], [25] while sociologists have taken an interest in the development of an entire branch dedicated to examine issues involving group decision making [26] – [27]. Epidemiologists also use game theory for analyzing immunization procedures and methods of testing a vaccine or other medication [28]. Economists and policy makers are generally concerned with the study of economic problems relating to public goods provisioning [29], efficient auctions for resource allocation [30], bargaining and negotiation [31], [32] etc. Game theoretic principles are likewise applied to analyze the outcomes of competition between firms and corporations [33] in business and the modeling of stock market [34] for financial institutions etc. In politics, outcomes of elections are closely studied by political scientists via the concept of voting [35]. Mathematicians and game theorists have also analyzed and devise good strategies for games like poker, chess and checkers.

Other than the classical form of game theory, analysis using variants of the theory has also provided useful insights. Biologists have used evolutionary game theory (EGT) [36] to explain numerous seemingly incongruous phenomena in nature e.g. altruism and kin selection [37], [38]. Behavioral game theory [39] – [41] is also linked to phenomenal works by psychologists and cognitive scientists that give us a better understanding of the complex human being.

1.5   Modeling and analysis

To be able to perform insightful analysis in game theory, the ability to construct feasible models which capture essential and realistic aspects of interaction among all players participating in the game constitutes an important prerequisite. As a means of determining a solution to the decision problem that each player faces e.g. deriving the optimal strategies which dictate how players should act in order to maximize their individual payoffs, models should allow researchers to incorporate sophisticated micro-models of reasoning and preference for individual players and flexibly replicate strategic interaction without a need to abstract away such details. Since the popularization of game theory, analytical, empirical and computational approaches have constituted the primary methodologies of performing modeling and analysis. These will be discussed in the following sub sections.

1.5.1 Analytical approaches

Traditionally, the modeling and analysis of game theoretic problems has always been done using analytical approaches, where rigorous theoretical proofs are used to obtain precise prediction for the existence and nature of dominant strategies and NE points – situations where every player chooses actions that are best responses to the best responses of all others. The heart of such approaches is based around the theory of n-player non-zero sum games - in which John Nash formulated and proved the existence of at least one equilibrium solution for every generic game that involves N preference-maximizing players. This important research finding provides a powerful theoretical framework to optimize an individual's strategy e.g. choosing a best response in an interaction of such nature, and predicting a likely combination of joint actions as the eventual outcome - NE.

Refinements have since been made along the way, leading to Harsanyi's concept of a Bayesian-Nash Equilibrium (BNE) [42] and Maynard Smith's theory of evolutionary games [36]. The prior deals with situations where payoffs in the game are dependent on some private unobservable properties of a player e.g. the cards which a player holds in a game of Poker. The latter generally overlays a dynamic model of gradual strategy-adjustment on top of the static equilibria of Nash's original formulation. Evolutionary dynamics and existence of evolutionary stable strategies (ESS) can then be studied using replicator equations [36].

Despite the desirability of such techniques, using mathematical treatments to model complex problems typically involves a need to impose multiple core assumptions and constraints such as homogeneous player types, use of symmetric information for basis of decision making, common strategy framework, perfect rationality etc for tractability reasons. The result is an inevitable scale down of the actual problems to their much simplified versions, of which, the intrinsic realism of the problems will be largely compromised for solvability. In essence, we will no longer be addressing the original problems which we ought to be solving. The large mismatch between what we meant to solve and what we are actually solving generally renders any analysis of results from rigorous mathematical derivations senseless with regards to their applicability to the associated real world context.

Moreover, the idealized context which we derive the optimum or dominant strategy solutions from theoretical proofs also casts a doubt with regards to the degree of reproducibility for such strategy usage in practical settings. For example, Goeree and Holt [43] give an overview of ten simple games where game theoretic solutions are easily obtainable but intuitively implausible. This is due to the likely fact that players tend to be boundedly rational with finite computation power and

limited knowledge of their environment of interaction. Given these imperfections of reality, it is unlikely that the solutions derived from analytical approaches will apply with absolute certainty even if they are rigorously proven to be theoretically sound. This is due to the fact that players do not necessarily adjust their behavior to the theoretical optimum strategy in the midst of their interaction.

1.5.2 Empirical approaches

To create models that mirror real world interaction to a more realistic degree, the corresponding methodologies for modeling and analysis should seek to preserve the characteristic features of the original problem as far as possible. This naturally leads us to think about and re-examine the usage of empirical approaches, where experimentation is conducted on actual human subjects. Such methodologies are widely employed by economists, psychologists and social scientists alike, to study behavioral interaction in game theoretic settings.

As opposed to the analytical approaches which are theoretically grounded, experimental observations to testable hypotheses are primarily used to guide the research study in empirical approaches. One obvious advantage of such means is the fact that a large supply of players - human subjects, is available off the shelf for experimentation. Ideal as it may seem, experiments are typically designed to be performed under laboratory controlled condition for ease of isolating the salient factors that will help contribute to the verification of pre-defined hypotheses. As such, information gained in the process is again limited in the scope of study and might not necessarily reflect the actual situation where interaction is meant to take place e.g. in an auction house with information flowing freely among numerous bidders. The study is incomplete in some sense as it is not always straightforward [44] to analyze the necessary cognitive mechanisms which are utilized during the

course of interaction. Moreover, coupling effects among different factors might be impossible to study if the highly constrained laboratory scenario setup involves a deliberate exclusion of any related factor in the experiment design.

1.5.3 Computational approaches

With the possibility of addressing the challenges encountered by prior approaches, computational approaches present yet another viable alternative to perform game theoretic modeling and analysis. This is usually realized via the use of simulation in agent-based computational models (ACMs) [45], which Axelrod [46] regards as a third way of doing science in addition to deduction and induction techniques. Following the tremendous increase in computing power and processing speed of computers in recent decades, the utilization of computation as a feasible problem solving paradigm is becoming more popular and increasingly relevant in today's context. Nonetheless, it is to be noted that computational methodologies are never conceived to replace the existing approaches but rather to complement them by offering alternative insights into the nature of game theoretic interaction through new perspectives of modeling and analysis.

The ACM methodology is similar, and in essence a subset of the empirical approaches as mentioned earlier, with the exception that human subjects are now replaced by computer agents [47] – intelligent software entities which are flexibly designed with the ability to perceive, evaluate and make independent decisions on the basis of current information and past experiences, and to act in accordance to their self-interests and preference-maximization behavior to satisfy internal goals. Equipped with limited knowledge and bounded rationality, the agents embrace learning and adaptation to their environment similarly to the way which humans locally cope with a changing world through scenarios of interaction. In this sense,

ACMs are particularly suitable to model and study systems which are composed of multiple interacting entities and exhibit emergent properties [48], [49] - those arising from interaction of different entities which cannot be deduced simply by aggregating the properties of each. By designing multi-agent systems (MAS) [50] and conducting controlled computational experiments where multiple autonomous agents interact simultaneously in setups that closely resemble the relevant contexts of study; observations and analysis on interaction outcomes will be able to provide us with increased understanding and useful insights into the problem of interest.

Similar to human-based empirical experiments where the test subjects are readily available in abundance, number of entities in ACMs can also be scaled up to investigate outcomes of interaction with large numbers. The added advantage is that the numbers, as a form of model parameter, can be flexibly adjusted with ease through a change of simulation settings. The scope of study for ACMs is also less restrictive since we are free to design computational experiments to considerable degree of complexity as we deem fit - something which is of great difficulty to replicate in the much constrained laboratory settings of human-based experiments.

As opposed to analytical approaches that usually entail simple closed form solutions, ACMs are also not bounded by issues of mathematical intractability, allowing complex scenarios with more realistic features to be studied. Moreover, given the fact that interaction of real world entities is generally contingent on past experiences, and entities continually adapt to those experiences, ACMs might be the only practical method of analysis as mathematical methods are typically very limited in its ability to derive dynamic consequences [46]. This is especially so in the context of repeated games, in which the iterative nature of interaction clearly highlights the suitability of ACMs for modeling and analysis.

1.6 Learning in agent-based models

In tandem with the application of ACMs to game theory, learning methodologies often form part and parcel of the implementation. As a crucial aspect of artificial intelligence [51], they define means by which agents are able to process, update and utilize current information and past experiences that are acquired from their environment to make intelligent decisions in a dynamic way. More importantly, learning methodologies facilitate positive strategy adjustments which help agents improve their payoffs or positions relative to their environment of existence and interaction over time, by drawing from available information and experiences.

By far, the ability to learn and improve constitutes an important element of human adaptation and is especially vital when it comes to modeling aspects of game theoretic interaction in the real world context – one that is characterized by a dynamically changing environment where multiple players are constantly adapting their strategies to one another within an underlying mechanism of game play that is possibly also changing as well. Without learning, modeling of agent behavior in computational models becomes unrealistic. Some popular examples of learning methodologies in ACMs include Q-learning [52], Bayesian learning [53], branch-and-bound [54], dynamic programming [55], temporal difference learning [56], gradient descent [57], and simulated annealing [58] among many others.

As much as learning is important in ACM, the incorporation of realistic modes of learning must also not be under-emphasized. For instance, we are not nearer to understanding the properties of systems if we simply compute outcomes of interaction by running experiments which we equip agents homogeneously with the same non-dominant strategy [44]. From the perspective of individual agents, learning methodologies should ideally take into account of realistic elements such

as the dynamism of learning process, probabilistic nature of decision making and notion of bounded rationality [59] – which includes limited information, imperfect cognitive processing and learning capabilities, and finite duration for decision making. The constraints of bounded rationalism are due to the fact that decision-makers usually lack the abilities and resources to arrive at optimal solutions in reality, and instead apply their rationality only after simplifying available choices substantially. To this note, many existing techniques fail to deliver the required sense of realism as most operate with core assumptions that agents are perfectly rational, embrace homogeneous forms of learning, or interact and make decisions which are clearly too deterministic.

On a wider note, learning in game theoretic interaction can be saliently viewed as a process where entities in ACMs evolve gradually and incrementally in response to a changing environment (which comprises of the game mechanism as well as all other evolving entities). For instance, agents do not instantaneously and simultaneously adjust their behavior to theoretical optimum strategies. Rather, the adoption of a new strategy may spread through a population of agents as word of its efficacy diffuses in a manner akin to mimetic evolution [44]. We can view each agent and its environment as coevolving counterparts where each undergoes *co-evolutionary learning* [60] as a form of adaptation to one another.

Finally, with appropriate learning mechanisms in place for each entity in an ACM, a paradigm is also required to discover eventual outcomes of the game theoretic interactions which we are seeking to analyze from different perspectives e.g. the nature of dominant strategies, existence of NEs and possibly different pathways of convergence to the outcomes - whose dynamism are typically not addressed by learning models in classical game theory. From the perspective of

18

analytical approaches, this ideally equates to solving multi-player optimization problems and deriving the solution outcomes where all players play out their best strategies. As far as ACMs are concerned, a dynamic and realistic computational framework, similar to that proposed in EGT, is needed to model and simulate co-evolutionary learning and adaptation in strategic environments.

1.7   Evolutionary Algorithms

To the above note, Evolutionary Algorithms (EAs) [61] present a simple and elegant framework to address challenges of modeling realistic learning experience and solution discovery in ACMs. Originally conceptualized based on Darwin's Law of Natural Selection, the paradigm's inherent capability for solving complex optimization tasks via stochastic, parallel searches makes it a suitable candidate for finding solutions to complex game theoretic problems, especially those which are mathematically intractable to analytical approaches and too extensive in scope to be covered by human-based experiments. For instance, in the attempt to assess the presence of strategy mixtures which constitute equilibria in any game theoretic interaction, it is necessary to evaluate the interaction between known strategies as well as the space of strategies which are yet to be considered. Given the very large strategy space, exhaustive search will prove infeasible. In comparison, population-based heuristic search methods like EAs clearly speed up the process of solution discovery and present possible avenues for studying interaction between different strategies by sampling the search space in a systematic manner [44].

Apart from being a search and optimization paradigm, EAs also accounts for realistic aspects of replicating learning experiences for agents. As opposed to deterministic, idealized learning models in which agents always choose the best

decision that maximizes payoffs, the use of stochastic elementary processes like selection, recombination and mutation in EAs introduces a probabilistic dimension to the process of agent learning and strategy discovery. This mode of evolutionary learning is more in sync with the nature of how humans learn in the real world context, which is essentially characterized by uncertainties and imperfections in decision making. For instance, making unintentional mistakes, bounded rationality in thinking, incomplete or imperfect knowledge about the situation of game play etc, can well result in outcomes where agents do not always make the best choices that are available to them. The list goes on. As a dynamic optimization framework, EAs, unlike many existing static methodologies also drives the process of learning and adaptation for the agent population on a continuous basis.

In addition, different agents are likely to embrace learning in diverse ways e.g. some might like to imitate or partially adopt the strategies of others while the rest might prefer a trial and error mode of learning. Instead of assuming that all agents will always adopt homogeneous learning styles and converge in a straight forward manner towards the adoption of optimal strategies, models should seek to accommodate mixing and blending of different learning methods, so that the final stable states, if there are any, can be attained via varied pathways of convergence. To some degree of flexibility, such assorted outcomes can be subtly captured by the process of evolutionary learning. Details will be furnished in Chapter 2.

Although some arguments have been staged against EAs with regards to its inconsistency in obtaining optimal solutions, the paradigm is nonetheless, easy to design and yield good, if not the best, solutions most of the time. This is crucial as we usually seek and settle for good enough or satisfactory solutions rather than the best solution in most of our real world encounters [62]. This is especially true

given the earlier stated facts that agents are imperfect in their process of making decisions. It makes not much sense to study situations of optimality when agents themselves might not even acquire the best strategies. Given the context of real world interaction, it is necessary to examine the attainability of solution outcomes given the existing strategic behavior of agents. Focusing our attention on good strategies with a greater likelihood of attainability e.g. large basin of attraction [63] in the strategy space is more realistic and pragmatic than mapping optimal ones that have low chances of adoption. The analysis of strategies should suffice as a useful guide for social scientists and policy makers alike to attest the effectiveness of mechanisms and policy decisions, as well as to design and formulate new ones.

EAs also provide the flexibility to incorporate input knowledge from users so that parameter optimization can be carried out within the bounds considered to achieve effective abstraction of the problem. This constitutes an important trait as designers of social experiments can flexibly include subsets of information that are useful, and exclude those that have little or no contribution to the outcome and whose inclusion might even complicate the search process. With input knowledge well represented in structured chromosomes, it also becomes easier to analyze the final strategy due to the explicit nature of solution representation in EAs.

1.8   Overview of this Work

From the afore-mentioned discussion, game theoretic modeling and analysis is a challenging research topic that requires much attention from social scientists and researchers. To achieve accurate and effective modeling which yields meaningful analysis and insights into game theoretic interaction, the difficulties in analytical and human-based empirical methods will have to be overcome; together with the

paramount need to facilitate solution discovery and integrate realistic and dynamic elements into the process of learning for individual entities during their interaction. Though EAs provide a feasible solution measure to address the above issues, it is however, very difficult or almost impossible to construct an absolute measurement of quality via which traditional evolutionary approaches and optimization-based search algorithms can be used; since the "goodness" of game strategies can only be evaluated when they pit themselves against one another. In view of the above challenges, co-evolutionary algorithms (CEAs) [64], a special variant of EAs, are used. Implementing the same general evolutionary framework as traditional EAs, CEAs are suitable to simulate learning in games as they do not require any explicit quality measurement of strategies in order to function - the search for increasingly better strategies are driven solely by strategic interactions among competing ones.

The prime motivation of this work is to provide a comprehensive treatment on co-evolutionary simulation modeling – the application of stochastic CEAs to simulate evolution and adaptation processes and further game theoretic analysis in ACMs. In particular, the thesis will assess the viability of using CEAs as a simple but complementary alternative to existing mathematical and experimental methods in the study of repeated games. The interest in repeated interaction is largely due to its extensive applicability in many real world situations and the added fact that cooperation may be easier to sustain in a long-term relationship than in a single encounter [65]. As opposed to the analysis of short-run games which is often too restrictive, the analysis of interaction in repeated games can probably provide us with interesting insights into how cooperation can be achieved and sustained.

This rest of the work is organized into four parts. Part one, consisting of Chapter 2 will cover some core concepts, advantages as well as some applications

of EAs, followed by a comprehensive review on CEAs, and then finally drawing parallels as a means of comparison between aspects of CEA and game theory. The second part of this work will attempt to verify the ability of using co-evolutionary and/or hybridized approaches to derive solutions and discover good strategies that are closely similar or comparable, if not better, than solutions which are proposed by existing methodologies, in two game theoretic test problems. Chapter 3 seeks to develop a competitive computer poker player that specialized in Texas Hold'em. This is achieved by means of exploring the possibility of applying CEA to evolve Nash-optimal poker strategies that are comparable in performance to those derived through traditional means [66]. Chapter 4 redirects the application of CEA to the classical IPD problem setup, where the comparative performance and adaptability of evolutionary, learning and memetic strategies is benchmarked against a list of existing IPD strategies [67]. The objective is to assess whether evolution, learning or a combination of evolution with learning can lead to formation of strategies that will adapt and thrive well in complex environments.

The third part concentrates on the use of co-evolutionary approaches for game theoretic modeling and simulation, from which we can analyze interesting emergent behavior and trends that will give us new insights into the complexity of collective interaction among diverse strategy types across temporal dimensions. Chapter 5 extends the IPD model discussed in Chapter 4 to a spatial version in an attempt to simulate and analyze the phenomenon of civil violence as the behavior of autonomous agents within a multi-agent social network [68] is co-evolved over time. Chapter 6 pursues the modeling and analysis of a multi-player public goods provision game, focusing specifically on the scenario where agents interact and co-evolve under asymmetric information [69]. In both chapters, simulated results

23

can be used to complement findings from existing game theoretic studies and to assess the validity of related social theories in theoretical and complex situations that often go beyond their original scope of assumptions. Finally, chapter 7 in the fourth and final part concludes the thesis with a broad summary of contributions and brief discussion on possible research works that can be embarked on in future.

1.9 Summary

In this chapter, we have covered the necessary concepts, definitions, scope and a survey of development and applications in game theory to appreciate this work. This chapter also presented the deficiency in some of the existing approaches with regards to the modeling and analysis of game theoretic interaction. Subsequently, the use of EAs, specifically CEAs, as a viable learning method, has been proposed to complement existing computational approaches of using ACMs for the purpose of addressing issues of mathematical intractability, constraints in scope of analysis, inherent realism of interaction and the dynamism of learning, solution discovery and strategy improvement in game theoretic modeling and analysis. Finally, the overview of this work is presented with a brief introduction to the chapters that are relevant to the context of the research work.

Chapter 2

Evolutionary Algorithms

Before we embark on the use of evolutionary approaches to model and simulate game theoretic interaction using ACMs, a core understanding of EAs is necessary. Originated as a branch of computational intelligence [70] techniques which also encompass fuzzy logic [71], artificial neural networks [72] swarm intelligence [73] - ant colony and particle swarm optimization, and artificial immune systems [74] etc, EAs are stochastic, population-based search algorithms that are inspired from Darwin's theory of evolution and use several stochastic processes like *selection*, *reproduction*, *crossover* and *mutation*, among many others, to develop generations of strategies that follow the basic principle of survival of the fittest.

To further elaborate, Darwin's theory states that all organisms have their own unique genetic make-up. During reproduction, their genes are passed on to the next generation, of which some are altered occasionally by variation processes. All organisms are tested by the environment and by one another, where only the fittest survive to propagate their genes to subsequent generations. Over time, only those with genes that are best suited for adaptation to the environment is left. EAs use precisely this concept to solve complex optimization tasks via a population of candidates - each being a possible solution to the problem. Candidates are tested and sorted according to their performance (or *fitness level*), and those that perform better will get a higher likelihood to "reproduce". A candidate may also be varied to widen the scope of search and avoid locally optimum solutions. Information exchange among population from one generation to another that is provided by selection and variation processes is used as an efficient guide to direct the parallel

search via the solution/strategy space. After substantial iterations, the algorithm should eventually evolve a solution that is optimal or near optimal for the problem.

Other than genetic operators, EAs also uses mechanisms which are absent in the natural world to improve its performance and efficiency during the course of search. Two such examples include elitism [75] and niching [76]. The prior clones the best candidates and replicates the exact genetic makeup in the next generation to ensure that good solutions found so far are not lost through evolution. The latter penalizes candidates with similar characteristics by reducing their likelihoods of reproduction. This has an effect of preserving population diversity and widening the search capability of EAs. There are four generic variants which are in used – the genetic algorithms (GA) [77], evolutionary strategies (ES) [78], evolutionary programming (EP) [79] and genetic programming (GP) [80], each of which differs in terms of the representation, genetic processes used or means of implementation. As far as the scope of thesis is concerned, GA will be used throughout different chapters unless otherwise stated explicitly. A brief pseudo-code of EA's operation is shown in Figure 2.1 and details of the various elements that comprises EAs is highlighted and described in section 2.1.

__Initialize__ population of individuals
__Evaluate__ each individual in the initial population
t :=0
Repeat
   __Niching__ to penalize like individuals
   __Select__ parents from population
   __Generate__ offspring from parents by genetic operators
   __Evaluate__ offspring population
   __Elitism__ to retain elite individuals
   __Select survivors__ for new population
   t := t + 1
Until some __terminating criteria__ is satisfied

Figure 2.1: Pseudo code of EAs

2.1 Elements of EAs

Several basic elements constitute the crux of the robust search and optimization evolutionary paradigm. These are highlighted in subsections 2.1.1 to 2.1.9.

2.1.1 Representation

Before EAs can be applied to a problem of concern, there must first and foremost be a way to represent an individual or entity of evolution, which is otherwise also known as a chromosome. This is inspired by the encoding of genetic inheritance in the DNA of every biological organism. Each chromosome essentially encodes a possible solution or set of solution parameters to be optimized. In the context of game theory, each chromosome will denote a possible strategy which players can use to interact with other players during the game. By structuring a chromosome appropriately in terms of the representation, effective evolution can take place to evolve the optimized solution or parameters eventually. Some commonly used representation includes real number, binary or even complex data structures such as finite state machines (FSM) and neural networks.

2.1.2 Fitness

Fitness represents the criteria to which nature selects individuals to survive on to subsequent generations. For a biological organism, its fitness is measured by the corresponding interaction with its environment e.g. its lifespan, the opportunities to reproduce, number of offspring etc. For EAs, an individual's fitness is likewise measured by the "goodness" of the solution which it represents. For instance, a dominant strategy will have higher fitness as compared to a dominated one. The fitness value is then used to determine the extent which an individual e.g. strategy is allowed to reproduce into the next generation.

2.1.3 Population and generation

As a population-based paradigm that conducts multiple concurrent searches, EAs are commonly initialized with a random pool of potential solutions as a start. After each solution has undergone fitness evaluation, the entire population is subjected to evolutionary processes as described from 2.1.4 to 2.1.8. This produces a new population of individuals and one generation or evolutionary cycle is said to have elapsed. EAs will typically require from several to many generations (depending on the complexity of the problem and size of search space at hand) before a good or optimal solution can be derived from the search process.

2.1.4 Selection

Selection is one of the most fundamental operations in EAs where individuals are selected to propagate to subsequent generations. Based on nature's law; the fitter individuals in the population should be given higher likelihoods to survive and reproduce but weaker ones should nonetheless be still given some finite chances of survival. It is crucial to implement a fair selection scheme so that balance can be maintained in the EAs e.g. the algorithm will not be overly biased towards the choosing of fit individuals at the expense of weaker ones as this will lead to the rapid population of like individuals and loss of genetic diversity. This can lead to premature convergence and danger of being trapped in a local optimum. Likewise, the algorithm should not give too much emphasis on preserving weak individuals as this leads to low selection pressure and rate of convergence. A fine balance between exploration and exploitation is typically required for good performance. Commonly used selection schemes include the fitness proportionate or roulette wheel selection [61], tournament selection [81], stochastic universal sampling [82], rank-based selection [61] etc.

2.1.5 Crossover

Crossover or recombination is the process where genetic characteristics from two individuals are blended together and passed down to their offspring. This is in the hope that at least some of the children will be fitter than either of their parents. In EAs, crossover will generally involve an exchange of chromosomal materials in the creation of offspring. Individuals chosen by selection e.g. parents will usually reproduce with a certain crossover probability which is typically set high, so as to facilitate the exchange of search information among individuals in the population from one generation to the next. This is one advantage which EAs have over other independent search schemes. Exchange of genetic materials is usually performed using a variety of crossover schemes and much is dependent on the chromosomal representation and problem nature. Some popular schemes [83] include single and multi-point, uniform, shuffle, arithmetic, selective and order-based crossovers.

2.1.6 Mutation

In the natural world, mutation denotes the random modification of some genetic material which is inherited by an individual. Though most mutation would appear harmful, they may be beneficial occasionally and result in increased fitness for the organism. Mutation is necessary in EAs to preserve diversity of individuals and maintain the exploration ability of the evolutionary search process. This is usually implemented with a low probability and involves randomly changing each of the bits of an individual's chromosome in a uniform or non-uniform manner.

2.1.7 Niching

Originally proposed by Goldberg [76] to promote population distribution, prevent genetic drift as well as to search for possible multiple peaks in single objective

optimization problems, niching is a mechanism which is implemented in EAs to maintain the diversity of individuals within the population pool. It works on the principles of speciation such that individuals who are too alike are penalized to reduce their fitness and chances of being selected. Though not directly linked to nature's evolution, this mechanism has an effect of spreading evolutionary search effort across the problem's search space, thereby increasing its search capability and subsequent chances of locating the global optimum, especially in complex or multi-modal problems. Niching is usually implemented on the basis of a sharing or crowding radius e.g. individuals within the radius are considered alike to one another and each is penalized in accordance to the number of individuals which share like characteristics and the proximity of the feature set to those individuals.

2.1.8 Elitism

First conceptualized by De Jong in [42] to preserve the best individuals found and prevent lost of good ones due to the stochastic nature of evolutionary processes, elitism, like niching, is a mechanism which does not see any parallels in nature's version of evolution. It is employed in EAs to ensure that the fittest chromosomes in the population are passed on to the next generation without alteration by genetic operators. Elitism ensures that the population's minimum fitness is never reduced from one generation to the next and this usually also entails a more rapid inherent convergence. Typical implementation involves replacing the weakest n_e% of the offspring population with the fittest n_e % of the parent population.

2.1.9 Stopping Criteria

The stopping criteria in EAs refer to the conditions which will stop the evolution process when met. This is important as problems are different in their own right

and computational resources and time are also limited. As such, it is important to set some criteria so that good, if not optimal solutions, can be derived within the constraints which we have to abide by in realistic circumstances so as not to allow the EAs to execute forever. Some typical criteria can involve setting a maximum number of generations, stopping once a certain level of convergence is reached etc.

2.2 Advantages of EAs

As stated earlier, EAs as a robust and generalized heuristical search method offers added advantages over the existing search paradigms when it comes to tackling complex problems; since we are mostly concerned not so much to find the global optimum solution, but rather a solution that is the best that can be achieved with available time and resources. In terms of representing solutions, EAs work with a coding of the problem's parameters and not the parameters directly. It operates directly using only objective function values without problem-specific information or even derivatives, giving it a considerable advantage in tackling a very broad range of problems successfully. Performing search via a population of individuals instead of a single independent search entity reduces the chance of getting trapped in a local optimum and also hastens the process of solution discovery. By virtue of its flexibility of implementation, EAs can also be hybridized easily with other methods to deliver added performance improvement. Moreover, the probabilistic elements in the heart of EAs, though far from indicating directionless search, are actually used to guide the algorithm to explore areas of the search space which are most likely to lead to improvement [84].

Since its popularization, EAs and other evolutionary methods have since made significant contributions to countless areas of research and applications such

as massive parameter optimization [85], scheduling [86], engineering design [87], analysis of social interaction [88] and complex multi-agent systems [89]. Many of the areas often traverse multi-disciplinary boundaries, thus allowing researchers to discover creative solutions, derive insights from cross-disciplinary contexts and understand existing problems from whole new perspectives.

2.3   Co-evolutionary algorithms

In the context of our study, it can be extremely difficult to formulate a fitness function that reflects the underlying properties of games as accurate measurement to determine "goodness" of solutions cannot be obtained in most cases. This is because the fitness of each strategy can only be evaluated through interaction with other evolving strategies who can be members of the same or different populations, depending on the search problem of concern. The deliberate use of any ill-defined fitness measure to suit the application of traditional EAs and optimization-based paradigms can well lead the search process towards the discovery of inferior/sub-optimal strategies, which is certainly not desirable.

Given this perspective of concern, CEAs constitute a special type of EAs whose nature offers a fitting and viable solution. Inherently, CEAs apply selection and variation processes iteratively to the competing population of strategies under the same general evolutionary framework as discussed previously but differs from conventional EAs with respect to how the fitness of a typical strategy is derived. For traditional EAs, fitness value of a solution is always invariant and independent of the population's composition at any point of the evolutionary time-line. The distinguishing feature in CEAs pertains to the notion of fitness inter-dependency among different individuals which the fitness evaluation process establishes. For

instance, each strategy's fitness is highly dependent and correlated to the fitness of his interaction partners, which in turn suggests that a strategy's chance of survival depends effectively on its fitness relative to the partners. Fitness is relative and dynamic and manifests itself as a function of the population composition which is subjected to change from one generation to the next. Regardless of the number of populations, the most conventional pattern of interaction - complete mixing [36] is to have every member interact with every other individual who can possibly serve as potential partners - a symmetric two-player game with a single population of n individuals yields $n(n-1)/2$ distinct interactions while an asymmetric two-player game with two populations of size m and n will derive mn distinct interactions.

CEAs are generally divided into competitive and cooperative [90] schools of differentiation where the latter aims to solve a difficult problem X – which is decomposable into a collection of easier sub-problems; by coevolving an effective set of solutions – each individual denoting a solution to one of the corresponding sub-problems, that can work together to form a complete solution to the original, larger problem. In cooperative CEAs, there is no sharing of genetic information between solutions in different sub-populations and fitness evaluations are made by forming collaboration between an individual of one population and representatives of other populations. In competitive CEAs, each individual represents a complete solution to the problem and competes with one another for the right to survive just like what is typically in place for conventional EAs.

As far as the scope of this thesis is concerned, subsequent chapters will focus on the use of competitive CEAs. Nonetheless, both types of CEAs, despite innate differences, share similar motivations with respect to the learning process and apply to problems in which formulation of explicit fitness function is difficult

or impossible e.g. evolving game playing strategies in the context of this work. There are generally four variations of fitness measures as Wiegand [91] defines:

Definition 1: Objective measure

A measurement of an individual is *objective* if it considers that individual independently from other individuals, aside from scaling or normalization effects.

Definition 2: Subjective measure

A measurement of an individual is *subjective* if it does not consider that individual independently from other individuals.

Definition 3: Internal measure

A measurement of an individual's is *internal* if it does influence the course of evolution in some way.

Definition 4: External measure

A measurement of an individual is *external* if it does not influence the course of evolution in any way.

From the definitions of the above four types of measure, it is clear that traditional EAs adopt an objective internal measure in its fitness evaluation. CEAs, on the other hand adopt a subjective internal measurement for fitness assessment and this pertains to the payoff which each player derives in the context of game theoretic interaction. In the co-evolutionary framework, two or more populations generally co-exist and co-adapt to one another over time. This is especially suitable when it comes to modeling game theoretic interaction and learning among asymmetric or different groups of strategies. The utility maximization behavior of each rational entity in the population via strategy improvement can be efficiently modeled as

EAs are naturally designed for optimization tasks. In such context, each player starts with an initial strategy and adapts to the dynamic environment by bettering its strategy over time by means of co-evolutionary learning where players of the same type evolve strategies collectively and independently of other types. As a subset of EAs, the dynamism in CEAs does provide an important element that is missing from the traditional theory of games, making it appropriate for analyzing scenarios with repeated interactions and modeling social systems. CEAs also maintain population diversity better than "classical" EAs [92].

2.4 Drawing parallels

To draw close parallel with game theoretic interaction in reality, it is vital for any model to possess a viable and realistic learning mechanism for players to improve their strategies over time. With the assumptions of bounded rationality, players do not have perfect information about the global environment and are not attributed with advanced information processing capacities to undertake strategy revision using Bayesian Learning or Nash optimization. Though the replicator dynamics in EGT seems viable, its applicability relies essentially on the core, but somewhat unrealistic assumption of an infinitely large player population [93]. A probabilistic element is also lacking as outcomes generated based on iteratively simulating the static replicator dynamics equations are by and large deterministic in nature.

CEAs is selected as the proposed learning mechanism in the series of work that are presented in the subsequent chapters of this thesis, as it is able to produce characteristics that are closely similar [94] to the replicator dynamics using only a finitely large population. Such characteristic of the co-evolutionary framework not only fulfills constraints of limited computational resource, but also allows us to

flexibly study situations that may not involve infinitely large number of players. The co-adaptation of entity populations over numerous evolutionary episodes also captures the essence of the population-based and temporal nature of EGT. In addition, co-evolution provides a stochastic learning framework which incorporate uncertainty and realistic imperfections into the process of simulation. In situations when theories only cover idealized scenarios where core assumptions are not violated, CEAs can allow flexibility to model realistic constraints like information asymmetry, bounded rationality, framing and other model imperfections. The elegant co-evolutionary framework captures three distinct aspects of learning [94] within each evolving population or strategy type - *learning by replication, social exchanges* and *experimenting.* These processes correspond to notions of selection, crossover and mutation respectively. The analogy between various components will be further elaborated in Chapter 6. In summation, the following parallels in Figure 2.2 can be drawn between CEAs on one hand and game theory on the other when attempting to employ co-evolutionary simulation modeling in the context of game theoretic modeling and analysis.

CEAs		**Game theory**
Fitness	←→	Payoff
Individual/Chromosome	←→	Strategy
Selection of fit individuals	←→	Selection of good strategies
Selection	←→	Learning by replication
Crossover	←→	Learning by social exchanges
Mutation	←→	Learning by experimenting

Figure 2.2: Drawing parallels between CEAs and Game theory

2.5 Summary

In this chapter, we have covered core concepts and fundamental processes that are involved in the implementation of EAs. An understanding of these basic building blocks leads on to a discussion on its advantages and potential applications. This is followed by a comprehensive review of CEAs – stating its salient characteristics and distinction from EAs. Finally, parallels are drawn as a means of comparison between aspects of CEAs and game theory.

Chapter 3

Evolving Nash Optimal Poker Strategies

Poker is a card game that is widely played by many around the world. In the recent decades, it has experienced an unprecedented surge in popularity owing to the prevalence of online poker which made it much more convenient for players to search for and join a poker game. In unison, the decreasing cost of computational power has also allowed the creation of strong computer players using artificial intelligence (A.I.). Much research had revolved around the game, not only to develop better strategies, but also using it as a viable means to study psychology, economics and the effectiveness of neural [95] and Bayesian networks [96].

Suitability of poker in such studies spans from a couple of factors. Firstly, it is a game of imperfect information as some information of the game state e.g. the opponent cards [97], is not known to players at any one time. This differs in contrast to games of perfect information e.g. Chess, where all game information is displayed on board. Secondly, poker is computationally less complex than other games of imperfect information e.g. bridge. Despite so, impact of this imperfect information trait is nonetheless not as negligible as that in scrabble [97]. Due to such dynamic nature of the game, no computer player has ever beaten the human poker champion, unlike what had been achieved in Chess [98]. Though computer players are getting better at the game from the current state of A.I. research, none is as yet, able to beat a human of grandmaster ranking in both the heads-up (one versus one) and multiple player version consistently.

With the goal of developing good poker strategies, CEAs present a viable means of evolving intelligence as it is able to create generations of strategies that

follow the basic principle of survival of the fittest via evolutionary processes. A foreseeable advantage of this technique is its ability to produce good strategies with minimal, if not without use of expert knowledge and explicit fitness measure. This allows the creation of objective strategies and possibly a discovery of those unthought-of before. Closely related to EAs, CEAs have the potential to "solve complex problems even their creators themselves do not fully understand" [77].

This chapter attempts to develop a poker A.I. that plays approximately at NE [3] using a CEA that employs offline competitive co-evolution e.g. [99] as the means of adaptation. The version of poker used is the heads-up pot-limit Texas Hold'em and the reason for aiming to achieve NE instead of merely maximizing winnings is due to the intransitive nature of poker. This implicates that attempts to develop players which win maximally against other players through any offline evolutionary means might not be possible, unless excellent opponent modeling is present. Being able to create players that play at NE e.g. at worst draw against any opponent [3], is crucial, at least as a start point for developing good poker players. Based on performance analysis of these players, insights on how well CEAs can be applied to full-scale Texas Hold'em can then be made.

The chapter is organized as follows. Section 3.1 discusses some prominent works in the existing poker literature. Section 3.2 provides an overview of Texas hold'em. Section 3.3 introduces the game theoretic fundamentals behind poker and Section 3.4 describes the game engine design. The proposed co-evolutionary model is elaborated in Section 3.5 and Section 3.6 highlights findings from a preliminary study. Section 3.7 presents and analyses the simulated results and efficiency of CEAs. Section 3.8 concludes with a broad summary of discussion on the result analysis as well as some possible future model improvements.

3.1 Background study

Numerous techniques had been used to develop poker A.I. The most successful of all are developed by the Department of Computer Science, University of Alberta. Poki is one such A.I. that specializes in multiple-players pot-limit Texas Hold'em. The system structure [100] (shown in Figure 3.1) is segmented into hand-strength assessment, hand-potential assessment, betting strategy, bluffing, unpredictability as well as opponent modeling [101] – [103]. The A.I. makes use of probabilistic knowledge and selective-sampling simulation [104] to implement betting. Every time it is to make a decision, it will do a selective-sampling simulation to look ahead and determine its best course of action. A probability triplet which consists of three probabilistic numbers representing the probability of it folding, calling or raising, is returned. One action is chosen randomly according to their probabilities. The opponent modeler is a component used to predict the next action of opponents. During the development of this component, neural network was applied [105] to improve it. The biggest strength of Poki is its ability to adapt to its opponents' style of play and exploit their weaknesses.

PSOpti, the most successful heads-up pot-limit Texas Hold'em A.I. is also developed by University of Alberta. It was the winner in both the Association for the Advancement of Artificial Intelligence (AAAI) Computer Poker Competitions in 2006 and 2007 [106]. The A.I. uses Game Theory to play poker [107]. Firstly, the game tree is simplified through abstraction of the game. Abstraction is done by reducing and eliminating the number of betting rounds, composing preflop and postflop models, and using bucketing techniques that group hands of similar value together. These reduce the complexity of solving the game tree from $O(10^{18})$ to $O(10^7)$. Linear programming is finally used to solve the smaller game tree. From

the games it had played, PSOpti managed to beat all computer players (Table 3.1) and most of the human players, except those of master ranking and above (Table 3.2). The strength of PSOpti lies in its ability to play close to the NE by using a pseudo-optimal strategy that displays almost no exploitable weakness. However, it employs no opponent modeling, which makes it less capable of exploiting much weaker opponents as compared to other poker A.I. systems.
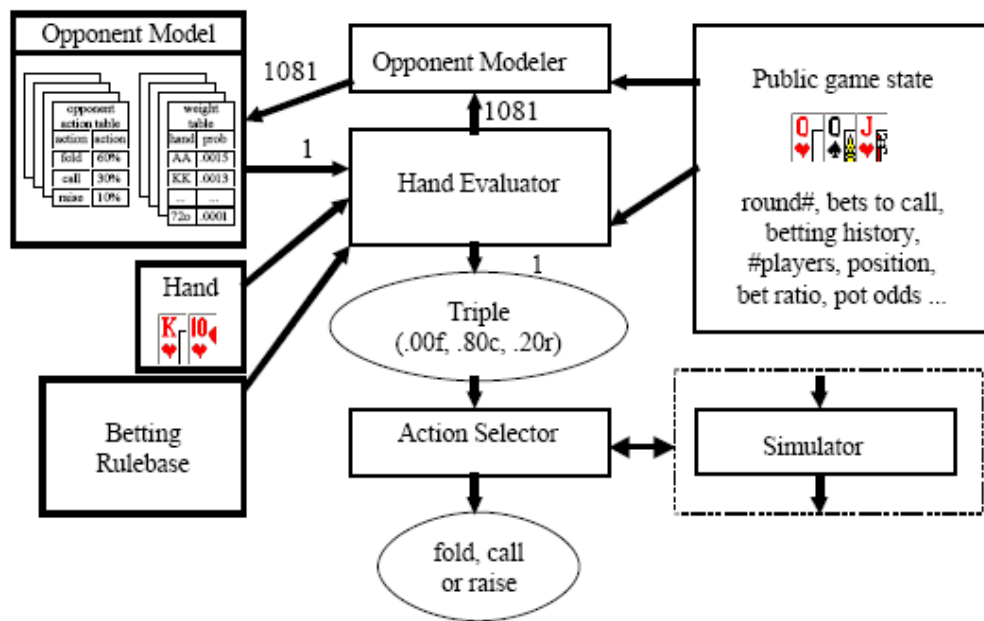
Figure 3.1: Overall architecture of Poki

Table 3.1: Performance of the various computer players against one another [107]

No.	Program	1	2	3	4	5	6	7	8
1	PsOpti1	X	+0.090	+0.091	+0.251	+0.156	+0.047	+0.546	+0.635
2	PsOpti2	-0.090	X	+0.069	+0.118	+0.054	+0.045	+0.505	+0.319
3	PsOpti0	-0.091	-0.069	X	+0.163	+0.135	+0.001	+0.418	+0.118
4	Aadapti	-0.251	-0.118	-0.163	X	+0.178	+0.550	+0.905	+2.615
5	Anti-Poki	-0.156	-0.054	-0.135	-0.178	X	+0.385	+0.143	+0.541
6	Poki	-0.047	-0.045	-0.001	-0.550	-0.385	X	+0.537	+2.285
7	Always Call	-0.546	-0.505	-0.418	-0.905	-0.143	-0.537	X	=0.000
8	Always Raise	-0.635	-0.319	-0.118	-2.615	-0.541	-2.285	=0.000	X

Table 3.2: Humans vs. PSOpti2 [107]

Player	Hands	Posn 1	Posn 2	sb/h
Master-1 early	1147	-0.324	+0.360	+0.017
Master-1 late	2880	-0.054	+0.396	+0.170
Experienced-1	803	+0.175	+0.002	+0.088
Experienced-2	1001	-0.166	-0.168	-0.167
Experienced-3	1378	+0.119	-0.016	+0.052
Experienced-4	1086	+0.042	-0.039	+0.002
Intermediate-1	2448	+0.031	+0.203	+0.117
Novice-1	1277	-0.159	-0.154	-0.156
All Opponents	15125			-0.015

Apart from conventional means, EAs have also been used in poker research. Barone and While [108] applied EAs to a simplified version of poker where they had a player with evolving strategies play many games against fixed opponents. The strategies for each situation are the ones undergoing evolution rather than the player itself. Experimental results indicate that the evolving player performs better against fixed opponents as generations elapse. However, the player takes many generations before it can fully exploit the opponents. This makes it infeasible for playing against real opponents, as games do not last many rounds and opponents are not fully static. In view of the complexity in poker, the authors also used EAs to find specialized intransitive countering strategies [109].

Another application of EAs in poker is postulated by Frans Oliehoek et al. to calculate NE using co-evolution [110]. The experiment was done on a simplified version of poker with only 8 cards. The objective is to verify if the CEAs can help to speed up the calculation to achieve an optimal strategy. Simulated results show that not many generations are required to achieve a strategy which plays relatively close to the NE in the 8-cards variant. This highlights the possibility of applying EAs, particularly CEAs, to larger scale of the game.

3.2 Overview of Texas Hold'em

Texas Hold'em is played with a standard 52-cards deck by 2 to 10 players. It is different from normal poker as community cards rules are included. This offers more strategic depth and less luck factor; making it one of the most popular [111] poker variant that is played today.

3.2.1 Game rules

A game round is divided into four stages - Preflop, Flop, The Turn and The River. Each stage is differentiated from one another by the number of community cards revealed. In the pot-limit version, stakes are determined by the small bet and big bet amounts, where the big bet is typically twice of the small bet.

- *Posting of blind*

Before every round begins, blinds are posted by the first two players – the dealer and player on his/her left. The dealer pays an amount equal to half the small bet while the second player pays a full small bet. These are called the big blind and small blind respectively. The cards will then be shuffled and two cards will be distributed to each player.

- *Preflop*

Preflop is the first stage of betting. Betting will start with the player left of the small blind, i.e. the third player. During his/her turn, a player can choose either to fold, call or raise. If the player chooses to fold, he/she will be out of the game immediately. If the action is to call, the player will have to bet as much money as needed to match the highest stake from any of the other players at that point in the game. If the player chooses to raise, not only does he/she need to match the

highest stake, he/she also has to add an additional amount that is equivalent to that of a small bet to the highest stake, thereby creating a new highest stake. After this, the turn goes in a clockwise manner around the table. Betting will continue until everyone that is still in the game calls, which will then conclude this stage. It is to be noted that the stake can only be raised three times during each stage.

- *Flop*

The Flop stage commences after Preflop ends. In this stage, three community cards will be dealt and revealed faced up on the table. These are cards which any player can use to form combinations of five cards with the cards on their hands. Quality of the combinations is used to determine the winner at the end of the game. After the three cards are revealed, betting will resume with players that are still in the game taking turns to choose their actions. Beginning with the dealer, this will proceed clockwise just as in the Preflop. Also, betting will continue until everyone still in the game calls and the stakes can only be raised three times.

- *Turn*

In the Turn that comes after Flop, an additional community card will be dealt face up. Betting proceeds just as it was done in the previous two stages. In the Turn and the River, the raise amount is increased to the big bet amount. Each time a player raises, the raise has to be the amount equivalent to the big bet, instead of the small bet like in the previous stages.

- *River*

The River marks the last of the four stages where a final community card is dealt face up to bring the total number of community cards to five. The raise amount

remains fixed at the big bet amount and betting proceeds just as it was done in the previous three stages.

- *Showdown*

If there is only one player left after the end of the River, he/she will be the automatic winner of that round. Otherwise, a showdown stage will occur where all contending players take turn to reveal the two cards on their hands or choose to withdraw without revealing the cards (called "muck"). Players will form the best possible combination of five-cards with the community cards and his/her two cards. The combinations are ranked and the player with the best combination wins the round and all the money in the pot. In the event of a tie, pot winnings will be shared among all tied players. Figure 3.2 shows the various combinations in poker. For details on the ranking of combinations, refer to Appendix A.

Figure 3.2: Name of poker card combinations

3.2.2 Playing good poker

Various forms of skills are required to master the game of poker. Some important ones include hand-strength evaluation, risk-rewards analysis, taking into account factors such as player position, bluffing, unpredictability and psychology.

- *Hand-strength evaluation*

The most important skill in poker is the ability to evaluate the goodness of one's cards. This informs a player of his chances of winning and subsequently helps him to decide on the action to take. Intuitively, a player should raise more often if his chances of winning is high to maximize winnings; and fold earlier if his chances of winning is low to minimize losses.

- *Risk-rewards analysis*

Despite occasions where a player's chances of winning are not particularly good, he should also call and stay in the game when the amount in the pot is very large as compared to the amount he has to bet. For example, paying a small call amount of $2 to get a chance at winning a potential reward of $50 in the pot does justify a good risk to take despite having a low chance of winning.

- *Player position*

It is known that players at later positions have greater advantages than those at earlier positions, owing to their privilege of observing the actions of most players before making choices. Such information reveals how confident other players are of their chances of winning. It is to be noted that player position, however, plays a lesser role in games with fewer players.

- *Bluffing*

As poker is a game of imperfect information, not only does a player not know of his exact chance of winning, his opponents are equally uncertain as well. In order to maximize winnings, players will have to play on this fact. At times, they have to make the opponents believe that they have a better hand than what they actually

have, so as to trick them into folding. To be effective, the art of bluffing has to be executed with great caution and good timing.

- *Unpredictability*

Unpredictability is necessary to make it difficult for opponents to find weaknesses in a player. A player who plays predictably will be exploited by his opponents in no time. Thus, a good player is one who will vary his style of game play in order to prevent opponents from forming an accurate model of his strategy.

- *Psychology*

Finally, a right interpretation of the opponents' psychological styles of game play is also crucial to play well in poker. An accurate opponent model, for instance, allows a player to predict his opponent's actions and hence, achieve maximum winnings against him.

3.3   *Game theory of poker*

Poker is a sequential, stochastic, zero-sum game of imperfect information. To devise a good evolution model for developing strategies that play approximately at NE for a game of such nature, a good understanding of the game theoretic fundamentals is essential.

3.3.1 Nash Equilibrium

NE is defined as the state where no player stands to gain anything by changing his/her strategy unilaterally. In games of perfect information, pure strategies [112] are used to identify the NE. A pure strategy is one where every scenario that is represented in the strategy space corresponds to a single action which is always

performed with a probability of 1. In contrast, a mixed strategy is one where a player's action to each scenario is determined by a probability distribution of all allowed actions. To achieve NE in games of imperfect information, players have to employ a mixed strategy. This can be reasoned via the following discussions.

3.3.2 Illustration of game theory for poker

As a full scale Texas Hold'em game is too complex to be solved theoretically, a simplified variation of poker is used in the following illustration. Consider a two players poker game which consists of only one stage and a betting round. At the start, each player will post $1 as blind. During his/her turn, the player can fold or call, but is not given any option to raise. If he calls, he pays an additional $1; if he folds, he is out of the game immediately. In this variation, there are only a total of three cards, numbered 1, 2 and 3 with a larger number value denoting a better card.
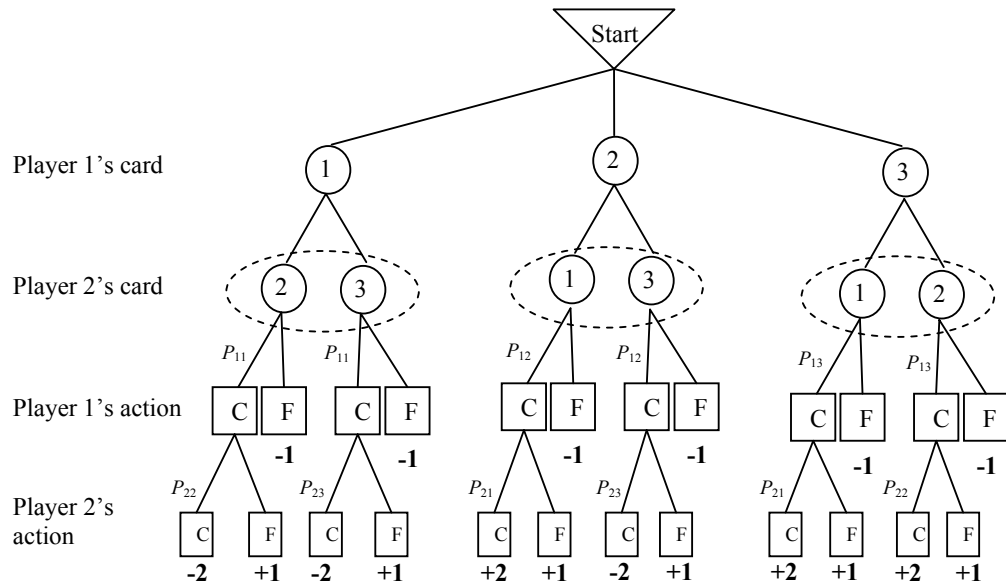
C: Call   F: Fold
P_{ij}: probability of player i calling when he has card j.
Numbers in bold are the pay-offs of player 1.
Pay-offs of player 2 are just the negative of player 1's pay-offs.
Dotted circles indicate ignorance of player 1, i.e. player does not know what card the opponent has.

Figure 3.3: Game tree of simplified poker variant from player 1's perspective

48

Figure 3.3 shows the game tree of the simplified poker variation, where P_{ij} is the probability that player i should call if he has card j. To achieve NE, all players must play the most possible way beneficial to themselves. From the game tree, it is observed that some actions are simply bad actions for player 2 e.g. calling when he has the card value 1. In some scenarios, it is also always good to call e.g. when the card value is 3. For player 1, it is always good to call if he has card 3. By observation, the solutions for p_{21}, p_{23} and p_{13} can be found to be

$$p_{21} = 0, \qquad p_{23} = 1, \qquad p_{13} = 1$$

The expected payoff, E for each player can be derived by using the Law of total probability as follows.

E(Player 1's payoff) $= E_1$

$$= \frac{1}{3} \times \frac{1}{2} \times \{(p_{11})[(p_{22})(-2) + (1 - p_{22})(1)] + (1 - p_{11})(-1)\}$$

$$+ \frac{1}{3} \times \frac{1}{2} \times [(p_{11})(-2) + (1 - p_{11})(-1)] \quad + \frac{1}{3} \times \frac{1}{2} \times [(p_{12})(1) + (1 - p_{12})(-1)]$$

$$+ \frac{1}{3} \times \frac{1}{2} \times [(p_{12})(-2) + (1 - p_{12})(-1)] \quad + \frac{1}{3} \times \frac{1}{2} \times 1$$

$$+ \frac{1}{3} \times \frac{1}{2} \times [(p_{22})(2) + (1 - p_{22})(1)]$$

$$= \frac{1}{6} p_{11}(1 - 3p_{22}) + \frac{1}{6} p_{22} + \frac{1}{6} p_{12} - \frac{1}{3} \tag{3.1}$$

E(Player 2's payoff) $= E_2 = - E_1 = -\frac{1}{6} p_{11}(1 - 3p_{22}) - \frac{1}{6} p_{22} - \frac{1}{6} p_{12} + \frac{1}{3}$ \hfill (3.2)

From (3.1), it is observed that the expected payoff maximization for player 1 is limited only to the adjustment of parameters p_{11} and p_{12}. As a larger p_{12} gives a larger E_1, we can set $p_{12} = 1$. This leaves us with only parameter p_{11} for player 1

whose value is to be determined. Likewise for player 2, value of parameter p_{22} is left to be worked out. Expected payoffs of both players can now be expressed as

$$E_1 = \frac{1}{6}p_{11}(1-3p_{22}) + \frac{1}{6}p_{22} - \frac{1}{6} \tag{3.3}$$

$$E_2 = -\frac{1}{6}p_{11}(1-3p_{22}) - \frac{1}{6}p_{22} + \frac{1}{6} \tag{3.4}$$

With equations (3.3) and (3.4), it is found that the expected payoff of either player is dependent on the strategy of the other player. As NE is a state where no player will be exploitable, the values of p_{11} and p_{22} can be found by solving (3.5)

$$\frac{\partial E_1}{\partial p_{22}} = 0 \quad and \quad \frac{\partial E_2}{\partial p_{11}} = 0, \tag{3.5}$$

$$\Rightarrow \frac{\partial E_1}{\partial p_{22}} = -\frac{1}{2}p_{11} + \frac{1}{6} = 0 \text{ and } \frac{\partial E_2}{\partial p_{11}} = \frac{1}{2}p_{22} - \frac{1}{6} = 0 \tag{3.6}$$

Therefore, we have $p_{11} = \frac{1}{3}$ and $p_{22} = \frac{1}{3}$. \tag{3.7}

The mixed strategy at NE (which is also the optimal strategy essentially) is shown in Table 3.3:

Table 3.3: Nash strategy for simplified poker

Probability of calling				
Card value		1	2	3
Player 1		1/3	1	1
Player 2	Player 1 called	0	1/3	1
	Player 1 folded	NA	NA	NA

With these strategies, it is found that:

$$E_1 = -\frac{1}{9}, \quad E_2 = \frac{1}{9} \tag{3.8}$$

3.3.3 Discussion on calculated results

From the calculated results obtained through game theoretic analysis in section 3.3.2, several observations can be made. Firstly, to achieve NE in the context for full scale poker, decision making must be modeled by a mixed strategy e.g. a probability triplet which uses separate probabilities to denote the tendencies to fold, call and raise. As a strategy consists of a set of rules for all decision nodes in a game tree and that a node is reached only via traversing branches, information that reflects the node which the game is currently on is necessary for a player to attain NE. Information that needs to be supplied to players are the cards (on both hand and community table) and history of opponent's and player's actions.

Secondly, it can be seen that there are three types of strategies in zero-sum games. NE or optimal strategies are those which will not lose nor exploit the weaknesses of other strategies. Intransitive strategies, in contrast, are those which are likely to draw with optimal strategies, but are not optimal themselves. They tend to beat some strategies by huge margins but are in turn counter-able by some other strategies. For instance, strategy $(p_{11} = 1, p_{12} = 1, p_{13} = 1)$ will achieve the same expected payoff $E_1 (= -\frac{1}{9})$ as the optimal strategy $(p_{11} = \frac{1}{3}, p_{12} = 1, p_{13} = 1)$ if player 2 is also playing his optimal strategy $(p_{21} = 0, p_{22} = \frac{1}{3}, p_{23} = 1)$. However, if player 2 changes his strategy to $(p_{21} = 0, p_{22} = 1, p_{23} = 1)$, player 1's pay-off will become $E_1 = -\frac{1}{3}$, which is worse; indicating that player 1's strategy is counter-able in this case. Poor strategies are those that will lose to the optimal strategies and probably also those of other types as well. An example of such a strategy is given by $(p_{11} = 1, p_{12} = 0, p_{13} = 0)$. These observations will help us to identify key features

51

to look out for when designing the co-evolutionary model. The model will need to be able to eliminate all poor strategies and discern the optimal strategy from those equally competent but exploitable ones e.g. intransitive strategies.

3.4 *Designing the game engine*

Several objects are necessary for the design of a poker game engine. They are the card, deck, player, a poker game, player AI, odds calculator and a Graphic User Interface (GUI).

3.4.1 Basic game elements

The card is an object that consists of a face value (A, K, Q, J, 10, 9, 8, 7, 6, 5, 4, 3, 2) of type integer and a suit of self-defined type (♠, ♣, ♥, ♦). An integer called the value is also included which is an enumeration of the face value and suit. A deck consists of an array of card objects in a particular arrangement. The deck needs to supply the function to shuffle the deck and to deal a card from the top of the deck. The player is an object that contains a hand of two cards, the player's status (playing, folded, etc.), amount of money, fitness, name, A.I. type, history, etc. The player object needs to provide the functions to draw a card from the deck and to make a decision. A poker game object will handle all proceedings of the poker game. It contains a deck object, a number of player objects equal to the defined number of players, the pot amount, the bet amount and community cards. It needs to provide functions to start the round and to retrieve the winner of the round. Finally, a player AI object is a decision making model for the player. A player, if initialized with a particular AI object, will invoke a decision making function from its AI class whenever it needs to take an action. For an evolving AI, the AI

object will also need to receive feedback from the poker game so as to implement evolutionary procedures. Upon completion of the design for the various elements, the program was coded with Microsoft Visual Studio 2005 Express Edition.

3.4.2 The odds calculator

The odds calculator is an important component for the implementation of AI as far as poker is concerned. It is used to calculate a player's chances of winning if the player is to reach showdown stage with the current objective information available. This encompasses information about the cards in the player's hand, community cards, game stage and number of players. The calculator will transform the above information into a probability value from 0 to 1, whose magnitude represents the player's chance of winning. With such means, a computer player will be able to interpret pieces of complex information. Due to its high usage, it is crucial to write an efficient odds calculator.

In terms of the actual implementation, a separate odds calculator is written for each stage of the game. This is due to different number of community cards at different stages of the game. In the Preflop stage, online calculation of odds is very intensive as very few cards are revealed. This would also mean that there are as much as 169 possible combinations which each player could have. Calculation through the enumeration technique is thus performed outside the program and the results are then hard-coded. When a player calls the Preflop odds calculator, a binary search is performed to find the odds corresponding to its hand cards from the table of odds. In the Flop stage, it is difficult to attain optimum memory-speed trade-off as pure online calculation is too slow and pure look-up table is too large. A mixture of the two techniques is used. For the turn and river stages, pure online calculation is used due to relatively faster computation.

3.4.3 Graphical User Interface

A GUI was also developed for the game engine. Its primary purpose is to test and debug the program. As poker is a game that is visual in nature, having a GUI also makes it much simpler to detect and debug errors. When the program first starts, it will appear as in Figure 3.4, with three buttons - run, pause and step. Initially, pause is disabled. When run is clicked, the simulation starts and proceeds without interruption. Once the simulation is running, pause becomes clickable. If pause is clicked, simulation pauses (Figure 3.5) and the current game state e.g. generation, round of this generation, cards etc is displayed. The step button can be clicked when the simulation is paused to advance the simulation by one event, such as when the player performs the action "call". With the completion of GUI, extensive testing was done on the game engine to ensure that it works correctly, efficiently and without errors using debugger program provided in Visual C++.
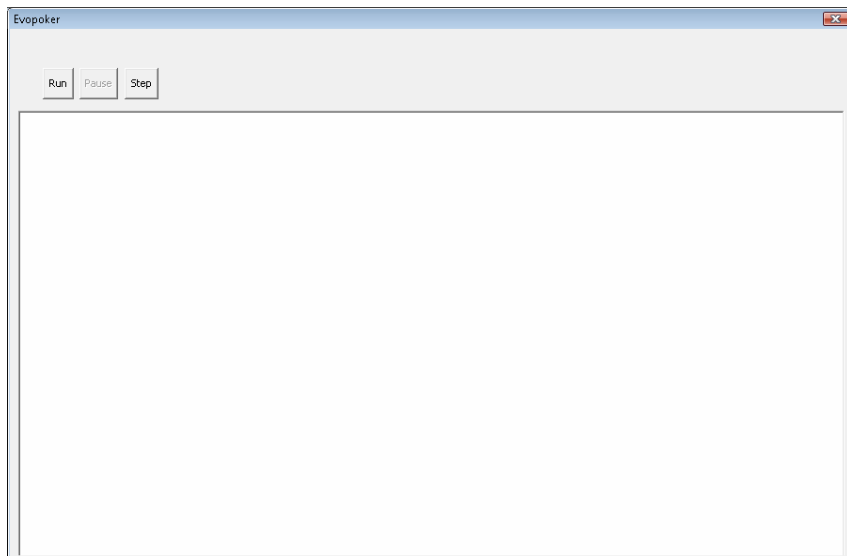
Figure 3.4: Initial state of the GUI
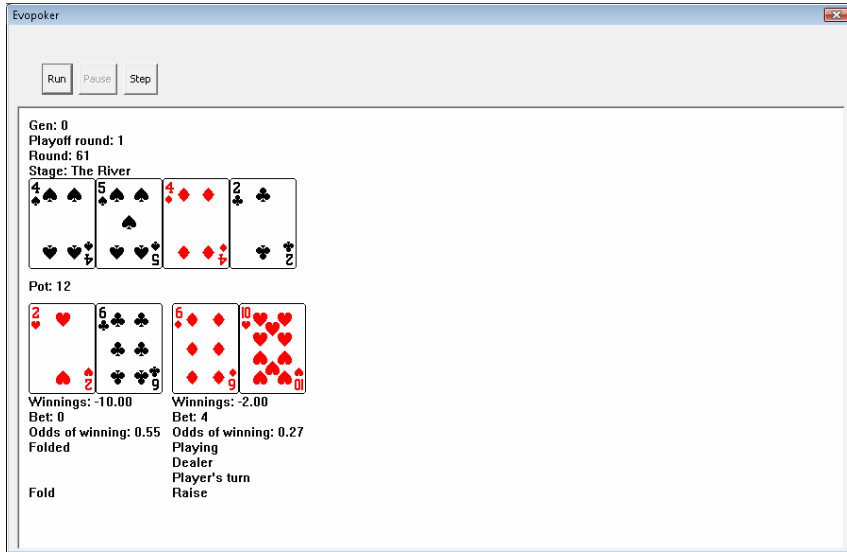
Figure 3.5: The GUI at a paused simulation

3.5   *The co-evolutionary model*

With game theory providing the necessary guidelines, a co-evolutionary model –

one where all the candidates play against and reproduce with one another after

every round-robin tournament is formulated to evolve strategies that play close to

the NE. In a round of tournament, each candidate in the population of 100 will

play against every other for 100 rounds of poker so that individual fitness can be

evaluated. Though a mere 100 rounds is considered small for eliminating the luck

element in a game of chance like poker, a necessary trade-off is needed to reduce

overall computational complexity. After fitness assessment, candidates are sorted

according to their fitness levels. The top 10% are cloned and replicated in the next

generation (Elitism). The remaining population is filled up with off-spring created

from the current generation via a sequential process of selection, crossover and

mutation. Selection is done using tournament selection, where pairs of candidates

are randomly picked from the parent population of which the fitter in each pair is

chosen to reproduce. From the pool of selected individuals, genetic variation is

introduced. In crossover, two randomly selected individuals will exchange traits such that the off-spring's genes – which comprises of a number pair {fold, raise} thresholds, will be chosen from one of its parents with a 50-50 probability. After crossover, each gene is mutated with a 20% probability. If mutation does occur, a normally distributed random number with mean equal to the original value and standard deviation equal to 0.1 will be generated to replace the old raise and fold threshold values of that gene. A generation is deemed to have elapsed in the evolution sense whenever a new population of offspring is formed.

3.5.1 Strategy model and chromosomal representation

In Section 3.3, it is known from game theory that the history of players' actions and their cards are two crucial pieces of information supplied to candidates. The ability to process such information becomes imperative for the candidates to make effective decisions. As both information types can well assume numerous values, it is practically impossible to consider all possibilities. Abstractions of information will have to be used instead. In the design of such abstractions, the ease of human interpretation is to be taken into consideration as well.

To abstract the card combinations, the hand strength (HS) which reflects the likelihood of winning with the cards on hand is used. This ranges from 0 to 1 and is computed using the odds calculator as described earlier. In contrast to HS, the history of actions is a very complex piece of information, which represents the sequence of events from the start of game to a player's turn, right down to every single detail. To make appropriate abstraction, some standard poker information such as the player's position in the game, total raise (TR) and the fraction of raise made by the opponent e.g. opponent's raise (OR) are used. As the poker variant used is for two players, a player's position is not extremely vital and is discarded

to reduce the data size. Conversely, both TR and OR abstract, to a certain extent, information on the branches of the game tree that the game is currently moving on. Moreover, both pieces of information are also fairly interpretable by humans. In typical games played amongst human players, TR actually determines the pot size. The larger the pot, the more likely players will call than raise. OR can be used to determine how confident the opponent is of his chance of winning. The higher the value, the more likely a player should fold. Via the above information abstractions, a strategy model can then be formulated (Figures 3.6 and 3.7).
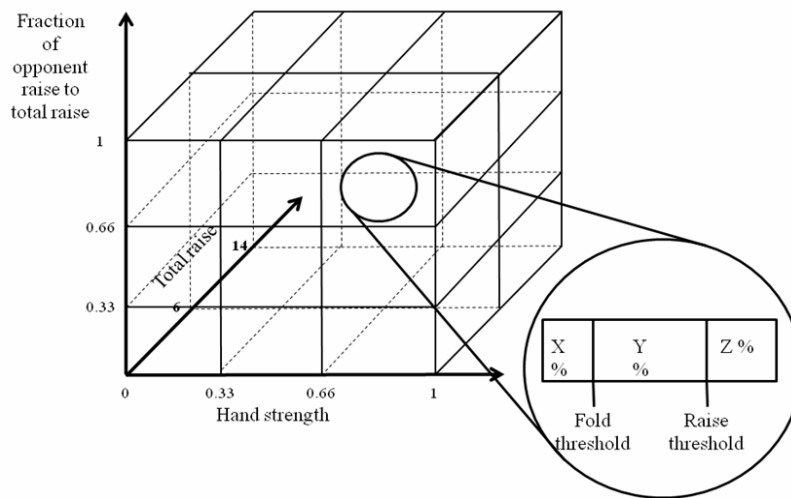
Figure 3.6: Strategy structure for Preflop/Flop
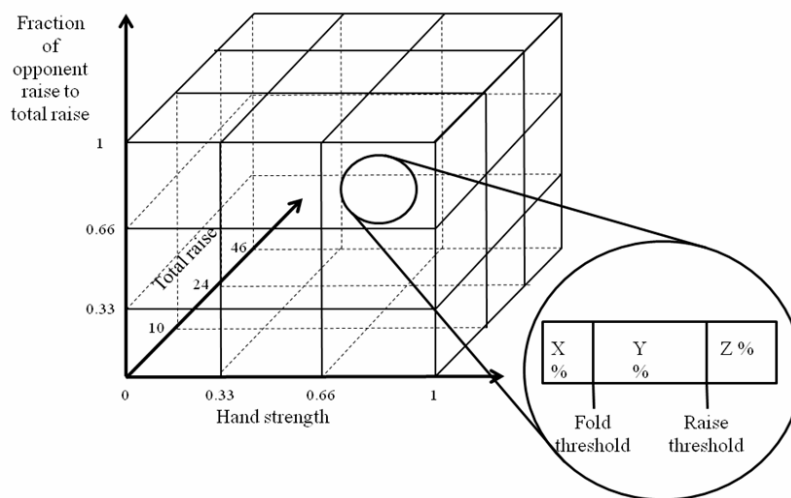
Figure 3.7: Strategy structure for Turn/River

57

Multi-dimensional arrays are used to represent the structure of strategies in the model. HS and OR are divided into three equal intervals - low, medium and high. As the raise amount during Preflop and Flop (e.g. $2) is different from that during the Turn and River (e.g. $4), distinct structures are used. Intervals for TR are not evenly distributed as TR is more often low than high in a game of poker. The intervals are thus made smaller at the low end but greater at the high end to ensure that all slots will be looked-up in a more evenly fashion. TR is divided into two intervals - low and high in Preflop and Flop stages; and three intervals - low, medium and high for the postflop stages. Decision making is based on probability triplets in order to implement a similar optimal strategy as discussed in Section 3.3. In each slot, there are two numbers which represent the fold and raise thresholds respectively, of which, the prior is always smaller than the latter. In totality, the array size for Preflop and Flop is 36 and that for Turn and River is 54.

Whenever a candidate is required to make any decision, it looks up his 3-D strategy table for the slot that contains the intervals which matches its HS, OR and TR. It then generates a random number from the uniform distribution U [0,1] and compare it with the fold and raise thresholds. The candidate folds if the number is smaller than the fold threshold, calls if it is between both thresholds and raises if otherwise. All fold and raise threshold values are randomly initialized at the start of simulation and subjected to changes during the course of evolution.

3.5.2 Fitness criterion

A fitness criterion is proposed to evaluate a candidate's goodness using guidelines from the game theoretic analysis in Section 3.3. Every candidate starts off with zero fitness and after every 100 rounds of game play between a candidate pair, fitness of each candidate in the pair is updated. The candidate who lost will have

its fitness reduced while the one who won will have its fitness left unchanged. Reduction in the loser's fitness is set to be the square of the amount of money it loses in the game. This is mathematically expressed in equation (3.9).

$$F_i = - \sum_{j=1}^{N} \left[U(-W_{ij}) \right]^2 \qquad (3.9)$$

where F_i is the fitness of candidate i,

$\quad$ N is the total number of candidates,

$\quad$ W_{ij} is the money won by candidate i against candidate j and

$$U(x) = \begin{cases} x, & x \geq 0 \\ 0, & otherwise \end{cases} \qquad (3.10)$$

In this way, conditions necessary to achieve NE could be satisfied. Though an optimal player will not lose and has practically no weakness, it could well still lose within a mere span of 100 rounds owing to bad luck, though not by much. To distinguish these players from those that will lose exceptionally heavily to certain strategies, square of the money lost, rather than just the money lost is deducted if a player loses. This ensures that candidates with weaknesses are penalized heavily whilst those who lost due to bad hands are not penalized as much. In conjunction, as the optimal strategy is not meant to be a counter to any specific strategy, the amount of money that the winner wins is not added to its fitness. This prevents a player who is only good in beating certain players from having its fitness pumped up when it meets opponents that are vulnerable to exploitation by its strategy.

This above is best verified by an example scenario. Let A denotes candidates that play with near-optimal strategies. Let B denotes candidates that draw with A but are very good against some other candidates and also has weaknesses against

others. Let C refers to players that are simply poor. In a population consisting of A, B and C, A will beat C and draw with B such that candidates of A will have fitness approximately equal to zero. Although B will beat C and also draw with A, some candidates from B will exploit others from B as well. Though the exploiter wins lots of money, its fitness is not increased as winnings are not added. The exploited, however, suffer a heavy drop in fitness due to a reduction in the square magnitude. Overall, fitness in order of the highest to lowest will be A, B and C.

3.6   *Preliminary study*

In this section, a preliminary study is conducted to verify the correctness of the proposed co-evolutionary model. In particular, the model is adjusted and applied to the simplified poker variant that is defined in Section 3.3. At the same time, several fitness criteria are also tested in order to determine the one which is most suitable for obtaining the NE.

3.6.1 Strategy model for simplified poker

The strategy model for the preliminary study consists of a two dimensional array, with one dimension denoting the card (1, 2 or 3) and the other representing the position (1st or 2nd). Inside each element, there is a real number between 0 and 1 denoting the fold threshold (Figure 3.8). Whenever a player makes a decision, a random number is generated in the range 0 to 1. The player will fold if the number is smaller than the fold threshold and call if otherwise. With tournament and co-evolutionary settings kept unchanged, several distinct fitness criteria are tested to examine their effects on the behavior of strategies which emerged. Comparison is done to ascertain the criterion that is most suitable to obtain the NE strategy.

Card value

	1	2	3
1	f_{11}	f_{12}	f_{13}
2	f_{21}	f_{22}	f_{23}

Position

Fold threshold

Figure 3.8: Strategy array of the strategy model for the simplified poker

3.6.2 Fitness criterion equivalent to winnings

The first fitness criterion to be tested is one where the winnings of a poker player correlate positively with its fitness level. This is expressed mathematically as

$$F_i = \sum_{j=1}^{N} W_{ij} \tag{3.11}$$

With this criterion, it is hypothesized that intransitivity will play a major role as a player with higher winnings will most likely be one who is able to counter the strategies of others. The experiment was carried out with a population of 100 and the following results are obtained after 500 generations (Figures 3.9 and 3.10). The figures depict plots of fold thresholds of the winners in each generation. As expected, f_{12} and f_{13} are 0, implying that a player holding card 2 or 3 at position 1 should call with probability 1. In accordance to theoretical calculations in Section 3.3, f_{21} and f_{23} are also found to be 1 and 0 respectively.
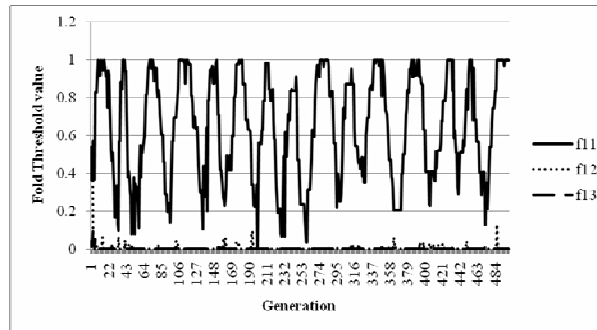
Figure 3.9: Plot of fold thresholds of winner in each generation for position 1, fitness criterion 1.
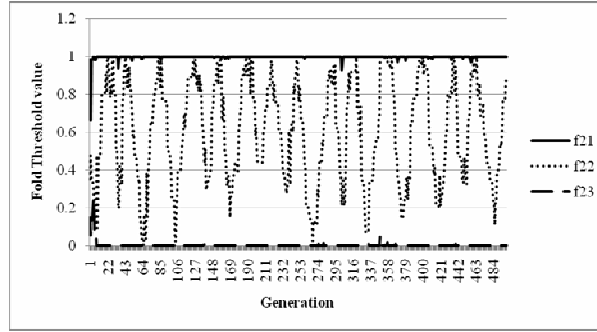
Figure 3.10: Plot of fold thresholds of winner in each generation for position 2, fitness criterion 1.

However, the values of interest e.g. f_{11} and f_{22} tend to exhibit fluctuating behavior as shown in Figure 3.11. From comparison, it is observed that both plots track one another closely. As f_{11} increases, f_{22} also increases several generations later, and as f_{11} decreases, f_{22} decreases likewise within the next few generations. This highlights the intransitive nature of poker. The average variance of f_{11} from generation 400 to 500 is 0.065341 while that of f_{22} is 0.066685. Also of interest is the mean of fluctuations - 0.66829 for f_{11} and 0.59689 and f_{22}. These values are rather close to the calculated optimal strategy of 0.6666 and 0.6666 respectively.
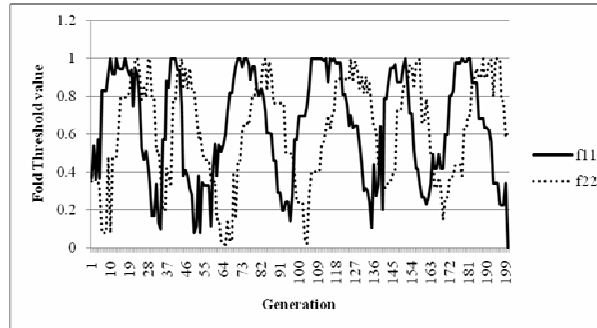
Figure 3.11: Comparison of plots of f_{11} and f_{22}.

3.6.3 Fitness criterion excluding winnings and deducting the squares of losses

The second fitness criterion to be tested is the one originally proposed during the design of the co-evolutionary model, that is:

$$F_i = -\sum_{j=1}^{N}\left[U(-W_{ij})\right]^2 \text{ where } U(x) = \begin{cases} x, & x \geq 0 \\ 0, otherwise \end{cases} \qquad (3.12)$$

The results, after 500 generations are shown in Figures 3.12 and 3.13. The values of f_{12}, f_{13}, f_{21} and f_{23} are similar to those of the previous fitness criterion. Although signs of intransitivity are still observed, the fluctuations are of smaller magnitude this time. The average variances of f_{11} and f_{22} from generation 400 to 500 are also smaller at 0.037467 and 0.034844, implying that this fitness criterion does reduce the intransitivity element of the co-evolutionary process. Mean of f_{11} is 0.67425 and that of f_{22} is 0.70171 in the same period of consideration.
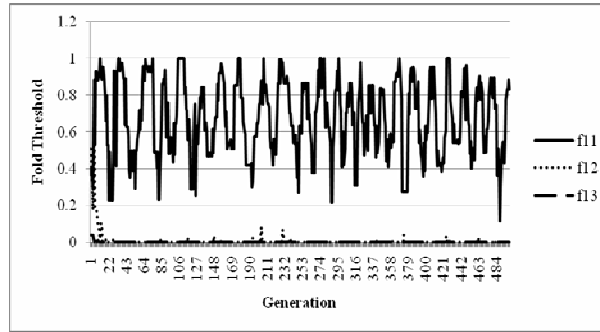
Figure 3.12: Plot of fold thresholds of winner in each generation for position 1, fitness criterion 2.
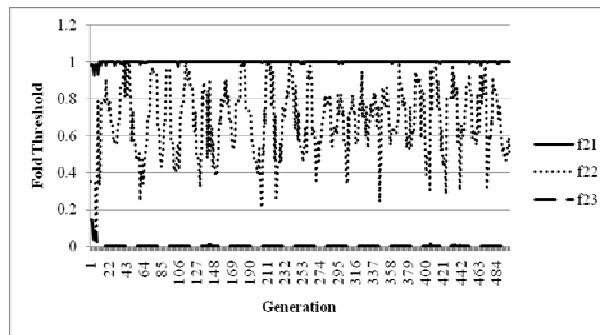
Figure 3.13: Plot of fold thresholds of winner in each generation for position 2, fitness criterion 2.

3.6.4 Fitness criterion with higher power

Due to encouraging signs from the previous fitness criterion, one of even higher power e.g. 20 is experimented. The fitness level equation is expressed as

$$F_i = - \sum_{j=1}^{N} \left[U(-W_{ij}) \right]^{20} \tag{3.13}$$

Results after 500 generations are shown in Figures 3.14 and 3.15. The magnitude of fluctuations is further reduced, but only insignificantly. Average variances of f_{11} and f_{22} are 0.028717 and 0.024055. Their means are 0.67205 and 0.73729.
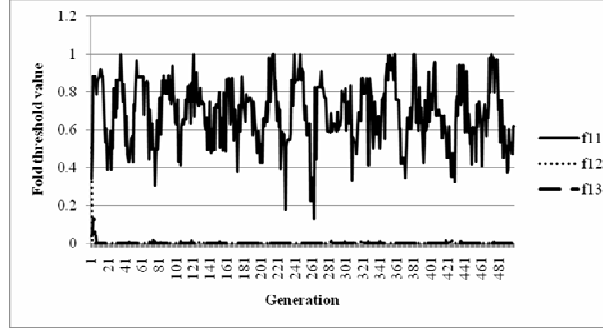
Figure 3.14: Plot of fold thresholds of winner in each generation for position 1, fitness criterion 3.
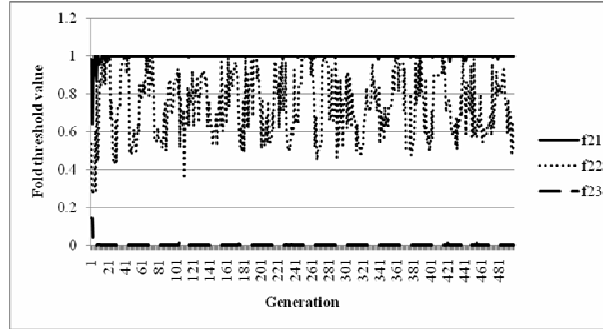
Figure 3.15: Plot of fold thresholds of winner in each generation for position 2, fitness criterion 3.

3.6.5 Discussion on preliminary findings

As far as reduction of intransitivity in the co-evolutionary process is of concerned, it is found that the criterion where the fitness level is determined by the power of losses is better than the criterion where fitness is equated with winnings. On a side note, it is also found that higher power leads to greater reduction in the fluctuation magnitude. However, this reduction is inconsequential if compared to the amount of complexity that is introduced. Considering all factors, the originally proposed

fitness criterion with the power of two will be used for the subsequent simulation studies. Finally, it can be deduced that it is still possible to obtain optimal values for f_{11} and f_{22} by finding the average of fluctuations after many generations.

3.7   Simulation results

Upon confirmation from preliminary studies, the experiment is conducted on full scale Texas Hold'em using the game engine and co-evolutionary model defined as before. 271 generations are simulated on a shared server with two Xeon dual-core 3.0GHz processors and 8GB memory. An attempt is made to actualize the Nash optimal strategy by averaging all winner strategies in the last 100 generations. The analysis of behavioral outcomes is presented in the ensuing subsections.

3.7.1 Verification of results

To verify the functionality of CEAs, some straight forward results are examined. From Figure 3.16, it can be observed that thresholds for high OR, low TR and low HS increases as the generation advances, indicating that it is best to fold in these situations, which is expected. If OR is high, the opponent is confident of winning. If TR is low, the reward for taking the risk of betting is poor. When HS is low, the chance of winning is bad. Combining all factors, we observe that CEA is accurate by folding in this situation. With high HS and zero TR, both thresholds decrease as the generation advances (Figure 3.17). This indicates that it is always desirable to raise in this situation. This is logically sound as a player with good HS would want to maximize its own winnings. With zero TR, the winning is very little and can only be maximized by raising the bet.
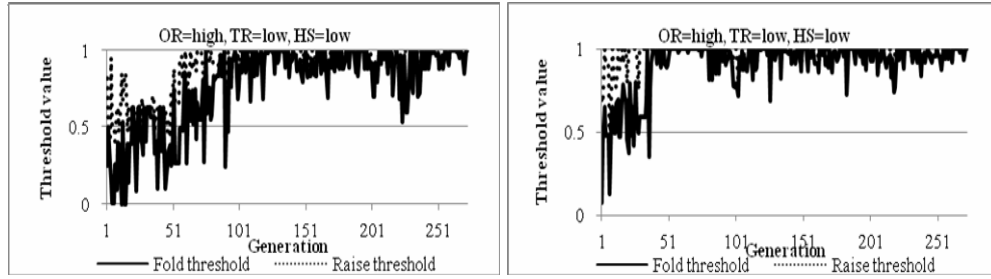
Figure 3.16: Plot of fold and raise thresholds against generation when "Opponent Raise is high, Total raise is low and Hand strength is low" for Preflop/Flop (left) and Turn/River (right).
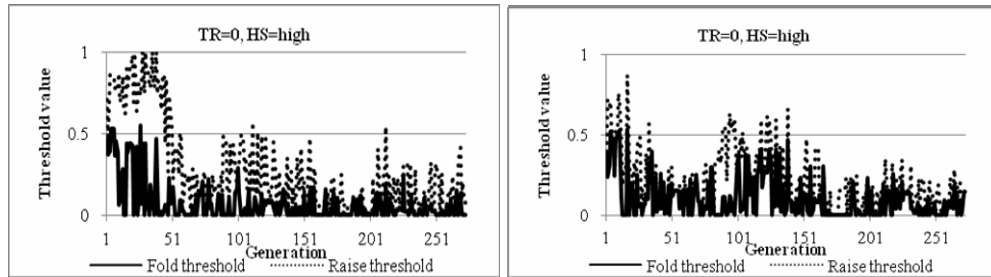
Figure 3.17: Plot of fold and raise threshold against generation when "Total raise is 0 and Hand strength is high" for Preflop/Flop (left) and Turn/River (right).

When TR is high, a player would want to call even if its chance of winning is low, as the amount that it could potentially win is worth the risk. However, the player would not want to raise in this situation to avoid losing even more. The strategy which is evolved by CEA also derives this accurately as seen from plots in Figure 3.18 e.g. a high raise threshold coupled with a low fold threshold signify that the player should call. It is to be noted that fluctuations can also be seen. This is most likely due to the uncertainty in the nature of this risk and the intransitivity factor.
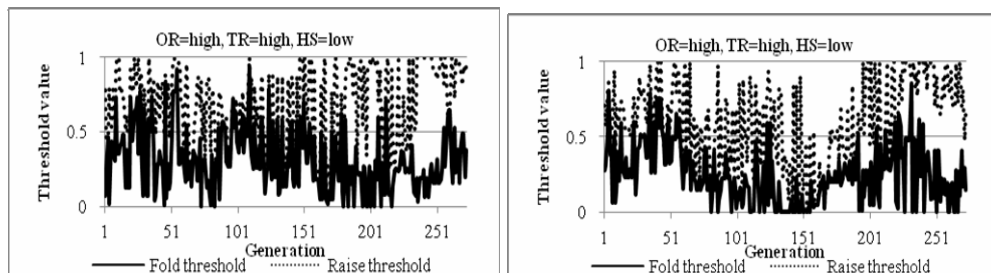
Figure 3.18: Plot of fold and raise threshold against generation when "High opponent raise, high total raise and low hand strength" for Preflop/Flop (left) and Turn/River (right).

66

The above results certify the credibility of the co-evolutionary model; that the population gets better as generation elapses. The next subsection will attempt to analyze and explore insights of the evolved CEA strategy.

3.7.2 Analysis of the evolved CEA strategy

After 271 generations, a final strategy is determined by finding mean thresholds of winners in the last 100 generations. This evolved strategy is then analyzed. When TR is 0 e.g. a player first starts off a round, his decision is made primarily using the HS information. As of Figure 3.19, the strategy proposes unequivocal usage of fold and raise for low and high HS respectively in all stages - as shown by similar fold and raise threshold values. For a medium HS, differences in threshold values indicate a lower tendency to call during Preflop/Flop as compared to Turn/River.
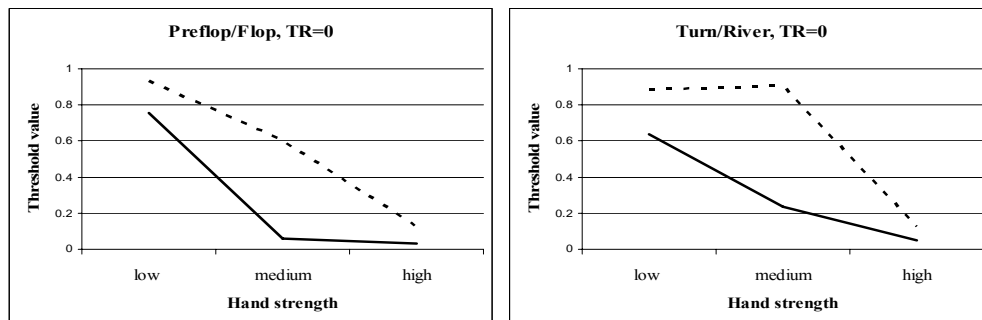
Figure 3.19: Plot of threshold value against hand strength for Preflop/Flop (left) and Turn/River (right). Dotted line: raise threshold. Solid line: fold threshold.

In the Preflop/Flop, the player alternates between calling and raise periodically since the probability of winning or acquiring a high HS in the subsequent stages, conditioned on a medium HS, is rather high. Such behavior could be spurred by a desire to boost the value of the empty pot. In contrast, winning in the Turn/River is mostly conditioned on having a high HS as much of the game would have been decided - most community cards are revealed by then. Based on independent use of HS information, it is intuitive that a player with medium HS and diminished

chance of winning should call if not fold, rather than raise, especially if the worth of the pot is not worth the effort to raise the bet

Apart from the above, a myriad of other scenarios with various levels {low, med, high} of OR and TR are also explored in both the Preflop/Flop (Figures 3.20 to 3.23) and Turn/River (Figures 3.24 to 3.29) stages. Strategy plots of threshold values against HS are shown below. A particular setting e.g. "OR is low and TR is med" is represented by a unique pair of lines where the higher and lower lines denote the raise and fold thresholds respectively. Pairs of lines that correspond to different settings are distinguished by dotted, solid or dashed lines.

From the collection of figures, it is observed that the addition of information like OR and TR to HS in the decision making process allows the CEA to evolve a multitude of strategy variants that exhibit much greater complexity when it comes to deciding whether to fold, call or raise in different situations. Unlike the case where TR is 0, it is no longer simply about folding if HS is low and raising if HS is high. Nonetheless, certain traits do remain unchanged in the considerably more sophisticated strategies. For instance, fold thresholds in all scenarios consistently display a decreasing trend as HS improves, indicating that the evolved strategy invariably folds less often as long as it acquires better chances of winning in any fixed scenario. The raise thresholds, on the other hand, undergo erratic variations across different scenarios e.g. the improvement in HS does not always entail a progressive decline in the raise threshold (Figure 3.23) and are apparently not correlated to any one decision information. Such irregularities would suggest that raising in poker is a decision that is complex to make in nature. Further insights into the behavioral aspects of the evolved strategy can be gained by analyzing its traits under different game stages.

3.7.2.1 Preflop/Flop strategies

It is observed that the strategy almost never folds as long as it has acquired at least a medium HS during Preflop/Flop, regardless of OR and TR. This "Call and see" nature – as a majority of scenarios in Preflop/Flop proposes calling with a medium HS, spans from great optimism towards a possibility of achieving an even better hand in future stages. Such view is unshaken even in disadvantageous situations where the opponent is perceived as having better chances of winning or when the pot size is just too low for fruitful contention. For low TR, strategies across all OR values generally behave in tandem to HS, similarly to the results observed when TR is 0 - {fold, call, raise} for {low, med, high} HS respectively. The only minor deviation occurs when OR is low - since the opponent is perceived to have a low chance of winning (Figure 3.22). Given a low HS, the evolved strategy justifies an action to call instead of fold as there is an equally probable chance of winning.

For high TR, the evolved strategy almost never folds in the face of a pot with potentially huge winnings. This is true across all possible combinations of HS and OR values. Another interesting point which is observable from the raise threshold is bluffing. For med OR, the evolved strategy tends to raise very often even if HS is low (Figure 3.23) - indicative of an attempt to bluff. However, such behavior is absent for the equivalent scenario when TR is low, since only a high TR warrants justification for taking the calculated risk of attempting to "scare" the opponent into folding with a raise. For high OR, the tendency to raise is greatly diminished at low HS in view of the larger perceived disparity in HS between the player and opponent as compared to med OR. From another perspective, the high possibility that the opponent will match up to any potential raise, owing to high confidence in his cards also deters the player from raising and risking a likelihood

of incurring more loses by doing so. In view of the large pot value, the player adopts a "Call and see" attitude instead of fold. The raise behavior is nonetheless shifted rightwards instead and exhibited for med HS. This signifies that the player is willing to adopt a bluff strategy only if his HS is not perceived to be too far off from his opponent's. This will at least give the player a fair chance of winning in the event that the opponent did not fall for the bluff. When the player has high HS, the proposed strategy is to call and follow the opponent's bet without signaling his confidence or weakness by raising or folding. This is similar for a player with med HS under med OR. The underlying stance is to get the opponent into subsequent stages before deciding whether to challenge him aggressively to the game.

From Figures 3.20 and 3.21, it is observed that fold thresholds are generally lower if TR is higher, indicating that a player folds less frequent given a larger pot of potential winnings. No obvious relation of OR with the thresholds is observable when TR is low as bluff is triggered only if TR is high. This suggests that TR is more dominant than OR as a factor for triggering bluff. Once bluff is in place, it is then OR which is seen to affect the actual position in which bluff is performed.

Figure 3.20: Plots of thresholds against hand strength for Preflop/Flop and medium OR.

Figure 3.21: Plots of thresholds against hand strength for Preflop/Flop and high OR.

Figure 3.22: Plots of thresholds against hand strength for Preflop/Flop and low TR.

Figure 3.23: Plots of thresholds against hand strength for Preflop/Flop and high TR.

3.7.2.2 Turn/River strategies

As the game proceeds into the Turn/River, the evolved strategies generally exhibit

behavioral traits which are rather different from those in the Preflop/Flop. Such

change is largely due to the huge reduction in the game's overall uncertainty since

the opening of three community cards after the Flop and an additional one on the

Turn. It is clear that the prior conviction of not folding with at least a medium HS

remains true only when TR is at least med during the "Turn/River". Unlike the Preflop/Flop, however, a high HS is now strictly required to pursue a strategy of non-folding that is independent of OR and TR; since it gets increasingly harder for a player to alter his chances of winning when only one or no card is left to be dealt. As observed, HS is no longer the sole factor that affects decision making if there is only a fair chance of winning e.g. med HS. In such cases, it would make ample sense to proceed on with the game only when the rewards of the venture are worth the risks which the player has to undertake.

Apart from TR, the decision making process in the Turn/River is also greatly affected by OR. As seen, the strategy almost never folds if OR is low, regardless of HS and TR (Figure 3.24). As it is improbable that the player's hand quality will undergo major changes upon uncovering another community card, the likelihood of winning is largely determined by the opponent's HS. This is inferred indirectly from the OR information which reflects the opponent's confidence and chances of winning. A low OR signifies a heightened chance to win irrespective of a player's current HS, which justify a strong tendency to carry on with the game. Another interesting behavior which is identified is the fact that the strategy always raises strongly if "OR is high and HS is high", regardless of TR (Figure 3.26). This is different from Preflop/Flop where the strategy will tend to bluff by calling in the same scenario if TR is high. Given that the showdown stage is drawing nearer, the evolved strategy finds it beneficial to signal his true HS rather than concealing it. The reason for such a move is that the strategy would want to raise aggressively as a final attempt to deter the opponent with high OR from continuing the game. This move might at the same time also trick an opponent who is bluffing with a high raising history into folding.

When TR is low, the player's decision is equally affected by both HS and OR, unlike the equivalent scenario in Preflop/Flop where HS exerts predominant effect. In general but with exception of "OR is med and HS is med", the evolved strategy tends to raise if the player's HS is higher or on par with OR and fold if otherwise. This shows that a relative comparison of the winning chances is crucial for decision making. As seen from a tendency to call under equivalent scenarios of med and high TR, the exception is due to the player's unwillingness to risk losing more over a pot with low potential winnings. Exploring further, the strategy raises very often when OR is low. This presents an interesting emergent behavior which depicts that the strategy has a tendency to raise as long as the opponent's chances of winning is perceived to be low (Figure 3.27). Whilst the desire to raise for med or high HS is justified as an action to boost the low pot value and acquire more winnings from the relatively weaker opponent, the same action clearly also has an ubiquitous element of bluff in the case of a low HS. By raising on a low HS, the player in fact tries to conceal his weak HS position by creating a confident image, in the hope of misleading the weak opponent to believe that the player has a good HS. The opponent, who in fact has comparable HS, might just be tricked to fold on account of the perceived image and a low pot size that is not worth to vie for.

However, the CEA does not find it desirable to attempt such bluff for med and high OR due to anticipation that the opponent might easily match the raise. This is shown by the large tendency to fold. Apart from low OR, the strategy also raises very often when HS is high (Figure 3.27). The motivation behind such a raise is however very different from bluffing as the player is actually revealing his true position and relative confidence indirectly and using it as the basis to scare the opponent into folding. Overall, there are considerably more raises when TR is

low as the strategy figures that it is not likely to lose much from raising, judging from the low contributions which it has made to the pot thus far. In addition, the potential winnings can certainly be increased by raising.

When TR is med, the tendency to raise when OR is low is weakened and the strategy calls more frequently in anticipation that the opponent will be tempted to match up to any raise on account of the higher pot value. Though the dominant strategy is still to fold when HS is low for OR is med or high (as in TR is low), the proposed action for a combination of HS is med and the same OR values is to call. In response to a higher pot size, the strategy is now willing to call more frequently for situations where its HS is on par ("OR is med and HS is med") or even lower ("OR is high and HS is med") than its opponent's perceived chance of winning. With an even higher TR (e.g. TR is high), the temptation of calling against an opponent with higher chances of winning is further extended at HS is med, turning the tendency to call into a raise. Although the strategy still raises often when HS is high, another situation of bluff is detected when OR is med and HS is high (Figure 3.28). As opposed to the previous bluff position when "TR is low, HS is low and OR is low", the strategy attempts to conceal its high HS by simply calling. The idea is to lead the opponent with OR is med into the subsequent stages or even the showdown, so that higher winnings can be reaped eventually. With TR increasing to high for high HS, such deceptive behavior of calling against a weaker opponent of med OR continues to dominate, and with even higher probability.

When TR is high, the proposed action is rather uncertain at HS is low, as seen from the close probabilities for fold, call and raise. This dilemma is probably due to conflicts between the rational action of folding against an opponent with higher OR on one hand and opposing action of calling on the game in view of the

74

large amount that has been contributed to the pot thus far. As opposed to the "OR is med and HS is low" scenario in Preflop/Flop where the temptation of a high TR induces the strategy to attempt bluff by raising, this is no longer so for Turn/River. This is because the opponent will tend to fold less often on account of the larger personal contribution and potential winnings that are at stake. However, consistent with Preflop/Flop, fold thresholds are generally lower if TR is high (Figure 3.29), indicating that a player folds less frequent given a larger pot of potential winnings.

On the whole, the Turn/River entails more complex strategy combinations for different scenarios as compared to Preflop/Flop. With more certainty revealed, strategies no longer adopt the "Call and see" approach" by postponing concrete decision making to the future but are more cautious in their decisions to fold, call, raise or even bluff. In combination, the pot's worth, opponent's perceived chances of winning, the player's HS etc all exert crucial impact on the decision making process as the game draws nearer to the showdown.
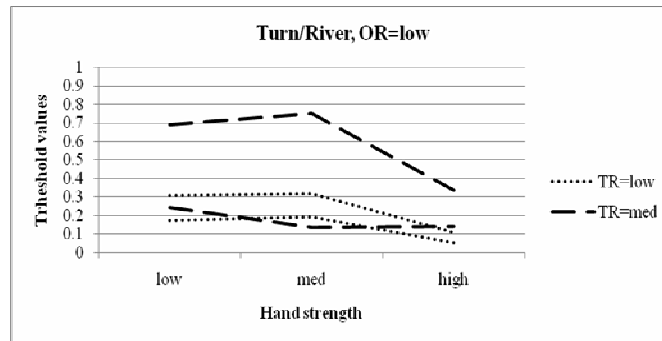
Figure 3.24: Plots of thresholds against hand strength for Turn/River and low OR.
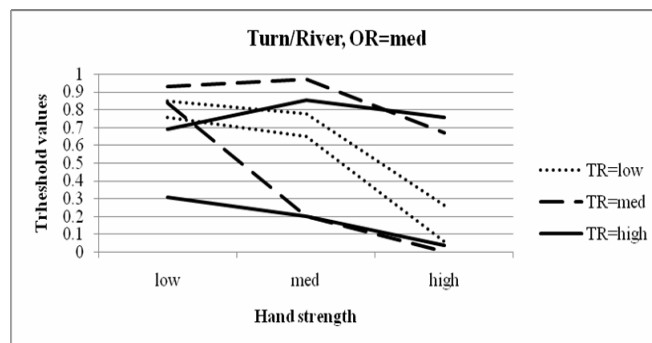
Figure 3.25: Plots of thresholds against hand strength for Turn/River and medium OR.
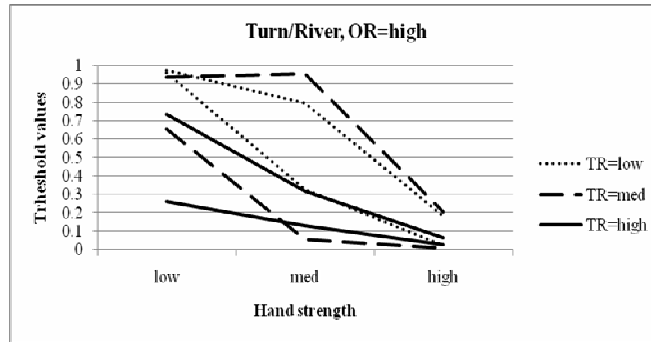
Figure 3.26: Plots of thresholds against hand strength for Turn/River and high OR.
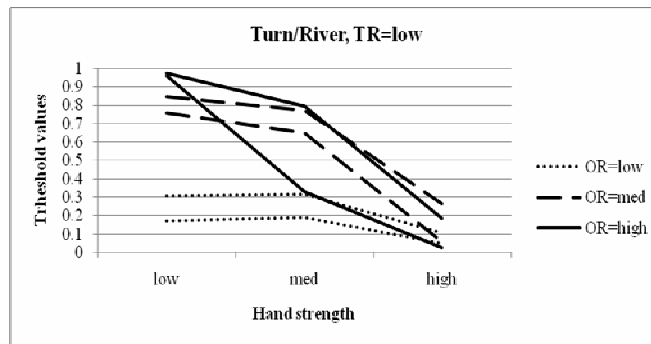
Figure 3.27: Plots of thresholds against hand strength for Turn/River and low TR.

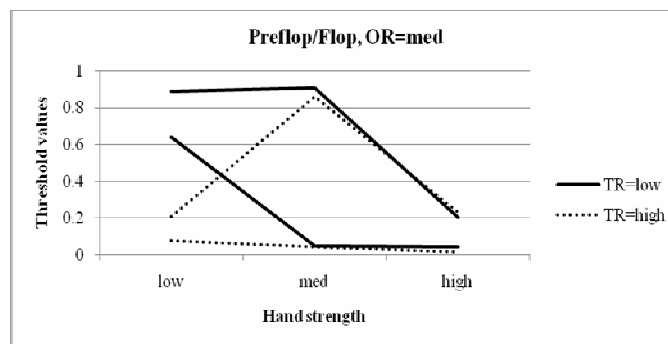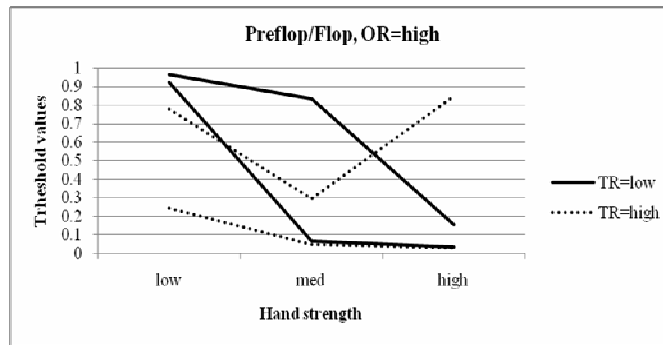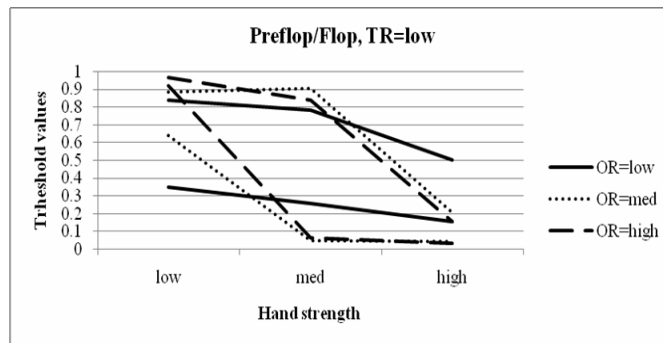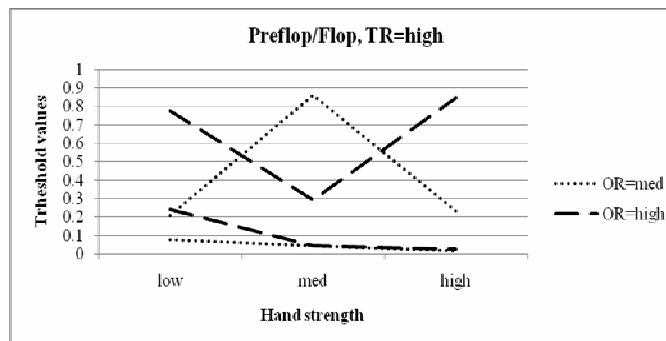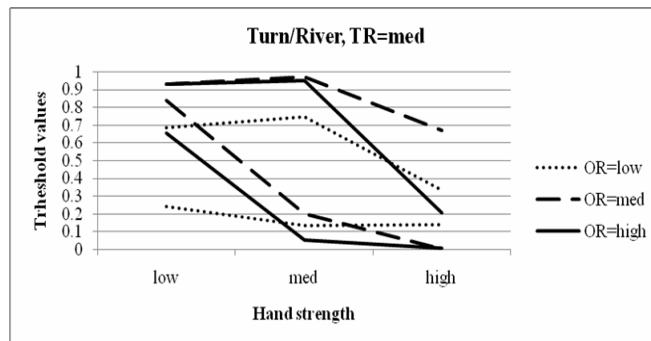Figure 3.28: Plots of thresholds against hand strength for Turn/River and medium TR.
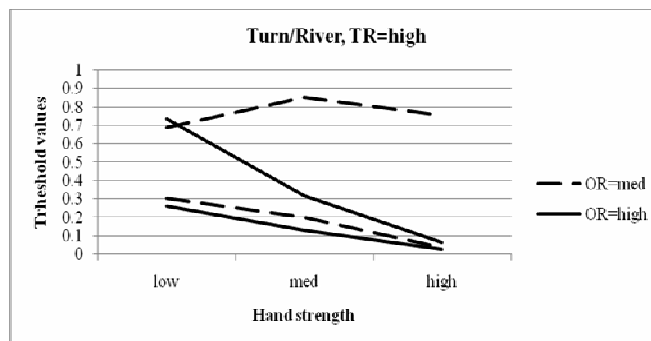
Figure 3.29: Plots of thresholds against hand strength for Turn/River and high TR.

76

3.7.3 Benchmarking

The strategies that were evolved by the CEA were benchmarked against the poker A.I.s from the University of Alberta, namely PSOpti and Poki. PSOpti is a poker playing agent which specializes in two-player Texas Hold'em. By formulating its strategy using a pseudo-optimal game-theoretic approach [107] that is non-exploitive, PSOpti plays close to NE. Poki, on the other hand, is an agent which specializes in multi-player Texas Hold'em and employs opponent modeling [101] – [103] during game play. As a means of comparison, the evolved CEA player (named Evobot) was setup to play against PSOpti for a game lasting 2000 rounds after every generation of evolution. Figure 3.30 shows the resultant winning trace of Evobot across generation. The unit of winnings used is the small bet per hand (sb/h), calculated by dividing the money won by the small bet amount, which is $2 in the program, by the number of rounds played e.g. 2000. From the plot, it is apparent that Evobot improves its overall performance and narrows the inter-strategy score margin as the generation advances. Relative to PSOpti, performance of Evobot is slightly lower owing to the mere 3 by 3 by 5 array that was used to represent the strategy chromosome. This is a foreseeable outcome as it is unlikely for any optimum strategy that surpasses PSOpti, if there is really one; to be fully represented within the bounds of the much constrained data structure. Attempts to improve performance by increasing the structure size is greatly hindered by huge computational space and time that are involved in the poker simulation. Even so, close eventual performance of Evobot to PSOpti is, nonetheless, an indicator that effective evolution is taking place via an efficient exploitation of strategy structure.
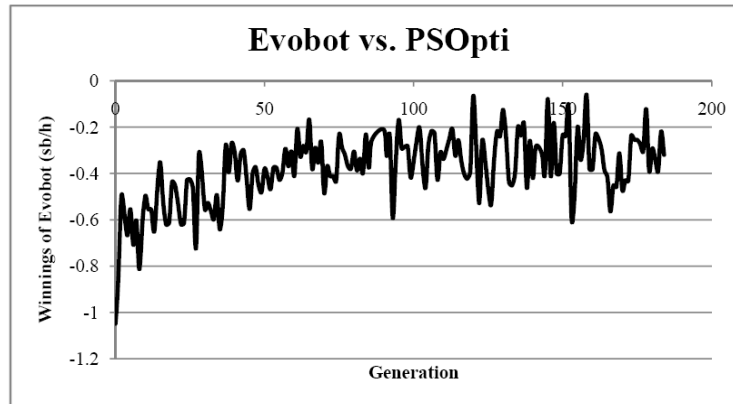
Figure 3.30: Winnings of Evobot vs. PSOpti against generation of evolution.

Evobot was also setup to play against Poki after every five generations, with each game lasting 4000 rounds. Despite higher starting losses (Figure 3.31), the performance of Evobot is almost on par and only trails behind slightly. As in the winnings trace against PSOpti, losses of Evobot decrease as generation advances. Results signify that CEA is able to evolve strategies which are similarly adaptable to players that specialize in two or multi-player Texas Hold'em alike.
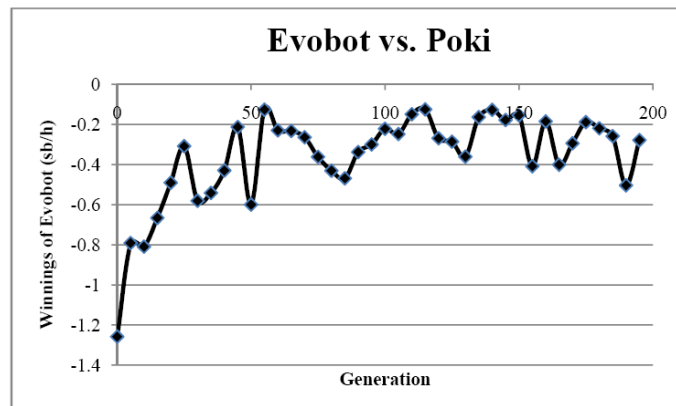
Figure 3.31: Winnings of Evobot vs. Poki against generation of evolution.

The evolved Nash optimal player is then played against both PSOpti and Poki for 10000 rounds each. Table 3.4 shows the overall winnings of the final average strategy. Though lower, the performance of Evobot is comparable to both its opponents. As a benchmark of comparison, it is known that a player who folds

78

every single hand will win by -0.75 sb/h. It is also found that a player who always call when playing against PSOpti and Poki respectively scores {-0.505, -0.537} sb/h while one who always raise scores a corresponding {-0.319, -2.285} sb/h [107]. In relative terms, Evobot's respective performance of {-0.2296, -0.1670} sb/h is significantly higher than the performance of these generic strategies. In fact, it is far from being bad when one considers the search limitations that are imposed by the constrained strategy structure on the CEA.

Table 3.4: Winnings of Evobot and several conventional strategies against PSOpti and Poki

	PSOpti	Poki
Evobot	-0.2296 sb/h	-0.1670 sb/h
Always Fold	-0.75 sb/h	-0.75 sb/h
Always Call	-0.505 sb/h	-0.537 sb/h
Always Raise	-0.319 sb/h	-2.285 sb/h

3.7.4 Efficiency

Figure 3.32 shows the plot of the time taken against generation. At the start of simulation, the time taken to complete all the games in one generation is relatively small. This is largely due to the random strategies that the candidates tend to adopt. As more generations elapse, the candidates start to adopt better strategies which inevitably cause a typical game to last much longer. The time taken per generation eventually stabilizes at around 9000 seconds. At this rate, it takes approximately 27 days to reach 271 generations where the evolution process stabilizes. This constitutes a limitation as to why a large data structure is not used to represent strategies. Despite the fairly long process time that is taken to evolve competent strategies which is typical for games of such nature, it is to be noted that no expert knowledge e.g. opponent modeling, is injected at all throughout the entire process of evolution. This is perhaps one aspect that the CEA can value-add to the existing methods of training good poker players.

79
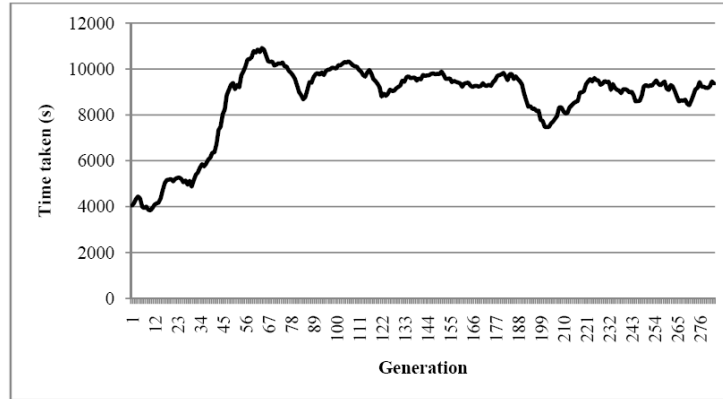
Figure 3.32: Plot of time taken against generation

3.8 Summary

This chapter demonstrated the possibility of applying CEAs to the development of a competitive computer poker player that specialized in Texas Hold'em. Game theory was first applied to analyze a simplified version of the game. Knowledge gained from the analysis was used as guidelines to design a co-evolutionary model for the purpose of achieving strategies that play at NE. From analysis, the player that was evolved by CEA not only displayed strategies that are logical, but also reveal insights that are not easily comprehensible. Some of these insights include bluffing indicators. An attempt to attain the Nash optimal strategy was made by finding the average of strategies from generation 172 to 271. This strategy, named Evobot, was benchmarked against existing poker A.I.s PSOpti and Poki. Despite the much constrained representation for the strategy chromosome, differences in score margins between Evobot and the opponents were low, signifying that CEAs are good at exploiting the structure of the problem to attain near NE solutions. Though CEA tends to take a fairly long time to evolve a stable strategy, no expert knowledge is required at all throughout the entire evolution process. CEA is able to adapt and develop good strategies by simply playing continuously over time.

Chapter 4

Adaptation of IPD strategies

Apart from its application in games like poker, game theory is widely used to study interaction in many social contexts. The Iterated Prisoner's Dilemma (IPD) [4], in particular, is an abstract mathematical game that is widely used to model numerous aspects of behavioral interaction in reality, when conflicts of interest arise between two or more groups of entities. Though conceptually simple, this classical problem in game theory presents a useful tool to study human behavior in various social settings and has contributed insights to areas like engineering, science, economics, analysis of social network structures [113] and psychology.

To this note, CEAs, apart from the ability to evolve poker strategies which perform close to Nash optimality, presents a useful tool for finding good strategies and observing interactions in the study of IPD as well. Adding on to contributions made in this realm of discipline, the core issue considered in this chapter pertains to the adaptation of IPD strategies to different environments. Existing works had showed that evolutionary schemes are highly successful in discovering effective adaptation methods to rich situations [114] – the evolved strategies adapt well to specific environmental settings and are capable of defending against defectors and cooperating with cooperators [89], [115], [116]. While most works are concerned with the generalization ability of evolved strategies [117], the focus of this chapter is on the adaptability of evolved strategies to diverse environmental setups.

Learning, similar to evolution, is another paradigm that is extensively used to adapt strategies in many game theoretic problems [93], [118], [119]. According to Hingston and Kendall [120], it is crucial for creating adaptive IPD strategies

which thrive well in competitive settings by exploiting non-adaptive strategies. Learning presents a scenario similar to one where knowledge accumulated from the past rounds of game play is used by IPD players in future. As the pattern of decision making is rarely constant [121] but highly dependent on the environment and complex interaction among competing strategies, adaptation is important to ensure good performance. Although evolution facilitates information exchange between strategies, it is limited by poor exploitation abilities. Creation of new individuals is more of a trial and error process [122], which often produces naïve strategies upon convergence. Learning, on the other hand, causes large variance in performance among strategies as a result of the diverse learning experiences [123] and is also prone to premature convergence [124]; although it allows strategies to make spontaneous decisions as the environment changes.

In view of the above, this chapter considers the development of a memetic adaptation framework [125] for IPD strategies to exploit complementary features of evolution via a CEA; and learning via a double-loop incremental learning (IL) methodology that incorporates classification, probabilistic update of strategies and a feedback learning process. Despite the widespread use of memetic adaptation in optimization [126] - [129], little work has been done to illustrate its applicability to IPD. Combining the two forms of adaptation schemes introduces a fair degree of realism, especially when it is used to model the behavioral aspects of players. Simulation is performed for us to gain insights into the complexity and intricacies of interaction between evolution and IL; and in the process demonstrates how adaptive strategies are created via the memetic scheme.

Organization of the chapter is as follows: Section 4.1 presents an overview of IPD and Section 4.2 formally introduces the adaptation models. Section 4.3

presents the proposed IL methodology. Section 4.4 highlights the implementation details of adaptation strategies while Section 4.5 describes the simulation tests and evaluates strategy performance via a series of case studies. Section 4.6 concludes the chapter with a summary of discussions on the simulated results and also areas where future work can be embarked on.

4.1   Background study

The IPD is a classical game-theoretic [93] problem which encompasses the study of complex decision making – balancing the short term rational decision for self-interest against the long term decision for overall interest. Each player has the option to COOPERATE (C) or DEFECT (D) in each round and the outcome of interaction is governed by a Payoff Matrix and its binding conditions (Table 4.1 and 4.2). The single-round PD has mutual defection as its unique NE while the IPD has a single NE only if the number of rounds is known in advance [4]. Though the potential incentive to defect is higher in the short run, Axelrod [4] had showed that mutual cooperation is a better solution in the long run. Folk theorem further verified that the cooperative solution is found in the set of NEs of infinitely repeated rounds [130]. Even so, unconditional defection or cooperation is not the optimal strategy for a player as much depends directly on his opponent's strategy [4]. Strategies can perform much better if they achieve cooperation with reciprocal cooperators, exploit unconditional cooperators, and resist defectors [131].

The standard IPD is played repeatedly among competing strategies, each with its own set of behaviors. Some are memory-less while others base their next move on a history of previous moves [132]. For simplicity and in accordance to the default settings used in Axelrod's tournaments [89], a finite memory of up to

three rounds is allowed. Always Defect (ALLD), Always Cooperate (ALLC) and Random (RAND) are examples of memory-less strategies while Tit-for-Tat (TFT), Pavlov and Tit-for-Two-Tats (TFTT) make use of past histories to make decisions. Some strategies are deterministic while others are stochastic. Each strategy has its own advantages and disadvantages; strong against some opponents but also weak against others. A brief description of the characteristics for some commonly used benchmark strategies is provided in Table 4.3.

There are two basic setups - round robin or evolutionary tournament [132]. In the former, every strategy has a chance to play against all other participating strategies in the tournament and the population size of each strategy type is fixed throughout. The latter is conducted on the basis of natural selection, where good strategies are favored by a proportionate increase in numbers while weaker ones experience a subsequent decrease in numbers. The process is repeated until certain stopping criteria e.g. substantial convergence in population sizes of all strategy types is attained. In this context, a good strategy is one that thrives for long duration and in large proportion. This setup is more applicable to situations where each strategy type has large number of players for effective evolution to occur. Complex issues like the group survival rate of strategies can then be investigated.

Table 4.1: Payoff Matrix for the Iterated Prisoner's Dilemma

| | | Player 1 | |
		COOPERATE	DEFECT
Player 2	COOPERATE	3,3	5,0
	DEFECT	0,5	1,1

Table 4.2: Conditions governing the construction of a valid payoff matrix

CONDITION 1	T>R>P>S
CONDITION 2	T+S<2R

Table 4.3: List of some commonly used benchmark strategies

Strategy	Full Name	Brief Description
ALLC	Always Cooperate	Cooperates indefinitely.
ALLD	Always Defect	Defects indefinitely.
TFT	Tit-for-Tat	Cooperates initially but then repeats opponent's previous move.
Pavlov	Pavlov (WSLS)	If previous move is successful (T or R), increase the probability of executing the same move. Reduce the probability otherwise.
TFTT	Tit-for-Two-Tats	Resembles TFT but forgives opponent for 1 defect. Strategy defects only after 2 consecutive defects by opponent.
RAND	Random	Defects or cooperates with probability ½.
STFT	Suspicious TFT	Defects initially but then repeats opponent's previous move.

4.2   Adaptation models

Other than evolution, two other adaptation schemes - learning as well as memetic learning (ML), are proposed for the purpose of driving strategy improvements in the IPD. Each equipped with its own unique characteristics, these methodologies are presented and described in more details in the following sub-sections.

4.2.1   Evolution

Evolution, as an optimization model, is widely used to evolve IPD strategies [114], [133]. By retaining fitter strategies and discarding weaker ones cyclically, this population-based learning technique [134] facilitates the eventual convergence towards robust and effective strategies [117]. Numerous variants of evolutionary implementations have existed as of today e.g. in terms of representation, evolving strategies exist in binary forms [134], neural networks [135], probabilistic [134] and real number [136] coded strings or even FSMs [137]. The method of selecting good individuals comes in many forms as well - truncation selection [138], [139], (μ, λ) and $(\mu + \lambda)$ selection [22], [140] fitness-proportional selection [116], [120], [141]. Choice of variation operators also differs across implementations. While most EAs use a combination of crossover and mutation to evolve strategies [134], [142], [143], pure mutation operators are used by Hingston and Kendall [120]; and

well as Chong and Yao [22] for their co-evolutionary [144], [145] framework to analyze various aspects of the IPD. Other evolutionary means involve the use of speciation or niching to maintain genetic diversity [146] and elitism to avoid loss of good individuals from the mating pool. Despite differences, the fundamental framework is essentially similar and can be broadly summarized as a sequential process of fitness assessment, genetic selection, and genetic variation (Figure 4.1).
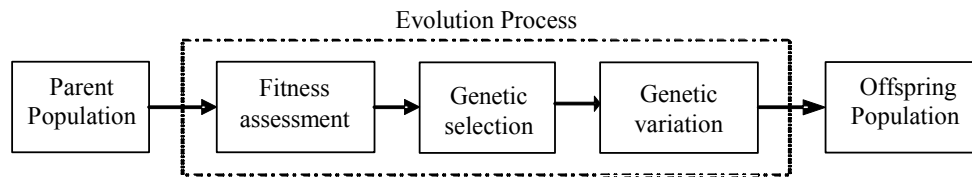
Figure 4.1: Overview of the evolution process

4.2.2 Learning

The learning methodology can be progressive [147] or reactionary [148]. The former includes hill-climbers and gradient-based searches which are commonly applied to static environments where conditions are fixed and notion of optimality is well defined. The latter is applicable to a dynamic setting where the notion of optimality is always changing or simply non-existent. Notable examples include probabilistic Pavlovian Learning [149] and stochastic searches. In its classical form, learning has no facility for any communication and only affects individual strategies. It functions as a local search operator and drives agents to traverse the direction which is deemed more "favorable". Learning agents typically exploit domain information available at hand to improve performance based on some form of heuristics. Since the pattern of decision making is rarely constant for any iterated game set but highly dependent on agent interaction, learning should be performed incrementally, with partial memory [150] of past experiences. This

86

more accurately models the IPD, where players are capable of making complex decisions spontaneously using some finite memory of past actions.

4.2.3 Memetic Learning

ML is a hybrid adaptation technique that unifies learning and evolution [151] in one algorithm. As far as the IPD is concerned, the notion of evolving strategies memetically is less studied compared to learning and evolution. Two widely used variants of ML are the Baldwinian [152], [153] and Lamarckian [129] models. In the former, offspring do not inherit learned abilities from their parents but merely experience an added ability to learn skills that are acquired by the ancestors [154]. In the latter, however, desirable traits acquired by parents via learning are passed down to the offspring who will inherit the traits directly [155]. Despite differences, the underlying framework is similar. As opposed to learning, evolution in ML facilitates information exchange among agents and allows the knowledge acquired through learning to propagate to future generations. This stabilizes and reduces performance variance across learning agents. Learning, on the other hand, is used as a form of directed search to guide evolution in attaining convergence towards an optimum strategy if it exists. The "Meta-Pavlov Learning" [156] that integrates evolution and Pavlovian Learning is an example of Baldwinian-based ML strategy. By harnessing the synergy between learning and evolution, ML strategies should be able to acquire better performance than evolutionary and learning strategies.

4.3   Design of learning paradigm

In this section, an IL scheme that integrates classification into the decision making process of strategies is presented. This is adopted by all IL and memetic players in

the IPD tournaments and is characterized by a framework that identifies nature of opponent strategies, conducts strategy revision and allows players to recover from mistakes through double loop IL. This breeds good strategies that can respond and adapt well to different opponents.

4.3.1 Identification of opponent strategies

According to Axelrod [4], discrimination of others is among the most important of abilities as it allows an individual to handle interactions with many individuals without having to treat them all the same. With the above motivation in mind, a simple classification heuristic is formulated based on the correlation between the received payoffs and opponent's likelihood to execute defection and cooperation. Opponents are classified into three broad categories, strategies with a tendency to exploit others (Exploiters), strategies which reciprocate cooperation (Reciprocals) and strategies that are likely to cooperate unconditionally (Cooperators). Nature of each new opponent is mapped out according to the sum of payoffs received in the first three rounds of game play. The range of scores (0-15) is divided into three equal intervals, each corresponding to an opponent class as shown in Table 4.4. This classification process acts as a basis for the player to gain a rough insight into the nature of unknown opponents, so as to facilitate the adoption of an appropriate strategy during the subsequent game play with that opponent.

Table 4.4: Identification of opponent strategies

SCORE RANGE	0 - 4	5 - 10	11 - 15
NATURE OF OPPONENT	EXPLOITERS	RECIPROCALS	COOPERATORS

4.3.2 Notion of "success" and "failure"

Inspired by John Nash's [3] idea of a NE, it is conceptualized that every pair of competing strategies, despite their complexity and nature, can have a desired state

at each round of game play, where both execute their relative best responses [2]. With the same argument, each opponent classification can give rise to an attached desired response. This is defined for Exploiters, Reciprocals and Cooperators by means of a Taxonomy Matrix in Table 4.5. This Matrix effectively dictates and maps out the direction for local search during the process of learning.

Each IL strategy is represented by a string of bits - each encoding the action to be taken when a distinct sequence of past three moves is made by both player and opponent. As a basis of IL, outcome of using a bit is classified into "success" (**S**) or "failure" (**F**) trials. An **S** trial occurs when interaction outcome from using a strategy bit coincides with the perceived nature of opponent as the desired reply is played while an **F** trial denotes otherwise. Rules for updating **S** and **F** trials are characterized by the Taxonomy Matrix. S count is incremented when the "desired" outcome is achieved for the opponent's assumed strategy type, and the F count is incremented when any other payoff is achieved. Overall, the process of updating counts indicates how good a strategy bit is from time to time and is used as a form of IL heuristics to determine whether the encoded action should be revised or left unchanged. The Taxonomy Matrix and underlying notion of **S** and **F** are proposed to refine the Performance Matrix used by Pavlov - **S** is defined as receiving the Temptation (T) or Reward (R) payoff and **F** as being awarded the Punishment (P) or Sucker (S) payoff. This is not necessarily a good way to define the matrix due to the following set of reasons:

1) Receiving P in the context of exploiters which defect perpetually is considered good as the player is not exploited.

2) Receiving R in the context of unconditional cooperators is considered bad as there are opportunities for exploitation.

3) Receiving T in the context of reciprocals is not the best policy as it can well lead to endless cycles of retaliation.

S and **F** hold a fuzzy meaning when payoff is P, R or T. As there is no knowledge about the opponent, uncertainty is involved during IL - strategies that are good against one opponent might be bad for another. As the notion of **S** and **F** is crucial for determining the "goodness" of IL and exerts great impact on strategy performance, it is not predetermined in advance but dynamically updated based on the opponent's perceived nature as the IPD game proceeds.

Table 4.5: Taxonomy Matrix for carrying out IL

	Player	
	COOPERATE	**DEFECT**
COOPERATE	RECIPROCALS	COOPERATORS
DEFECT	-	EXPLOITERS

Opponent

4.3.3 Strategy Revision

In the proposed IL scheme, replacement of weaker strategy traits with stronger ones is devised via the **S** and **F** counts accumulated over the entirety of the game. Fitter bits have larger **S** counts to indicate that they are performing desirably against opponents while weaker bits have larger **F** counts. Unused bits contain a zero for both counts. The action that a strategy bit encodes is updated (changing C to D or vice versa) only when the following conditions are met

$$\text{Swap (True) iff } \mathbf{F_T} / (\mathbf{S_T} + \mathbf{F_T}) > \mathbf{Ps} \text{ AND } \mathbf{S_T} + \mathbf{F_T} > \mathbf{L} \qquad (4.1)$$

where $\mathbf{F_T}$ and $\mathbf{S_T}$ are total **F** and **S** counts respectively. **Ps** is a threshold that can be adjusted to suit the desired level of failure tolerance – amount of failure that an IL player is willing to take on before strategy revision. This affects the willingness to learn implicitly e.g. a higher **Ps** makes a player less likely to revise its strategy.

The minimum learning threshold, **L** - defined by the minimum number of rounds which a bit is played before strategy revision is considered, affects the sensitivity of a player's response to environment changes e.g. a large **L** delays IL but in turn allows the goodness of a bit to be assessed from a wider observation window. An inherent tradeoff arises between the need to react spontaneously to changes in opponent's pattern of game play (so that payoff is maximized regularly) and the need to maintain a sizable window of past experiences before performing strategy update (so that well-informed choices can be made). Prior simulation tests are conducted using the proposed IL strategy and an empirical set of opponents to select appropriate values of **Ps** and **L** for the update criteria. **S** and **F** counts of a bit are reset to zero when updating of that bit occurs and also upon meeting new opponents. This prevents past histories from affecting the current performance of the bit. The above ensures that strong and desirable strategy bits are more likely to remain intact while weaker ones are more susceptible to change.

4.3.4 Double-loop Incremental Learning

The double-loop IL is a reclassification and relearning process that draws parallel to human's way of learning through perceiving, reasoning, self-evaluating and readjusting. A scenario when the inferred opponent is perceived incorrectly during classification is remedied via a separate IL cycle which involves reclassification of opponent and re-mapping of best response. This cycle is triggered when the accumulated **F** counts for all bits used within the game exceed a value of **f**; of which the notion of **S** and **F** will be changed and a new perceived best response adopted. If this corresponds to the actual best response, increase in **F** counts will be reduced to indicate that the right strategy has been used against the opponent. Otherwise, reclassification and relearning continues until the perceived and actual

best responses coincide. While inner-loop IL allows players to form perception models of the opponents, outer-loop IL facilitates evaluation and readjustments of each model. The feedback mechanism allows players to learn, adapt and realign their strategies dynamically to changes in the opponent nature through formation, evaluation and revision of perceptions. Via the process of learning and relearning [157], strategies learn to perceive each opponent accurately despite insufficient knowledge initially. Unlike absolute reactionary IL, the integrated double-loop IL can prevent chances of entering into a loop of endless updating and subsequently lowers the possibility of being trapped in a local maximum (Figure 4.2).
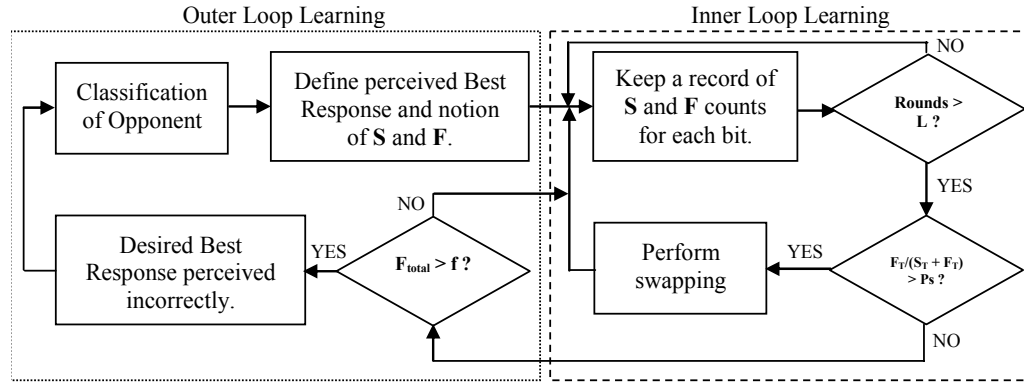
Figure 4.2: Overview of the double-loop learning process

4.4 Implementation

Performance of the adaptation frameworks; evolution, IL and ML are investigated. They are represented by a GA, incremental learning strategy (ILS) and memetic algorithm (MA) respectively in the context of IPD. Each strategy is represented by a 67-bit binary chromosome (Figure 4.3). The first three bits encode the condition for triggering the first three moves at the start of every iterated game set while the remaining bits denote 64 (2^6) possible histories of 6-bit memory configurations that correspond to different combinations of previous three moves of both player

92

and opponent. For ILS and MA players, accumulated **S** and **F** counts for each bit are recorded as the basis of IL. Each strategy also encodes an independent 6-bit memory of round histories that are used to decide the player's next move. As each derives its fitness from the scores attained by playing with others, GA is basically a CEA which evolves strategies by co-evolutionary learning. As far as the chapter is concerned, MA adopts a Lamarckian learning scheme - retains genome changes from mid-game learning at the point where the next generation is formed.
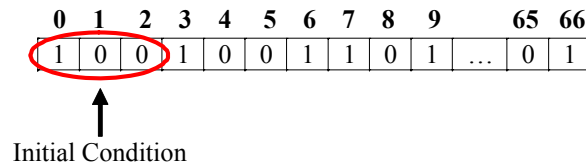
Figure 4.3: Strategy representation of a typical player

The initial populations of all the adaptation schemes are randomly generated. Fitness, given by the sum of payoffs accumulated throughout the game play, is assigned after each complete tournament. Niching with sharing distance (defined by the genotypic similarity between players) of **r** is used to encourage growth of a diverse repertoire of strategies. Two strategies are alike if the number of identical genes between them exceeds **r**. By and large, the mechanism evaluates similarities between individuals and penalizes fitness of those that are too alike. This is crucial so that the evolving strategies would not, due to inability to adapt after substantial level of strategy convergence is reached, lose out in terms of performance. This also ensures fairer comparison among GA, ILS and MA. Results from preliminary simulation have verified that GA performed worse off in the absence of niching as premature convergence does set in at a very fast rate, compromising GA's search capability unduly. From experimentation, **r** is selected to avoid an overly fast or slow convergence rate in the midst of preserving the search capability of GA.

After niching, elitism is implemented so that the strategies which are above average in both fitness and niche counts are selected into the next generation without genotypic changes. Depending on the proportion of individuals that are above average, number of elites varies but is limited to a maximum of the top 10% of the population. After which, tournament selection is performed to select the remaining population based on overall fitness that is accounted for by niche scores:

$$\text{Overall Fitness} = \textbf{(game scores)/(1+niche scores)} \qquad (4.2)$$

From above, it can be garnered that the larger the degree of similarities between an individual and others in the population, the larger is the penalty on its original fitness. Selected individuals will then undergo uniform crossover and binary bit-flip mutation. The new generation of strategies is formed jointly by the elites and individuals that have gone through the variation process. To summarize important procedures involved, the general overview of the flowcharts for GA, ILS and MA are presented respectively in Figures 4.4, 4.5 and 4.6.
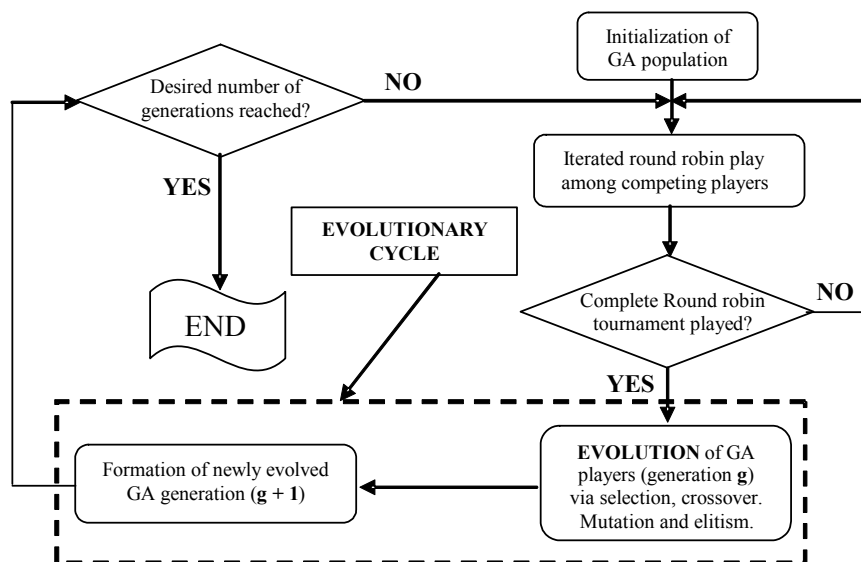
Figure 4.4: Simple flowchart depicting the operations of the GA strategy
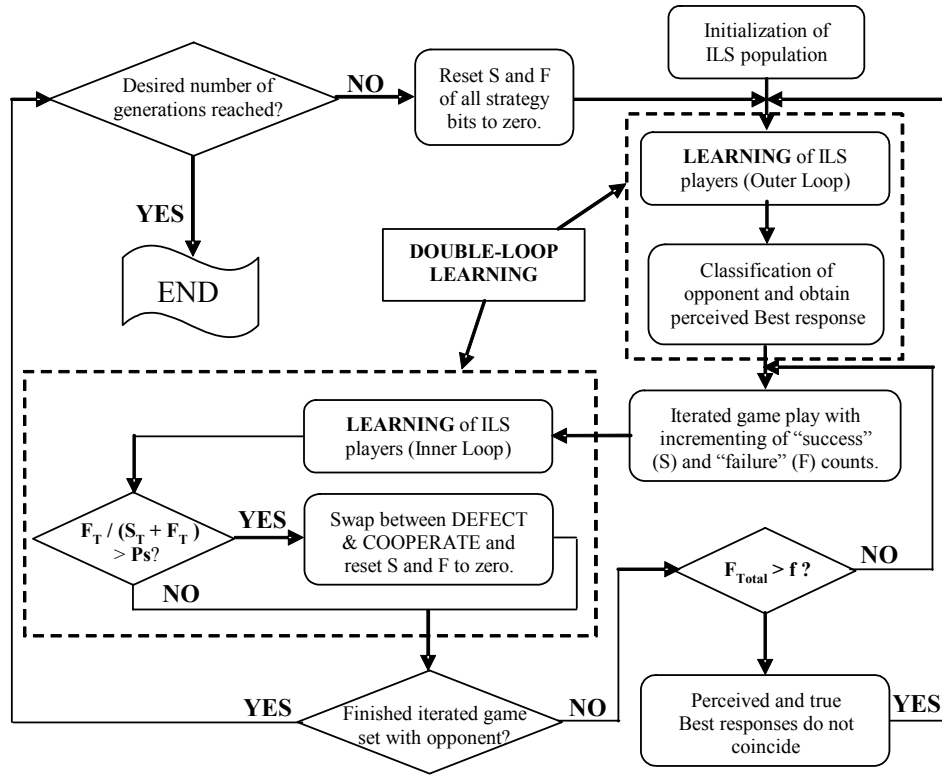
Figure 4.5: Simple flowchart depicting the operations of the ILS algorithm
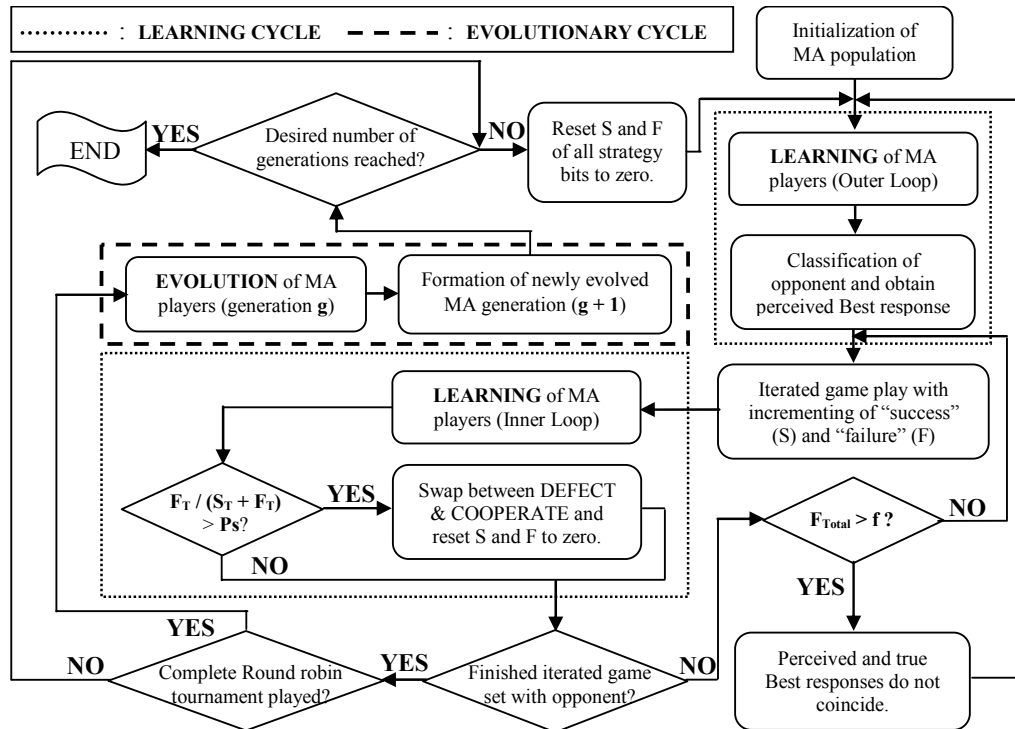
Figure 4.6: Simple flowchart depicting the operations of the MA algorithm

95

4.5   Simulation results

Simulations are carried out using the Visual C++ development kit. A summary of the parameters used in simulations is shown in Table 4.6. Three case studies are simulated to examine the synergy between evolution and IL in different settings. For all case studies, number of players for each strategy type, is always set as **p**, to avoid any bias towards any type. The population size, **n** thus varies for different test cases, according to the number of strategy types used e.g. for a case study that involves three strategies, **n = 3*p.** 200 rounds [4] is played in a pair-wise manner amid the competing players in each iterated game set. A round robin tournament is completed when each player has played against all others.

The order of opponent strategies during round robin tournament is decided randomly e.g. a strategy will not know the nature of opponent it will be facing in the next game as this is picked randomly from the pool of strategies that have yet to play. This will allow the performance and adaptability of IL and evolution to be assessed in a more generalized setup, where overall performance is independent of the order of appearance of opponents. In all runs, evolution is triggered after each round robin tournament (which denotes one generation) while IL is performed throughout the course of game play. Based on Chong and Yao [135], each run is conducted for 600 generations to ensure convergence of result and to track the rate of improvement for all adaptation strategies. Parameter values are aptly chosen based on good players' performance after several rounds of preliminary runs. A brief summary of the case studies to be simulated is depicted in Table 4.7.

Table 4.6: List of parameter values used in the simulation runs

Tournament Parameters	Values
No. of rounds in an iterated game set between two players, α	200
No. of generations used to carry out each simulation run, **g**	600
Population size of all players in the tournament, **n**	Variable
Size/Number of players for each strategy type, **p**	30

Evolution parameters	Values
Size for carrying out tournament selection, **s**	2
Rate for performing uniform crossover, **c**	0.8
Rate for performing binary bit-flip mutation, **m**	0.05
Niche radius in terms of number of identical strategy bits, **r**	30
Incremental Learning parameters	**Values**
Bit failure ratio, **Ps**	0.3
Minimum learning threshold, **L**	10
Threshold of failure counts to trigger reclassification, **f**	15

Table 4.7: Brief summary of case studies to be conducted

Case Study	Tests	Strategies Used	Objective
1	A	MA, GA, ILS, TFT, ALLD, ALLC	Assess performance and adaptability of MA, GA and ILS when each plays against different combinations of opponent strategies.
	B	MA, GA, ILS, TFT, ALLD, ALLC, PAV, RAND, STFT, TFTT	
2	C	MA, GA, ILS	Assess performance and adaptability of MA, GA and ILS when playing among themselves and other opponent strategies.
	D	MA, GA, ILS, TFT, ALLD, ALLC, PAV, RAND, STFT, TFTT	
3	E	MA, GA, ILS, opponents strategies that are changing every {1, 10-20, 100-150} generations	Assess performance and adaptability of MA, GA and ILS in dynamic environments, where nature of the opponents is changing from time to time.

4.5.1 Case Study 1: Performance against benchmark strategies

The first case study compares individual performance of MA, GA and ILS as each

plays with fixed pools of benchmark strategies (Table 4.3). The chosen strategy

subsets contain a mix of cooperators, defectors and reciprocals, to which a player

has a distinct, best response to maximize its payoff against each opponent type.

Two different tests (A and B) are setup and simulated to evaluate the performance

of the adaptation strategies in small and large strategy subsets using generation

payoffs, cooperation ratios and performance box plots. In both test scenarios, each

unique strategy type has a total of **p** players.

4.5.1.1 Test A: Performance against ALLC, ALLD and TFT

To assess performance against a common group of opponents, MA, GA and ILS,

each in a separate milieu, is set to play against ALLC, ALLD and TFT in the same

tournament. The normalized average generation score per round (AGS) – sum of

payoffs in one generation over the total rounds played; and average cooperation ratio (ACR) – proportion of total rounds for which cooperation is played; of each strategy type are plotted for different test configurations (Figure 4.7). Results are averaged over 20 runs to minimize effects of large statistical deviation due to any one run. On average, AGS plots showed that all adaptation strategies are robust against reciprocals, defectors and cooperators as indicated by their ability to outperform deterministic benchmark strategies (Figure 4.7a). Despite the use of diverse adaptation mechanisms, both GA and ILS are comparable in their eventual performance, except for the path taken to reach convergence. Both AGS and ACR for ILS are constant but gradually increasing and decreasing respectively for GA.

Whilst the GA strategies are evolved with more room for exploration, the exploitative nature of ILS accounts for the fact that TFT fares slightly better than ALLD with GA but worst off with ILS. The similarities in performance between GA and ILS do not, in any way, constraint the performance of MA to the AGS attained by either of them. In contrast, inheriting the synergetic blend of evolution and IL allows MA to achieve significant score advantage over GA, ILS and the benchmark strategies. The existence of this notable score disparity between MA, GA and ILS, despite uniformity in their ACRs around 0.4 (Figure 4.7b), illustrates that the key to attain good performance depends not so much on the overall extent to which a strategy cooperates or defects, but more on its ability to do so at the right time, according to the nature of his interacting opponent.

The strategy specific AGS and ACR plots in Figure 4.8 reaffirm the claims made above. The fluctuating dynamics introduced by co-evolutionary learning in MA and GA are conspicuously dissimilar from the fairly stable AGS and ACR traces of ILS. Inherent consistency or stagnation in the performance of ILS across

generations arises due to premature convergence. Exploitation is emphasized over exploration as players react spontaneously to the opponent strategies by seeking incremental improvement in a certain direction via a common set of IL heuristics. Co-evolutionary learning allows more opportunities for players to explore and attain continual improvement, but possibly at the expense of a slower learning rate and larger dynamics - owing to disparity in AGS of evolving population between successive generations. The potential advantage of exploration is clearly depicted in Figures 4.8c and d, where GA surpasses ILS eventually, despite starting with a lower AGS as a result of random initialization of strategies. The collection of plots showed that ILS and GA perform well in different settings. ILS players are able to attain higher payoffs than their GA counterparts by being more cooperative with players of their own type (Figure 4.8a) and defecting against ALLC (Figure 4.8d). Conversely, GA surpasses ILS players by cooperating and defecting to a larger extent against TFT and ALLD respectively (Figures 4.8b and c). The sum of each score advantage constitutes their equivalence in overall performance (Figure 4.7a).
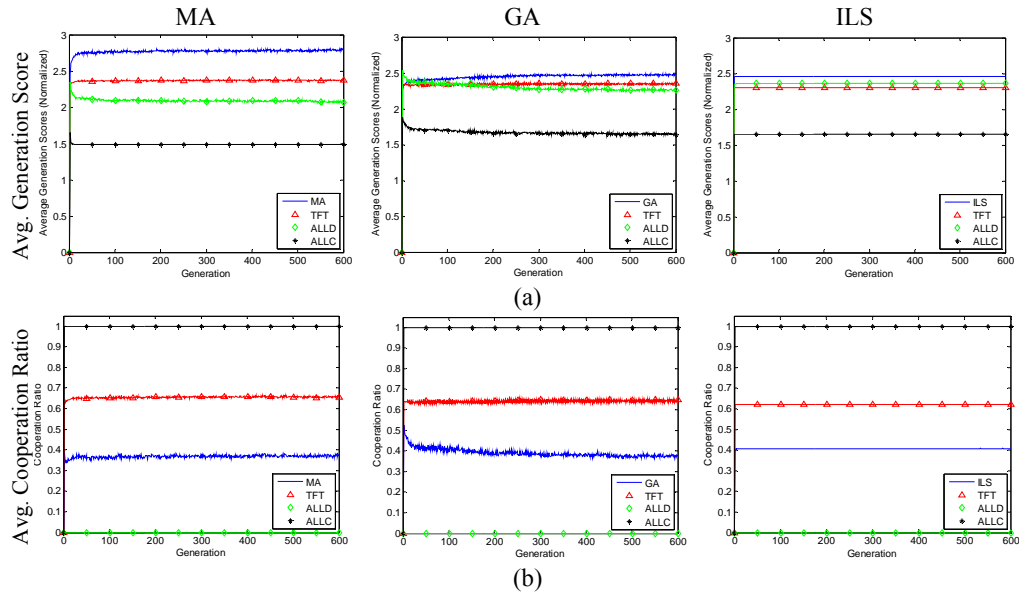
Figure 4.7: (a) AGS and (b) ACR for MA, GA and ILS when each plays with TFT, ALLD and ALLC over 20 runs

99

With both IL and evolution employed, MA's performance outdoes GA and ILS for all strategy types. Despite similarity in ACR (Figure 4.7b), MA is able to secure a considerable lead over GA and ILS, by cooperating and defecting aptly with the right opponents - attaining AGS {2.7, 2.5, 0.95, 5} and ACR {0.8, 0.65, 0.05, 0} for {itself, TFT, ALLD, ALLC}. In entirety, MA maintains the closest performance to a hypothetical ideal player who attains AGS {3, 3, 1, 5} and ACR {1, 1, 0, 0} if playing against the same corresponding strategy types in that order.

On top of the statistical plots, an alternative performance measure to assess the goodness of a strategy pertains to the score deviation between members of the same population. For example, even when the top player belongs to a particular strategy, it is unfair to claim that the strategy is good when there is large score variance among players using that strategy; since not everyone is doing as well. To have a convincing claim, performance should be measured on the collective rather than individual basis. With this yardstick, a strategy is considered good only if those playing it are able to perform uniformly well throughout the tournament.

Performance box plots which depict the distribution of mean, variance and minimum AGS within each strategy group are presented across 20 runs (Figure 4.9). MA has the highest mean AGS followed by GA and ILS – both of which are comparable (Figure 4.9a). As compared to ILS, It is evident that the mean AGSs of MA and GA vary over a much wider range. The stochasticity and differences in the environments across distinct runs signify that the mean performance of an evolving population in any run is closely correlated to the settings in that run e.g. strategy initialization, sequence of opponents encountered etc. By adhering to a fixed set of heuristics, the performance of ILS tends to be more consistent and less affected by such stochastic difference across runs.
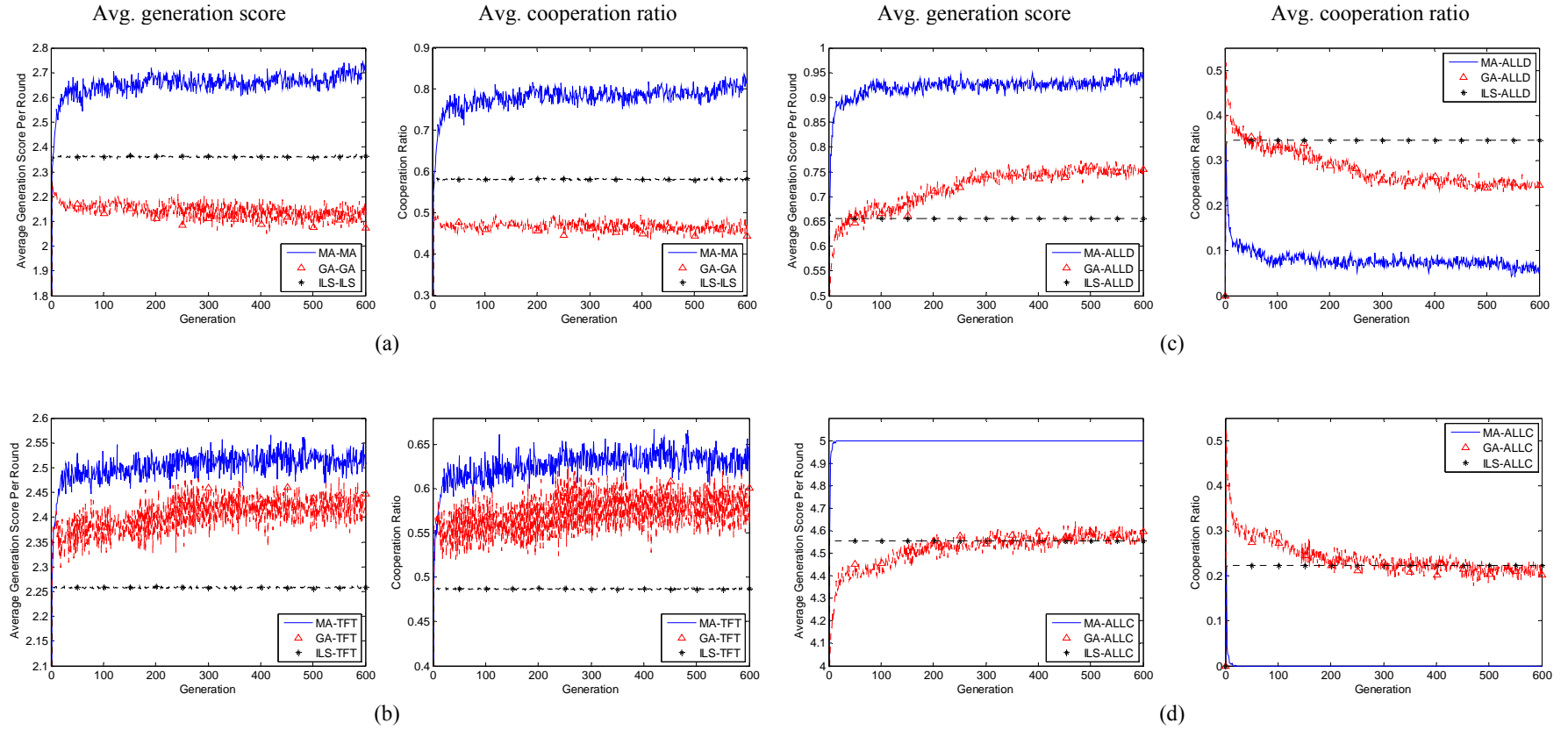
Figure 4.8: Strategy specific AGS and ACR for MA, GA and ILS as each plays with (a) itself, (b) TFT, (c) ALLD and (d) ALLC over 20 runs

ILS, however, suffers from a much larger mean score variance among its players as compared to GA and MA (Figure 4.9b). A low variance of variance implies that the performance disparity across ILS players is consistently large in every run. Compared to GA which attains similar mean AGS, a likely reason for the large variance is due to the diverse experience of different ILS players as each embraces independent IL. Unlike GA players, knowledge acquired by each ILS player is not conveyed or shared among other players. Due to lack of information exchange, improvement as a group is much harder for ILS as players differ widely in both performance and nature of their learnt strategies.

Evolution reduces score variance in the GA population as weaker strategies adopt good traits from stronger ones by means of collective learning and periodic exchange of information during each evolution cycle. This ensures that all GA players progress and improve on a collective basis. These observations shed new light to the advantages of evolution and IL. The former evens performance of a population of players within each run; while the latter ensures consistency and stable performance across different runs. Fusing the benefits of both mechanisms allows MA to attain the highest mean and minimum AGS (Figure 4.9c) without compromising its ability to perform well as a cohesive group e.g. attains score variance that is comparable to GA.
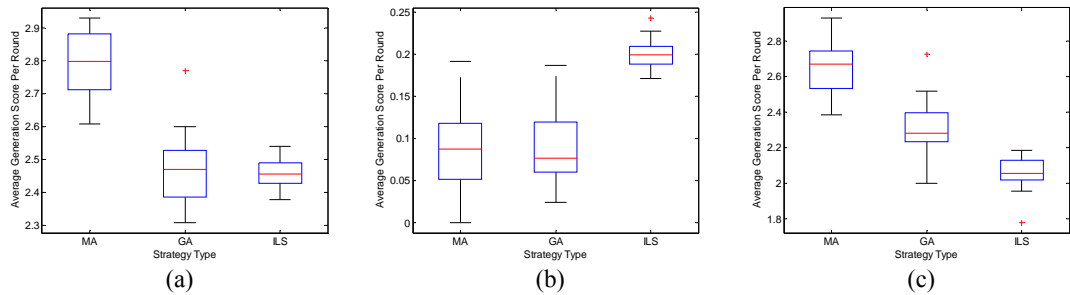
Figure 4.9: Box plots depicting distribution of (a) mean, (b) variance and (c) minimum AGS in the MA, GA and ILS populations as each plays with TFT, ALLD and ALLC over 20 runs

4.5.1.2 Test B: Performance against seven different benchmark strategies

Extending from test A, MA, GA and ILS, each in a separate tournament, is setup to play against the set of all seven strategies {TFT, ALLD, ALLC, PAV, RAND, STFT, TFTT}. The complexity in the nature of interaction is increased as more strategy types are added. This is reflected by a reduction in the AGS difference among various strategy types (Figure 4.10a). It is more difficult for any strategy to maintain a large score advantage, on average against other strategies as each is required to score well against more opponents of diverse nature. Tradeoffs are probably incurred as a strategy that is good against a certain opponent type might not be necessarily good against all others.

It is apparent that the performance of GA and ILS has degraded - no longer the best in their respective environments; with the former overtaken by PAV, and the latter by both PAV and TFT. AGS of ILS has decreased slightly to 2.44, though it still remains relatively stable across generations. Despite starting with a lower AGS, the two phase improvement of co-evolutionary learning - exponential followed by gradual increase in fitness, allows GA to overtake ILS and TFT in 20 and 200 generations respectively; only to be marginally surpassed by PAV. This shows that evolution; with its ability to explore and adapt on a collective basis, allows players to perform well against a fixed pool of opponents that are largely deterministic. Fusing IL with evolution enhances the players' ability to perform well against a large pool of unknown strategies. Corrective actions are made in the course of game play when evolved strategies are not aligned with the opponent's nature. With the Lamarckian mode of social IL, all the learnt traits and beneficial changes are preserved and acquired directly by offspring during evolution, thereby

enhancing the rate of co-evolutionary learning. This form of guided evolution allows MA to gain a substantial score margin above GA and ILS.

Inference can be drawn from the traces about the fact that IL predominates in the initial stages as it provides evolution with a substantial boost during the phase when the MA population is experiencing an exponential improvement in fitness. This gives MA an early score advantage to start with. Subsequently, IL plays the role of maintaining the population's overall performance against a fixed set of opponents. Further improvement in performance is gradual and largely due to evolution, whose effect sets in much later and lasts throughout the generations. As MA's ACR is between GA and ILS (Figure 4.10b), results suggest that being too cooperative or defect-oriented does not ensure good performance, but rather the ability to strike a balance by cooperating and defecting suitably.

Comparing the strategy specific AGS and ACR (Figure 4.11) in the current setting, it is perceptible that GA tends to perform better relative to ILS on average; attaining higher AGS when playing against itself, TFT, PAV, STFT and TFTT. This indicates that the GA strategies are fairly good at reciprocating cooperation while ILS on the other hand performs better against cooperators, defectors and random strategies, due to its exploitative nature. Though co-evolutionary learning is a slow process of improvement – players learn on a collective basis only when each evolution episode is triggered; the ability to explore constantly is a valuable asset to GA as it enables a possibility for further improvement. More importantly, it allows cooperation to evolve when players are adapting against a large pool of reciprocals; which is less likely to arise for the more exploitative ILS population.

MA · GA · ILS
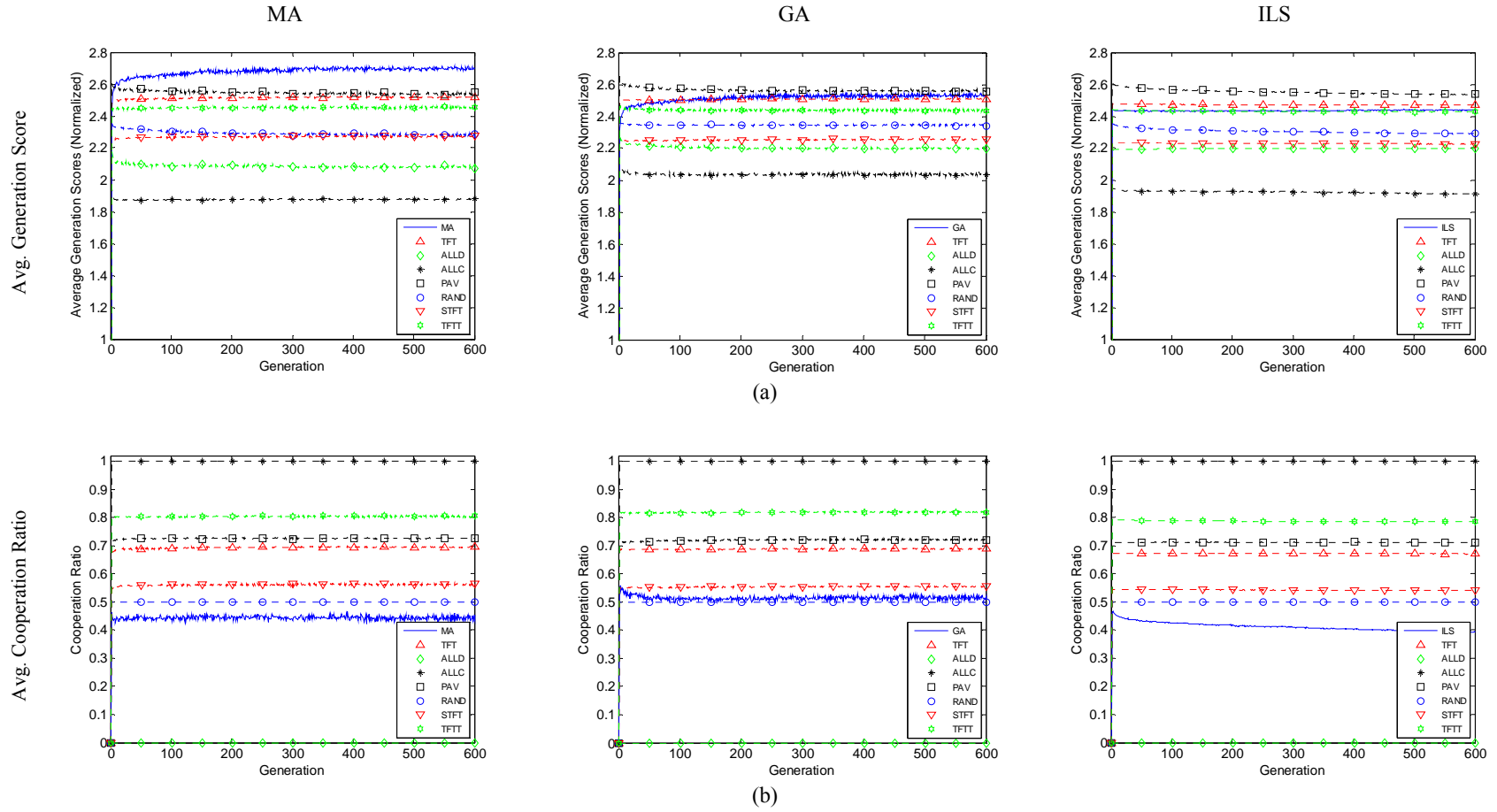
Figure 4.10: (a) AGS and (b) ACR for MA, GA and ILS when each plays with TFT, ALLD, ALLC, PAV, RAND, STFT and TFTT over 20 runs

Strategies used in previous setting/ New strategies introduced in current setting

Avg. Generation Score

Avg. Cooperation Ratio

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)
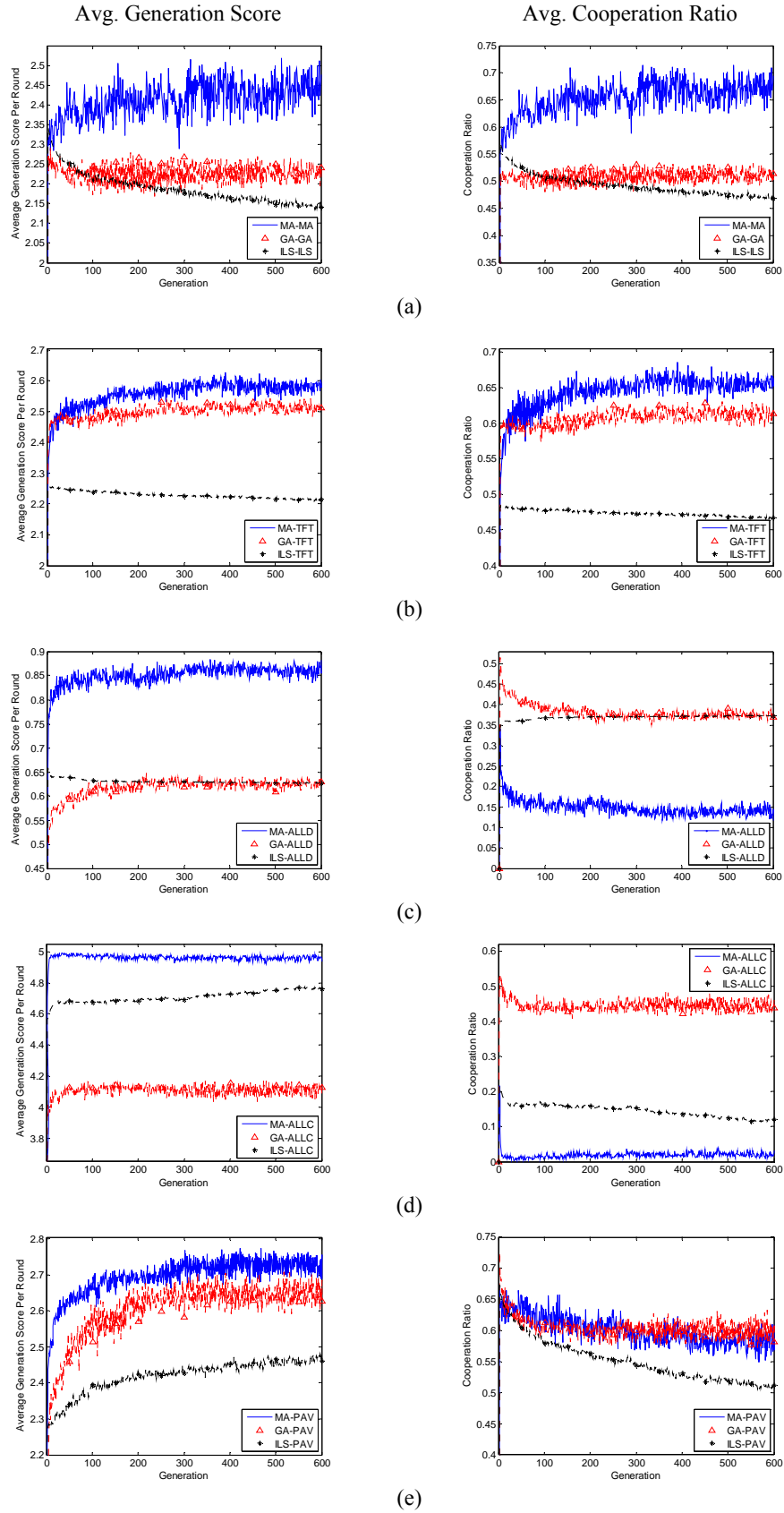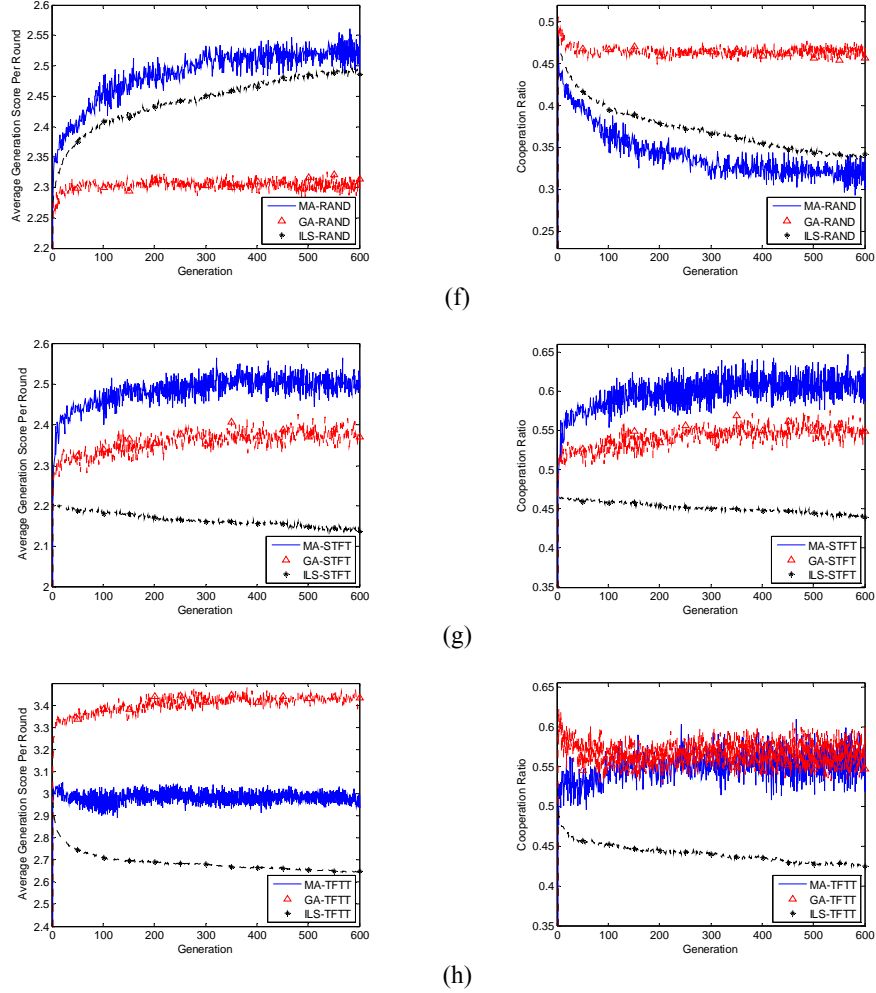
Figure 4.11: Strategy specific AGS and ACR for MA, GA and ILS as each plays with (a) itself, (b) TFT, (c) ALLD, (d) ALLC, (e) PAV, (f) RAND, (g) STFT and (h) TFTT over 20 runs

A balance between exploration and exploitation allows MA to preserve its trend of dominance over both GA and ILS for all strategies except TFTT - to which GA replicates strategies that closely resemble the ideal strategy of alternating between cooperation and defection. This is probably due to the tradeoff involved when IL disrupts the actual usage of evolved strategies by altering them from time to time; as strategy fine tuning is most probably triggered when vastly different opponent types are encountered in consecutive game play. Save for this isolated case, MA essentially still exhibit robust performance.

Even so, comparison between the current and previous settings indicates that the performance of all adaptation strategies are adversely affected following an increase in complexity within the environment e.g. AGS of MA against itself and ALLD has decreased despite improvement against TFT. Similarly, performance of GA has improved when playing against itself and TFT but declined for ALLD and ALLC. ILS does better against ALLC but worst off against itself, TFT and ALLD.

Figure 4.12: Box plots depicting distribution of (a) mean, (b) variance and (c) minimum AGS in the MA, GA and ILS populations as each plays with 7 benchmark strategies over 20 runs

It is clear from the box plots (Figure 4.12) that MA still secures the highest mean and minimum AGS, followed by GA then ILS. The advantages of evolution actually outweigh those from IL. Variance of mean AGS across the runs remains larger for MA and GA (Figure 4.12a). Exploration though beneficial, entails score deviation between evolved populations across runs. This highlights the inherent stability that IL introduces. Absence of information sharing, however results in large score variance among ILS players, which inevitably lowers the performance of the population (Figure 4.12b). In all, evolution coupled with IL is crucial for stabilizing the performance across a population of strategies and to attain good score advantage over the opponents.

4.5.2 Case Study 2: Performance against adaptive strategies

The second case study assesses the effectiveness and adaptability of MA, GA and ILS as they are set up to play against one another in absence (Test C) and presence (Test D) of other benchmark strategies across 20 runs. As strategies are constantly adapting to one another, interaction is actually more complex than the previous case study where strategies are largely fixed. The relative strategic dominance of evolution, IL and ML are evaluated using generation payoffs, cooperation ratios, niche counts, learning ratio - proportion of total rounds with IL taking place, box plots and statistical tests. In both tests, each unique strategy type has **p** players.

4.5.2.1 Test C: Relative performance of MA, GA and ILS

To assess relative performance, MA, GA and ILS are configured to play against one another within the same tournament, but in the absence of other benchmark strategies. Figure 4.13a depicts that MA maintains a sizeable score margin above GA and ILS, attaining a mean AGS of ~2.7 and ACR of 0.52 (Figure 4.13b). As opposed to previous test cases, ILS – with an AGS of 2.1, took the lead over GA in this setup. Since IL tends to be exploitative, defect-oriented traits of ILS are clearly still present. Adaptive opponents also results in a substantial amount of fluctuation in GA's ACR as there is a tendency for the best responses to these strategies to be constantly shifting until a stable equilibrium is reached after some mutual adaptation phase. The huge performance disparity among strategies despite similarities in ACR between MA and GA; and uniformity in niche counts of MA, GA and ILS reiterates the importance of cooperating and defecting at appropriate times (Figure 4.13c). Learning traces showed that evolution disrupts IL only in a minor way (Figure 4.13d) - by undoing some of the learnt changes; so much so that MA is seen to exhibit similar level of IL as ILS from time to time.
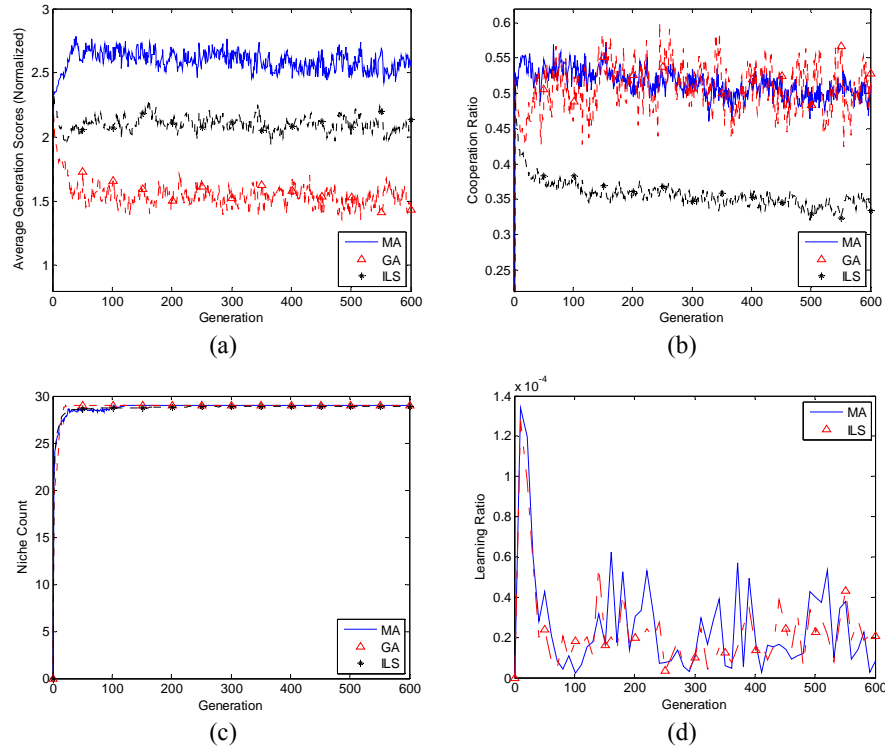
Figure 4.13: (a) AGS, (b) ACR, (c) niche count and (d) learning ratio as MA, GA and ILS play against one another over 20 runs

The strategy specific AGS plots (Figure 4.14a) show that both MA and ILS are able to attain scores that are close to the mutually rewarding payoff with GA; whereas GA is unable to do so vice versa – achieving an AGS that is as low as the punishment payoff. This indicates that the efficacy of IL takes precedence over evolution for a setup with fixed pool of adaptive strategies. In this context, GA is exploited by MA and ILS through their ability to learn, revise strategies and react spontaneously to changes amid the evolutionary episodes. This is the prime cause for the low performance of GA, other than its inability to cooperate exceptionally well with other GA players. The difference in the performance of MA and ILS is distinguished by the added ability of MA to cooperate better with players of the same type, though both perform just as well against each other. Quantitatively, MA attains AGS of nearly 3.0 compared to a mere 1.8 for ILS. Evolution opens

up opportunities for exploration and helps to propagate the evolved cooperative traits throughout the population via strategy exchange. Both steer the MA players towards mutual cooperation with one another.

Some distinct characteristics of each strategy are evidently shown in the compositional ACRs (Figure 4.14b). The fixed nature of the GA strategies amid evolutionary episodes is visibly marked by consistent ACR traces for all strategies. Absence in dynamism and flexibility in performing strategy revision in the course of game play compromises the adaptability of GA extensively, especially when strategies are dynamically changing. In contrast, ILS players have the ability to discriminate among diverse opponent types and revise their strategies accordingly. Whilst players perform better against the adaptive opponents than their evolving counterparts, IL tends to channel more efforts towards exploitation, apart from the lack of knowledge exchange and performance comparison on a collective basis. This result in an independent set of learnt strategies which are largely defect-oriented against all opponents; indirectly reduces the innate capability to adapt. MA population depicts the best performance among the group by exhibiting an ability to cooperate well with like players and defect against players of other types.

As opposed to the previous case study - where the adaptation strategies play against deterministic ones, a setup with adaptive strategies favors IL to evolution, as reflected by the higher mean AGS of ILS over GA (Figure 4.15a). A mixing of evolution and IL still ranks MA as the best in terms of the average and worst performances (Figure 4.15c) among GA and IL. Even so, the possible conflicts between IL and evolution have led to large variance in mean AGS of MA across the runs. Though IL is still marked by large mean score variance among players, variance of score disparity across runs is much larger (Figure 4.15b) than before.

111

Lower group variances of MA and GA highlight the importance of collective learning and information exchange in ensuring the consistency of intra-group performance across runs. The mean score variance among players is particularly low when both IL and evolution are used concurrently.

To reinforce our conclusion that MA performs better than GA and ILS, a paired T-Test is conducted between all paired combinations of the three strategies. The null hypothesis denotes the proposition that the mean AGS of two matched populations are equal while the alternative hypothesis denotes otherwise. The test determines how different or alike two strategy populations are. Using a statistical significance level of five percent, results of T-Test for strategies in each individual run, computed over the span of all simulated runs, confirm that strategies shaped independently by evolution, IL or ML are substantially different from one another. GA and MA evolve to be similar like ILS in just 1 out of the 20 runs (Table 4.8), indicating that any performance differences among the three adaptation strategies are indeed significant and non-trivial.

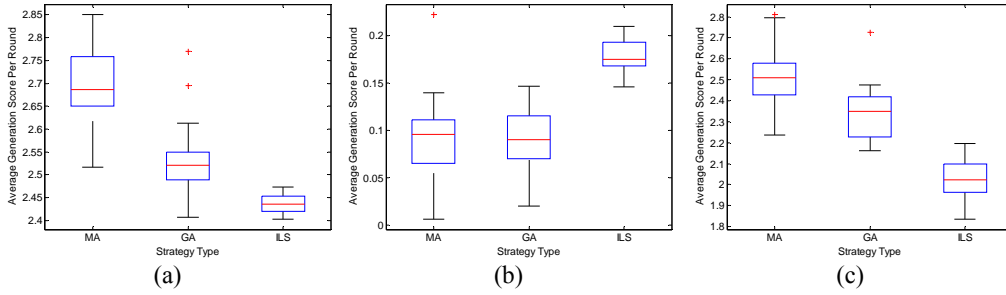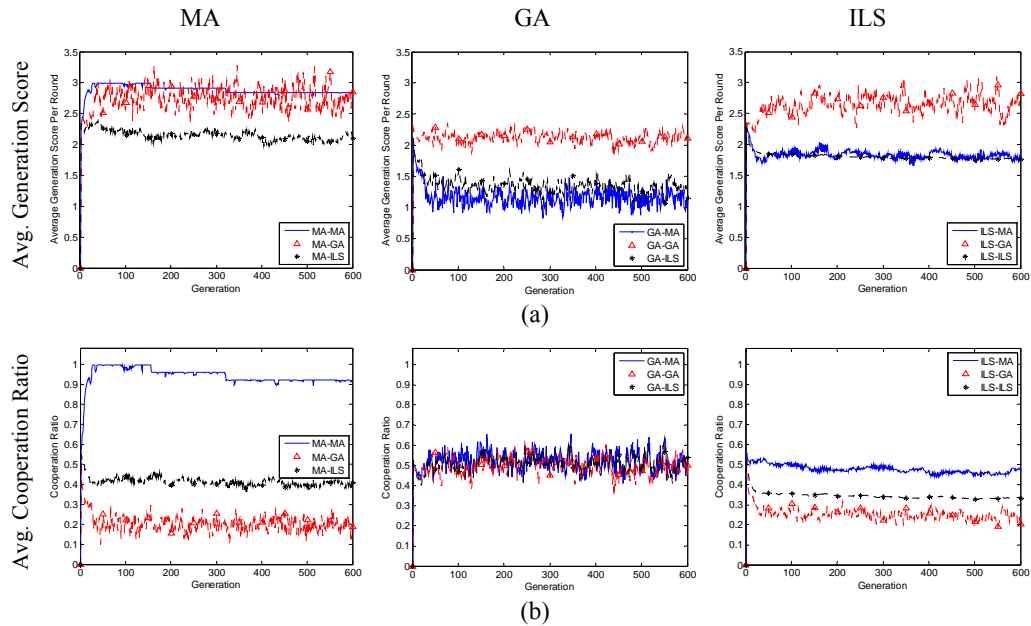Figure 4.14: Strategy specific (a) AGS and (b) ACR for MA, GA and ILS over 20 runs

112
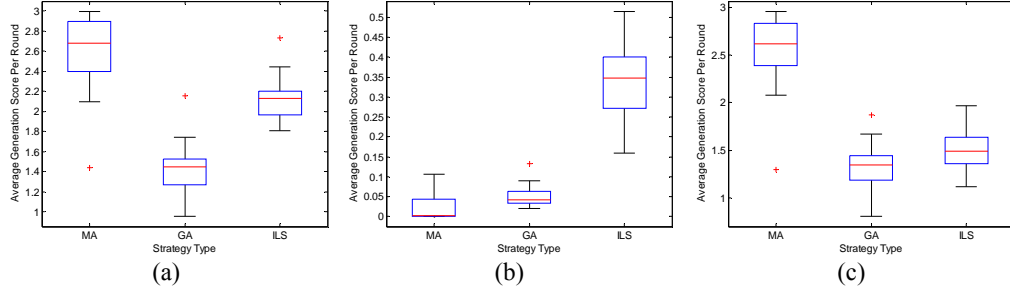
Figure 4.15: Box plots depicting distribution of (a) mean, (b) variance and (c) minimum AGS in the MA, GA and ILS populations when each plays against one another over 20 runs

Table 4.8: Proportion of runs that a row-wise strategy is better, similar and worse than a column-wise strategy

	Better (>)			Similar (≈)			Worst (<)		
	MA	**GA**	**ILS**	**MA**	**GA**	**ILS**	**MA**	**GA**	**ILS**
MA	NaN	0.95	0.90	NaN	0.00	0.05	NaN	0.05	0.05
GA	0.05	NaN	0.05	0.00	NaN	0.05	0.95	NaN	0.90
ILS	0.05	0.90	NaN	0.05	0.05	NaN	0.90	0.05	NaN

4.5.2.2 Test D: Performance of MA, GA and ILS in setup with 10 strategy types

To verify results from test C, performance of MA, GA and ILS are now evaluated in the presence of other benchmark strategies. The tournament comprises of MA, GA, ILS and seven benchmark strategies. Figure 4.16a shows that MA continues to preserve a large score margin above all strategies - duly achieved when IL and evolution complement and compensate each other's strengths and weaknesses. Compared to MA's AGS of 2.7, ILS attains an AGS of 2.4, ranked 3rd on the overall - just below TFT and above PAV. In the same order as test C, evolved GA strategies, with an AGS of 2.0, are strategically inferior to MA and ILS when exposed to both adaptive and deterministic strategies. Compared to test B, large fluctuation in GA's AGS is due to a need to adapt recurrently to the changing MA and ILS strategies. As part of the environment, they influence how GA strategies evolve and play a contributory role in lowering GA's performance with respect to benchmark strategies. Increase in complexity with the addition of seven strategies is indicated by a considerable rise in learning ratios for MA and ILS. Compared

against the results of test C – the frequency of learning and relearning increases as more diverse strategy types are encountered during game play (Figure 4.16b). Possible conflicts between evolution and IL also cause a distinctly higher learning ratio for MA over ILS during strategy improvement.
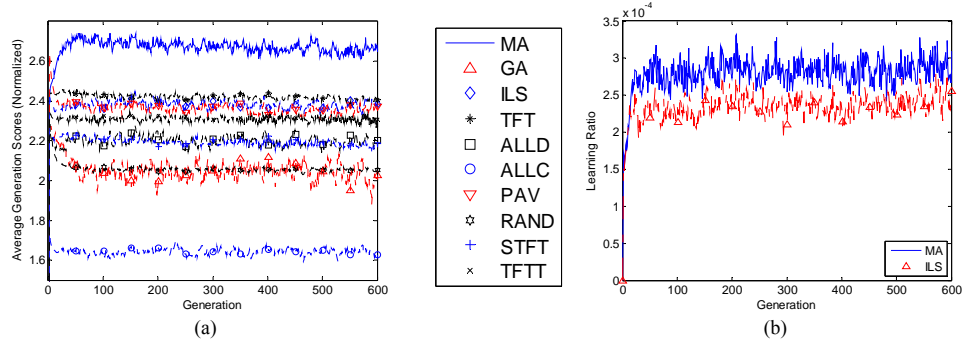
Figure 4.16: (a) AGS and (b) learning ratio obtained when MA, GA and ILS play with one another over 20 runs

Similarities between results observed in the performance box plots and those from test C ascertain the strategic dominance in descending order of MA, ILS and GA for a setup with adaptive strategies. Though GA also suffers large variance in mean and minimum AGS (Figures 4.17a and c) across runs, players still maintain a low intra-run AGS variance, suggesting that evolution is indeed important to smooth out score differences across players in the population. Results show that MA and ILS are still the dominant sources of influence, as the added benchmark strategies only exert trifling impact on GA's performance. As the role of evolution is to adapt strategies to their environment, traits of GA strategies are closely tied to the domain of interaction. As varying strategies emerge in distinct runs, diverse inter-run performance in the GA population is inevitably entailed.

ILS attains on average, a consistently huge intra-run score variance among players across the runs, suggesting that the ILS players perform differently despite learning under the same learning framework. The performance of IL strategies is

not significantly different from that of several other strategy types according to the results of paired T-test (Table 4.9 and 4.10). With the exception of ALLC and RAND, ILS perform in close similarity to strategies {MA, GA, TFT, ALLD, PAV, STFT, TFTT} respectively for {10, 15, 60, 15, 40, 5, 20} % of the runs. This huge intra-run disparity in performance is due largely to differing learning experiences that shape the traits of ILS strategies independently from one another. The absence of evolutionary pressure in correcting the large score differential between the ILS players presents a likely reason why ML surpasses IL in performance uniformity among players in the population.

(a)   (b)   (c)
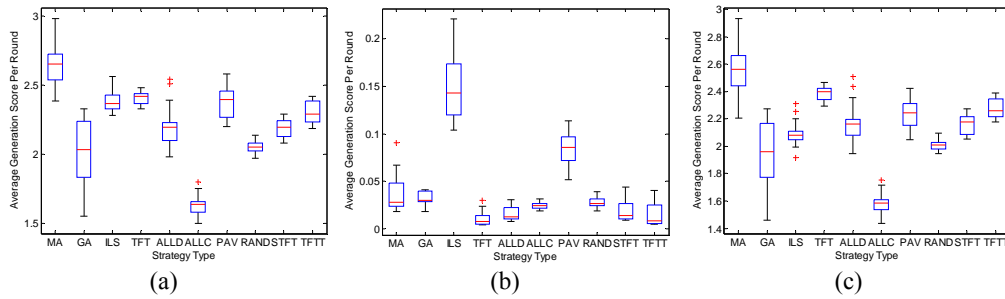
Figure 4.17: Box plots depicting distribution of (a) mean, (b) variance and (c) minimum AGS in MA, GA and ILS as each plays in the presence of benchmark strategies over 20 runs

Despite so, the higher mean and minimum AGS of ILS with consistently low variance across different runs indicates that IL is crucial for preserving good performance in the presence of adaptive strategies (Figure 4.17). An ability to perform strategy revision makes players more adaptable to the changing strategies of opponents. Observations show that MA attains the highest mean and minimum AGS and lowest intra group score variance (Figure 4.17). Variances for the above are also middling between GA and ILS. The sets of T-test results further verify the significance in the scale of difference across strategy scores. In combination with the box plot results, MA is notably different to other strategies at 5% significant

level. Overall, the collection of results depicts that ML breeds good strategies with fairly consistent performance across runs.

Table 4.9: Proportion of total runs that row-wise strategy is better than column-wise strategy

	MA	GA	ILS	TFT	ALLD	ALLC	PAV	RAND	STFT	TFTT
MA	NaN	1.00	0.90	0.95	0.95	1.00	1.00	1.00	1.00	1.00
GA	0.00	NaN	0.00	0.00	0.30	1.00	0.10	0.45	0.25	0.20
ILS	0.00	0.85	NaN	0.15	0.85	1.00	0.35	1.00	0.95	0.65
TFT	0.05	1.00	0.25	NaN	0.85	1.00	0.45	1.00	1.00	0.95
ALLD	0.00	0.65	0.00	0.15	NaN	1.00	0.05	0.90	0.50	0.20
ALLC	0.00	0.00	0.00	0.00	0.00	NaN	0.00	0.00	0.00	0.00
PAV	0.00	0.70	0.25	0.35	0.80	1.00	NaN	1.00	0.85	0.50
RAND	0.00	0.50	0.00	0.00	0.10	1.00	0.00	NaN	0.00	0.00
STFT	0.00	0.70	0.00	0.00	0.50	1.00	0.10	0.95	NaN	0.15
TFTT	0.00	0.80	0.15	0.05	0.80	1.00	0.20	1.00	0.85	NaN

Table 4.10: Proportion of total runs that two strategies are similar according to paired T-test

	MA	GA	ILS	TFT	ALLD	ALLC	PAV	RAND	STFT	TFTT
MA	NaN	0.00	0.10	0.00	0.05	0.00	0.00	0.00	0.00	0.00
GA	0.00	NaN	0.15	0.00	0.05	0.00	0.20	0.05	0.05	0.00
ILS	0.10	0.15	NaN	0.60	0.15	0.00	0.40	0.00	0.05	0.20
TFT	0.00	0.00	0.60	NaN	0.00	0.00	0.20	0.00	0.00	0.00
ALLD	0.05	0.05	0.15	0.00	NaN	0.00	0.15	0.00	0.00	0.00
ALLC	0.00	0.00	0.00	0.00	0.00	NaN	0.00	0.00	0.00	0.00
PAV	0.00	0.20	0.40	0.20	0.15	0.00	NaN	0.00	0.05	0.30
RAND	0.00	0.05	0.00	0.00	0.00	0.00	0.00	NaN	0.05	0.00
STFT	0.00	0.05	0.05	0.00	0.00	0.00	0.05	0.05	NaN	0.00
TFTT	0.00	0.00	0.20	0.00	0.00	0.00	0.30	0.00	0.00	NaN

In all, test D validates that MA players with the ability to evolve and learn at the same time, indeed perform better than GA, ILS and all other strategy types. Problems pertaining to the inability of GA players to cope with the increasing complexity of the environment and large score variance among ILS players have been aptly addressed by the synergy between evolution and incremental learning.

4.5.3   Case Study 3: Performance Assessment in Dynamic Environment

After assessing the performance of MA, GA and ILS in setups with a fixed pool of deterministic and random benchmark strategies in case study 1 and also adaptive strategies - case study 2, the final case study investigates the performance profile which arises when strategies are subjected to opponents that resume varying traits and characteristics constantly. The dynamic nature of the environment constitutes a good testing ground to validate whether strategies are resilient enough to cope

well on the whole. Relative adaptability of MA, GA and ILS in a dynamic setting can then be aptly addressed. In this study, MA, GA and ILS are setup to play with an opponent that changes its type probabilistically after every 1, 10-20, 100-150 generations. The sequence of change is made identical for MA, GA and ILS so as to ensure the existence of a common platform for comparison of results. Similar to MA, GA and ILS, the pool of dynamic opponents has **p** players.

4.5.3.1 Test E: Performance of MA, GA and ILS against dynamic opponents

From Figure 4.18a, the average performance of GA and ILS are actually on par despite the fact that GA does not have the luxury of altering the strategies in the course of each round robin tournament. MA's AGS is consistently higher than GA or ILS, demonstrating its superior adaptability in the dynamic setting. In order of increasing duration length between strategy changes, the trend of progression in AGS varies from rapidly fluctuating profile – when distinct strategy types are encountered every generation, to one that undergoes mild variation – when the opponent resume a certain strategy type for at least 100 generations. Complexity in the setup is correlated with an increase in the frequency of change in opponent strategy type and degree where IL is embraced (Figure 4.18b).

Zooming into a randomly selected run where the opponent changes every 50-100 generations, it is clear that the synergic blend of evolution and IL still confers the best performance, followed by IL and evolution (Figure 4.19a). As opposed to a fixed pool of opponents, good performance in the current setup comes with the ability to react spontaneously and appropriately to drastic changes in the nature of the environment. The poor adaptability of GA is due largely to a slow improvement rate and reaction of evolving players to the diverse nature of changing opponent types. In the absence of a guiding force like IL, it is unlikely

117

that evolution will find the optimum strategy to the changed opponent type in a short time amidst the trial and error search process. Innate ability to explore and adapt closely to changes in the environment across generations – as illustrated by a more fluctuating AGS profile, however, allows GA to surpass ILS when playing against some opponent types. Poor adaptability of ILS is due to onset of premature convergence and over-exploitation. Disparity in performance of IL players also contributes, in large parts, to the low performance of the population. Its ability to adjust dynamically to changing opponents through double loop IL allows players to assume strategies that are vastly different from those acquired when playing against previous types. This, in a way, makes them less dependent on opponents' traits and ensures smooth adjustments and performance consistency even when the opponent population transits between two widely different strategy types.

On the whole, MA depicts the most promising results by being able to attain full cooperation - AGS of 3.0, with players of its own type (Figure 4.19b) and performing best against all distinct opponent types that appeared in succession (Figure 4.19c). In summary, IL allows MA players to reevaluate their performance constantly during the game play. Any drastic change in the opponent's nature is captured by double loop IL, which introduces adequate variation to allow players to derive significantly different strategy traits, even when many have acquired considerably similar genotypes after numerous evolution cycles. As compared to random mutation in GA, IL makes changes more explicit. The search space to hunt for better strategies is expanded via the varying learning experiences of each MA player. Evolutionary pressure then comes into play to allow weaker players to adopt traits from those which have developed good strategies against the new opponent population. As long as a sizeable portion of MA population has acquired

118

the right perception about the desired response against the opponent strategy, this information is propagated quickly to other members of the community and overall strategy traits of the entire population are adjusted almost instantly. In summary, evolution supplements IL to improve the performance of each player over time, and eventually helps the entire MA population to adapt well on a collective basis.
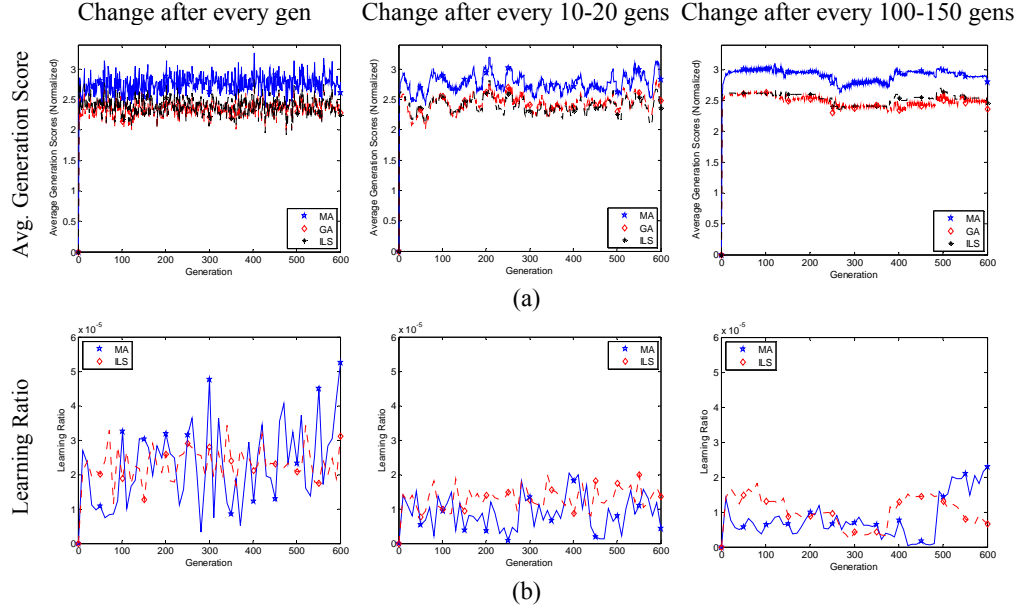
Figure 4.18: (a) AGS and (b) learning ratio for MA, GA and ILS as each plays with an opponent that changes dynamically every 1, 10-20 and 100-150 generations
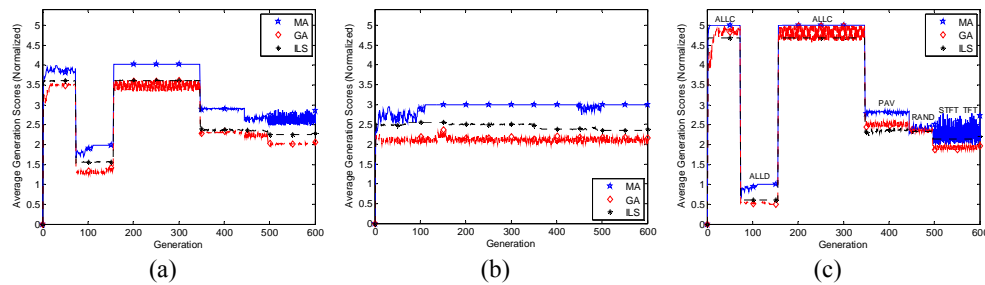
Figure 4.19: AGS attained as MA, GA and ILS play separately against (a) itself and opponent, (b) itself and (c) the opponent, when opponent's nature changes every 50-100 generations

4.6   Summary

In this chapter, the performance and adaptability of evolutionary, learning and memetic strategies are assessed in various IPD settings. Evolutionary strategies

are realized by GA based on co-evolutionary principles and learning strategies by a double-loop incremental learning scheme, ILS that incorporates classification, probabilistic update of strategies and a feedback learning mechanism. A memetic adaptation framework, MA is also developed to harness the synergy of evolution and learning. In this framework, learning assists the evolving strategies to acquire good strategy traits and react spontaneously to changes in the environment, while evolution provides an avenue for knowledge exchange between players so that the disparity in performance between learning strategies is minimized.

The comparative case studies that are conducted for different environmental conditions showed that the players adapted by MA exhibit superior performance relative to GA and ILS. GA is found to be slow to react to environment changes and its performance also deteriorates against adaptive opponents. ILS, on the other hand, suffers from diverse learning experiences among individuals and a tendency to over-exploit which undermine the performance of the entire population. The combination of incremental learning and evolution, however, allows MA players to balance the tasks of exploration and exploitation of diverse strategies while preserving its trend of dominance consistently. It is gathered from the chapter that both incremental learning and evolution are essential elements of adaptation in the IPD game. Their concurrent interaction is crucial for the formation of strategies that will adapt and thrive well in complex, dynamic environments.

Chapter 5

Modeling Civil Violence

After verifying the ability of CEAs to evolve competent strategies in both poker and IPD, and give insights about the respective problems via meaningful analysis, we shift our attention to model and analyze interesting social phenomenon. Civil violence, in particular, is widely used in the context of modern society to describe associated acts of violation and destruction, which are carried out as a sign of defiance against a central authority or between opposing groups. It is manifested in many forms and categorized according to the nature, degree of involvement and severity of conflict. These can range from small-scale riots and demonstrations to large-scale revolutions such as civil and ethnic wars. Researchers have sought to interpret the causes and effects from various perspectives. In social conflict theory [158], sociologists consider unrest as the result of socio-economic instability [159]. Economists adopt an opportunist's viewpoint by relating rebellion to profits [160]. Political scientists question the motives and attribute upheavals as the result of resource or political deprivation [161], [162].

The perspective of associating civil violence with pent-up grievances has varied widely. Collier and Hoeffler [163] have looked at possible economic causes while Regan and Norton [164] have regarded it as a function of mass mobilization. Substantive differences are also identified in both ethnic and non-ethnic motivated violence [165], [166]. Since "each war is as different as the society producing it" [167], causes should be analyzed using the nature of conflict. Despite compelling differences in views, the underlying structure of conflict remains similar with respect to widespread, collective, random movement of crowds and interactions

between people [168]. Empirical models of social conflict in the form of riot games [169], game theoretic models [170], [171] and social networks - simulated to offer statistical, spatial-temporal analysis [172] of conflict and its role playing dynamics [173] in crowds, generate emergent social phenomena such as behavior clustering [174], mass-mobilization [175] and massive conflicts [159], which are indisputably a reflection of real-life conflicts. Although these models provide a good avenue to study strategies for managing civil violence [176], none actually accounted for the autonomous behavioral evolution of agents, which is consistent with the fact that humans learn and adapt.

The chapter focuses on the design and development of a spatial Evolutionary Multi-Agent Social Network (EMASN) to simulate and study the macroscopic-behavioral dynamics of civil violence, as a result of microscopic game-theoretic interactions between goal-oriented agents in various situational settings. Inspired by evolutionary computation, agents modeled from multi-disciplinary perspectives [177] have their strategies evolve over time via co-evolution [64], [178], [179] and learning. Experimental results reveal some fascinating emergent phenomena and interesting patterns of agent movement and autonomous behavioral development [180]. The results analysis establishes micro-macro [159], [181] interconnections between the attributes of conflict and provides new insights into the rich dynamics which arise from unrest. Collectively, the EMASN framework facilitates the study of autonomous emergent behavior and serves as an avenue to gain a more holistic understanding of the fundamental nature of civil violence.

The organization of this chapter is as follows: section 5.1 presents a short literature review of existing works and the general framework of EMASN. Section 5.2 introduces the model specifications. Section 5.3 focuses on the discussion of

the evolutionary and learning mechanisms that drive the autonomous behavioral changes in agents as they move and interact in the model. Section 5.4 evaluates the series of simulation results based on different model extensions to analyze the effects of various parameters on the behavioral response of the model. Section 5.5 presents a broad summary of discussion, highlighting the significant results while Section 5.6 concludes the paper with an overview and some comments on areas where future work can be embarked on.

5.1 Evolutionary multi-agent social network

5.1.1 Overview

The popularization of game theory saw a widespread usage of agent-based [177] approaches to model human entities as rational utility maximizers. Applications can range from the study of cooperation in the IPD to the modeling of investors' behavior in stock markets [182]. There is also a paradigm shift from the traditional top-down approach to a bottom-up approach [183] of studying emergent system behavior through the collective microscopic interaction among individual agents within complex multi-agent systems.

Numerous empirical-based computer simulations have been constructed over the years to model complex dynamic systems [184], [185] across many disciplines. The procedure involves decomposing complex verbal theories and then translating them into semi-mathematical equations which are integrated systematically into the model design. Developed models are simulated over time to create meaningful trends and patterns for the purpose of analysis and drawing of conclusions. The same approach was also taken by Epstein [173] to model decentralized rebellion against a central authority and communal violence between two ethnic groups. By

virtue of its simplicity, the elegant agent-based computational model was able to replicate salient features of violence dynamics through simple empirical rules and equations. The MANA model [176] then extended Epstein's model by introducing specific movement strategies which were aimed at correcting the purely random agent movement. Situngkir [159] also modeled the phenomena of massive conflict by invoking its analogy with the macro-micro link in Sociological Theory.

5.1.2   EMASN Framework

Inspired by existing models [173], [175], [176], and concepts in EGT [36], [186], the proposed EMASN framework consists of a civil violence model (CVM) and an evolutionary engine (EE) (Figure 5.1).
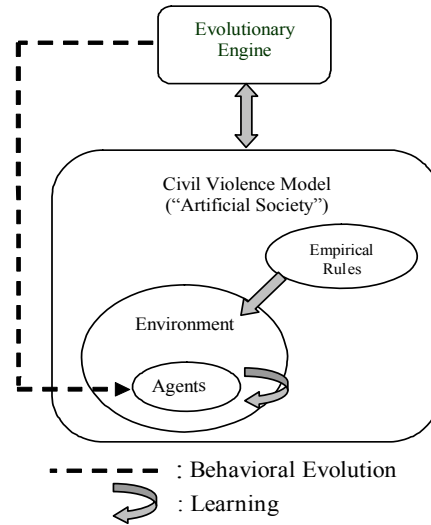
Figure 5.1: Framework of the Evolutionary Multi-Agent Social Network

As agents interact game theoretically, their strategies are evolved via independent learning and collective co-evolution using a CEA. Overall, EMASN attempts to model events of conflict by integrating the complexity of human behavior and random nature of crowd movement. Emphasis is placed on the modeling of agents and their interactions through behavioral rules at the microscopic-level, so as to

124

recreate the macroscopic-emergent outcomes that agree well with contemporary views. Descriptions of components and functionalities of the CVM and EE will be provided in the subsequent sections.

5.1.3 Game theoretic interaction

The CVM models interaction using features of a spatial IPD game [89], [187]. Though simplistic, this approach deals with complex issues of decision-making and self-interest [89], where the underlying concepts are subtle but far-reaching. Analogy between the CVM and IPD frameworks can reveal interesting dynamics on how agents maximize their benefits in view of situational changes [188].

1) General game play: A total of three different agent types are specified in the model. They are specified and denoted as follows:

- Quiescent Civilians - ◯
- Activists - 🔴
- Cops - ✪

At any time instance, movement is subjected to the spatial constraint of a 2D-Grid and only cops and activists are allowed to interact. Each agent establishes game play with every other opposing agent (e.g. cops and activists) within the vision radius in a pair-wise manner. No interaction will take place between agents of the same type and isolated agents (Figure 5.2).

| (a) | (b) | (c) |
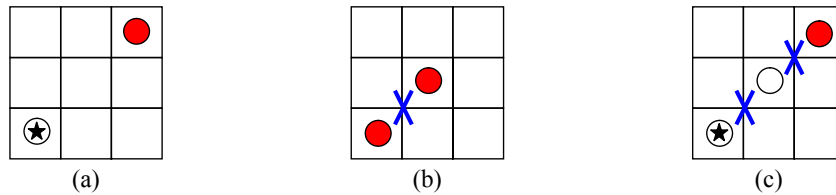
Figure 5.2: No interaction between (a) isolated agents, (b) like agents and (c) quiescent and other agents.

2) Payoff matrix: In view of the possible strength differences between opposing agents in any spatial neighborhood, the CVM adopts a payoff scheme that varies in accordance to the situation of encounter. A set of three payoff matrices which correspond to the scenarios where the number of cops is equal to, more than or less than the activists are shown respectively (Table 5.1a-c). Each player has the option to Cooperate or Defect. An analogy is drawn between the meaning of Cooperate and Defect in the IPD and CVM. This differs for each agent type. Adapting from the MANA model [176], activists cooperate by not challenging cops, revolting aggressively or instigating civilians to revolt, and accept peaceful settlements; and defect otherwise. Cops cooperate to protect civilians and defect to pursue rebels. Each group has conflicting goals that involve tradeoffs. Activists aim to create havoc and garner support from the quiescent civilians to fuel the ongoing unrest while avoiding arrest. On the contrary, cops aim to maintain order within the regime while fulfilling their role as protectors by minimizing casualties.

Table 5.1: Payoff matrix when number of cops is (a) equal to, (b) greater than or (c) less than the activists in sight

		Activists	
		COOPERATE	DEFECT
Cops	COOPERATE	3,3	0,5
	DEFECT	5,0	1,1

(a)

		Activists	
		COOPERATE	DEFECT
Cops	COOPERATE	1,4	2,2
	DEFECT	3,3	4,1

(b)

		Activists	
		COOPERATE	DEFECT
Cops	COOPERATE	4,1	3,3
	DEFECT	2,2	1,4

(c)

3) Rationale for different payoff: The matrices are constructed based on the goal of each agent group to maximize its benefit and minimize casualties in different

situations of contact. Consider the scenario when cops outnumber activists in a particular spatial area of interaction.

a. If both groups Defect, cops will gain the upper-hand due to their superiority in numbers. Their successful intervention to stem the unrest should be rewarded the temptation payoff (T). Activists, due to huge casualties should be rewarded the sucker payoff (S).

b. If both groups cooperate, payoffs reverse in favor of the activists as cops have missed a good opportunity to make arrest while deciding to protect the general population. For the activists, Cooperate paid off as they successfully avoided conflict with the massive cop population.

c. If cops Cooperate and activists Defect, both groups get the Punishment payoff (P) as cops should have Defect to confront activists while activists should have Cooperate in order to avoid challenging the domineering law enforcers openly and inviting casualties.

d. If cops Defect and activists Cooperate, a logical equilibrium is attained as the majority exerts dominance over the minority whilst the latter avoids direct conflict. This is the best situation as the benefits of both groups are accounted, justifying reward payoff (R) for both.

Assuming symmetric game-play, the above settings are considered in the reverse manner when activists outnumber cops. Despite the addition of two new matrices to account for strength disparity between cops and activists, the goal of agents in the CVM is similar to players in the IPD as each seeks to maximize the eventual payoffs through its interaction in different setups.

5.2 Civil violence model

The CVM consists of multiple agents interacting and coexisting in an artificial society; a computational structure where new social theories can be verified or developed [189]. It is composed of three distinct inherent components: The agents, environment and a set of empirical rules.

5.2.1 Agents

Agents form the crux of the CVM. Accurate modeling of their attributes is crucial for a close-life depiction of human behavior in situations of civil upheaval. In the CVM, the quiescent civilians are neutral members of the community who thrive amidst unrest and hardship. They pose no danger to the central authority but do respond to internal and external stimuli from time to time. They remain peaceful and law-abiding but turn active if conditions are favorable to express their anger and frustration publicly. Cops maintain order in the regime by arresting activists and play a crucial role in determining the success of violence control strategies.

1) Basic attributes: In line with Berdal and Malone [190], CVM models grievance and greed [191] as the two idealized components which collectively measure the tendency of joining the revolt. Apart from the heterogeneous perceived hardship [173] that is modeled endogenously as $H_{endo} = U(0,1)$, the definition of hardship is extended to account for its correlation with the level of unrest via an exogenous, homogenous attribute. The rationale is that civilians face the added burden and psychological trauma of fear due to looting, pillaging and repression from the authority in an effort to root out potential rebels. This is mathematically expressed:

$$H_{exo} = \frac{A_{\Sigma}}{A_{\Sigma} + \overline{A_{\Sigma}}}$$

(5.1)

128

where A_Σ and $\overline{A}_\Sigma$ refer to the number of activists and quiescent civilians present in the CVM. The overall hardship experienced by a typical civilian is formulated as:

$$H_{overall} = 0.5 \cdot H_{endo} + 0.5 \cdot H_{exo} \qquad (5.2)$$

Legitimacy $L = U(0,1)$ refers to the perceived legality of the central authority [173] and is uniform for all agents. Grievance (G) is defined as a function of $H_{overall}$ and L in the form shown in (5.3) where H_{exo} accounts for the possible changes in G as the condition of unrest changes from time to time.

$$G = H_{overall} \cdot (1 - L) \qquad (5.3)$$

Greed $Gr = U(0,1)$ is the perceived opportunity to gain wealth. In economic perspectives, Gr will be much dependent on how lucrative the revolt is. When viewed psychologically, a greedy agent will have a high tendency to rebel even if opportunistic gain is small. According to Collier and Hoeffler [191], G triggers a revolt while Gr sustains it. Tendency to revolt (Rev) is formulated as

$$Rev = T_f \cdot G + (1 - T_f) \cdot Gr \qquad (5.4)$$

where $T_f = [0,1]$, as shown in (5.5) is a time factor that is inversely related to the active duration T_{ad} of a rebel:

$$T_f = \exp(-0.5 \cdot T_{ad}) \qquad (5.5)$$

Besides G and Gr, the decision to revolt depends on the net risk that an agent is exposed to. This is modeled from three dimensions – the inclination to take risk, probability of getting caught and the jail term to serve upon arrest. Risk aversion, $R_A = U(0,1)$ is an agent's willingness and capacity to subject itself to

129

danger. Likelihood of arrest, P_a depends on the agent's vision radius, $VR(Ag)$ and the ratio of cops to activists within sight of an 8-neighbor radius (Figure 5.3).
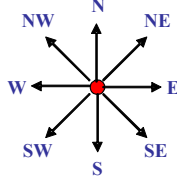
Figure 5.3: 8-Directional Agent Vision Radius

Information is local in the CVM and decisions regarding revolt, direction to move etc are governed by information available within the agent's field of vision. Formally, P_a is given by

$$P_a = 1 - \exp\left(-k_b \cdot \frac{\overline{A}_{VR}\left(C_p\right)}{\overline{A}_{VR}\left(A\right)} \cdot (2 \cdot VR(Ag)+1)^2\right) \tag{5.6}$$

$$k_b = -\frac{\ln 0.1}{(2 \cdot VR(Ag)+1)^2} \tag{5.7}$$

$\overline{A}_{VR}(C_p)$, $\overline{A}_{VR}(A)$ and $\dfrac{\overline{A}_{VR}\left(C_p\right)}{\overline{A}_{VR}\left(A\right)} \cdot (2 \cdot VR(Ag)+1)^2$ denote the number of cops, activists and cop-to-activist ratio within the vision radius. k_b introduces a bias such that cops have 90% of making a successful arrest in a one-one situation with an activist. Jail term, J is defined as the number of time episodes that an agent is put out of action and is expressed as

$$J = \begin{cases} 1 + J_{max} \cdot \dfrac{J_H}{J_{H\,max}} & \text{if } J_H \leq J_{H\,max} \\ I_{max} & \text{otherwise} \end{cases} \tag{5.8}$$

J_{max} is the maximum jail penalty, J_H is the number of times an agent is caught formerly while J_{Hmax} is the maximum number of times tolerable for the repetition

130

of crime in society's view. In a variable jail term formulation, the fixed part is the minimum sentence while the variable part accounts for the increasing penalty for repeated offenders. I_{max} is a large number denoting life imprisonment. In a fixed jail term formulation, J is constant. The net risk N, perceived by an agent with an intention to revolt is modeled by (5.9) where J_α determines the deterrent effect of the jail component.

$$N = R_A \cdot P_a \cdot J_{\max}^{J_\alpha}$$ (5.9)

2) Game theoretic attributes: Apart from basic attributes, a different set of traits is set out in Table 5.2 to account for the results of game-theoretic interactions.

Table 5.2: Summary of Game Theoretic Agent Attributes

Parameters	Description
$Ag_{strategy}$	14-bit binary string that encodes the behavior of an agent.
Pay_{acc}	Effectiveness of an agent strategy over the course of simulation.
Pay_{gen}	Effectiveness of an agent strategy over the previous generation.
GH_{acc}	Number of games played over the course of simulation.
GH_{gen}	Number of games played over the previous generation.
GS_{lost}	Number of lost game sets over the last generation.
R_{co}	Ratio of cooperative games to the total games played.
R_{def}	Ratio of defection games to the total games played.
SH	Number of "successful trials" accumulated over the previous generation.
FH	Number of "unsuccessful trials" accumulated over the previous generation.

5.2.2 Empirical rules

Empirical rules govern agent interaction and ensure proper functioning of CVM. They are crucial to the formation of desired simulation outcomes that depict close replicas of unrest vividly described in numerous literatures on revolution and wars.

1) State Transition Rule: Civilians turn active when $NAI = \text{Rev} - N$ is larger than a predefined threshold ($A_{threshold}$) and stay quiescent otherwise. A summary of state transitions is presented (Table 5.3). Cops are assumed to be loyal to the cause of

the regime and insusceptible to any form of bribery e.g. state transitions will only occur between activists and quiescent civilians.

Table 5.3: Summary of State Transition

Current State	NAI - AT	State Transition
Quiescent	> 0	Quiescent → Active
Quiescent	≤ 0	Quiescent → Quiescent
Active	> 0	Active → Active
Active	≤ 0	Active → Quiescent

2) Jail Release Rule: In contrast to the MANA model [176] where jailed agents revert to the active state after release, the CVM allows transition to the quiescent or active state with a certain probability. This takes into account the possible onset of rehabilitation – high chances of converting jailed agents back to law-abiding citizens and also the curse of the minority – a low possibility that persistent rebels will continue in their old ways upon release (Figure 5.4).

$P_{convert} = 0.9$ $P_{revert} = 0.1$

Quiescent ⟷ Active ⟶ Jailed
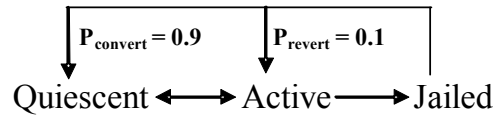
Figure 5.4: State transition flow diagram between different agent states

3) Movement Rule: The movement of agents between consecutive episodes is modeled using a set of simple update rules, such that the position of each agent in the next time episode is determined by its current position and existing states of all cells within its vision radius. In particular, these rules are adapted from John Conway's Game of Life, which states that a cell with one or no neighbor dies of loneliness while one with four or more neighbors dies due to overpopulation. Only those with two or three neighbors survive. By treating dynamic agents in the CVM as subjects of concern e.g. cells; and drawing an analogy between loneliness and isolation as well as overpopulation and overcrowding, simple rules (Table 5.4) are

devised to govern movement on the grid. The underlying rationale is that isolated agents are likely targets of attack by opposing agents and they prefer to move to safer spots if like agents in the vicinity is low. Likewise, densely packed agents tend to move towards sparsely filled regions to avoid the danger of losing sight of the situation. The destination cell that an agent moves to is randomly chosen from the set of vacant ones that are adjacent to it. On top of these basic rules, each agent type also has its own preference movement strategies as shown in Table 5.5.

Table 5.4: Basic movement rules in the CVM

Movement Rules	Description
Rule 1	An agent will remain in its original position at the next time instance if the number of neighboring agents within its vision radius is **2**, **3** or **8**.
Rule 2	An agent will move to a new position at the next time instance if the number of neighboring agents within its vision radius is **1**, **4**, **5**, **6** or **7**.

Table 5.5: Preference movement strategies for different agent types

Movement strategies	Agent Type	Description
Avoid the Cops	Activists	Activists attempt to minimize contact with cops in order to lower the chances of arrest.
Stay if favorable	Activists	Activists prefer to stay put rather than venture out into the unknowns if the current location is safe.
Eradicate the Civilians	Activists	Activists take initiative to root out and eradicate any unarmed civilian in sight.
Pursue Activists	Cops	Cops take initiative to arrest activists.
Protect Civilians	Cops	Cops take initiative to protect the general population from the threats of activists.
Run from Activists	Quiescent	Quiescent civilians run for their lives when activists are on a killing spree.

4) Arrest Rule: An arrest is made at any episode when a cop wins an IPD game set against an activist in the neighborhood of interaction. Taking into account the onset of group effect, cops will have higher chances of apprehending activists if they are pursuing the same target.

5.2.3 Environment

The environment defines an *N x M* space where all agents move and interact. Coupled with global and situational parameters, it allows access to information

about the state of unrest and an overview of the spatial interaction between agents on both global and local scales.

5.3   Evolutionary Engine

Following the CVM formulation, it is crucial to devise channels for agents to better their performance in the stochastic model. This is achieved by improving their strategies over time through evolution and learning. The two processes allow agents to shape their behaviors and react aptly to unforeseen circumstances. This ability gives rise to interesting behavioral dynamics and provides insights into the autonomous behavioral development of different agent types.

5.3.1   Evolution of Agent Behavior

At the start of each generation, CVM passes agents to the EE where co-evolution of strategies takes place (Figure 5.5). Groups are evolved independently by the EE (Figure 5.6). This framework has since been used to implement search heuristics [192], generate pursuit and evasion behavior [193], analyze population dynamics [194] and study cooperation in EGT models [195], [196]. In recent literatures, this approach gained further significance via a finding where co-evolution improves computational performance in fitness prediction [197] as well as a successful attempt to measure generalization performance in co-evolutionary learning [198]. In the proposed model, co-evolution is conceptualized in analogy to the exchange of ideas, in reality, between members of the same group. Through the multi-directional flow of information, agents with weaker strategies learn from stronger ones by adopting some of their better traits. Overall fitness of each group is raised as more competent strategies are discovered with each elapsed generation. This

134

added dimension of realism will enable the analysis of interesting outcomes in agent interaction across different scenarios as strategies co-evolve over time.
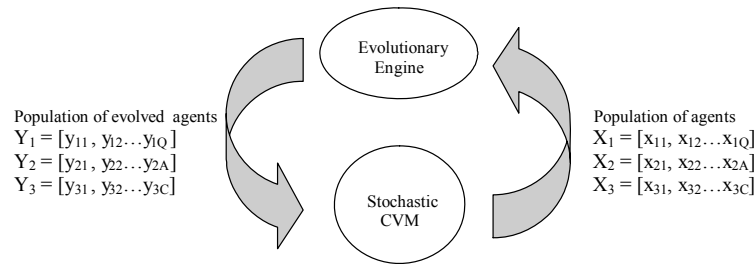
Figure 5.5: Relationship between EE and CVM
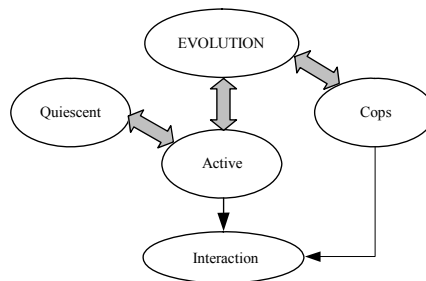
Figure 5.6: Co-evolution of different agent groups

1) Chromosomal Representation: Each agent is defined by a 14-bit binary string which encodes the strategy bits ("1" – Cooperate or "0" – Defect) used in different situations (Figure 5.7). The next move will be decided based on previous moves made by both the agent and his opponent. In the absence of move history, this is determined by bits encoding initial conditions.
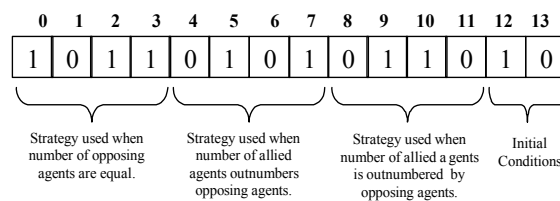
Figure 5.7: Binary encoded genotype for agent strategy

2) Fitness Representation: Effectiveness or fitness of each strategy is defined in terms of the normalized payoff per game over the previous generation. This is mathematically given by (5.10). To ensure that fitness evaluation is meaningful, only agents that have played at least one game set in the previous generation are included in the respective sub-populations for evolution.

$$\left\langle Pay_{gen} \right\rangle = \frac{Pay_{gen}}{GH_{gen}} \tag{5.10}$$

3) Elitism: Elitism is employed to ensure that strategies that perform substantially well are adopted by agents in the next generation. The average generation fitness (AGF) of each population is used as the level for implementing elitism. This is given mathematically in (5.11). Strategies with payoffs larger than AGF are given priority for reuse so that the population can continue to benefit from them in the next generation. These above average strategies serve as a benchmark for others to learn and evaluate themselves with.

$$\left\langle Fit_{gen}(population) \right\rangle = \frac{\sum_{i=1}^{Agents} \left\langle Pay_{gen}(i) \right\rangle}{Ag_{total\ valid}} \tag{5.11}$$

4) Selection: The EE performs a dual-stage selection. The first stage uses binary tournament selection without replacement to avoid multiple selection of a strategy at expense of others. The second stage is analogous to a local search operation where agents are selected arbitrarily and subjected to slight perturbation [199] to allow preservation of good genes of inferior strategies.

5) Genetic Operators: Uniform crossover is performed to simulate knowledge exchange between strategies which are propagating to the next generation. This ensures that desired traits of good strategies are passed on to the offspring in an

attempt to create even better strategies in the subsequent generations. Binary bit-flip mutation is also implemented to introduce diversity into the population.

6) Algorithmic flow: The new offspring population will constitute a collection of evolved strategies that will be used in game play as agents move and interact over the next generation. A generalized algorithmic workflow of the evolution process is shown in Figure 5.8.
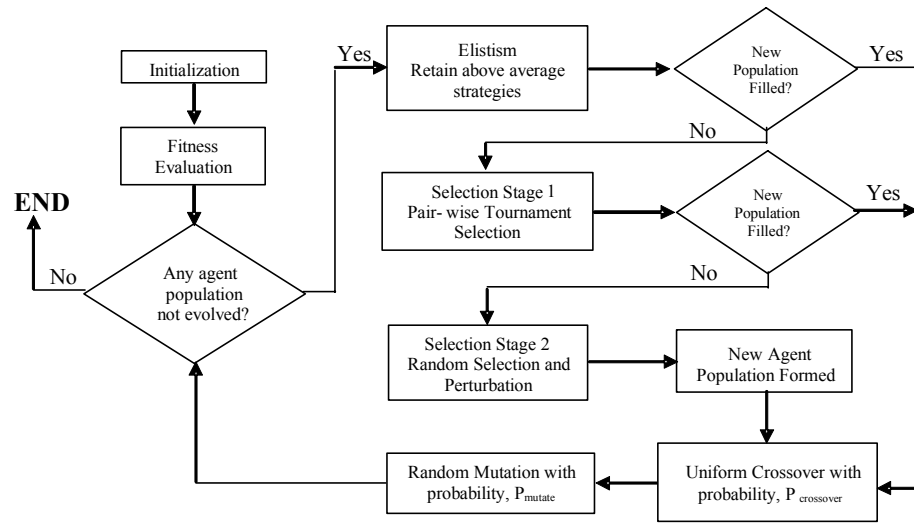
Figure 5.8: Workflow for evolution of agent strategies

5.3.2   Learning

Learning is a form of local search operation that is carried out by agents to better their performance based on some form of heuristics which uses domain-specific information available on hand. In contrast to evolution, learning is performed independently without exchange of information.

1) Significance of Learning: In the absence of information sharing, man is a good instance of an entity that is capable of acquiring knowledge and making complex, independent decisions. In a dynamic CVM, the analogy applies as agents learn on

the go and react to unforeseen circumstances by altering their strategies on a timely basis. Learning is typically performed based on partial memory [150] of past experiences. In addition to performance enhancement, learning also facilitates the autonomous behavioral development of agents and brings out the interesting dynamics that will be insightful to the study of human behavior in different setups.

2) Design of Learning Heuristics: In the absence of information exchange and collaboration, agents are simply entities which are required to learn independently by making use of information they perceive. In view of the variegated situations encountered by agents, diverse learning experiences will inevitably be entailed. Good learning strategies in general, are those which respond and adapt well to opponent strategies, and in the process achieve maximization of overall payoff. The basis of learning adopted in the paper is devised by keeping a record of the number of successful and unsuccessful trials accumulated by all strategy bits in the course of game play. This is determined by a set of three performance matrices which corresponds to the three payoff matrices used in the CVM (Tables 5.6–5.8).

Table 5.6: Performance Matrix when number of agents is equal to opposing agents in sight

		Player 2	
		COOPERATE	DEFECT
Player 1	COOPERATE	Success	Failure
	DEFECT	Success	Failure

Table 5.7: Performance Matrix when number of agents outnumbers opposing agents in sight

		Player 2	
		COOPERATE	DEFECT
Player 1	COOPERATE	Failure	Failure
	DEFECT	Success	Success

Table 5.8: Performance Matrix when opposing agents outnumbers number of agents in sight

		Player 2	
		COOPERATE	DEFECT
Player 1	COOPERATE	Success	Success
	DEFECT	Failure	Failure

The heuristics are conceptualized based on Pavlovian Learning, where T is considered "success" while S is considered "failure". R and P are considered as "success" and "failure" respectively. The number of "success" and "failure" trials indicates whether a strategy bit should be revised. Intuitively, better strategy bits have larger "success" than "failure" count while the weaker ones have a larger "failure" count. Bits which are not used have zero for both counts. The search process only revises a strategy bit if the following criterion is met.

$$SH < FH + 10 \cdot GS_{lost} \qquad (5.12)$$

The number of Lost Game Sets (GS_{lost}) is introduced as a means to penalize the strategies which are losing consistently. This ensures that the strong and desirable strategy bits are likely to remain intact and weaker ones are susceptible to change. Learning is performed incrementally in an iterated game set after every k games, where k is a variable learning parameter. In general, a lower k value results in a larger number of learning cycles and higher learning rate. For simplicity, k is fixed in CVM. Incremental learning is used to correct undesirable traits of the current strategy and allows better adaptation to the environment of opponent strategies.

5.4 Simulation results

Simulation runs are carried out using Microsoft Visual C++. A summary of the parameter values used is depicted in Table 5.9. In all runs, size of 2D environment is fixed to ensure uniformity in the result analysis across different case studies. A mean of 200 games is played in a pair-wise manner between any two opposing agents which are one square radius from each other at any instance. Simulation duration is set at 1000 or 5000 episodes while the number of agents and initial

population composition vary according to test scenarios. All agents have a vision radius of one and quiescent civilians turn active beyond a threshold of 0.2. On jail release, agents turn quiescent and active with probabilities 0.9 and 0.1 respectively.

Evolution is triggered every 5 episodes and learning every 20 games. Model extensions of varying complexity are introduced in the following sections to track behavioral development in each agent population. Due to the stochastic nature of CVM that arises from the disparity in interaction pattern of agents over different episodes, it is crucial to conduct several simulation runs to obtain more consistent depiction of the outcome. This serves to minimize stochastic variation and verify consistency of any behavioral trends that are observed across the runs.

Table 5.9: List of parameter values used in the simulation runs

CVM Parameters:	Values
Size of 2D Grid, Sz_{grid}	20 x 20 squares
No. of games in an iterated game set, α	200
Total number of simulation time episodes, T_{max}	1000 or 5000
Total number of agents, Ag_{Σ}	variable
Agent vision radius, $VR(Ag)$	1 square radius
Active Threshold, $A_{threshold}$	0.2
Probability of successful rehabilitation, $P_{convert}$	0.9
Probability of reverting back to active state, P_{revert}	0.1
Tournament selection size, $Sz_{tournament}$	2
Perturbation probability, $P_{perturb}$	0.02
Crossover rate, $P_{crossover}$	0.8
Mutation rate, P_{mutate}	0.05
No. of time episodes per evolution cycle, g	5
No. of games played per learning cycle, k	20

5.4.1 Basic CVM Dynamics

This section validates the correctness of the CVM by comparing its basic response with the dynamics that were observed in Epstein's model. This is paramount as it lays the foundation for the development of more complex extensions and serves as a standard to which the results in subsequent models can be compared with.

Simulations are performed when number of cops, $N_C = 60$ and $J =100$ for a civilian population of 180. Agents are interacting vis-à-vis a game theoretic setup and subjected to periodic cycles of evolution and learning. Success of arrests in a spatial neighborhood depends on the relative competency of the cop and activist strategies. With activists adopting *Avoid the Cops* and *Stay if favorable* and cops using *Pursue Activists*, the resulting active ratio - denoting ratio of activists to civilians, is plotted in Figure 5.9. The characteristic plot depicts a fluctuating waveform which is persistently plagued by short term instances of unprecedented up-shoots. This coincides with Epstein's notion of a "Punctuated Equilibria" - which states that long periods of relative stability are punctuated by outbursts of rebellious activities. From this, it can be clearly inferred that peace and stability is a dynamic equilibrium which emerges from the interaction between agents rather than a static equilibrium itself. Any factor that alters the mode of agent interaction or movement will cause changes in the macroscopic temporal response. Presence of such features serves as a preliminary step towards verifying the validity of the CVM. More intuitive substantiation will be done in the upcoming sections.
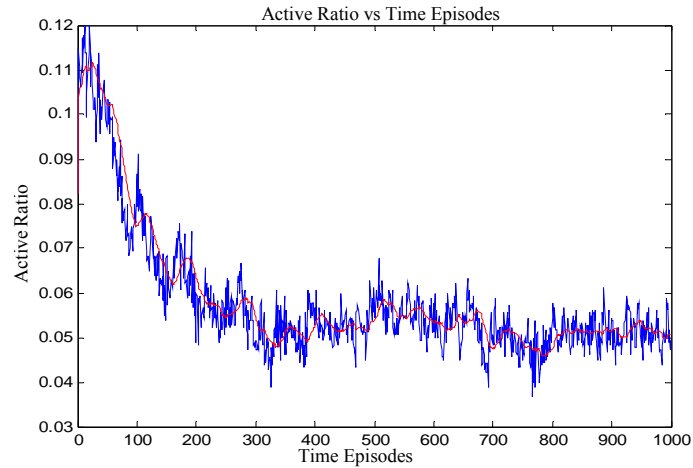
Figure 5.9: "Punctuated Equilibria" in temporal response of CVM

5.4.2 CVM Response under varying N_C

Besides verifying the consistency and soundness of the CVM, its response can be studied from other perspectives. This section investigates the effect of varying N_C on aspects of its temporal and spatial dynamics. Interesting emergent behaviors are explored, followed by an analysis and discussions on salient CVM attributes.

1) Impact of N_C on basic temporal dynamics: The active ratios for N_C = [0, 20, 60] are plotted. The long term profile with 0 cops is analogous to a step response that starts with a steep rise in rebel activities and ends with 58% of the population on revolt - maximum amount of rebel activities that can occur with the pre-defined $A_{threshold}$ (Figure 5.10a). The short term profile fluctuates about the long term mean with considerable regularity. As N_C is increased to 10, the plot undergoes milder long term variations, encompassing a steady drop in active ratio followed by some notable changes and transitions (Figure 5.10b). The system response is unstable due to insufficient cops to suppress the rebel activities within a stable equilibrium level. With 60 cops, peak and settling ratios are reduced and a substantial degree of stability is attained since added cops are able to perform the task of arrest more effectively (Figure 5.10c). The characteristic waveform, nonetheless, gets more fluctuating as discerned by occurrences of instantaneous outbursts with significant peaks occurring over shorter intervals. The resulting equilibria are "punctuated" with frequent alternation between periods of unrest and stability.

(a) (b) (c)
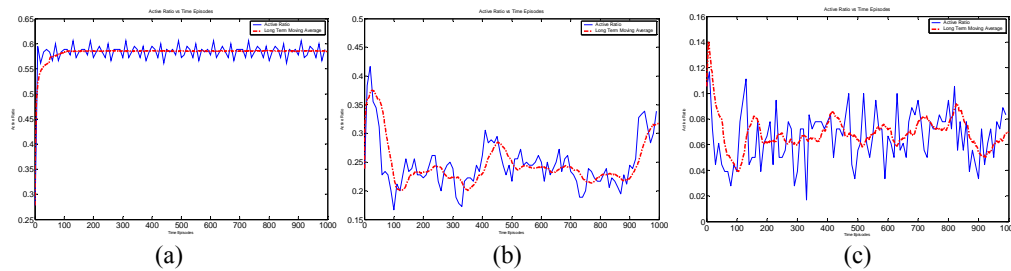
Figure 5.10: Temporal response for (a) 0, (b) 10 and (c) 60 cops

Figure 5.11 shows the overview of simulated profiles corresponding to different N_C. An inverse relationship between active ratio and N_C is correctly established as verified from the collection of temporal response curves. The decline in both peak and settling ratios gets less significant with more cops, indicating presence of a saturation level – where help rendered by each cop addition decreases as N_C gets larger. The drop in marginal contribution implies that further reduction in active ratio will have to come from other aspects of improvement.
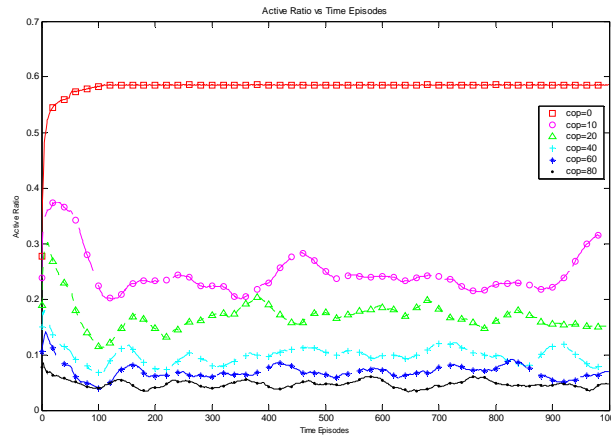
Figure 5.11: Family of temporal response curves for different N_C

2) Impact of N_C on basic spatial dynamics: The spatial response of the CVM provides an overview of the 2D environment with its interacting agents, allowing observations to be made from both global and local viewpoints. This facilitates the task of tracking movement patterns of agents and through the process, uncovers fascinating emergent phenomena. Analysis of the microscopic interaction can also offer micro-macro explanations of temporal responses at the macroscopic scale. This is examined for configuration setups with different N_C.

- *Local Outburst:* A closer examination of the spatial interaction suggests that the presence of unprecedented up-shoot is largely due to spontaneous occurrence of outburst in regions of low cop density (Figure 5.12a). In these areas, cop-to-

activist ratio is low enough for mildly aggrieved agents to turn active concurrently (Figure 5.12b). These initial spatial correlations of activists act as a catalyst to elicit further outbursts by drawing more quiescent civilians into the revolt (Figure 5.12c). As stated [173], "when the mob is already very big relative to the cops, the level of grievance and risk taking required to join the revolt is far lower". This can be explained by the "Seeding" effect – an initial pool of rebels is able to seed out potential but less aggrieved activists in the vicinity to add to the severity of unrest.

A quiescent agent will feel safer to join a rebellion and display its discontent publicly if there are already a lot others doing it. This reiterates Mao Tze Tung's view that "a single spark can cause a prairie fire" [200], [201] and reinforces the concept of mass-mobilization – a crucial mechanism for triggering civil violence and fuelling growth of small-scale protests into larger ones [175], [202]. Once the outburst of rebel activities gets very large, a resulting mob usually fuels its growth and is by and large self-sustaining. A proposed plan [173] is to curb freedom of assembly by imposing curfews that restrict chances of collation among activists.
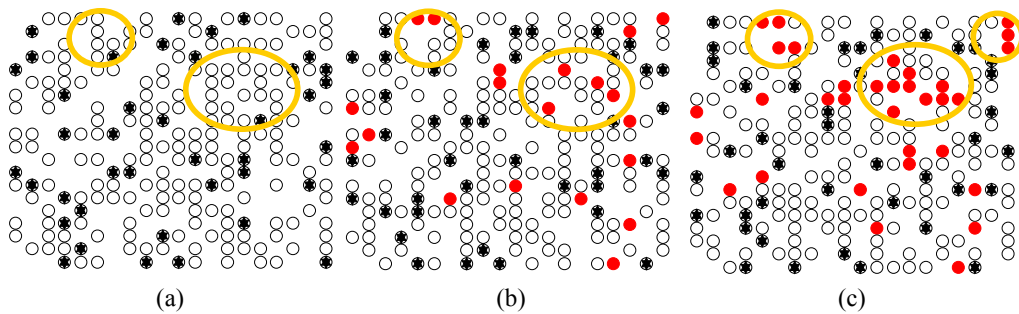
Figure 5.12: Spatial response depicting local outburst with 40 cops at episode (a) 1, (b) 2 and (c) 3

- *Group Clustering:* Spatial interaction of agents also depicts group clusters - collection of connected agents in 8-connectivity space, amidst evolution of agent movement (Figure 5.13). Activists collate as a form of collective behavior to create regions of low cop-to-activist ratio e.g. "safe havens" [173] as this reduces

the chances of arrest and prolong the unrest considerably. Le Bon proposes in his theory that "crowds seem to be governed by a collective mind, and that contagion causes members to experience similar thoughts and emotions" [203]. Sigmund Freud [204] further reinforces the fact that "individuals are able to satisfy basic needs for membership, hostility and so on by joining crowds." In reality, activists seek collective identity and group belongingness, the direction to vent their anger, protesting strength and safety, all of which are present among crowds in areas of low cop density. These psychological and behavioral needs often account for the conglomeration of scattered activists into small groups and amalgamation of small clusters into large ones (Figure 5.13). Since it gets harder to arrest rebels in huge clusters, duration of unrest is lengthened unless effective strategies are used.

A common challenge posed to the cops in crowd management is to disperse clusters before they turn into massively large mobs that are beyond control. An effective cop strategy is one that places cops in strategic positions (Figure 5.14). As seen, a burgeoning cluster between two activist groups is dispersed by cops in an attempt to form a line of defense that cuts through possible points of assembly. Similar to the partitioning of two warring groups [205], the above crowd control strategy seeks to lessen the severity of civil unrest by thwarting attempts made by activists to crowd together. The breakdown of large clusters facilitates the process of arresting activists, which would tend to appear very much in scattered numbers.

(a)   (b)
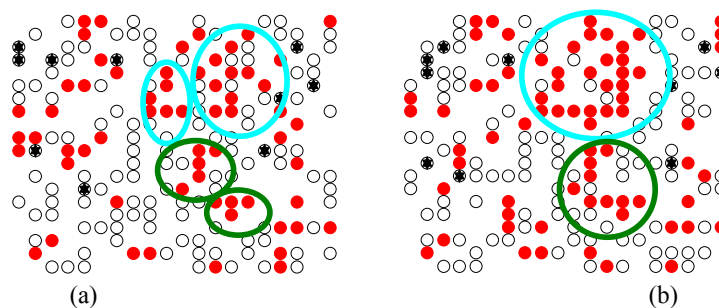
Figure 5.13: Spatial response depicting group clustering with 10 cops at episode (a) 3, and (b) 4

145
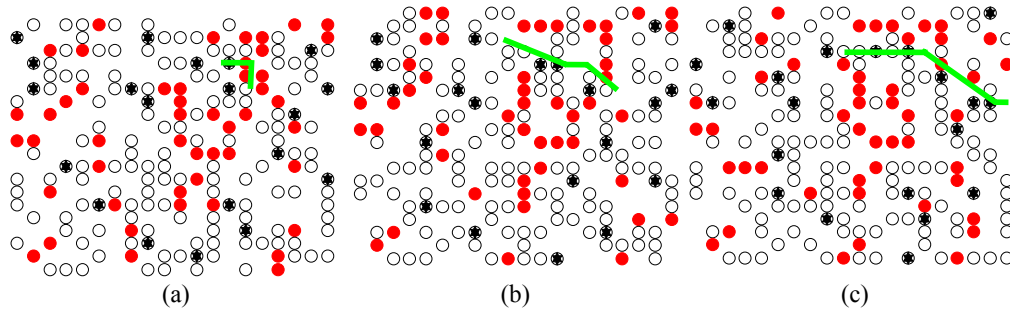
Figure 5.14: Spatial response of crowd dispersing with 20 cops at episode (a) 4, (b) 5, and (c) 6

- *Deceptive Behavior:* One salient feature of human behavior lies in the ability to put on a false front when dangerous encounters are imminent. Such deception is illustrated where two privately aggrieved agents appear ostensibly quiescent in the presence of cops but turn active if adjacent cops move away (Figure 5.15). This is due to a fall in cop-to-activist ratio within the local spatial neighborhood, which results in the reduced risk of arrest and higher tendency to revolt. Though subtle in nature, the display of such behavior is paramount to the study of human entities.

Famous military strategist, Sun Tzu [206] quoted that "Warfare is the art of deceit" [207]. An element of surprise which precedes any attack e.g. in Guerilla Warfare, is a pertinent factor that led to the success of numerous revolutions and uprisings. Deception is simply one of the most vital tools of biological survival [208] that is used by living entities in their adaptation to different environments. Many forms of military tactics are also mirrored and portrayed in nature - decoys, camouflages, diversions, disinformation, dazzles, disruptive coloration, disguise. By practicing the art of deception, activists conceal their emotions by appearing to be law abiding. This allows them to avoid any detection and arrest while waiting patiently for the right opportunity to strike. This has far-fetching repercussions as it lengthens the duration of the actual unrest and is a causal factor that makes the task of apprehending active remnants increasingly difficult.

146
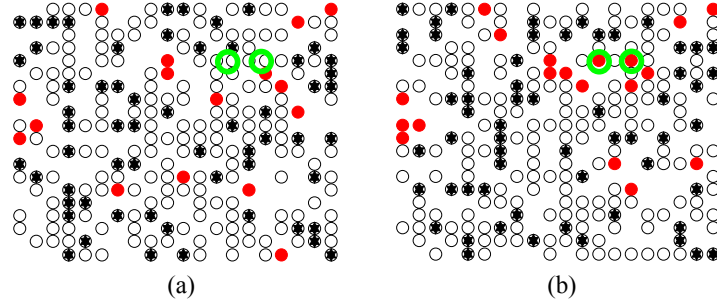
Figure 5.15: Spatial responses illustrating deceptive behavior with 80 cops at episode (a) 1 and (b) 2

3) Impact on active ratio: To extend the discussion, the emergence of deceptive behavior in the proposed co-evolutionary framework is examined in greater depths. Deceptive acts are revealed in the decision to stay quiescent despite favorable conditions to rebel. It is fascinating to understand the onset of such behavior for different N_C. In Figure 5.16, perceived active ratio denotes the typical active ratio while actual active ratio refers to the perceived ratio after all deceptive activists are accounted. This reflects the true state of unrest, which can be more severe than what cops perceived, as they can only spot active rioters.

Interestingly, Mao Tse Tung noted that revolutionaries "swim like fishes in the sea" [209], making them indistinguishable from the quiescent population. A full blown revolt may actually be brewing despite seemingly mild state of unrest. As the unrest follows its natural progression through time, deviation between the two ratios is observed as deceptive behavior emerges after substantial interaction between the activists and cops. Following the initial arrest of numerous activists, the remaining minority started to hide their discontent so as to avoid detection and arrest, causing a rise in deception level. The profiles of actual and perceived active ratio tend to be shaped similar, thus implying that deceptive behavior is exhibited largely by a small and unswerving group of activists.

147

(a)   (b)
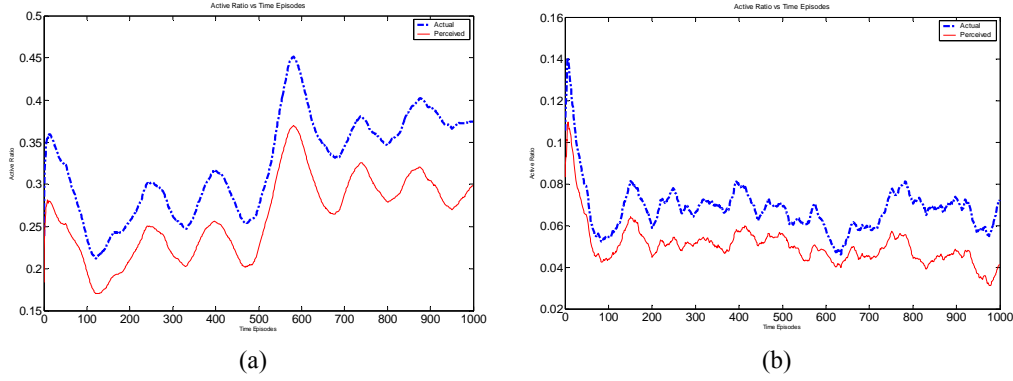
Figure 5.16: Actual and perceived active ratios for (a) 10 and (b) 60 cops

4) Impact on population dynamics and cooperation ratio: The variation of N_C also has sizeable impact on the population composition and cooperation profile of agent groups. Strong correlation is seen between the two (Figures 5.17, 5.18). For low N_C, the population dynamics of activist experiences large fluctuation (Figure 5.17a). The observed peaks are due to the mass release of jailed agents where a considerable number revert to the active state almost instantaneously due to low number of cops on patrol. This further dampens the cop-to-activist ratio in these neighborhoods, making the ambience superseding for activists to express their anger publicly. This in turn gives rise to a defect-oriented profile for the activist population (Figure 5.18a). With more cops and arrests, the activist population size became more stable as tension built up across the group is reduced (Figure 5.17b).

(a)   (b)
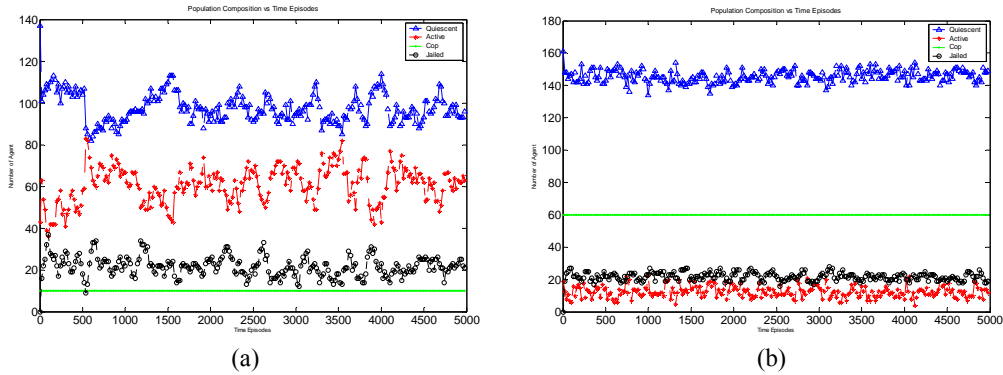
Figure 5.17: Population dynamics for (a) 10 and (b) 60 cops over 5000 episodes

148
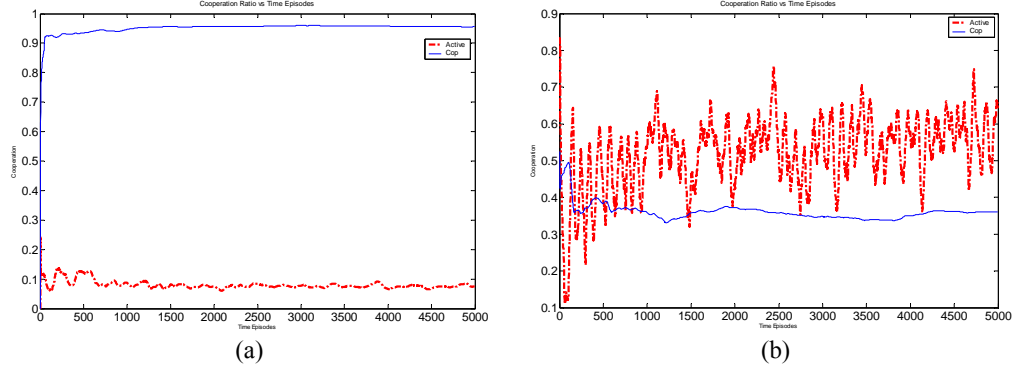
Figure 5.18: Cooperation ratio for (a) 10 and (b) 60 cops over 5000 episodes

Higher cooperation ratio is seen as activists contemplate more before deciding to riot (Figures 5.18b). There is also a concurrent dip in the cooperation level of cops as there are more occasions where the sizeable N_C justifies a pro-active strategy to pursue activists. A large N_C thus serves as an indirect form of deterrence.

5) Impact on average grievance and greed: The effects of N_C on both grievance and greed (Figure 5.19, 5.20) are also explored. Grievance is notably higher for activists across the plots since it is vital for transition to the active state. Overall, grievance profile is limited to a fairly low vacillation level while greed variation is contrastingly more pulsating (Figure 5.20). Although activists have high grievance in general, greed levels vary widely. Rapid fluctuation is due to the cyclical arrest and release of greedy and persistent activists. The quiescent greed profile is less susceptible to variation due to absence of exceptionally greedy agents.

Since an increase in N_C places the persistent activists in captivity and also reduces the mean population greed concurrently, it can be deduced that subversive activists are largely greedy. This effectively validates the nature of grievance as a primary, stable component that sparks off unrest and the nature of greed as a crucial factor to fuel the continual willingness of persistent activists to revolt.

149
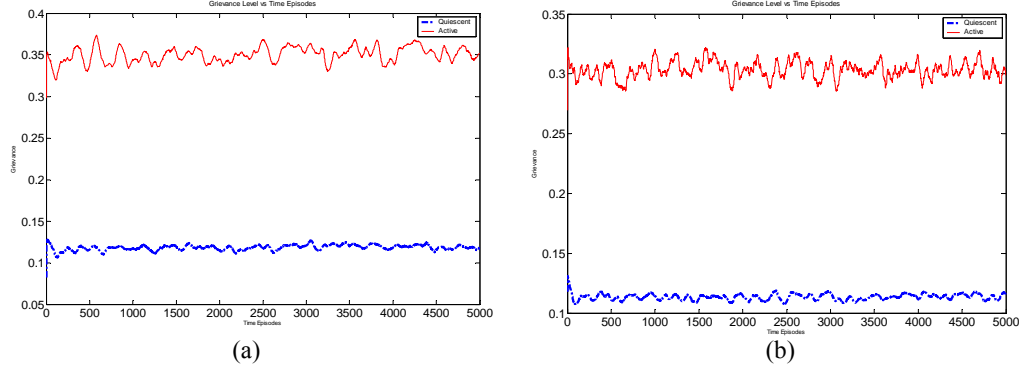
Figure 5.19: Average grievance level for (a) 10 and (b) 60 cops over 5000 episodes
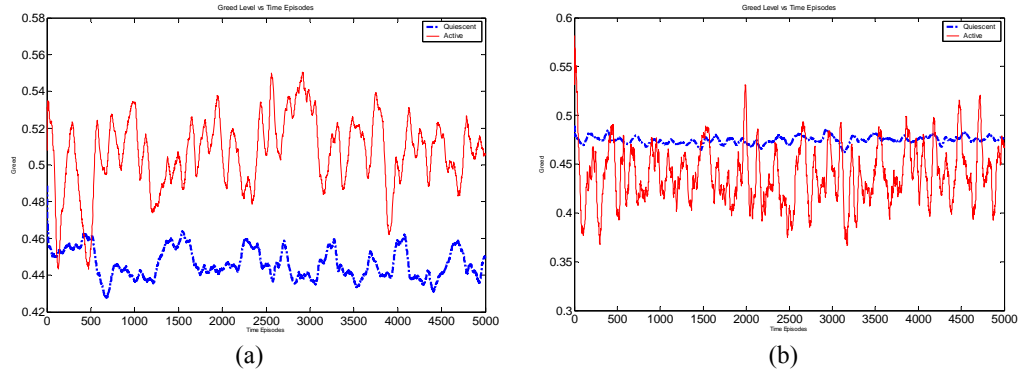
Figure 5.20: Average greed level for (a) 10 and (b) 60 cops over 5000 episodes

6) Impact on active history and duration: Active history depicts the degree of behavioral switching – the frequency that each agent type switches between the quiescent and active states. Active duration tracks the distribution of activists across time episodes that they have rioted since the last release. As N_C increases (Figure 5.21), a rising active history is observed for activists owing to the large degree of switching from the quiescent to active state following more arrests. This indicates that activists comprise persistent rioters who refuse to learn from their old ways. In contrast, the quiescent group undergoes less behavioral switching. Observations are consistent with the active duration plots (Figure 5.22). For a small N_C of 10, variance in active duration is large e.g. [0, 4700] – more activists riot for long periods. With $N_C = 60$, the variance is reduced e.g. [0, 2000] with concentration of activists in low frequency bins.

150
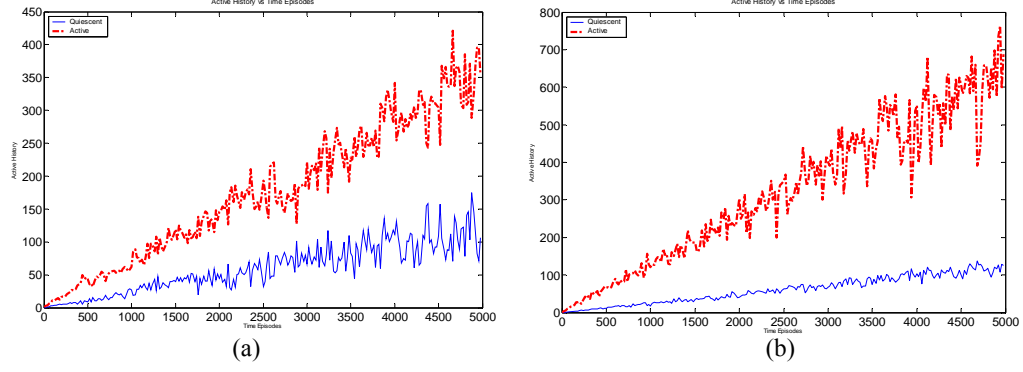
Figure 5.21: Active history for (a) 10 and (b) 60 cops over a span of 5000 episodes
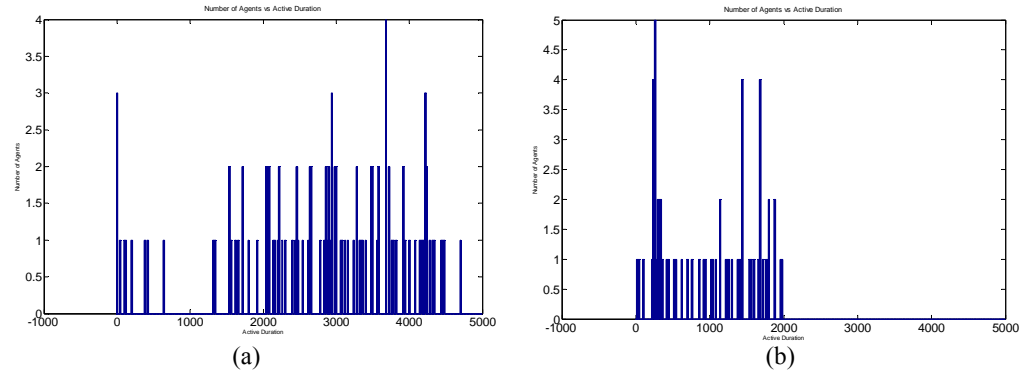
Figure 5.22: Active duration distribution for (a) 10 and (b) 60 cops over 5000 episodes

5.4.3 Active defectors and charismatic leaders: Effects on quiescent civilians

In this case study, the CVM is extended to account for the influence of activists on their neighboring quiescent civilians. Two classes of activists – the defectors and leaders are defined. Defectors are those that exert influence by demonstrating bold acts of defiance while leaders influence the crowds charismatically in more subtle ways. It will be interesting to explore the effect of each group on the actual unrest. In order to model the effect of influence, *NAI* is rewritten in (5.13), where $\overline{A}_{VR}(A_d)$ and $\overline{A}_{VR}(L_d)$ denote the number of defectors and leaders respectively within the vision radius of an agent. Each actively demonstrating defector will contribute 0.02 while each leader adds 0.05 to the *NAI* of each neighboring agent, increasing the probability to revolt.

151

$$NAI = \mathrm{Re}\, v - N + \overline{A}_{VR}(A_d) \cdot 0.02 + \overline{A}_{VR}(L_d) \cdot 0.05 \qquad (5.13)$$

Effects of incorporating influence are analyzed using active ratio, population dynamics, cooperation ratio and active duration. Active ratio variations, with and without influence, are plotted in Figure 5.23 for $N_C = 40$. The dynamics of the first 1000 episodes are superimposed. As shown, influence tends to cause more severe outburst of rebel activities in initial stages and gives rise to higher peak, mean and settling active ratios, since defectors and leaders induce more quiescent agents to revolt. Assimilation of these activities on the micro-scale led to outbursts on the macro-scale. There is more deviation between actual and perceived active ratios in the short run, indicating that influence promotes deceptive behavior (Figure 5.23b). The rationale is because the increased interaction between the activists and cops creates fundamental behavioral changes in the activist community as the members learn of the high opportunity cost for revolting against an overpowering police force. Intuitively, an act of deception - staying dormant to avoid detection presents itself as the best alternative. As seen in Figure 5.24, influence also creates greater fluctuation in the population dynamics for quiescent civilians and activists.

A momentous change in the cooperation profile of activists is also shown in Figure 5.25. Following an upsurge of rebel activities in initial stages, the activist population, by virtue of its sheer size breeds defect-oriented behavior as reflected by a low initial cooperation ratio. In the absence of influence, cooperation ratio of the activists rises rapidly and fluctuates about a mean level (Figure 5.25a). With influence, more civilians are instigated to revolt. This dampens cop-to-activist ratio in various localities, translating extensively to a tendency to adopt defect-oriented strategies - slow rise in cooperation and a lower mean when the dynamics stabilizes (Figure 5.25b). The gap between mean cooperation levels of cops and

activists is larger as well. Influence also reduces the duration that activists can roam about freely. A leftward skew of the active duration distribution histogram (Figure 5.26) is induced by the increased number of less aggrieved agents who turn active momentarily and possibly also due to more frequent arrests.

(a)   (b)
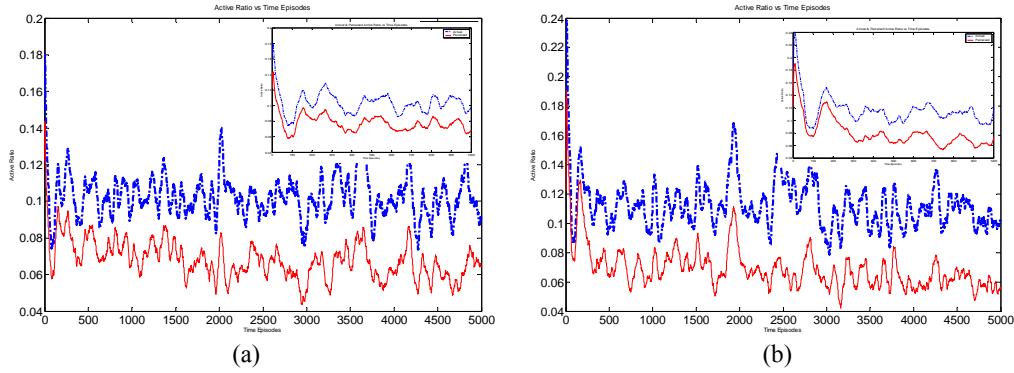
Figure 5.23: Actual and perceived active ratios (a) without and (b) with influence over 5000 episodes

(a)   (b)
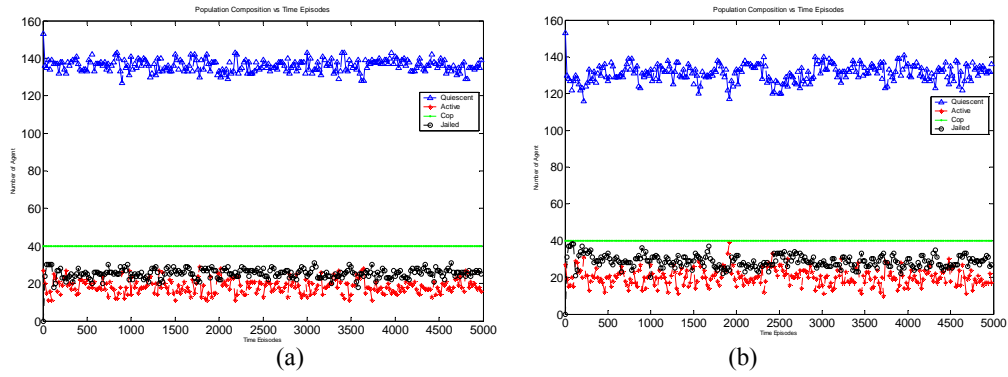
Figure 5.24: Population dynamics (a) without and (b) with influence over a span of 5000 episodes

(a)   (b)
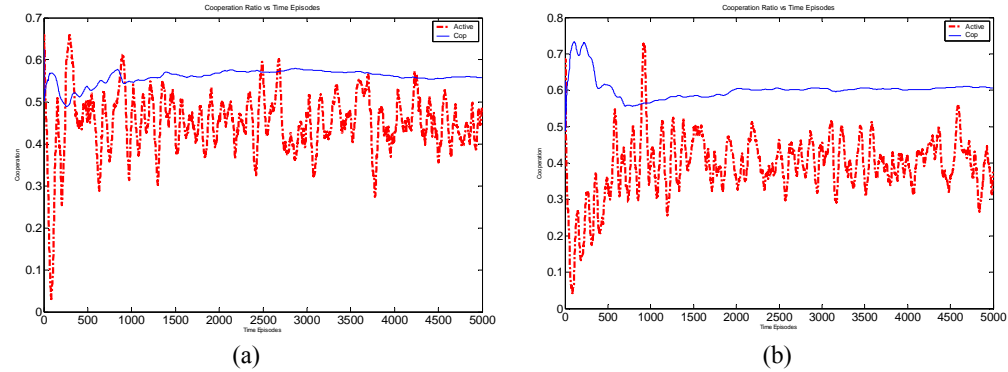
Figure 5.25: Cooperation ratio (a) without and (b) with influence over a span of 5000 episodes

153
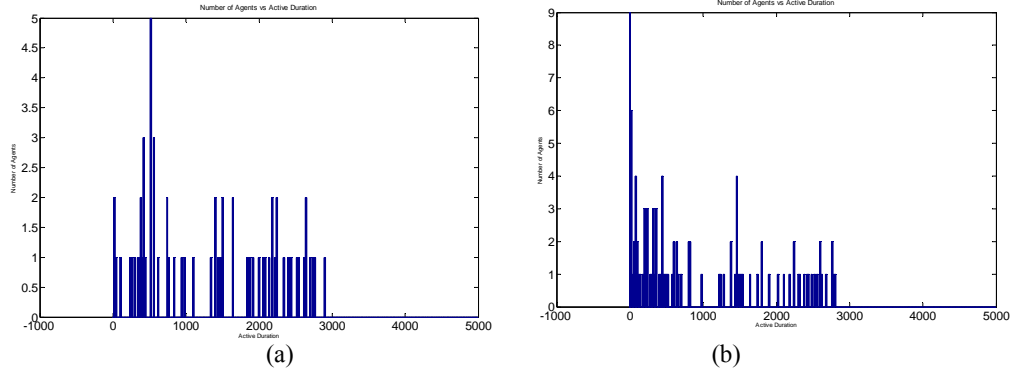
Figure 5.26: Active duration distribution (a) without and (b) with influence over 5000 episodes

Changes in the dynamics of unrest can also be explored when influence is introduced at different times. It is clear that early influence causes a sharp dip in the perceived active ratio (Figure 5.27a). Even if defectors and leaders are able to stir up emotions of hatred towards the central authority, such sentiments tend to subside after the cops initiate pursue and arrest. Instead of inciting more agents to revolt, early influence actually invokes massive arrests, creating strong deterrent effect indirectly which spurred more activists to exhibit deceptive behavior. This effect wanes off when influence is introduced much later (Figure 5.27b).
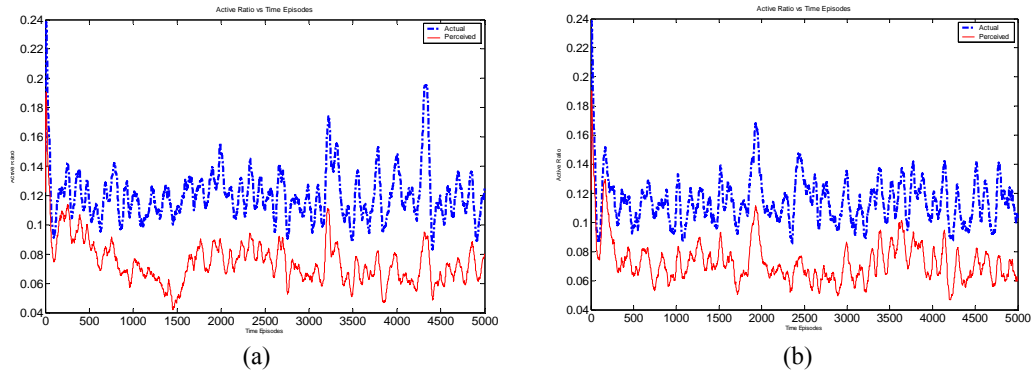
Figure 5.27: Actual and perceived active ratios of introducing influence at (a) 20th and (b) 2500th episode

Injecting influence in the early stages also introduces greater dynamics to the long term active profile by allowing a gradual build up of tension across the population. This translates to sudden outbursts (Figure 5.27a) that are absent when influence is injected at later stages (Figure 5.27b). Findings are fairly consistent to

reality as it is easier to kindle feelings of hatred if emotions are still vacillating and the spirit of revolution is high amidst the activist regime. Behavioral changes occur over a considerable period and early instigation does assist in encouraging the development of rebellious behavior within the active community.

5.4.4 CVM Response under varying jail terms

Besides deterrence from cops, jail duration also constitutes a major vitiating factor to the willingness to revolt. The nature of punishment is crucial in shaping the behavioral profile of activists. This section examines the effects of both fixed and variable jail terms on the dynamics of unrest. In the prior, a jail sentence of fixed magnitude is imposed on any arrested activist. The latter entails an increasing penalty for repeated offenders until life imprisonment is reached.

1) Impact on active ratio: The active ratio for fixed jail terms of [5, 500] and variable jail term of $1 + (J_H / J_{H \max}) \cdot J_{\max}$ for $J_{\max} = 300$, $J_{H \max} = 3$ are plotted in Figure 5.28. A large fixed jail term lowers the mean level of rebel activities (Figures 5.28a-b), similar to a large N_C. The decline is due to long periods which the activists spent in captivity. Accompanying this is a reduction in the scale of fluctuation and outbursts, as the long jail term makes it less likely for activists to be released simultaneously. Deceptive behavior is also reduced. Compared to a large fixed penalty, imposing a low jail penalty and increasing it incrementally (Figure 5.28c) for repeated offenders achieves a low and more stable profile. As employed in societies, the variable penalty balances the punitive aspect – desire to punish offenders and rehabilitative aspect – aim to reform criminals, of law. It denotes a fair system of justice where mistakes of initial offenders are tolerated while heavier punishments are meted out for repeat offenders.

155
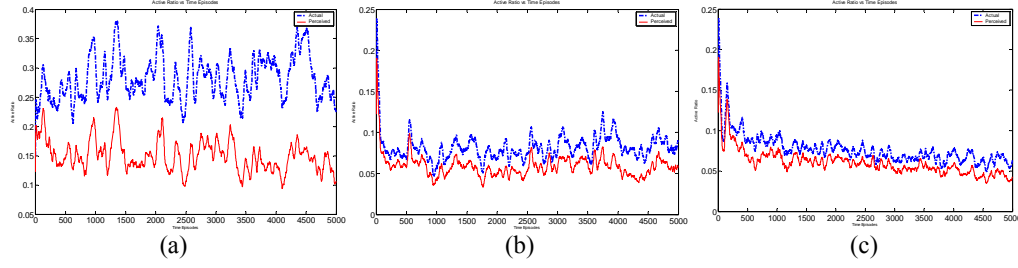
Figure 5.28: Actual and perceived active ratios for fixed jail terms of (a) 5, (b) 500 and (c) variable jail term

2) Impact on cooperation ratio: Changes in conviction period also affect the cooperation profile of activists. Low jail terms breed low cooperation (Figure 5.29a) owing to ever presence of activists. Higher fixed jail terms (Figure 5.29b) places the jailed activists in captivity for longer periods, causing the outnumbered remnants to favor more cooperative strategies after continual interaction with cops. Excessive penalty however reduces the cop-to-activist contact and in turn impedes autonomous behavioral development as jailed agents have less chance to interact, exchange knowledge, learn and evolve. Many remain persistent activists, unlike the case of a large N_C. As depicted in Figure 5.29c, cooperation is still increasing initially when most of the activists are convicted for short periods. As more and more are sentenced to life imprisonment, defection sets in for the residual activists, invoking further aggression and arrests. As rationalized, the increasing jail term is thus an efficient and effective means of isolation and tends to lower the unrest by minimizing contact among activists.
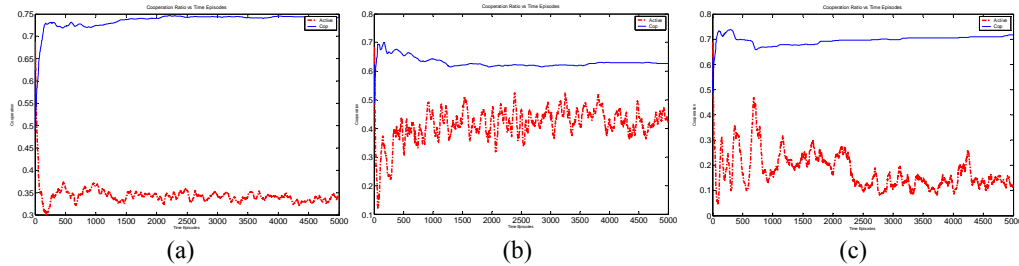
Figure 5.29: Cooperation ratio for fixed jail terms of (a) 5, (b) 50 and (c) variable jail term

3) Impact on active history and duration: Active history increases as higher jail penalty is imposed (Figure 5.30). With more activists convicted for considerable periods, residual ones are mildly aggrieved and sensitive to changes in the state of unrest. By virtue of this nature, a large degree of behavioral switching is portrayed. Consistent with this, the active duration is lowered as seen by a leftward skew of the histogram distribution (Figure 5.31). This reflects the lengthier period which activists spend behind bars. The dynamic range of active duration is less affected as opposed to the varying of N_C since any reduction in the range of active duration is largely due to the higher efficiency which is introduced when more cops are deployed to apprehend activists.

(a)   (b)
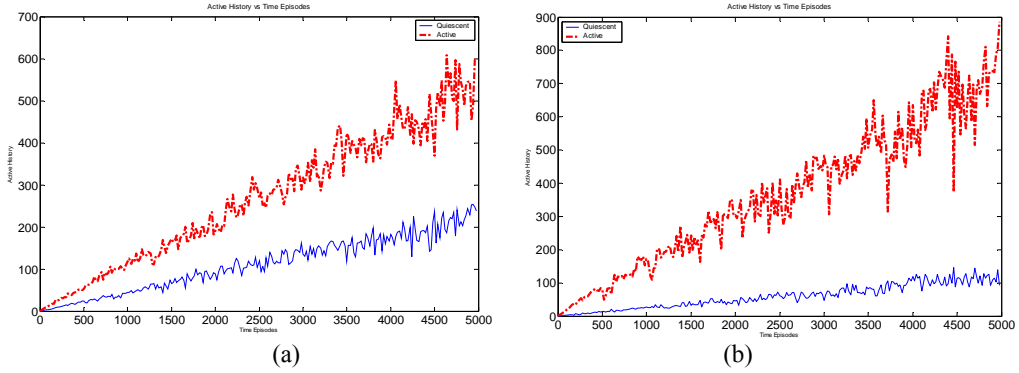
Figure 5.30: Active history for fixed jail terms of (a) 5 and (b) 500 over 5000 episodes

(a)   (b)
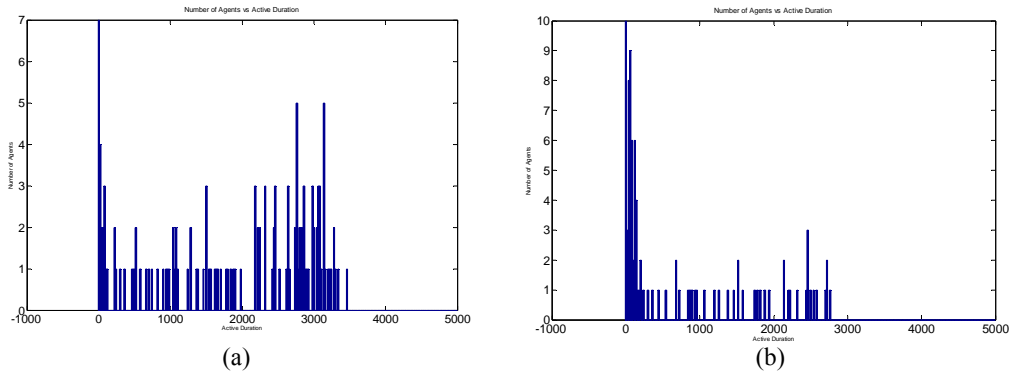
Figure 5.31: Active duration distribution for fixed jail terms of (a) 5 and (b) 500 over 5000 episodes

157

5.4.5 Casualty Model

The final section integrates the knowledge and insights gained from prior sections to create an empirical casualty model to investigate a scenario case study where harm is inflicted by one group on another. This encompasses the deliberate and often systematic elimination of an entire national, racial, political or cultural group [210] by distinct ethnic groups or coalitions [211] due to hatred [212] and distrust for one another. Examples include the Nazi-Jew holocaust, Hutu-Tutsi Rwanda genocide etc. The fundamental nature of such events can be better understood by probing into the underlying emergence dynamics.

In this model, cops assume the role of peacekeepers while activists are perpetrators. There is no state transition e.g. perpetrators and quiescent agents do not cross their own ethnic boundary. Quiescent agents adopt *Run from Activists*, perpetrators use *Eradicate the Civilians* while the peacekeepers espouse *Pursue the Activists* to arrest perpetrators and minimize casualties. Interaction only takes place among peacekeepers and perpetuators. Arrest is made if the prior wins; else, a randomly sited civilian is removed. The learning heuristics of perpetrators is altered to include GS_{win}- game sets won. Tactics revision occurs if

$$SH < FH + 10.GS_{lost} - 10 * GS_{win} \qquad (5.14)$$

Each run takes 1000 episodes and the objective is to investigate effectiveness of increasing peacekeeping size and jail penalty in minimizing number of casualties after each simulated window.

1) Situation without peacekeepers or jail terms: The situation of unrest, in the absence of peacekeepers or jail term, is simulated with perpetuators constituting 10% of the total civilian population. Both active ratio and population dynamics

158

(Figure 5.32) show rapid annihilation of the quiescent group when no intervening force is present to manage the unrest. The escalating increase in active ratio and sharp plunge in civilian population indicates that a small pool of perpetrators is capable of eliminating a much larger group within a short time. A spatial overview (Figure 5.33) depicts the exponential drop in the quiescent group. Almost half the population had suffered casualty by the 10th episode. Total annihilation is seen by the 57th episode, which denotes the extinction time – time taken to wipe out an entire group with a distinct identity.
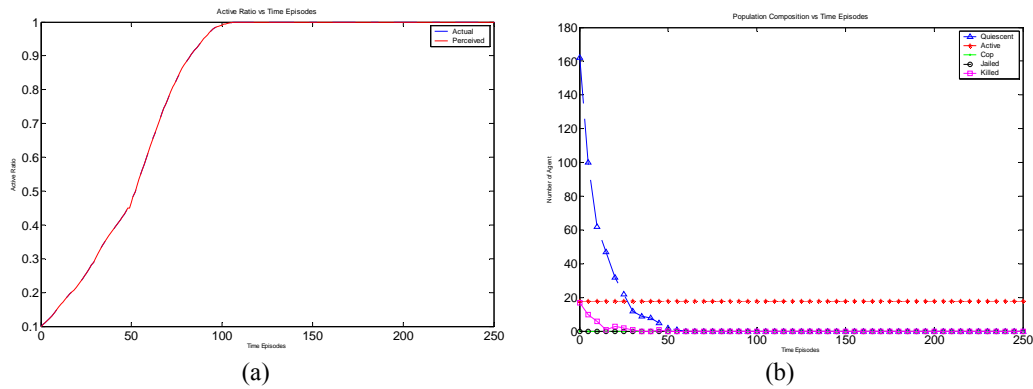
Figure 5.32: (a) Active ratios and (b) population dynamics for the first 250 episodes
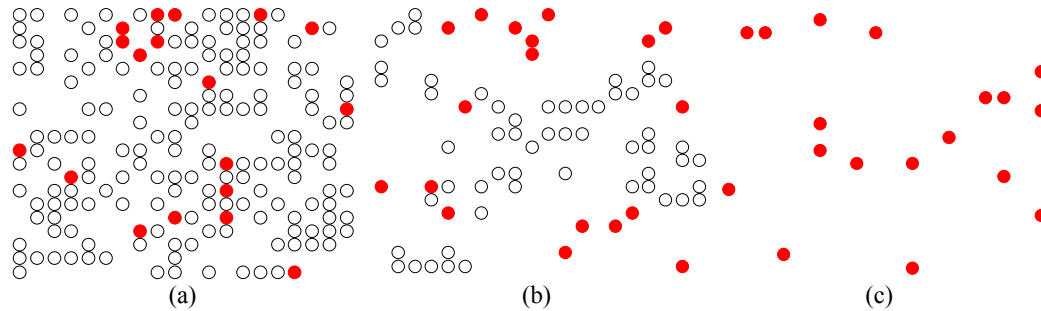
Figure 5.33: Spatial interaction between perpetrators and civilians for episode (a) 0, (b) 10 and (c) 57

2) Impact of varying peacekeepers: Peacekeepers are now added to alleviate the severity of unrest. Fixing jail term at 100 episodes, Figure 5.34c showed that the presence of more peacekeepers raise survivals after the first stable point ([83, 136, 140]) but excessive peacekeepers cause a sharp dip in quiescent group due to

overcrowding. The effectiveness to track, pursue and make arrest is lowered as peacekeepers impede each other's movement. This is analogous to the concept of "carrying capacity" [213], where a system can only accommodate limited number of agents. Excess ones die off or introduce some form of inefficiency to its innate workings. The results indicate that more peacekeepers are needed to manage the simultaneous presence of perpetrators and minimize casualty initially. Once the unrest stabilizes, excessive peacekeepers proved to be a con more than a pro. This claim is further substantiated in Figure 5.35c as perpetrators start to roam freely beyond 100 time episodes when the peacekeeping force gets too large. Introducing excessive peacekeepers to a constrained environment in this context thus hinders progress of arrest and results in the rapid elimination of the quiescent group.

3) Impact of varying jail terms: The jail term is now varied to analyze its effects on the unrest for a fixed peacekeeping size of 40. From the drastic dip in survivals, it can be garnered that isolating perpetrators is crucial in the short run to prevent excessive eradication of the quiescent group. As observed in Figure 5.36, a long period of captivity reduces the number of downward stepwise transitions in the quiescent population and preserves more survivals. Nonetheless, altering the jail term does not affect the initial decrease in the quiescent group, unlike the variation of the peacekeeping size. Survivals after the first stable point are similar for both jail terms of 100 and 500 episodes, as only a fixed number of peacekeepers are present to carry out pursuits and arrests. This places a fundamental limitation on the ability to curb the unrest as a time lag is present where sizeable casualties can occur before ample perpetuators are eventually arrested.
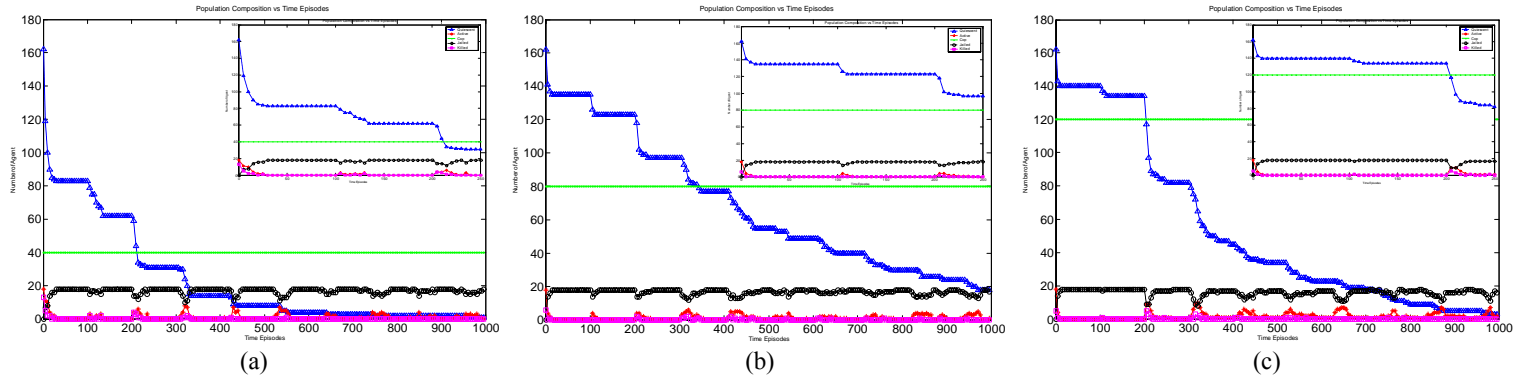
Figure 5.34: Population dynamics for peacekeeping force of size (a) 40, (b) 80 and (c) 120
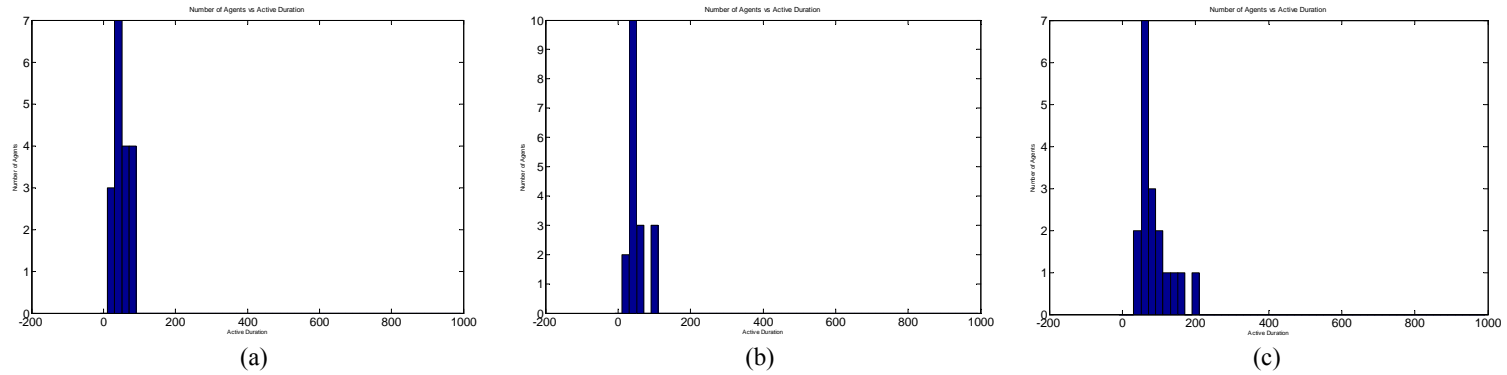
Figure 5.35: Active duration distribution for peacekeeping force of size (a) 40, (b) 80 and (c) 120

161
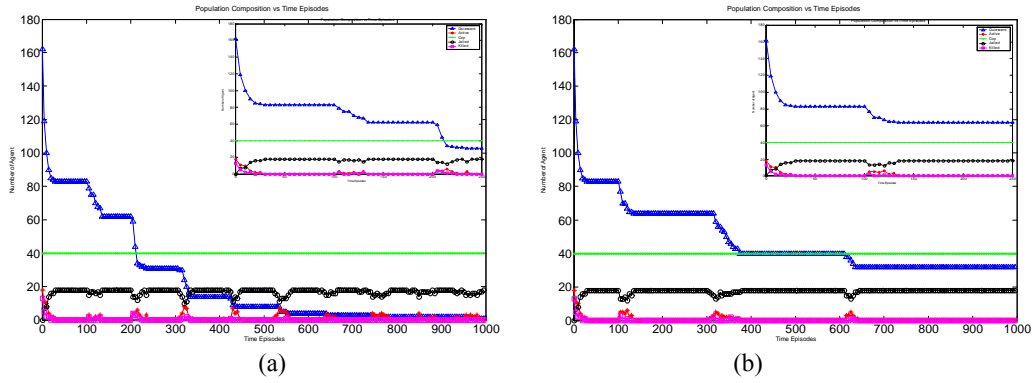
Figure 5.36: Population dynamics for fixed jail terms of (a) 100, and (b) 500 episodes

5.5   *Findings and discussions*

Interesting findings are revealed via the co-evolutionary simulation. The temporal response of the model showed the presence of "Punctuated Equilibria" and affirms peace and stability as a dynamic equilibrium that emerges from agent interaction. The spatial responses portray spontaneous local outbursts, group clustering but notably, a display of deceptive behavior. Increasing the number of cops has shown to promote deceptive behavior and drives activists to embrace cooperation. High grievance in activists is found to be the primary cause of triggering unrest while high greed levels is responsible for fueling the continual willingness of persistent rebels to revolt. Greater behavioral switching is also exhibited with more arrests.

The addition of influence triggers a severe upsurge of activists in the initial stages and high mean active ratio as mildly aggrieved agents are incited to revolt. The tendency to use defect-oriented strategies is increased, which paves the way for the development of deceptive behavior over time. Introducing influence in the early stages of unrest yields greater dynamics by allowing the gradual build up of tension across the population, which translates into unprecedented occurrence of outbursts at a later stage. Increasing the jail penalty reduces the mean active ratio and dampens both the scale and frequency of outbursts. An excessive jail term

minimizes the chances of contact for arrested agents and triggers defect-oriented behavior. The dynamic range of active duration is less affected, however, as the efficiency of arrest depends largely on the number of cops. Finally, the casualty model deduces that peacekeeping size and jail term affect the long term and short term profiles of unrest respectively. Though a large peacekeeping size is typically desired, excessive peacekeepers, nonetheless, worsen the prevailing state of unrest. Balance of both an adequate peacekeeping force and jail penalty is essential to achieve the lowest possible casualty rate.

5.6   *Summary*

The chapter showed that interesting macroscopic emergent dynamics are obtained through the microscopic autonomous behavioral development of agents under a co-evolutionary inspired framework which encompasses a hybrid combination of evolution with learning. Studying how the underlying behavioral dynamics evolve under different situational setups is crucial for the holistic understanding of the fundamental nature of civil violence.

Chapter 6

Public Goods provision under asymmetric information

Though interesting, the modeling of civil violence in the previous chapter still adopts a pair-wise scheme of interaction among agents, similar to that in classical IPD. Nonetheless, much of the interaction in the real world occurs simultaneously among multiple parties, giving rise to a situation of dilemma which is commonly known as "The Tragedy of the Commons" [214]. Similar in essence to a multi-player IPD, the dilemma - which finds its presence in scenarios ranging from the overgrazing of land to overconsumption of public resources, is typically attached to situations that involve the provision of public goods (PG) [29], [215]. We will shift our attention to focus on PG provision for the last chapter of this work.

The challenge of PG provision has always been a core economic issue in societies throughout the changing times. Unlike private goods [216], common pool resources [217] and club goods [218], the intrinsic characteristics of non-excludability and non-rivalry [29], [219] in the consumption of PG ascertain that its provision can confer positive externalities [220] which are collectively shared; but for which there is practically no efficient way of excluding non-contributors from enjoying. Coupled with the fact that PG provision involves the joint action of many individuals [221], misalignment of individual contribution with collective welfare [222] entails low incentives to contribute. This explains why the voluntary provision of PG [223], both local [224] and global [225], is extremely susceptible to market failure [226], so much so that supply of PG often falls short of Pareto Optimal [227] in the absence of government intervention [228].

Across contexts that are as diverse as assurance contracts [229], peer-to-peer networks [230], retail [231], drug imports [232] and welfare economics [233], the emergence of free-riders [234] has been identified as a prevalent cause of inefficiency. A major purpose of experimental literature on PG provision is then to assess the magnitude of free-riding and variables that affect it [235]. In game theory, the study of free riding is approached from the behavioral perspective of rational agents [236], by means of a PG game [237] where players form groups to decide how much to contribute to a PG using available information. In the iterated PG game (IPGG) [238], the game is played over many rounds. The level of PG provisioned is decided by the collective contribution [224]. Benefits derived are distributed evenly among all participating users, regardless of effort; but costs are born solely by those who played a part in provision according to efforts expended. This notion suggests that the expense of individual effort does not translate to a sole enjoyment of benefits but improves welfare indiscriminately.

Intuitively, a group does best if everyone contributes, as the eventual level of PG will be higher, with greater remuneration for all. However, this does not arise as individually rational players tend to free ride on others' contributions - social loafing [215]. In the same line of thinking, players are likely to dismiss a decision to contribute to avoid exploitation by free-riders. Players thus, do better on the whole by contributing zero regardless of the others' actions [29]. This is the Social Dilemma [222] – a paradox in social decision making where joint contribution is needed to attain shared goals, but an individual's rational choice is simply to free-ride. As considerable benefits can be enjoyed by all for every additional unit of contribution, there is potential for huge Pareto improvement in welfare if everyone embraces cooperation. While some may argue that this problem can be solved by

using an intermediate regulatory body to fund PG provision indirectly via taxes; such involuntary means can be inefficient at times since hidden costs are typically involved. It is in the interest of policy makers and economists alike to design mechanisms and functional models that provide insights into how the prevalent effects of Social Dilemma can be alleviated in diverse settings so as to allow the efficient voluntary PG provision to take place.

This chapter presents a co-evolutionary framework [239] to simulate and analyze the outcomes of PG provisioning under asymmetric information [5], [240] - [243]. Via an ACM, boundedly rational agents are conceptualized to interact in an N-player IPGG. They adapt to the dynamic environment by co-evolutionary learning in the course of game play similar to that of an N-player IPD [244]. The impact of information type, population and group sizes, rate of interaction, the number of available choices, the nature of provision and selection schemes, are studied under various settings. Simulated results reveal interesting dynamics in the strategy and usage profiles, welfare plots and evolution of cooperation. Analysis of these results offers a holistic understanding of collective action and insights of how the predicament of Social Dilemma can be mitigated, if not averted, in favor of the efficient voluntary provision of PG.

The chapter is organized as follows. Section 6.1 presents preliminaries of the IPGG and overview of the model design. Section 6.2 highlights the game theoretic fundamentals that are essential to formulate and appreciate the IPGG. Section 6.3 formally introduces a list of asymmetric agent types and their respective genotypic representations while Section 6.4 focuses on the significance of co-evolutionary learning and simulation. Section 6.5 evaluates and analyzes outcomes of simulated interaction in different settings of PG provision. Section 6.6 summarizes major

findings of the simulation study. Finally, Section 6.7 will conclude with a broad summary of discussions and areas for future research.

6.1 Iterated public goods game

Originating from experimental economics [245], the IPGG has striking similarities as the IPD [246]; with parallelism closely drawn between the study of contribution and cooperation [89] respectively. Mutually beneficial cooperation is threatened by unilateral strategic behavior as players are individually rational but collectively irrational. The IPGG also encompasses variants like the optional PG games [247], evolutionary games [248] with replicator dynamics [247], as well as games with punishment [238] and commitment [249]. Mechanisms like voting, peer effects and mobility [224], reputation and penalty [250], signaling and trust [251] can then be explored; together with their impact on contribution.

Most models have concentrated on the ideal scenarios where agents exhibit unbounded rationality [252] and interact under complete and symmetric [236] information e.g. using Nash [253] and non-Nash [254] inferences, and Bayesian Learning [255] to determine an optimal agent strategy set. Such approaches are unrealistic since individuals do not actually possess perfect information about the environment in reality. Even in the case where information is readily available, it will not be used entirely for decision making as players have clear preferences for particular information types [256] e.g. those which are most relevant and in line with their contribution strategies. To some extent, players are also not attributed with advanced information processing capacities – which traditional theoretical analysis would require them to have; to capitalize and take into account all the information available to them.

6.1.1 IPGG with Asymmetric information

In order to address the above modeling deficiencies and incorporate behavioral imperfection, players in the proposed IPGG will use their preferred information type for the basis of decision making. By doing so, the effects of framing [257] – in which a scenario can be interpreted differently by players in accordance to the perspectives they adopt, are accounted for e.g. contributions are likely to be higher if a PG game is framed as a community social event than when it is framed as an economic investment [258]. Framing an option as a cost versus an uncompensated loss also affects whether that option is chosen [259]. Other than a more realistic and interesting way to model PG provision, studies have also shown that different information types do affect the inherent dynamics of cooperation [249], [256]. With players formulating different contribution strategies, the proposed model offers another perspective to analyze the IPGG via the assessment of strategies and interaction outcomes that emerge from the use of diverse information types.

6.1.2 Mathematical formulation

The agent-based IPGG models an artificial society, $\boldsymbol{E} = \langle N, \boldsymbol{I}, \boldsymbol{S}, \boldsymbol{A}, U \rangle$, which comprises a population of N players, set of information, $\boldsymbol{I}$ about the global state of game play, set of N decision strategies, $\boldsymbol{S} = \{\boldsymbol{S}_1(t_1), \boldsymbol{S}_2(t_2), ..., \boldsymbol{S}_N(t_N)\}$ that dictates the players' responses to stimuli in the external environment, set of all possible actions, $\boldsymbol{A} = \{\boldsymbol{A}_1(t_1), \boldsymbol{A}_2(t_2), ..., \boldsymbol{A}_N(t_N)\}$ for players and a utility function, U which determines the payoffs awarded to players in all interaction outcomes. For all the above attributes, components of the set $\{t_1, t_2, ..., t_N\}$ refer to types of players 1 to N respectively. $\boldsymbol{E}$ can refer to any generic organization where PG is provisioned collectively by N players. As opposed to a typical IPGG [230] where all players in

the population are participating in the provision of a common PG, the proposed IPGG models a situation in which the task of PG provision is decomposed and assigned in parts to $M \in [M_{\min}, M_{\max}]$ smaller groups, of size $n \in [n_{\min}, n_{\max}] \subseteq N$ [256], such that sum of players in M groups totals N. The portion of task allocated to each group scales proportionally to its size. Players assigned to the same group will only contribute to the portion of total PG that they are tasked to provision. This setup can be drawn in analogy to a global PG that is funded collectively by several communities via taxation schemes. Entities in each community will decide whether to contribute or to evade taxes. The scenario can also be employed in a corporate setting where large projects are split into distinct parts and assigned in fair proportions to different teams of personnel. Each person in effect contributes to the part of project which his team has been assigned to.

In every iterated round of game play, players are required to decide the contribution amount or cooperation level towards PG provision within the group. Decision output of player i, in group g_j with type t_{ij}, is derived via strategy $\boldsymbol{S}_{ij}(t_{ij})$. This maps the type-dependent information subset, $\boldsymbol{I}_i(t_{ij}, g_j)$ - selectively chosen by player i from the information superset, $\boldsymbol{I}(g_j)$ - accessible by all in group g_j; to an action $C_{ij}(t_{ij})$, denoting the desired effort level out of all the possible choices in action space $\boldsymbol{A}_{ij}(t_{ij})$. This mapping is represented symbiotically by:

$$\boldsymbol{I}_i(t_{ij}, g_j) \in \boldsymbol{I}(g_j) \xrightarrow{\boldsymbol{S}_{ij}(t_{ij})} C_{ij}(t_{ij}) \in \boldsymbol{A}_{ij}(t_{ij}) \tag{6.1}$$

Since the nature of IPGG involves the collective action and decision making of multiple parties, outcome of PG provision is not determined unilaterally by the contribution of any one player, but rather by the sum of individual contribution from each member. The level of PG provisioned within a group, g_j, of n_j players,

$P(g_j)$, is mathematically expressed in the form:

$$P(g_j) = \sum_{i=1}^{n_j} C_{ij}(t_{ij})$$ (6.2)

Collectively, the total amount of PG which is provisioned by M groups of players is thus given by (6.3):

$$P_{Total}(g_1, g_2, ..., g_M) = \sum_{j=1}^{M} P(g_j) = \sum_{j=1}^{M} \sum_{i=1}^{n_j} C_{ij}(t_{ij})$$ (6.3)

The payoff of PG provision, which is derived by a player i in a group, g_j that consists of n_j players is specified by the equation for the Voluntary Contribution Mechanism (VCM) [260] as given in (6.4):

$$Payoff_j(i) = U(C_{ij}(t_{ij}), C_{\Sigma-ij}(t_{-ij}), n_j \mid i, -i \in g_j)$$
$$= U_{Basic} + P(g_j)/n_j - Cost_j(i)$$ (6.4)

where $C_{\Sigma-ij}(t_{-ij}) = C_{1j}(t_{1j}) + .. + C_{i-1j}(t_{i-1j}) + C_{i+1j}(t_{i+1j}) + .. + C_{n_j j}(t_{n_j j})$ is the collective contribution of all in group g_j, less the contribution of player i, $C_{ij}(t_{ij})$. U_{Basic} is the payoff which a player gets if no public good is provisioned. This occurs when everyone in the group decides to free-ride. In the context of each player, all efforts would be channeled solely to provision a private good that yields a default, non-zero payoff which is higher than that attained if all others free-ride on the player's contribution. $P(g_j)/n_j$ is the payoff that a player derives when the welfare from PG provision is evenly distributed among all within the group. $Cost_j(i)$ denotes the cost that is incurred by player i for contributing. Assuming that cost correlates positively with contribution, the more a player contributes, the larger is the effort expended and the higher will be the resulting cost incurred. For simplicity, cost is expressed mathematically as follows:

170

$$Cost_j(i) = C_{ij}(t_{ij}) \cdot (C_{factor} / R_{factor}) \tag{6.5}$$

where C_{factor} / R_{factor} is the proportionality ratio which denotes the value of a unit of contribution cost, C_{factor}, relative to a unit of the provisioned resource, R_{factor} e.g. $C_{factor} / R_{factor} = 0.5$ meant that 50% of any effort that a player commits to provision will be expended as personal cost; thus effectively only generating a net collective welfare value equivalent to the residual 50% of effort. Overall welfare enjoyed by society – measured by total payoff that is derived collectively for all players is:

$$W_{Society}(1,2,...,N) = \sum_{j=1}^{M} \sum_{i=1}^{n_j} Payoff_j(i) \tag{6.6}$$

The average welfare e.g. amount of PG enjoyed by a typical player is given by:

$$W_{Individual} = W_{Society}(1,2,...,N)/N \tag{6.7}$$

It is to be noted that $W_{Society}(1,2,......,N)$ is different from $P_{Total}(g_1, g_2,..., g_M)$ as the prior refers to the net benefits derived by society after accounting for contribution costs. In essence, $W_{Individual}$ denotes the net average welfare that each individual player benefits from the provisioned PG.

6.1.3 Environment

- Players are boundedly rational and have finite computation power [62]. They do not have full and perfect knowledge about their environment of interaction - types of players in the group, NEs in the game etc. Each uses limited, local information [261] which is selected according to type and preference to decide the amount to contribute in the next round.

- Every action is available to all players e.g. homogeny of action space where $A_{1j}(t_{1j}) = A_{2j}(t_{2j}) = ... = A_{Nj}(t_{Nj})$ regardless of group or type. However, the

171

actual action taken by a player depends on the nature of its type and strategy.

- All players have the same capacity to generate PG - with the same amount of effort put in; the quantity of PG generated will be the same for any player.

- Value of the provisioned PG is identical, in terms of its worth, to all players.

- No contribution amount can saturate the total maintenance benefit which is derived from the PG e.g. each unit of provisioned PG will always yield much higher returns than the unit of contribution that is put in towards its creation.

- Similar to the IPD, a situation where players free ride on the contribution of another is considered worst off, from the perspective of the exploited player, as compared to the situation where there is totally no provision of PG.

- The PG of concern in the proposed study is deemed finitely and discretely decomposable e.g. it can be split into smaller parts for easy delegation and segregation of provision tasks among groups.

- To an individual player, it is not the total benefits derived collectively by the team that is important, but the welfare solely enjoyed by himself ultimately.

- The effect of framing is assumed in the context of the IPGG. All players do not change their types or beliefs over time and tend to look at all scenarios with the same perspective e.g. information type. Constraining players to the same beliefs allows the flexibility to study which information types are more dominant in promoting cooperative strategies.

- All players will only choose to adopt strategies that are aligned with their own beliefs, of which, they will maximize their payoffs given the constraint of the fixed strategy structure.

- All players improve their strategies constantly so as to seek an eventual increase in their welfare over time.

6.2 Game theoretic fundamentals

The IPGG in essence, is similar in spirit to an N-Player IPD game where players have a temptation to defect (D) – free ride at the expense of other players. This depicts a situation where individual interest is in conflict with group interest - that is for players to cooperate (C) and contribute towards PG provision. Although D is the dominant strategy [89] when approaching from the perspective of individual rationality, it becomes collectively irrational if all players in the group choose to free-ride, since no PG is provisioned and no welfare is derived. Everyone can be better off by playing the dominated strategy C, which explains the existence of dilemma. To preserve the essence of this dilemma in the context of a 2IPD game ($n = 2$), two conditions [262] must be satisfied. Firstly, the temptation payoff (T), reward payoff (R), punishment payoff (P) and sucker payoff (S) are assigned in descending order of their values ($T > R > P > S$). Secondly, alternating between T and S does not reward each player as much as if both players embrace repeated cooperation ($(T + S)/2 < R$) between themselves.

Similarly, the payoff function of the proposed IPGG, where ($n > 2$), is formulated such that the following equivalent conditions are satisfied during the actual game play. In all three conditions, C_{ij} and $C_{ij}^{'}$ are two distinct levels which individual i in group g_j can contribute at.

1) Bounded individual rationality – Defection is better

- Given that $C_{ij}(t_{ij}) < C_{ij}^{'}(t_{ij})$ and a fixed $C_{\Sigma-ij}(t_{-ij})$,

$$U(C_{ij}(t_{ij}), C_{\Sigma-ij}(t_{-ij}), n) > U(C_{ij}^{'}(t_{ij}), C_{\Sigma-ij}(t_{-ij}), n) \qquad (6.8)$$

2) Bounded collective rationality – Mutual Cooperation is better than mutual Defection

- Given that $C_{ij}(t_{ij}) \leq C'_{ij}(t_{ij})$ and $C_{\Sigma-ij}(t_{-ij}) < C'_{\Sigma-ij}(t_{-ij})$,

$$W_{Society}|_{(C_{ij}(t_{ij}),C_{\Sigma-ij}(t_{-ij}))} < W_{Society}|_{(C'_{ij}(t_{ij}),C'_{\Sigma-ij}(t_{-ij}))} \tag{6.9}$$

3) Coordinated alternation between Defection and Cooperation does not pay

- Given that $C_{ij}(t_{ij}) < C'_{ij}(t_{ij})$ and $C_{\Sigma-ij}(t_{-ij}) < C'_{\Sigma-ij}(t_{-ij})$,

$$U(C'_{ij}(t_{ij}),C'_{\Sigma-ij}(t_{-ij}),n) > 0.5\{U(C_{ij}(t_{ij}),C'_{\Sigma-ij}(t_{-ij}),n) + U(C'_{ij}(t_{ij}),C_{\Sigma-ij}(t_{-ij}),n)\}$$

$$\tag{6.10}$$

6.3   Information asymmetry

One main objective of this paper is to model, simulate and analyze the outcome of IPGG interaction under information asymmetry. This seems to be a more realistic representation for many real world situations [5], [241], [263], [264]. Players are fundamentally driven by different beliefs when making decisions on the extent to contribute e.g. one may prefer to use a certain information type over another. Asymmetry also accounts for the fact that information may be incomplete e.g. players are allowed access to different pieces of information for the same scenario.

6.3.1 Asymmetric player types

To capture the notion of asymmetry in the IPGG, a collection of $N_{Types} = 4$ player types $\boldsymbol{T} = \{T_1, T_2, T_3, T_4\}$, each differing in the type-dependent information used, is conceptualized in the proposed IPGG. For instance, a player i of type T_1 will only select information subset $\boldsymbol{I}_i(T_1, g_j)$ from superset, $\boldsymbol{I}(g_j)$ in group g_j as the basis of decision-making during game play. In a population of size N, type t_{ij} of player i is such that $t_{ij} \in \boldsymbol{T} \ \forall \ i = 1, 2, ..., n_j$ where n_j is such that $N/10 \leq n_j \leq N/2$. Possible

types which a player can assume are shown in Table 6.1. All information listed is obtained from the previous round of game play. As stated earlier, the type of a player does not change over time.

Table 6.1: Asymmetric Information Types used in the IPGG

Type	Information used	Value	Symbol	Notation
T_1	Number of players in group	n	(—)	NP
T_2	Average contribution of group	$P(g_j)/n$	(Δ)	AC
T_3	Total contribution of group	$P(g_j)$	($\Diamond$)	TC
T_4	Payoff received in the previous round of game play	$Payoff_j(i)$	(*)	PR

6.3.2 Genotypic representation

Chromosomal representations of all the possible types are shown in Table 6.2. The genotype of player i, in essence, represents its strategy, $S_{ij}(t_{ij})$, that creates a non-linear mapping from the set of information that it uses, to a possible contribution level, $C_{ij}(t_{ij})$ within the totality of its action space $A_{ij}(t_{ij})$ e.g. it will encompass a set of rules that informs the player on the amount to contribute for all possible scenarios of interaction. The first gene in each of the genotypes encodes the initial cooperation level, C_I that a player adopts when he interacts with others for the first time. This occurs when new groups are formed – either at the start of simulation or when players switch groups in the course of game play. C_I provides insights about the propensity that each type is willing to initiate cooperation on the first move against an unknown opponent. The remaining genes depict outcome-action pairs, where the next action at any round is determined solely by the outcome of one preceding it e.g. information relevant to a player's type is extracted from the previous round and used to map onto a contribution level which he will thus play in the current round. Accordingly, the inclusion of memory can result in higher

175

average cooperation even under asynchrony [265].

As seen from Table 6.2, genotype representations of players differ according to their corresponding information types e.g. given that $n \in [2,10]$, genotypes of type T_1 are structured such that all discrete possibilities in the information space, which in this case denote every possible values of n ranging from 2 to 10, are covered. Each of these unique possibilities will then map on independently to a contribution level which the player will adopt for use in PG provision.

Table 6.2: Genotypic Representation for Different Information Types

Type	Genotypic Representation											
T_1	n	**I.C.**	2	3	4	5	6	7	8	9	10	
	$C_{ij}(t_{ij})$	C_I	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	
T_2	$P(g_j)/n$	**I.C.**	0.0	0.1	0.2	0.3	…	4.7	4.8	4.9	5.0	
	$C_{ij}(t_{ij})$	C_I	$C_{0.0}$	$C_{0.1}$	$C_{0.2}$	$C_{0.3}$	…	$C_{4.7}$	$C_{4.8}$	$C_{4.9}$	$C_{5.0}$	
T_3	$P(g_j)$	**I.C.**	0	1	2	3	…	47	48	49	50	
	$C_{ij}(t_{ij})$	C_I	C_0	C_1	C_2	C_3	…	C_{47}	C_{48}	C_{49}	C_{50}	
T_4	$Payoff_j(i)$	**I.C.**	0.0	0.1	0.2	0.3	…	6.2	6.3	6.4	6.5	
	$C_{ij}(t_{ij})$	C_I	$C_{0.0}$	$C_{0.1}$	$C_{0.2}$	$C_{0.3}$	…	$C_{6.2}$	$C_{6.3}$	$C_{6.4}$	$C_{6.5}$	

6.3.3 Action spaces

To explore how the number of available choices affects the nature and outcome of decision making among players in the asymmetric setting, two distinct types of action spaces, $A_{ij}^2(t_{ij})$ and $A_{ij}^6(t_{ij})$, each with varied degrees of granularity in decision making, are considered. The prior allows players to contribute only at two extreme levels – full contribution or complete free-riding; while the latter splits the player's contribution into six possible discrete levels (Table 6.3). Whilst $A_{ij}^2(t_{ij})$ is widely used in numerous theoretical studies as a means of simplifying

the analysis of obvious dynamics; $A_{ij}^6(t_{ij})$ accounts for the more realistic fact that

players can actually choose among multiple contribution levels to provision PG in

the practical context. In tone with an earlier assumption, all actions that players

make in the course of game play will always be drawn from action spaces that are

identical in both the cardinality and range. Implicitly, this also assumes that the

availability of choices is homogeneous throughout the entire population.

Table 6.3: Types of Action Spaces used in the IPGG

Type	Possible discrete cooperation levels
$A_{ij}^2(t_{ij})$	Full Defection (0) and Full cooperation (5)
$A_{ij}^6(t_{ij})$	Full Defection (0), Medium Defection (1), Mild Defection (2), Mild Cooperation (3), Medium Cooperation (4) and Full cooperation (5)

6.4   Co-evolutionary learning mechanism

In the proposed IPGG, each player starts off with an initial strategy and betters it

over time through learning. Adaptation to the dynamic environment is by means

of co-evolutionary learning as players of the same type will evolve their strategies

collectively, and independently of other types. This is because framing constrains

the evolution of player strategies within the bounds of their corresponding strategy

structures. During revision, players only switch to strategies which are in line with

their beliefs. It is natural for players of the same type to collate and undertake

group learning - exchanging of ideas. To remain relevant in the game, strategies of

the same type compete in terms of their performance as evaluated against all other

types. This enhances intra-type adaptation, and ensures that good strategies are

constantly adopted in favor of the weaker ones. As more competent strategies

emerged over time, players of asymmetric types also serve as harder opponents

for one another. This aptly accounts for inter-type adaptability. As introduced in

Chapter 2, this elegant co-evolutionary framework captures three distinct aspects of learning within each evolving type, namely.

1) *Learning by replication*

This learning type is analogous to selection in CEAs whose purpose is to ensure that good strategies are adopted in favor of the weaker ones when players decide to revise their current strategies. Implemented using binary tournament selection of size T_S in the IPGG, this comes in two forms:

- Strategy preservation – Among the players whose strategies are selected for propagation to the next generation; those who have derived outstanding welfare are likely to retain their strategies without modification - leaders. In reality, this can also be applied to players who are confident about their strategies or simply those change-adverse ones as well.

- Elitism – There is a strong tendency for a random pool of Z players to revise their strategies by imitating the strongest Z players. As opposed to the prestige-based transmission [266], this is undertaken by followers - players who do not devise their own strategies but merely perform a full-scale adoption of the strategies which are used by those who enjoyed the most welfare at the end of each cycle, encompassing N_{Games} games of N_{Rounds}. In the IPGG, Z constitutes a fixed proportion of the evolving population.

2) *Learning by social exchanges*

- This learning type is analogous to crossover in CEAs whose purpose is to create variations that will differentiate the adopted and original strategies; through the amalgamation of traits between the parent strategies. As opposed to leaders and followers who adopt their strategies wholesale from the previous

cycle, only parts of the original strategies are preserved when players learn by exchanging strategic knowledge and expertise. Such collective social learning – uniform exchange of strategy bits, occurs with probability $P_{crossover}$ in the IPGG and simulates the creation of possibly new, hybrid strategies.

3) Learning by experimenting

- This learning type is analogous to mutation in CEAs where players experiment with small adjustments to strategies to create new ones. Unlike social learning, players fine tune their strategies independently by infusing their own discretion by trial and error. In the IPGG, there is a small probability P_{mutate} that players will revise their strategies by switching randomly to new contribution levels for each possible outcome of interaction.

After each successful phase of co-evolutionary learning – marked as one complete generation; the new set of evolved strategies will be adopted by players in the next interaction cycle. The process of co-evolutionary learning will continue until the maximum of N_{Gen} generations have elapsed.

6.5   Simulation results

Simulations for the IPGG are carried out using Visual C++ development software kit. A summary of the important parameter settings used are shown in Table 6.4.

Table 6.4: List of Parameter Settings used in the Simulation Runs

Symbol	Parameters	Values
N_{Types}	Number of different information types	4
R	Number of simulation runs	20
N	Number of players in the population	{240, 960}
M	Number of groups	{[24, 120], [96, 480]}
n	Size of each group	[2, 10]
N_{Rounds}	Number of rounds played - duration of a game	[1, 200]

N_{Games}	Number of games played – Duration of each interaction cycle	{50, 100}
P_{End}	Probability of ending after each iterated round of game play	0.00346
U_{Basic}	Basic utility derived from private component when no PG is provisioned	2.0
C_{factor}	Value of contribution cost	2.5
R_{factor}	Value of PG provisioned for a player when full cooperation is embraced	5.0
N_{Gen}	Number of generations simulated per run	600
T_S	Tournament size	2
$P_{crossover}$	Probability of crossover or knowledge exchange between players	0.8
P_{mutate}	Probability of mutation or independent learning by each player	0.02
Z	Elitism size or number of imitating players	$0.05*N$

To ensure consistency and eliminate errors due to stochastic variation, the simulation results are averaged over 20 runs, lasting 600 generations each. Every generation will last an interaction cycle of {50,100} games. In a game, players will be randomly collated in groups of 2-10 to play for 200 rounds. Groups of all sizes are equally likely – there will be approximately the same number of groups of each size. Taking into account that players can pull out of a group or a team project can end at any one time, there is a small probability that a game will end after each round. After a game ends, players will reshuffle to form new groups - similar to migration [267] and team switching, N_{Games} times before strategies are revised through co-evolutionary learning. This allows time for players to assess their strategies over an accumulated window of experiences, so that well-informed strategy choices can be adopted eventually.

6.5.1 Homogeneous vs Asymmetric game-play

Simulation is carried out to compare interaction outcomes in both homogeneous [268] and asymmetric settings. The prior solely involves the interaction between players of the same type; while the latter may involve interaction between diverse types. The asymmetric case resembles reality to a much closer degree as being

dissimilar in beliefs and ideologies, heterogeneous players make different sense of the same situation and are likely to make varied decisions accordingly. An action space of $A_{ij}^6(t_{ij})$ is used for all types unless otherwise stated. Interesting insights of the player strategy and usage profiles, cooperation dynamics and welfare level are revealed in the following case comparisons.

1) Performance of different player types: As depicted, the differences in mean welfare to each player - normalized average generation score per round (AGS) – Figure 6.1a and average cooperation level (ACL) – Figure 6.2a exist between groups that use different information types. For $N = 240$, $N_{Games} = 50$, evolutionary traces for the homogeneous setting showed that (AC, NP) attained the highest and lowest (AGS, ACL) respectively after 600 generations. The disparity is due to the inherent differences in structure and nature of information e.g. amortization of group effort subjects AC to less variation and makes it much easier for players to forge cooperative relationships via strategies which are fairly stable to frequent changes in *n*. In comparison, it is difficult for NP players to sustain high (ACL, AGS) as their actions change in direct relation to *n*. AC also provides a clear indication of the effort which an average other contributes and thus the eventual welfare that one is likely to derive. Such knowledge helps to elicit cooperation and facilitates reciprocity – contributing as a positive function of others' contributions [256], among players; as a player is generally willing to contribute conditioned on beliefs that others are doing so at similar levels [249]. This link is not as direct when it comes to other types e.g. TC does not signal about the expected welfare as much depends on *n* while PR varies with one's contribution relative to the group's average. Though it was claimed that evolution provides a simple and effective means to maintain cooperation in a group-structured population, depending solely

181

on population dynamics [269], the above suggests that the nature of information used does affect the tendency to cooperate. Similarities between the AGS and ACL plots indicate that the welfare enjoyed by players in a homogeneous setting is positively correlated to their cooperation levels.

On the contrary, plots (Figures 6.1b, 6.2b) reveal a distinct reversal in the trend of performance for player types under the asymmetric setting. As illustrated by the negative correlation between the convergence traces of ACL and AGS, cooperative player types like AC are worst off in welfare than types adopting pro-defection strategies e.g. NP; as the common environment of interaction effectively opens up opportunities for the latter types to reap a larger share of total welfare by exploiting the prior. This conjures the notion that it does not pay to contribute if others may not be subscribing to similar information types for decision making.

The interdependency between types, as each evolves and adapts its strategies to those of more diverse nature, brings contributions and the ensuing welfare of different types closer together. Unlike the homogeneous setting where traces show signs of recovery after an initial dip, the monotonically declining temporal trends for all types signify further difficulties in achieving voluntary PG provision under asymmetry. Benefits of mutual cooperation do not appear explicit since beliefs of asymmetric players tend to be are misaligned and actions mis-coordinated [270]. Coupled with the prevalence of Social Dilemma, the inability to realize potential gains from contribution fuels development of defect-oriented traits and composes a bleak picture towards provision. Whilst NP players are perceived to enjoy higher welfare from free riding, this is clearly insufficient to offset the drop in welfare for other types. From a collective point of view, AGS under asymmetric interaction is compromised and lower than before.
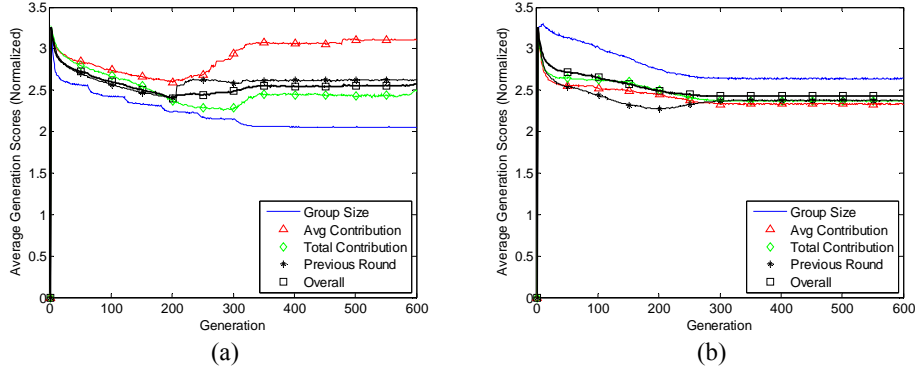
Figure 6.1: AGS of various types for (a) homogeneous and (b) asymmetric game play
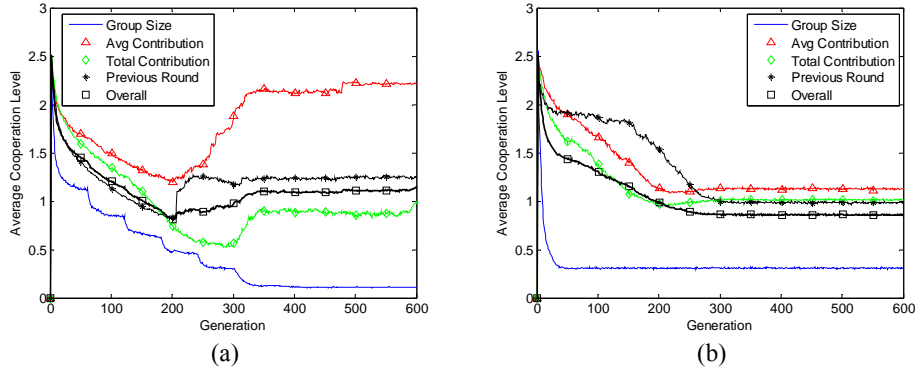
Figure 6.2: ACL of various types for (a) homogeneous and (b) asymmetric game play

2) Duration of interaction and number of players: Studies have shown that the amount of repeated interaction is an important factor which facilitates reciprocal play [271] and population size can also affect the emergent behavior of a group [272]. It is interesting to examine the impact of N_{Games} and N on the dynamics of cooperation for different types. Following an increase of N_{Games} to 100 (Figures 6.3a, 6.4a), only slight changes are detected. Save for these, the traces remained closely similar; indicating that N_{Games} is not a major factor which affects (AGS, ACL). Prolonged periods of contact do not induce significant behavioral change in this case as a substantial amount of interaction is already in place.

However, increasing N to 960 (Figures 6.3b, 6.4b) showed otherwise. In the homogeneous setting, disparity in (ACL, AGS) among types widens as there is a distinct rise in (AGS, ACL) for (AC, PR); whose traces are notably higher than

(NP, TC). These marked differences are due to the presence of a large player pool that emphasizes and accelerates the adoption and propagation of existing traits. For (AC, PR), pressure is exerted for a more cooperative milieu in intra-group interaction as there are more players subscribing to information that induces cooperation. On the same note, the increase in (NP, TC) players highlights pro-defection traits, causing (ACL, AGS) to fall. In contrast, disparity in (ACL, AGS) across types in the asymmetric case narrows. Given a proportionate rise in number of players for each type, a larger pool of less cooperative players accentuates and propagates the free riding cultures throughout the population by compelling the cooperative types that are freely exposed to dangers of exploitation to withhold contribution considerably to near full defection. Overall, players are worse off.

3) Analysis of strategy and usage profiles: Strategy profiles give complete action plans of how players on average, contribute under various outcomes of interaction while usage profiles record the mean frequency of occurrence for all outcomes. Together, the two reveal interesting blueprints of frequency distribution for each action-outcome pair and offer insights into the contribution patterns of types.

With homogeny, strategy profile of NP is by and large most defect-oriented (Figure 6.5a) as seen by the rapid decline of ACL from its highest at $n = 2$ to effectively zero at $n \geq 3$ and the rightward skewing of usage profile. Besides the complexity in multi-player games which comes with expansion in the breadth of possible strategies [271], a study which similarly explored the effect of n on group cooperation [273] via an evolutionary framework [244] verified that the inherent dynamics for the outcomes where $(n = 2)$ and $(n > 2)$ are contrastingly different; given that full, eventual cooperation is achieved for the prior but not the latter. Although it is widely conceived that more PG is generated with large groups, this

184

does not equate to the enjoyment of more welfare per individual. Profiles of (AC, TC, PR) reveal more potential of attaining higher contribution as the players on average, adopt a diverse continuum of ACLs. Whilst (AC, TC) possess upward sloping strategy profiles that associate high contribution with high ACL (Figures 6.5b, c), ACL does not peak where contribution is highest, indicating the existence of a desired effort level where players find most rewarding to contribute. Any contribution below it is too low to sustain stable collaboration while that above it will be high enough to tempt potential defectors to free ride. Alignment of long discrete lines in AC's usage space to peaks in the strategy space denotes high recurrent contributions. Despite similarities, the usage profile is skewed leftwards if TC is used as a gauge of input effort, indicative of the pro-defection traits.

PR's strategy profile is interestingly U-shaped – the players react with high ACL when the previous payoff is low or high (Figure 6.5d). Other than a result of panic response from those who seek to raise AGS by raising ACL, the prior is attributed to the fact that good strategies do practice forgiveness even if previous contribution is exploited. The latter is possibly due to indirect reciprocity [274]-[276] e.g. players may choose to repay the welfare derived from others' efforts circuitously by maintaining high contribution levels towards PG provision. Unlike direct reciprocity [37], altruism can be possible among N-persons [277], [278]. Concentration of usage outcomes in the mid region of the strategy space - where ACL is low, implies a low tendency to contribute when the decision is made from the perspective of personal gains from other players. Many are tempted to free-ride as a means to yield high personal payoff.

In asymmetric interaction, the range of non-zero contribution for NP in groups beyond $n = 3$ is widened surprisingly (Figure 6.5a). This is in retrospective

185

of the fact that cooperative types do induce a willingness to contribute in types that are inclined to free-ride. Players tend to reciprocate contribution when shifted from a pro-defection environment to one where the likelihood of contribution is higher. Likewise, AC's sparsely distributed usage profile is switched to one which concentrates usage in regions of low ACL (Figure 6.5b). The evolution of pro-defection strategies arises in similar principle to NP but with the difference that it involves an exposure of cooperative types to defect-oriented ones. Pro-defection behavior for (TC, PR) is seen by the respective skews of their usage profiles towards regions of low ACL (Figures 6.5c, d).

Usage distribution over a collection of outcome-dependent frequency bins for (AC, TC, PR) indicates that the players transit across a wider spectrum of contributions instead of just a few dominant ones (Figures 5b, c, d), suggesting possibilities for more diverse outcomes. Adding on to the strategy misalignment that occurs as n changes, absence of consistent/Nash contribution level is largely due to incoherent beliefs among types. Overall, asymmetry induces higher and lower ACL for types which are respectively less and more cooperative. There are also more varieties and assortment to the interaction outcomes as well as the ACL of various player types that led to their occurrence.

6.5.2 Varied degrees of decision making and nature of PG

After comparing PG provision under homogeneous and asymmetric information, this section explores effects of varied degrees in decision making and the nature of PG provision on asymmetric interaction outcomes. Scenarios where players can choose from two or six contribution levels are examined for the effects of coarse and fine granularity in decision making. The latter is studied in the previous

186

simulations while the prior involves a scale down of action space to $A_{ij}^2(t_{ij})$ - where choices are restricted to only full contribution or complete free-riding. Subtle as it may be, this constraint has profound repercussions on the willingness to contribute as well as the effective welfare derived.

The impact of free-riding is also explored in two forms of PG – VCM and provision point VCM (PPVCM) [279]. The prior denotes the scenario where PG available for consumption is scaled proportionally by the collective contributions of all within the group e.g. higher aggregate contributions entail greater welfare. In the latter, however, such correspondence does not apply with continuity as PG is provisioned only when a minimum threshold, T is met by the group's mean contribution. Above which, characteristics of a continuous PG prevails but under which, no PG will be provisioned and players derive zero welfare. As an example of a real world analogy, commission is awarded to a team only if a minimum sales target is met. Below T, effort expended for PG provision is wasted and incurred as uncompensated cost. Using $T = 3$, payoff function for player i when provisioning a PG in group g_j of size n_j under PPVCM can be expressed mathematically as:

$$Payoff_j(i) = \begin{cases} U_{Basic} + \dfrac{P(g_j)}{n_j} - Cost_j(i) & if \quad \dfrac{P(g_j)}{n_j} > 3 \\ U_{Basic} - Cost_j(i) & if \quad otherwise \end{cases} \qquad (6.11)$$

Similar in spirit to the case where input choices of players are constrained to two contribution levels, the threshold T restricts the eventual interaction outcome at certain levels of aggregate contributions, to one that involves provision or absence of PG. The prior influences inputs – action spaces while the latter affects the interaction outputs – derived welfares. Under the complex asymmetric interaction, the modifications considered may improve cooperation or worsen the prevailing

effects of Social Dilemma unknowingly. Fusing the above factors, four provision schemes $\{S_1, S_2, S_3, S_4\}$ are devised in Table 6.5, simulated and analyzed to compare their relative benefits in promoting cooperation and improving welfare. Insights gained can then be used to understand how good the provision schemes are in alleviating the effects of free-riding.

Table 6.5: Combinations of Different Settings for Varied Degrees of Contribution and Nature of Provision

Scheme	Degrees of Contribution	Nature of PG Provision	Action Space	Payoff Function
S_1	6	VCM	$A_{ij}^6(t_{ij})$	(4)
S_2	2	VCM	$A_{ij}^2(t_{ij})$	(4)
S_3	6	PPVCM	$A_{ij}^6(t_{ij})$	(11)
S_4	2	PPVCM	$A_{ij}^2(t_{ij})$	(11)

1) Analysis of AGS and ACL for different provision schemes

a) Homogeny of welfare distribution and contribution

Different schemes showed contrasting (AGS, ACL) plots (Figures 6.6, 6.7). Comparing S_1 and S_2, truncating the action space causes further convergence in the traces of different types; as the players effectively focus on just two radically distinct choices - contribute and free-ride. This eases the task of coordinating actions and raises the likelihood that players will contribute at similar levels despite type differences. In S_1, intermediate options accentuate welfare disparity among various types by allowing the freedom to select diverse actions. In S_3, homogeny of (AGS, ACL) traces for various types is also achieved (Figures 6.6c, 6.7c) by constraining the interaction outcomes - as welfare variation only persists when contribution exceeds T. Players are motivated to contribute close to T – minimum ACL required to yield non-zero welfare, as players can avoid deriving zero welfare and yet prevent added contribution beyond T from being exploited.

Overall, imposing choice restriction or structuring a PG by PPVCM translates to greater strategic uniformity and equity in welfare distribution across player types.

Combining the benefits of S_2 and S_3; S_4 nonetheless, does not ensure close resemblance of ACL across types (Figures 6.6d, 6.7d). A liable reason lays in the limitations imposed on both the action and outcome spaces simultaneously. With choice restriction in S_2, the exposure to risks of full exploitation is amplified. Cooperators tend to withhold contribution as non-zero payoffs can still be attained for contributions below T. Prevalence of such traits reduces the overall ACL but enhances uniformity across types. As for outcome restriction in S_3, intermediate choices similarly allow players to reap non-zero payoffs by adjusting ACLs to levels close to T; so that all can benefit without contributing overly.

With both constraints in place, free-riding is however not encouraged as the only way that players can derive non-zero payoffs is through full contribution. By virtue of the strong free-riding effects working against a need to contribute, NP's ACL experiences an inevitable drift from those of other types, towards T. Despite diverging ACL traces, S_4 entails the best welfare distribution as evident from the concurrence of AGS throughout the simulation. While it may be the goal of most provision schemes to ensure an even distribution of welfare by coercing dissimilar types to contribute at similar degrees, S_4 goes a step further to attain the same goal by accommodating differences in contribution patterns. Such robust trait allows each type to preserve its own distinctiveness amid the pursuit of mutual fairness. This is one important aspect of a good provision scheme that is much overlooked.

b) Overall welfare and contribution level

Besides comparing homogeneity in (AGS, ACL), schemes discussed thus far also differ in overall (AGS, ACL) - ranked $\{S_2, S_1, S_3, S_4\}$ in ascending order.

Though fewer choices [145], [262] is typically preferred to more choices in the 2IPD, this is however, less definite as far as the multi-player IPGG is considered under asymmetric information. The situation of analysis is complex - depending on the provision setup, a restriction of choices can work both ways: encourage contribution or breed free-riders. In VCM, multiple contribution levels actually promote higher (AGS, ACL). While it may seem commonsensical to improve the overall welfare by coercing players to execute full cooperation through choice restriction - S_2, immense risk is involved as defectors will free-ride fully. Due to its structure, changes in individual contribution translate only to inconsequential change in AGS especially for large n. Even if one risks exploitation, the group will only benefit marginally. Free-riding is a better choice as players stand to gain if a cooperator subsists in the group. As contribution is not sustained by incentives and its disincentives are not duly compensated by any counter-active measures; players will inevitably choose D over C, leading to a drop in (AGS, ACL). With multiple choices, S_1 offers more opportunities to contribute at levels beyond full defection. This is imperative to facilitate the increase in (AGS, ACL) from S_2.

PPVCM schemes – (S_3, S_4), in contrast depict a clear trend of dominance in (AGS, ACL) over (S_1, S_2). Players, regardless of types, are more contributive and derive higher welfare. This is attributed to nonlinearity of introducing a provision point. With the notion that "*either you get the PG or you don't*", the opportunity costs of not contributing are increased. Players are spurred to raise contribution above T so that efforts expended will not be in vain. For a switch from S_1 to S_3, players are instilled the message "*if an adequate level of contribution is not met, no PG is provisioned and everyone will get no share of the benefits*". This entails higher efficiency in staging the provision task, as players are factually compelled

to raise their contribution stakes or face the adverse outcome of deriving zero PG. This creates an upward pull in the overall ACL to levels where PG is provisioned in non-zero amount. Even so, ACL remains in the range [3, 3.5] as players do not have supporting incentives to embrace full cooperation as VCM sets in beyond T. To avoid over contributing, many players will find it more viable to contribute at mid levels below full cooperation.

S_4 addresses this issue via restricting choices. As the most stringent scheme, it expects all to "*contribute to their best or risk provisioning no PG*". The message that conveyed a need to contribute is a much stronger one-given the same number of cooperators in S_3 and S_4, those in the latter will be restricted to play only full contribution. This accounts for the momentous rise in ACL for most types to [4, 4.5], though NP still chooses to free-ride sporadically. Overall AGS is raised to a significant 3.3. Switching to PPVCM confers more benefits than simply restricting choices. The former changes the entire structure of what is provisioned [222], not just a tweak in the inner settings. Combination of both features is the best setting to encourage contribution and achieve efficient PG provision.

c) Overall trend and slope characteristics/dynamics

Eventual convergence of ACL traces (Figure 6.7) signifies the presence of evolutionary stable welfare levels that players of each type are willing to play so as to derive from. Even so, relation between ACL and AGS is no longer explicit - it cannot be ascertained whether a higher or lower ACL will yield higher AGS. AGS of each type will depend much on the nature of its information and strategies of others. VCM and PPVCM schemes are differentiated by dissimilarities in the trend of progression for (AGS, ACL) e.g. (S_1, S_2) exhibit a declining trend while (S_3, S_4) depict an upward moving one with generation. This is because for the

191

same strategy pool, ACL falls when the temptation to free-ride sets in for VCM, which fuels a drop in AGS for all types. For PPVCM, players are motivated to raise contributions above T or risk deriving no welfare. This translates to a rise in AGS over time. Overall, the PPVCM schemes are more effective in mitigating, if not eliminating the effects of Social Dilemma among varied types.
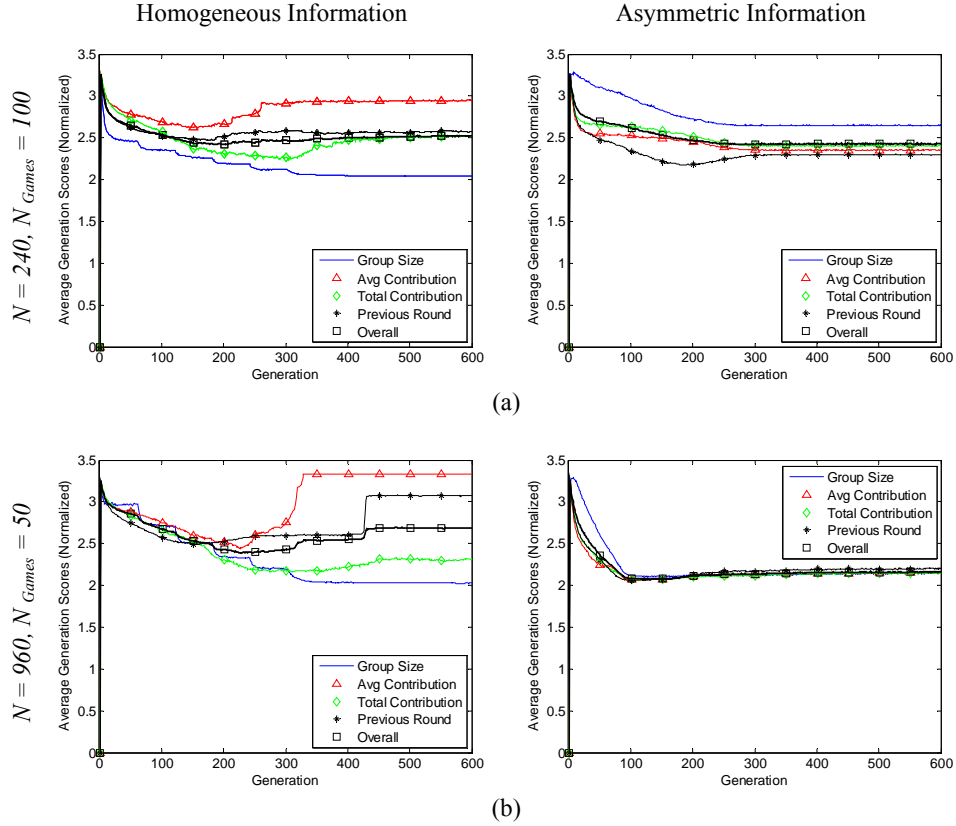
Figure 6.3: AGS of different player types for changes in (a) N_{Games} and (b) N
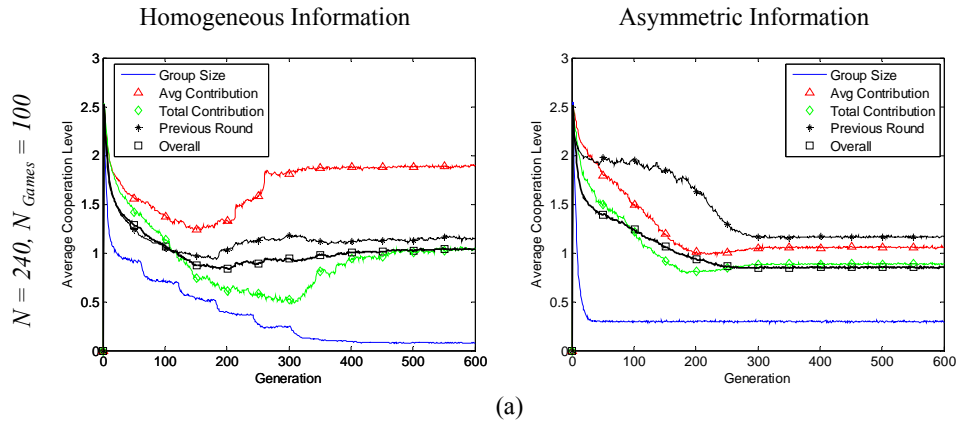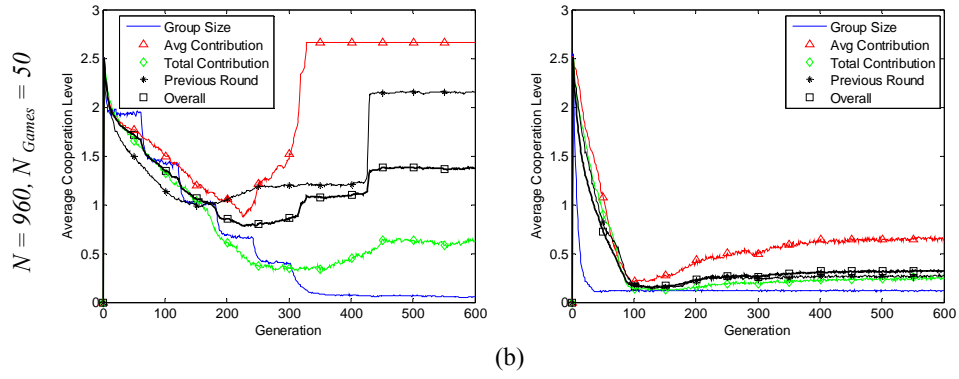
Figure 6.4: ACL of different player types for changes in (a) N_{Games} and (b) N

(a)

193
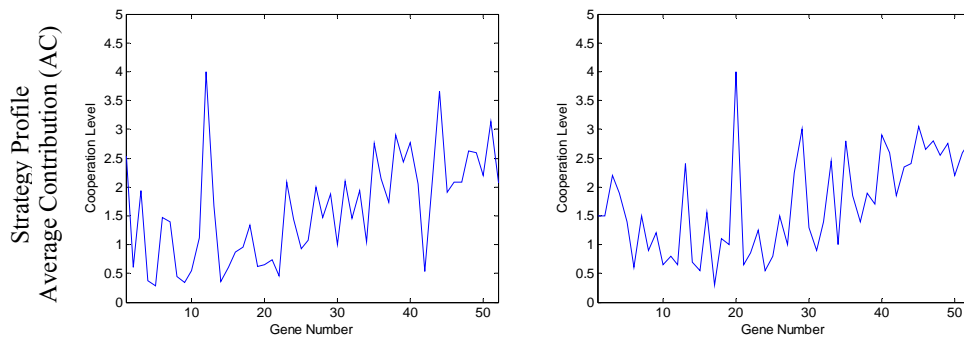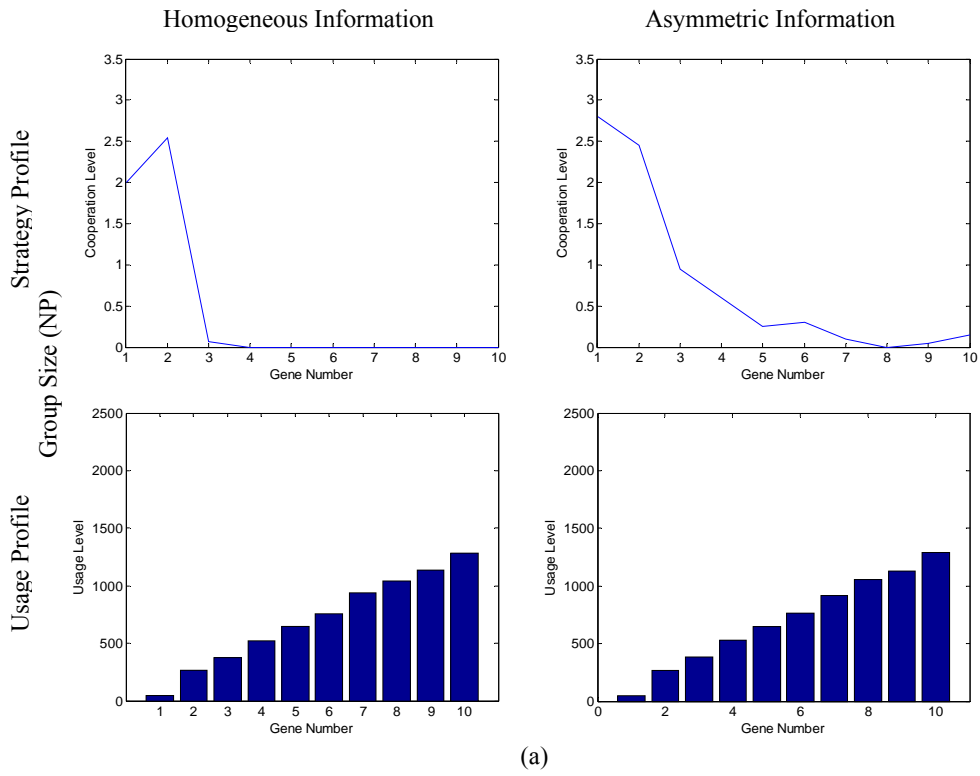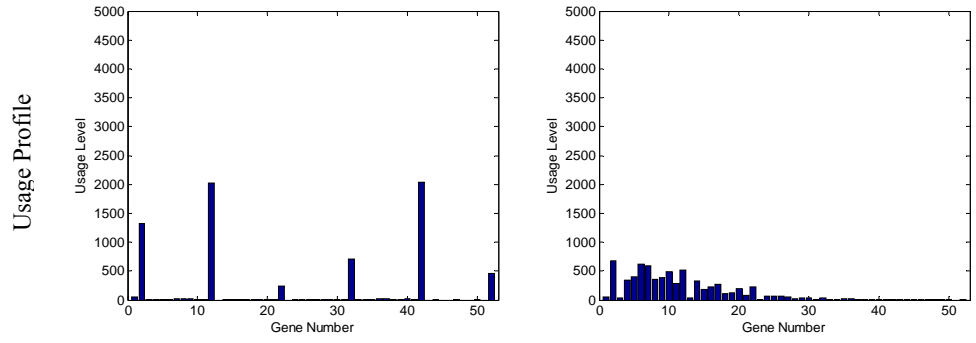
(b)
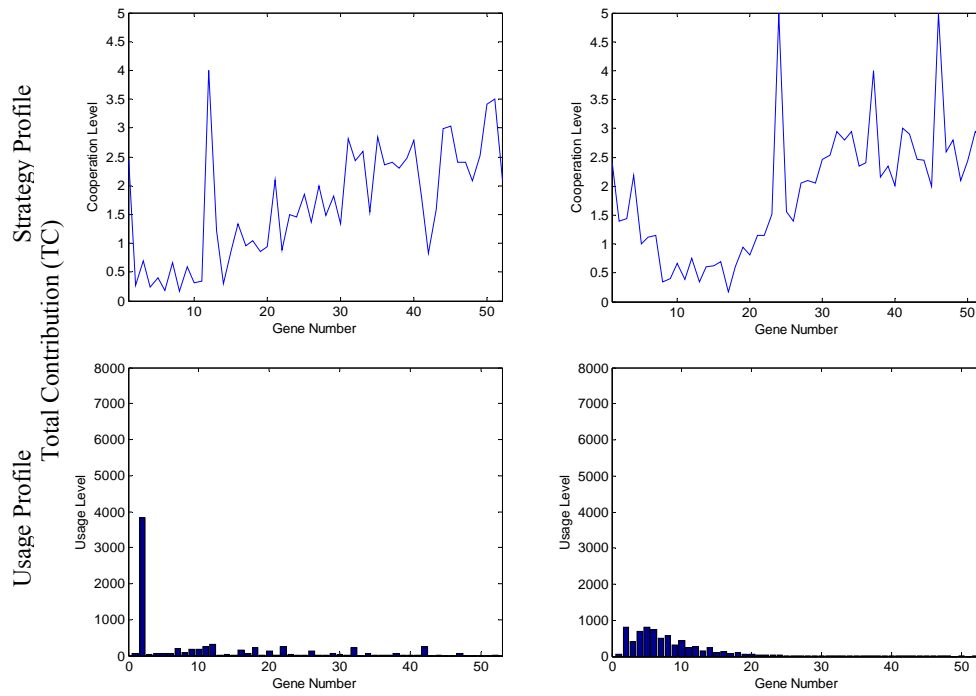
(c)
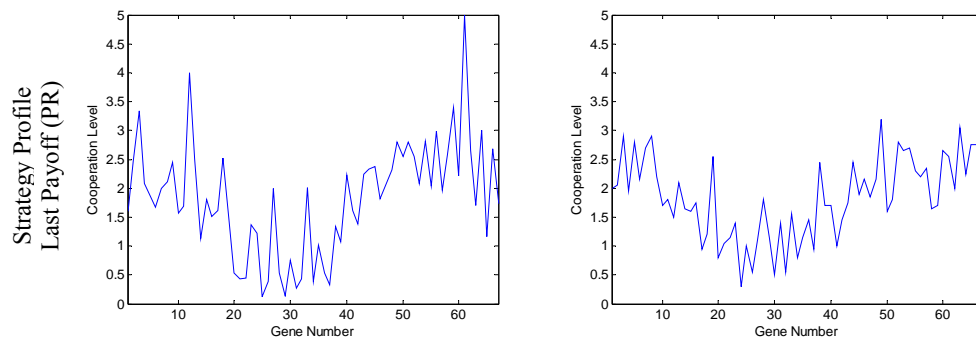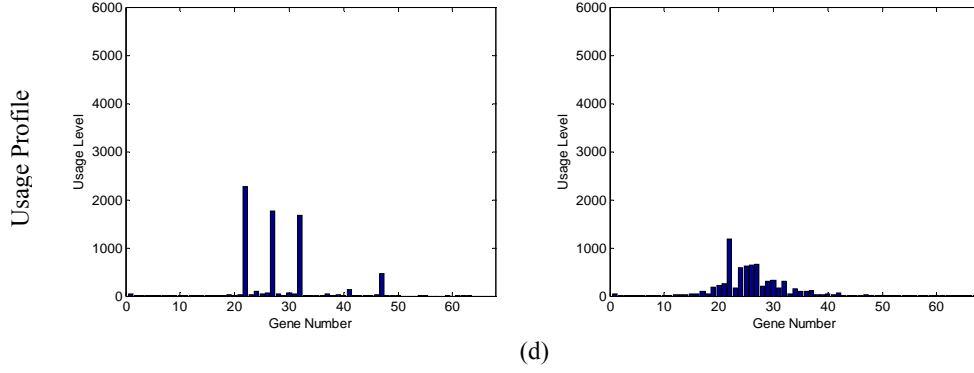
Figure 6.5: Strategy and usage profiles for type (a) NP, (b) AC, (c) TC and (d) PR under homogeneous and asymmetric information

3) Does a higher threshold trigger higher contribution?

After ascertaining that the PPVCM schemes do work better in promoting contribution via a minimum threshold, T, that pulls mean contribution up; the next task is to verify how the size of T affects welfare and extent of free riding [235]. Does the influence pattern differ with number of available choices? These queries can be answered by studying (AGS, ACL) of (S_3, S_4) for $T = \{2, 3, 4\}$ (Figures 6.8, 6.9). At $T = 2$, both traces are lowest due to the low incentive to contribute. It is reasonably foreseeable that more players are driven to free-ride at low Ts when choices are limited, as it makes no sense to match maximum effort with a low contribution goal. One can enjoy more benefits by leveraging on cooperators to realize contribution levels that are just enough to fulfill the provision task. From $T = 2$ to 3, overall (AGS, ACL) for (S_3, S_4) rose. Increase in ACL for S_3 is mild and translates only to a slight rise in AGS. In contrast, S_4 experiences a large shift in ACL that pushes AGS beyond the level in S_3, indicating greater sensitivity to changes in T. Probability that individual contribution determines provision or non-provision of PG is raised [222]. More players switch to full contribution as it is no longer enough to depend on others for PG provision - it takes three cooperators at full contribution to cover two free-riders as compared to two cooperators at full

195

contribution to cover three free-riders previously. The fact that (AGS, ACL) of S_4 supersedes S_3 from below signifies the potential to attain higher contribution via the synergistic blend of PPVCM and choice restriction, despite starting off low.

At $T = 4$, S_4 continues to experience a steady but smaller step up in (AGS, ACL), but S_3 a fall in AGS for corresponding rise in ACL. In conjunction with the less than proportionate rise in AGS from $T = 2$ to 3, the drop in AGS for S_3 is primarily due to the assorted actions that varied types undertake in each group. Amplified by multiple choices, action coordination is much harder as many may prefer to contribute at intermediate levels. This can cause ACL of a typical group to fall short of T. Thus, despite the overall increase in ACL, a decline in AGS is observed as costs are expended without achieving benefits. S_4 can realize the trend of increasing AGS as choice coordination is much clearer and players do have a tendency to contribute when deciding between the two extreme alternatives.

Presence of a period for players to coordinate their actions exists under S_4, as illustrated by the intersection for the family of S-curves (Figure 6.8b) – 150th generation. At $T = 2$, players can attain high AGS with their initial strategies but the inclination to free-ride tends to reduce overall AGS as actions are fully coordinated. For $T = 4$, overall AGS starts low as players are unable to fulfill the provision point requirement initially. The trend reverses when all are driven to coordinate their actions to attain higher AGS. The distinct S-shape is due to differences in the starting and ending AGS. The higher the value of T, the lower the initial AGS but the greater is the potential of attaining high eventual AGS. Overall, increasing T raises welfare by setting a high provision target to induce players to contribute, but this does not always hold as high T values are only beneficial when the players face limited choices. With multiple choices, the goal to

attain high contribution is likely to be hindered by a dilution of mean contribution. This can lead to a negligible increase or even a decline in overall AGS as T gets larger (Figure 6.8a).

6.5.3 Multi-level selection: group vs individual reward

After considering the means of improving contribution and welfare by restricting choices and altering the nature of PG provision, it is also of interest to explore the impact of varied selection schemes on the outcome of asymmetric interaction. Motivated by concepts of multi-level selection [280], [281] from evolutionary biology [282], it is known that different selection schemes can entail diverse outcomes. With supplement from the kinship theory [38] and reciprocal altruism [37], making a group the unit of selection [283] can provide explanation for the evolution of altruism [284], [285] and cooperation [286].

Consider a case where a firm wishes to reward its staff for past achievements and contributions. To ensure that corporate cultures of voluntary contribution and mutual cooperativeness continue to spread throughout the workforce in future; is it desirable to reward on an individual or group basis or a combination of both? Such a decision is crucial as it impacts the underlying organizational dynamics and sets the course that staff should work towards. When implemented correctly, a good reward scheme can raise the morale of deserving personnel, improves the overall efficiency of the work crew which leads on to ease of completing big-scale public projects by highly contributive individuals and groups. With the above objectives in mind, the last case study seeks to compare three different selection schemes and analyze their relative advantages. The schemes $\{S_I, S_G, S_M\}$ are

197

1) S_I : *Individual selection based on group effort*

Taking into account the fact that individuals are likely to switch groups from time to time, one is selected for reward according to mean accumulated effort per individual that is channeled by all groups that he has previously participated in for the PG provision. Since PG can refer to any community assignment or large-scale project where the payouts are not directly correlated to the effort expended by an individual, this performance measure helps quantify individual effort, which is usually hard to assess in a group context, owing to loafing and moral-hazard issues [287]. In simpler terms, the higher the amount of PG generated by groups that one was formerly a part of e.g. the more successful the past projects which one took part; the higher the chances of reward.

2) S_G : *Group selection based on group effort*

Individuals are selected for rewards on a group basis, in accordance to the efficiency in generating PG. As opposed to S_I, the unit of selection is the group. In essence, groups that can harness higher mean contribution per player will be in favor of being chosen; and upon successful selection, all individuals in the group are rewarded.

3) S_M : *Multi-level selection based on group effort*

As the name suggests, reward selection is done at both the individual and group levels. Selection criteria are identical to both S_I and S_G, so as to verify whether a combination of the previous two schemes delivers the best mechanism for reward than when either of them is considered separately.
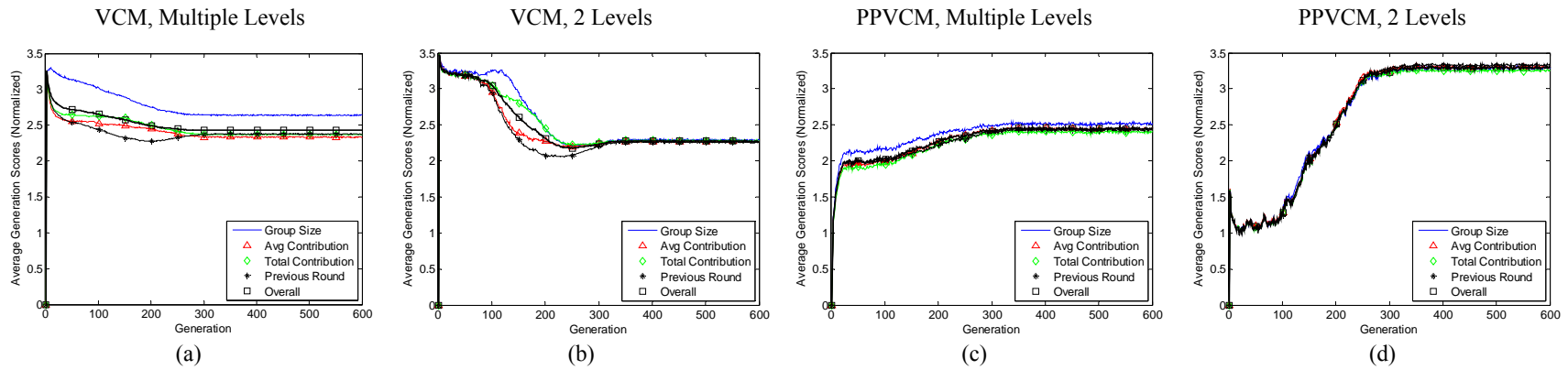
Figure 6.6: AGS for (a) S_1, (b) S_2, (c) S_3 and (d) S_4 with $N = 240, N_{Games} = 50$
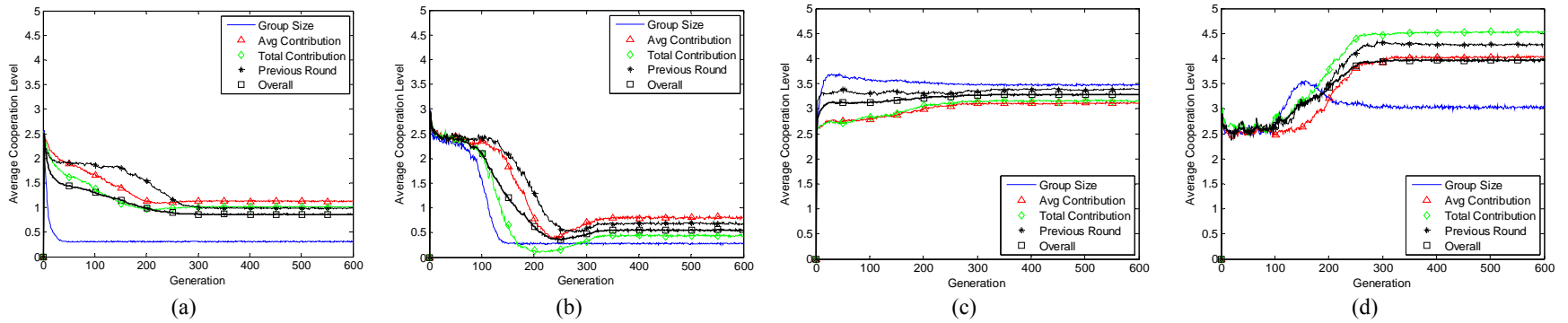
Figure 6.7: ACL for (a) S_1, (b) S_2, (c) S_3 and (d) S_4 with $N = 240, N_{Games}=50$
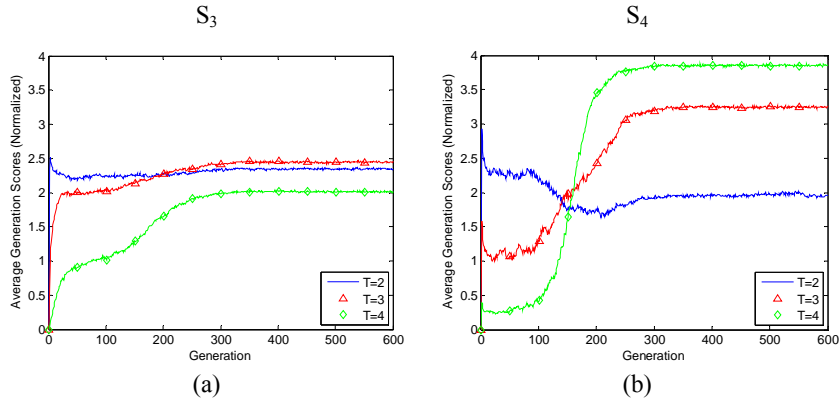
Figure 6.8: Overall AGS for (a) multiple and (b) two levels of contribution
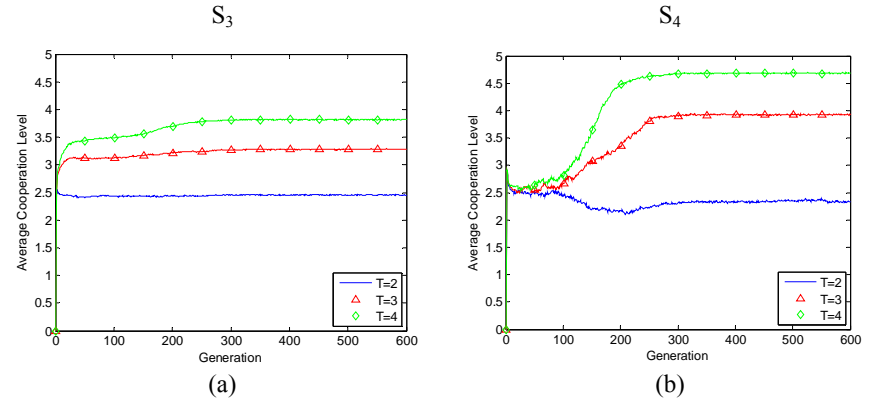
Figure 6.9: Overall ACL for (a) multiple and (b) two levels of contribution
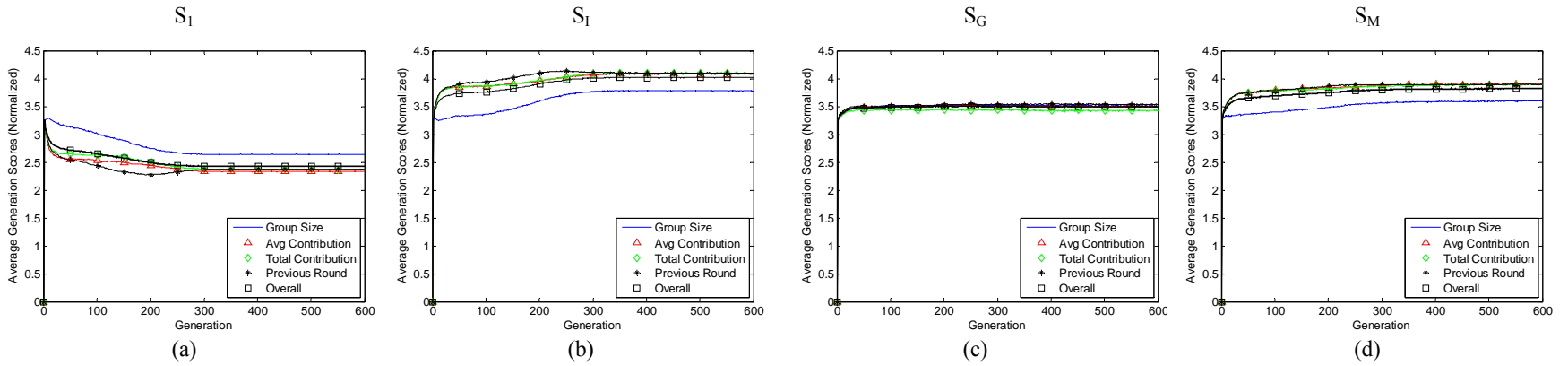
Figure 6.10: AGS for (a) S_1, (b) S_I, (c) S_G and (d) S_M with $N = 240$, $N_{Games} = 50$
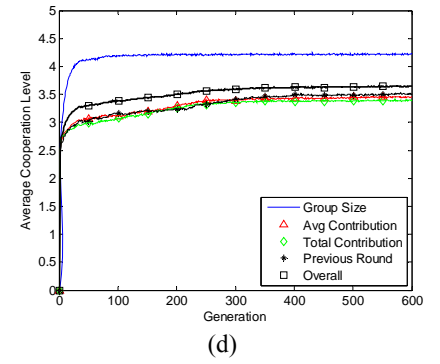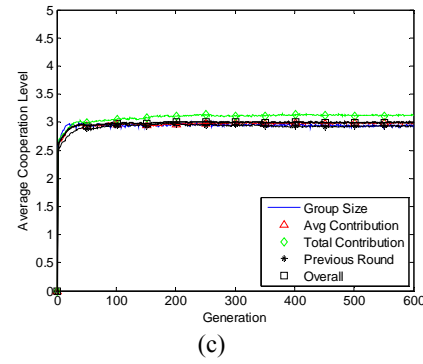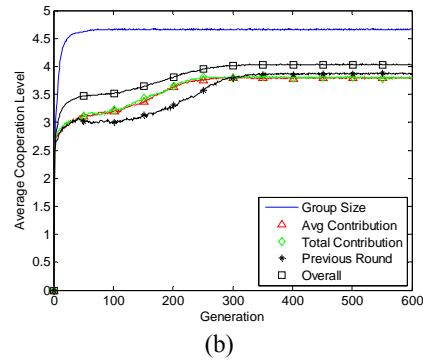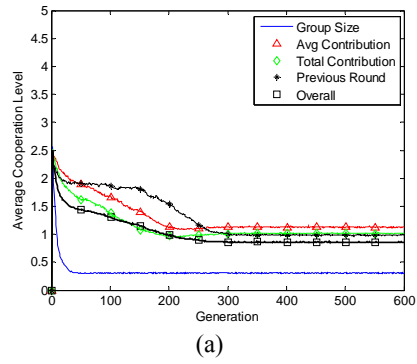
Figure 6.11: ACL for (a) S_1, (b) S_I, (c) S_G and (d) S_M with $N = 240$, $N_{Games}=50$

201

The schemes are simulated and outcomes of PG provision are analyzed (Figures 6.10, 6.11). Previous results for S_1 which select individuals based on individual welfare in each group are used as the common basis of comparison. It is clear that all schemes that select using group contribution supersede S_1. This is due to cost factoring when selecting based on individual welfare. A player that is selected based on high welfare level does not necessarily imply a cooperator by nature. In contrast, it turns out more frequently that he is either a free-rider that exploits the others' contribution successfully or a weak cooperator that withholds contribution tacitly by leveraging on others' effort. Selecting these players, to some extent, results in propagation and adoption of free-riding traits by players for the subsequent strategies. This inevitably imposes a limitation on the extent to which (AGS, ACL) can reach eventually.

Comparing among schemes which reward on the basis of group performance, S_I presents the highest (AGS, ACL) at the end of 600 generations, followed by S_M, then S_G. Although group selection can promote intra-group cooperation by raising the inter-group competition [287], this is not so in the context of IPGG. The downside rests on the fact that S_G has no way of distinguishing effectively amid cooperators and free-riders. This is an essential point of consideration as a group that does well in generating high mean contribution per individual may comprise a mixture of very cooperative players as well as mediocre free-riders. An attempt to reward all players in a group regardless of individual contribution is bound to admit free-riders for an equal share of the reward pie; even if they did not play any substantial part in realizing the PG provision- e.g. benefits from success of project becomes public to all. By handing out rewards to free-riders on top of benefits that they have already gained by exploiting contributions, S_G is clearly designed with

come elements of unfairness. Similar to S_1, there is a possibility for the continual proliferation of free-riding traits in group selection. Apart from fairness, S_G is also inefficient as iterated use of the scheme limits willingness to contribute at 3 and restricts mean welfare that each individual ultimately gets.

Unlike S_G, S_I projects a fairer and more efficient reward system. Although it is hard to capture individual contribution from group performance, amalgamation of achievements from all past coalitions that an individual had joined does piece up to provide a good clue and indicator. The entirety of an individual's history of group contribution implicitly captures a good perception of his effort level e.g. the more group achievements accumulated over his history of participation in teams within a specific time frame; the greater will be the likelihood that the player is an important and substantial contributor to the success of all his teams. Using this form of performance measure as a basis to select individuals for rewards clearly provides a fairly good means of differentiating between efforts put in by each member. This not only achieves fairness by excluding exploitative personnel from a share of the reward, but more importantly, it provides an exceptional driving force to motivate existing free-riders to contribute to avoid losing out in future reward opportunities. (AGS, ACL) of S_M is middling as it possesses the properties of both S_G and S_I e.g. the disadvantages of S_G somehow dilutes the advantages introduced by S_I. On the whole, S_I leads to a clear dominance of contribution and welfare over all other schemes.

6.6   *Findings and discussions*

Interesting findings are presented in the course of the co-evolutionary simulation, which aids in the understanding of differences in PG provision in homogeneous

and asymmetric interaction. Verified from a myriad of settings, players' beliefs are more aligned in the former, where certain strategies and outcomes tend to be strikingly more dominant. In the latter, a more diverse spectrum of strategies and outcomes is entailed due to misaligned beliefs. Though interaction amid dissimilar types leads to homogeny in welfare distribution, contribution certainly does not pay as efforts of cooperators are exploited by free-riders. This lowers mean contribution and welfare to all. Overall, free-riding is more pronounced in the asymmetric setting and players are generally worse off. While research has shown conclusive evidence that less choices is preferred to more in enhancing cooperation in pair-wise interaction, results of simulation studies have ascertained that this notion is less definite for an asymmetric setting with multiple players. Much depends on the structure of the provision scheme.

Restricting choices truncates action spaces and helps to align contributions among similar types; while the PPVCM coerces players to contribute above a minimum provision point or risk deriving no welfare. A combination of limited choice and PPVCM – where large thresholds induce more contribution; offers an effective means of mitigating the Social Dilemma. However, this fact does not hold true for multiple choices; as the goal of attaining higher contributions is hindered by a dilution of average contribution when players contribute at different levels. Finally, although it seems fairer to reward individuals on a group basis - as only group performance can be accurately assessed; such scheme suffers an inherent drawback of not being able to discriminate among cooperators and free-riders. Results have shown that contribution and welfare can be increased by rewarding on an individual basis, using the collection of group achievements for which the individual was previously attached to.

6.7 Summary

In conclusion, this chapter presents a co-evolutionary approach to model and implement an IPGG using ACM such that collective outcomes of PG provision under asymmetric information can be effectively simulated and analyzed. The simulated results reveal interesting interaction dynamics and added difficulties in achieving cooperation when information asymmetry is present among players. In general, the proposed framework provides a very useful platform to gain a better understanding of collective action and some insights into how the effects of Social Dilemma can be mitigated. This might in turn offer some ideas on how efficient public goods provision can be achieved in the practical context.

Chapter 7

Conclusion

Co-evolutionary simulation modeling is the application of stochastic CEAs to simulate the process of evolution and adaptation in ACMs. It has been found to be an efficient and effective framework to model, simulate and further the analysis of strategic interaction from numerous perspectives, especially when conventional analytical and empirical approaches fail under their intrinsic constraints. Inspired by Nature's evolutionary principles, where uncertainty is a common and inherent phenomenon, CEAs become a natural candidate to model realistic imperfections which mirrors and constitute real world interaction. As an optimization paradigm which functions primarily based on probabilistic and population-based searches, CEAs provide a dynamic framework that drives co-evolutionary learning and strategy improvement when agents interact in game theoretic settings – in which an absolute fitness measurement that reflects the underlying properties of games is extremely difficult, if not impossible to formulate. Equipped with a myriad mix of desirable characteristics, it will be interesting to examine the use of CEAs as a viable alternative and complementary avenue to existing approaches, particularly as a means to facilitate the discovery of good game strategies, analyze collective interaction outcomes of and gain better insights into the underlying dynamics that leads naturally to those outcomes.

7.1   Contribution

This work contributes towards to the application of CEAs to model, simulate and analyze game theoretic interaction in several interesting contexts of study. Chapter

3 focuses on the development of a competitive computer player for the one versus one Texas Hold'em poker using CEAs. A Texas Hold'em game engine is first constructed where an efficient odds calculator is programmed to allow for the abstraction of a player's cards, which yield important but complex information. Effort is directed to realize an optimal player which will play close to the NE by proposing a new fitness criterion. Preliminary studies on a simplified version of poker highlighted the intransitivity nature of poker. The evolved player displays strategies which are logical but reveals insights that are hard to comprehend e.g. bluffing. The player is benchmarked against Poki and PSOpti, which is the best heads-up Texas Hold'em A.I. to date and plays closest to the optimal NE. Despite the much constrained chromosomal strategy representation, the simulated results verified that CEAs are effective in creating strategies that are comparable to Poki and PSOpti in the absence of expert knowledge.

Chapter 4 examines the comparative performance and adaptability issues of evolutionary, learning and memetic strategies in different environment settings in the IPD. Evolutionary strategies are realized by GA based on co-evolutionary principles and learning strategies by a double-loop incremental learning scheme, ILS that incorporates a classification component, probabilistic update of strategies and feedback learning mechanism. A memetic adaptation framework is developed for IPD strategies to exploit the complementary features of evolution and learning. In the framework, learning serves as a form of directed search to guide evolving strategies to attain eventual convergence towards acquiring good strategy traits while evolution helps to minimize disparity in performance among the learning strategies. A series of simulation results verify that the two adaptation techniques, when employed concurrently, are able to complement each other's strengths and

compensate for each other's weaknesses, leading to the formation of strategies that will adapt and thrive well in complex, dynamic environments.

Chapter 5 focuses on the development of a spatial evolutionary multi-agent social network to study the macroscopic-behavioral dynamics of civil violence that culminates as a result of the microscopic game-theoretic interactions between the goal-oriented agents. Agents are modeled from multi-disciplinary perspectives and their strategies are evolved over time through collective co-evolution and independent learning. Spatial and temporal simulation results reveal fascinating global emergence phenomena as well as interesting patterns of group movement and autonomous behavioral development. Extensions of varying complexity are also used to investigate the impact of various decision parameters on the outcome of unrest. Analysis of the results provides insights into the intricate dynamics of civil upheavals and serves as a good avenue to gain a more holistic understanding of the fundamental nature of civil violence.

Chapter 6 presents a co-evolutionary, game theoretic approach to simulate and study the collective outcome of public goods provisioning in an agent-based model. Using asymmetric information as the basis for decision making, distinct groups are configured to interact in an iterated N-player public goods game, where co-evolutionary learning is used as the mechanism of adaptation to the dynamic environment. The impact of information type, number of players, group size, rate of interaction, number of available choices, nature of PG provision and selection schemes are studied over a variety of settings. Simulation results reveal interesting dynamics of strategy and usage profiles, level of derived welfare and the evolution of cooperation. Analysis of these attributes offers a more holistic understanding into the nature of collective action and some insights of how the effects of Social

Dilemma can be mitigated. This might provide a good guide to achieve efficient public goods provision in the practical context.

7.2 Future works

Although we have successfully applied CEAs to model game theoretic interaction and examined the outcomes of agent-based co-evolutionary simulation in different contexts, the series of works presented in this thesis barely scratched the surface of what is potentially left to be addressed.

The current poker model presented in Chapter 3 can be further improved from several perspectives. Better strategies could be evolved by simply increasing the precision of strategy parameters e.g. splitting hand strength information into finer intervals, or incorporating more parameters like position information in the model e.g. so as to account for scenarios with multiple players in a poker game. Such are, however, subjected to the availability of computational resources. The co-evolutionary process can also be sped up by injecting expert knowledge in the form of fixed non-evolving opponents. Though these players do not evolve, they do affect the fitness of evolving players and play a crucial role in shaping their strategies. On a side note, a better fitness criterion or tournament model can also be devised so that fluctuations due to intransitivity can be further reduced.

As far as the IPD study in Chapter 4 is concerned, possible improvements can encompass experimental simulation of the IPD game in the presence of other sophisticated benchmark strategies, deriving efficient learning methodologies as well as applying memetic learning to complex test settings through adding noise, devising complex payoff matrices and conducting evolutionary tournaments to analyze the interaction between strategies in terms of their growth rate and group

performance. A thorough research study to investigate complicated situations like the above-mentioned would be useful in providing us with greater insights into the intricacies and complexity involved in the IPD.

In the aspects of civil violence modeling presented in Chapter 5, additional research work can be carried out to study how the specific movement strategies of various agent groups are evolved over time, impact of vision radius and situational awareness on the performance of agents as they negotiate their way through the environment of interaction, extending the proposed spatial IPD model by adopting an N-player mode of game theoretic interaction, and exploring interesting areas of behavioral development. Maturity of such models will not only serve as a form of verification for complex social theories but more importantly, present a feasible avenue to simulate realistic scenarios of civil violence; in the hope to formulate violence management measures that are paramount to the mitigation of casualties.

For the multi-player IPGG in Chapter 6, future works can be embarked on to investigate models which incorporate behavioral elements such as punishment, reputation and mutual expectation; as well as those where players can realistically adopt beliefs that vary from time to time. Though interesting, complexity of the inherent model dynamics must be well managed for simulation outcomes to be effectively and meaningfully analyzed. Apart from just contributing or free-riding, the action spaces of players can be extended to include an added option of non-participation. Assessing possible impacts of the above-mentioned model attributes can provide avenues that will aid in constructing more cooperative landscapes.

Bibliography

[1] M. M. Gibbons, *An introduction to game-theoretic modeling*, Student Mathematical Library vol. 2. AMS, 2001.

[2] R. Gibbons, *A Primer In Game Theory*. Prentice Hall: Great Britain, 1992.

[3] J. F. Jr. Nash, "Equilibrium Points in n-Person Games," in *Proceedings of the National Academy of Sciences* 36, U.S.A., 1950, pp. 48-49.

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

[5] G. A. Akerlof, "The Market for Lemons: Quality, Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, vol. 84, no. 3, pp. 488–500, 1970.

[6] S. Parsons, M. Wooldridge, "Game Theory and Decision Theory in Multi-Agent Systems," *Autonomous Agents and Multi-Agent Systems*, vol. 5, no. 3, pp. 243-254, 2002.

[7] R. K. Dash, D. C. Parkes, and N. R. Jennings, "Computational Mechanism Design: A Call to Arms," *IEEE Intelligent Systems*, vol. 18, no. 6, pp. 40-47, 2003.

[8] J. von Neumann, and O. Morgenstern, *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

[9] J. C. Harsanyi, and R. Selten, *A General Theory of Equilibrium Selection in Games*. MIT Press, 1988.

[10] A. M. Spence, "Job Market Signaling," *Quarterly Journal of Economics*, vol. 87, no. 3, pp. 355-374, 1973.

[11] J. E. Stiglitz, and A. Weiss, "Credit Rationing in Markets with Imperfect Information," *The American Economic Review*, vol. 71, no. 3, pp. 393–410, 1981.

[12] R. J. Aumann, and M. B. Maschler, *Repeated Games with Incomplete Information*. MIT Press, Cambridge, MA, 1995.

[13] T. Schelling, *The Strategy of Conflict*. Harvard University Press, 1980.

[14] L. Hurwicz, "The design of mechanisms for resource allocation," *American Economic Review*, vol. 63, no. 2, pp. 1-30, 1973.

[15] E. S. Maskin, "Mechanism Design: How to Implement Social Goals," *American Economic Review*, vol. 98, no. 3, pp. 567-76, 2008.

[16] R. B. Myerson, "Mechanism Design by an Informed Principal," *Econometrica*, vol. 51, pp. 1767–1797, 1983.

[17] W. Poundstone, *Prisoner's Dilemma*. Doubleday, NY, 1992.

[18] J. Weesie, "Asymmetry and Timing in the Volunteer's Dilemma," *Journal of Conflict Resolution*, vol. 37, no. 3, pp. 569-590, 1993.

[19] A. Diekmann, "Cooperation in an Asymmetric Volunteer's Dilemma Game: Theory and Experimental Evidence," *International Journal of Game Theory*, vol. 22, no. 1, pp. 75-85, 1993.

[20] D. P. Myatt, and C. Wallace, "An Evolutionary Analysis of the Volunteer's Dilemma," *Games and Economic Behavior*, vol. 62, no. 1, pp. 67-76, 2008.

[21] J. P. Delahaye, P. Mathieu, and B. Beaufils, "The Iterated Lift Dilemma: how to establish meta-cooperation with your opponent," in *Computational conflicts: conflict modeling for distributed intelligent systems*, H. J. Müller, and R. Dieng, Eds. NY: Springer-Verlag, 2000, pp. 202 – 223.

[22] S. Y. Chong, and X. Yao, "Behavioral Diversity, Choices and Noise in the Iterated Prisoner's Dilemma," *IEEE Transactions on Evolutionary Computation*, vol. 9, no. 6, pp. 540-551, 2005.

[23] E. A. Stanley, D. Ashlock, and M. D. Smucker, Iterated prisoner's dilemma with choice and refusal of partners: Evolutionary results, in *Lecture Notes in Computer Science* Vol. 929. Springer Berlin, Heidelberg, 2006, pp. 490-502.

[24] B. O'Neill, Game Models of Peace and War: Some Recent Themes, in *Diplomacy Games*. Springer Berlin, Heidelberg, 2007, pp. 25-44.

[25] J. M. Smith, "Theory of games and the evolution of animal contests," *Journal of Theoretical Biology*, vol. 47, pp. 209-221, 1974.

[26] D. Lee, "Game theory and neural basis of social decision making," *Nature Neuroscience*, vol. 11, pp. 404 – 409, 2008.

[27] P. R. Kleindorfer, H. Kunreuther, P. J. H. Schoemaker, *Decision Sciences: An Integrative Perspective*, Cambridge University Press, England, 1993.

[28] M. Land, O. Gefeller, "A game-theoretic approach to partitioning attributable risks in epidemiology," *Biometrical journal*, vol. 39, pp. 777 – 792, 1997.

[29] P. A. Samuelson, "The Pure Theory of Public Expenditure," *The Review of Economics and Statistics*, vol. 36, no. 4, pp. 387-389, 1954.

[30] D. Dolgov, E. Durfee, "Computationally-efficient combinatorial auctions for resource allocation in weakly-coupled MDPs," in *Proceedings of the fourth international joint conference on autonomous agents and multiagent systems*, 2005, pp. 657 – 664.

[31] S. J. Brams, *Negotiation Games: Applying Game Theory to Bargaining and Arbitration*. New York: Routledge, 2003.

[32] S. Kraus, Automated negotiation and decision making in multiagent environments, in *Multi-agents systems and applications*. Springer-Verlag New York, 2001, pp. 150 – 172.

[33] J. D. Miller, *Game Theory at Work: How to Use Game Theory to Outthink and Outmaneuver Your Competition*. McGraw-Hill, 2003.

[34] B. Lebaron, "A builders guide to agent-based financial markets," *Quantitative Finance*, vol. 1, no. 2, pp. 254–261, 2001.

[35] C. Boix, "Setting the Rules of the Game: The Choice of Electoral Systems in Advanced Democracies," *American Political Science Review*, vol. 93, no. 3, pp. 609-624, 1999.

[36] J. M. Smith, *Evolution and the Theory of Games*. Cambridge University Press, 1982.

[37] R. L. Trivers, "The evolution of reciprocal altruism," *Quarterly Review of Biology*, vol. 46, pp. 35-57, 1971.

[38] W. D. Hamilton, "The genetical evolution of social behaviour I and II," *Journal of Theoretical Biology*, vol. 7, pp. 1-16 and pp. 17-52, 1964.

[39] C. F. Camerer, *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press, 2003.

[40] C. F. Camerer, "Behavioural studies of strategic thinking in games," *Trends in Cognitive Sciences*, vol. 7, no. 5, pp. 225-231, 2003.

[41] C. F. Camerer, T. H. Ho, and J. K. Chong, "Models of Thinking, Learning and Teaching in Games," *American Economic Review*, vol. 93, pp. 192–195, 2003.

[42] J. C. Harsanyi, "Games with Incomplete Information Played by 'Bayesian' Players, Parts I, II and III," *Management Science*, vol. 14, pp. 159-182, 320-334 and 486-502, 1967-1968.

[43] J. K. Goeree, and C. A. Holt, "Ten little treasures of game theory and ten intuitive contradictions," *American Economic Review*, vol. 91, no. 5, pp. 1402 – 1422, 2001.

[44] S. Phelps, "Evolutionary mechanism design," Ph.D. dissertation, University of Liverpool, Liverpool, UK, 2007.

[45] P. Humphreys, *Extending Ourselves: Computational Science, Empiricism, and Scientific Method*. Oxford: Oxford University Press, 2004.

[46] R. Axelrod, and L. Tesfatsion, "A Guide for Newcomers to Agent-Based Modeling in the Social Sciences," in *Handbook of Computational Economics*, Vol. 2 edn, L. Tesfatsion and K. Judd, eds., Elsevier, 2006, pp. 1647-1659.

[47] H. S. Nwana, "Software Agents: An Overview," *Knowledge Engineering Review*, vol. 11, no. 3, pp. 1–40, 1996.

[48] T. O'Connor, "Emergent properties," *American Philosophical Quarterly*, vol. 31, pp. 91 – 104, 1994.

[49] J. Fromm, *The emergence of complexity*. Kassel University Press, 2004.

[50] Y. Shoham, and K. L. -Brown, *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, 2008.

[51] S. J. Russell, and P. Norvig, *Artificial Intelligence: A Modern Approach* (2nd ed.). Upper Saddle River, NJ: Prentice Hall, 2003.

[52] J. Leezer, Y. Zhang, "Simulating human intuitive decisions by Q-learning," in *Proceedings of the ACM symposium on Applied Computing*, 2009, pp. 2077-2081.

[53] M. Zhang, Z. Tan, J. Zhao, and L. Li, "A Bayesian Learning Model in the Agent-based Bilateral Negotiation between the Coal Producers and Electric Power Generators," in *Proceedings of the International Symposium on Intelligent Information Technology Application Workshops*, 2008, pp. 859-862.

[54] A. H. W. Chun, and R. Y. M. Wong, "Optimizing Agent-Based Negotiations with Branch-and-Bound," *Lecture Notes in Computer Science*, Vol. 2252, pp. 235-243, 2001.

[55] I. Watanabe, N. Yamaguchi, T. Shiina, I. Kurihara, "Agent-based simulation model of electricity market with stochastic unit commitment, in *Proceedings of the International Conference on Probabilistic Methods Applied to Power Systems*, 2004, pp. 403-408.

[56] S. -L. Huang, and F. –R. Lin, "Using temporal-difference learning for multi-agent bargaining," *Electronic Commerce Research and Applications*, vol. 7, no. 4, pp. 432-442, 2009.

[57] V. Könönen, "Gradient descent for symmetric and asymmetric multiagent reinforcement learning," *Web Intelligence and Agent Systems*, vol. 3, no. 1, pp. 17-30, 2005.

[58] C. Baydar, "Agent-based modeling and simulation of store performance for personalized pricing," in *Proceedings of the Simulation Conference*, vol. 2, 2003, pp. 1759 – 1764.

[59] G. Gigerenzer, and R. Selten, *Bounded Rationality*. Cambridge: The MIT Press, 2002.

[60] H. Juillé, and J. B. Pollack, "Coevolutionary Learning: A Case Study," in *Proceedings of the Fifteenth International Conference on Machine Learning*, 1998, pp. 251 – 259.

[61] T. Bäck, *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms*. Oxford University Press, 1996.

[62] H. A. Simon, "A behavioral model of rational choice," *The Quarterly Journal of Economics*, vol. 69, no. 1, pp. 99–118, 1955.

[63] H. E. Nusse, and J. A. Yorke, "Basins of Attraction," *Science*, vol. 271, no. 5254, pp. 1376 – 1380, 1996.

[64] S. G. Ficici, "Solution concepts in coevolutionary algorithms," Ph.D. dissertation, Brandeis Univ., Department of Computer Science, Waltham, MA, 2004.

[65] The Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel 2005. *Scandinavian Journal of Economics*, vol. 108, no. 2, pp. 183 - 184, 2006. Available: http://ssrn.com/abstract=924487

[66] H. Y. Quek, C. H. Woo, K. C. Tan, and A. Tay, "Evolving nash-optimal poker strategies using evolutionary computation," *Frontiers of Computer Science in China*, vol. 3, no. 1, pp. 73-91, 2009.

[67] H. Y. Quek, K. C. Tan, C. K. Goh, and H. A. Abbass, "Evolution and Incremental Learning in the Iterated Prisoner's Dilemma," *IEEE Transactions on Evolutionary Computation*, vol. 13, no. 2, pp. 303-320, 2009.

[68] H. Y. Quek, K. C. Tan, and H. A. Abbass, "Evolutionary Game Theoretic approach for modeling Civil Violence," *IEEE Transactions on Evolutionary Computation*, accepted 2008, and forthcoming.

[69] H. Y. Quek, K. C. Tan, and A. Tay, "Public Goods Provision: An Evolutionary Game Theoretic Study under Asymmetric Information," *IEEE Transactions on Computational Intelligence and AI in Games*, accepted 2009, and forthcoming.

[70] A. Engelbrecht, *Computational Intelligence, An Introduction* Second Edition. Wiley & Sons, 2007.

[71] B. Kosko, "Fuzzy Logic," *Scientific American*, vol. 269, no. 1, pp. 76–81, 1993.

[72] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 1999.

[73] J. Kennedy, R. C. Eberhart, and Y. Shi, *Swarm Intelligence*. Morgan Kaufmann Publishers: San Francisco, 2001.

[74] L. N. de Castro, and J. I. Timmis, *Artificial Immune Systems: A New Computational Intelligence Approach*. Springer-Verlag, London, 2002.

[75] K. A. De Jong, "An analysis of the behaviour of a class genetic adaptive systems," Ph.D. dissertation, University of Michigan, 1975.

[76] D. E. Goldberg, "Sizing populations for serial and parallel genetic algorithms," in *Proceedings of the Third International Conference on Genetic Algorithms*, pp. 70-79, 1989.

[77] J. H. Holland, *Adaptation in Natural Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. MIT press, 1992.

[78] I. Rechenberg, *Evolutionsstrategie*. Frommann-Holzboog, 1994.

[79] L. J. Fogel, A. J. Owens, M. J. Walsh, *Artificial Intelligence through Simulated Evolution*. John Wiley, 1966.

[80] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, MIT Press, 1992.

[81] B. L. Miller, and D. E. Goldberg, "Genetic Algorithms, Selection Schemes, and the varying Effects of Noise," *Evolutionary Computation*, Vol. 4, no. 2, pp. 113-131, 1996.

[82] J. E. Baker, "Reducing Bias and Inefficiency in the Selection Algorithm," in *Proceedings of the Second International Conference on Genetic Algorithms and their Application*, Erlbaum, 1987, pp. 14-21.

[83] T. D. Gwiazda, *Genetic Algorithms Reference Vol.1 Crossover for single-objective numerical optimization problems*. Tomasz Gwiazda, Lomianki, 2006.

[84] M. C. Sinclair, "Evolutionary Algorithms for Optical Network Design: A Genetic-algorithm/heuristic hybrid approach," Ph.D. dissertation, University of Essex, 2001.

[85] T. Biick, and H. P. Schwefel, "An Overview of Evolutionary Algorithm for Parameter Optimization," *Evolutionary Computation*, vol. 1, no. 1, pp.1-24, 1993.

[86] H. Y. Quek, and K. C. Tan, "A Discrete Particle Swarm Optimization Approach for the global Airline Crew Scheduling Problem," *in Proceedings of the International Conference on Soft Computing and Intelligent Systems and International Symposium on Advanced Intelligent Systems*, Nagoya University, Nagoya, Japan, September 17-21, 2008, accepted.

[87] K. Deb, and T. Goel, "Multi-objective evolutionary algorithms for engineering shape design," I. Sarker, M. Mohammadian, and X. Yao (Eds.), *Evolutionary Optimization*. Kluwer Academic Publishers, USA, 2001.

[88] F. Alkemade, H. L. Poutré, and H. M. Amman, "Robust Evolutionary algorithm design for socio-economic simulation," *Computational Economics*, vol. 28, pp. 355-370, 2006.

[89] R. Axelrod, *The Complexity of Cooperation*. NJ: Princeton University Press, 1997.

[90] M. A. Potter, and K. A. D. Jong, "Cooperative coevolution: An architecture for evolving coadapted subcomponents," *Evolutionary Computation*, vol. 8, no. 1, pp. 1–29, 2000.

[91] R. P. Wiegand, "An analysis of Cooperative Coevolution Algorithm," Ph.D. dissertation, George Mason University, Fairfax, VA. 2004.

[92] R. Dreżewski, and L. Siwik, "Agent-Based Co-Operative Co-Evolutionary Algorithm for Multi-Objective Optimization," *Lecture Notes in Computer Science*, Vol. 5097, pp. 388-397, 2008.

[93] D. Fudenberg, and D. K. Levine, *The Theory of Learning in Games. Cambridge*, MA: MIT Press, 1998.

[94] C. Clemens, and T. Riechmann, "Dynamic voluntary contribution to a public good: Learning to be a free rider," Universität Hannover, Germany, Tech. Rep. Nr. 240, 2001.

[95] "Using artificial neural networks to model opponents in Texas Hold'em," Machine Learning in Games Undergraduate Research Course, CMPUT 499, Computer Poker Research Group, University of Alberta, November 1999.

[96] K. B. Korb, A. E. Nicholson and N. Jitnah, "Bayesian Poker," in *Proceedings of the Uncertainty in Artificial Intelligence (UAI)*, 1999, pp. 343-350.

[97] D. Billings, "Computer Poker," M.Sc. research essay, Computer Poker Research Group, University of Alberta, Alberta, Canada, 1995.

[98] Cosmic Log, MSNBC, "Human Beat Poker Bot … Barely," July 25, 2007,

http://cosmiclog.msnbc.msn.com/archive/2007/07/25/289607.aspx.

[99] C. S. Ong, H. Y. Quek, K. C. Tan, and A. Tay, "Discovering Chinese Chess Strategies through Coevolutionary Approaches," in *Proceedings of the IEEE Symposium Series on Computational Intelligence, Computational Intelligence and Games*, Honolulu, Hawaii, USA, April 1-5, 2007, pp. 360-367.

[100] D. Billings, A. Davidson, J Schaeffer and D. Szafron, "The Challenge of Poker," *Artificial Intelligence Journal*, vol. 134, no.1-2, pp 201-240, 2002.

[101] D. Billings, D. Papp, J. Schaeffer and D. Szafron, "Opponent Modeling in Poker," in *Proceedings of the fifteen AAAI conference*, 1998, pp. 493-499.

[102] D. Billings, D. Papp, J. Schaeffer and D. Szafron, "Poker as a Testbed for Machine Intelligence Research," in *Advances in Artificial Intelligence*, R. Mercer and E. Neufeld, Eds. Springer Verlag, 1998, pp. 1-15.

[103] D. Papp, "Dealing with Imperfect Information in Poker," M.Sc. thesis, Department of Computing Science, University of Alberta, Alberta, Canada, 1998.

[104] D. Billings, L. Pena, J. Schaeffer and D. Szafron, "Using Probabilistic Knowledge and Simulation to Play Poker," in *Proceedings of the sixteenth AAAI conference*, 1999, pp. 697-703.

[105] A. Davidson, D. Billings, J. Schaeffer and D. Szafron, "Improved Opponent Modeling in Poker," in *Proceedings of the 2000 International Conference on Artificial Intelligence (ICAI'2000)*, 2000, pp. 1467-1473.

[106] Express News, University of Alberta, "U of A researchers win computer poker title," August 9, 2006,

http://www.expressnews.ualberta.ca/article.cfm?id=7789.

[107] D. Billings, N. Burch, A. Davidson, R. Holte, J. Schaeffer, T. Scauenberg and D. Szafron, "Approximating game-theoretic optimal strategies for full-scale Poker," in *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI-03)*, 2003, pp. 661-668.

[108] L. Barone and L. While, "An adaptive learning model for simplified Poker using evolutionary algorithms," in *Proceedings of the Congress on Evolutionary Computation (GECCO-1999)*, 1999, vol. 1, p. 153-160.

[109] L. Barone and L. While, "Adaptive Learning for Poker," in *Proceedings of the Genetic and Evolutionary Computation Conference 2000 (GECCO-2000)*, 2000, pp. 566-573.

[110] F. Oliehoek, N. Vlassis and E. d. Jong, "Coevolutionary Nash in Poker Games," in *Proceedings of the seventeenth Belgian-Dutch Conference on Artificial Intelligence (BNAIC-2005)*, 2005, pp. 188–193.

[111] The Early Show, CBS News, "Poker's Popularity Surging," December 24, 2004,
http://www.cbsnews.com/stories/2004/12/25/earlyshow/living/main663053.shtml.

[112] H. W. Kuhn, "Extensive games and the problem of information," in *Contributions to the Theory of Games, Annals of Mathematics Studies*, vol. 2, no. 28, pp. 193-216, 1953.

[113] M. D. Smucker, E. A. Stanley, and D. Ashlock, "Analyzing social network structures in the iterated prisoner's dilemma with choice and refusal," Department of Computer Sciences, University of Wisconsin-Madison, Tech. Rep. CS-TR-94-1259, 15 December 1994.

[114] R. Axelrod, "The evolution of strategies in the iterated prisoner's dilemma," in Lawrence Davis (ed.) *Genetic Algorithm and Simulated Annealing*, London: Pitman, 1987, pp. 32-41.

[115] M. Nowak, and K. Sigmund, "A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game," *Nature*, vol. 364, pp. 56-58, 1993.

[116] J. Golbeck, "Evolving strategies for the prisoner's dilemma," *Advances in Intelligent Systems, Fuzzy Systems, and Evolutionary Computation*, pp. 299-306, February 2002.

[117] P. Darwen, and X. Yao, "On evolving robust strategies for the iterated prisoner's dilemma," in *Progress in Evolutionary Computation*, ser. Lecture Notes in Artificial Intelligence, 1995, vol. 956, pp. 276-292.

[118] S. O. Kimbrough, D. J. Wu, and F. Zhong, "Computers play the beer round: can artificial agents manage supply chains?*" Decision Support Systems*, vol. 33, issue 3, pp. 323 - 333, July 2002.

[119] S. Thrun, "Learning to play the round of chess," *Machine Learning*, vol. 40, issue 3, pp. 243-263, September 2000.

[120] P. Hingston, and G. Kendall, "Learning versus evolution in iterated prisoner's dilemma," in *Proceedings of the Congress on Evolutionary Computation 2004 (CEC'04)*, Portland, Oregon, vol. 1, June 2004, pp. 364-372.

[121] G. Eoyang, *Genetic Algorithm as Decision Support Tool*. Chaos Limited, 1996.

[122] R. S. Michalski, "Learnable evolution model: evolutionary processes guided by machine learning*," Machine Learning*, vol. 38, issue 1-2, pp. 9-40, February 2000.

[123] E. Lamma, F. Riguzzi, and L. M. Pereira, "Belief revision via Lamarckian evolution*," New Generation Computing*, vol. 21, no.3, pp. 247-275, May 2003.

[124] C. Fonlupt, P. Preux, and D. Robilliard, "Preventing premature convergence via cooperating genetic algorithms," *in Proceedings of Mandel'97*, Brno, Czech Republic, June 1997.

[125] C. G. Carrier, "Unifying learning with evolution through Baldwinian evolution and Lamarckism: a case study," in *Proceedings of the Symposium on Computational Intelligence and Learning (CoIL-2000)*, June 2000. pp. 36-41.

[126] E. K. Burke, P. Cowling, and P. De Causmaecker, "A memetic approach to the nurse rostering problem," *Applied Intelligence*, vol. 15, no. 3, pp. 199-214, 2001.

[127] P. M. Franca, A. Mendes, and P. Moscato, "A memetic algorithm for the total tardiness single machine scheduling problem," *European Journal Of Operational Research*, vol. 132, no. 1, pp. 224-242, 2001.

[128] H. Ishibuchi, T. Yoshida, and T. Murata, "Balance between genetic search and local search in memetic algorithms for multiobjective permutation flowshop scheduling," *IEEE Transactions on Evolutionary Computation*, vol. 7 no. 2, pp. 204-223, 2003.

[129] Y. S. Ong, and A. J. Keane, "Meta-Lamarckian learning in memetic algorithms," *IEEE Transactions on Evolutionary Computation*, vol. 8, no.2, pp. 99-110, 2004.

[130] P. Bodo, "In-class simulations of the iterated prisoner's dilemma round," *Journal of Economic Education,* vol. 33, no. 3, pp. 207-216, 2002.

[131] D. B. Neill, "Optimality under noise: higher memory strategies for the alternating prisoner's dilemma," *Journal of Theoretical Biology,* vol. 211, pp. 159–180, 2001.

[132] B. Beaufils, P. Mathieu, and J. P. Delahaye, "Complete classes of strategies for the classical iterated prisoner's dilemma," in *Proceedings of Evolutionary Programming VII,* LNCS 1447, Springer-Verlag, 1998, pp. 33–41.

[133] D. B. Fogel, "Evolving behaviors in the iterated prisoner's dilemma," *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 77-97, 1993.

[134] T. Gosling, N. Jin, and E. Tsang, "Population based incremental learning versus genetic algorithms: iterated prisoners dilemma," Department of Computer Science, University of Essex, England, Tech. Rep. CSM-401, March 2004.

[135] S. Y. Chong, and X. Yao, "The Impact of Noise on Iterated Prisoner's Dilemma with Multiple Levels of Cooperation," in *Proceedings of the Congress on Evolutionary Computation 2004 (CEC'04)*, Portland, Oregon, vol. 1, June 2004, pp. 348-355.

[136] D. Hales, "Change Your Tags Fast! - A Necessary Condition for Cooperation?," in *Proceedings of the Joint Workshop on Multi-Agent and Multi-Agent-Based Simulation*, July 2004, pp. 89-98.

[137] D. B. Fogel, "On the relationship between the duration of an encounter and the evolution of cooperation in the iterated prisoner's dilemma," *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 3, pp. 349-363, 1996.

[138] D. B. Fogel, G. B. Fogel, and P. C. Andrews, "On the instability of evolutionary stable states," *Biosystems*, vol. 44, pp. 135-152, 1997.

[139] D. B. Fogel, and G. B. Fogel, "Evolutionary stable strategies are not always stable under evolutionary dynamics," in *Evolutionary Programming IV,* J. McDonnell, R. Reynolds, and D. B. Fogel, Eds. Cambridge, MA: MIT Press, 1995, pp. 565-577.

[140] N. Meuleau, and C. Lattaud, "The artificial evolution of cooperation," in *Lecture Notes in Computer Science,* J.-M. Alliot, E. Lutton, E. Ronald, M. Schoenauer, and D.Snyers, Eds. New York: Springer-Verlag, 1996, vol. 1063, in *Proceedings of the European Conference of Artificial Evolution*, pp. 159-180.

[141] B. Skyrms, "Chaos and the explanatory significance of equilibrium: Strange attractors in evolutionary game dynamics," in *Proceedings of the Biennial Meeting of Philosophical Science Association,* vol. 2, 1992, pp. 374-394.

[142] B. Hosp, "The genetic algorithm and the prisoner's dilemma," in *Consortium for Computing Sciences in Colleges (CCSC): Southeastern Conference*, January 2004, pp.135-146.

[143] H. Ishibuchi, and N. Namikawa, "Evolution of Iterated Prisoner's Dilemma Game Strategies in Structures Demes Under Random Pairing in Game Playing," *IEEE Transactions on Evolutionary Computation,* vol. 9, no. 6, pp. 552-561, 2005.

[144] P. Darwen, and X. Yao, "Does extra genetic diversity maintain escalation in a co-evolutionary arms race," *International Journal of Knowledge Based Intelligent Engineering Systems*, vol. 4, no. 3, pp. 191-200, July 2000.

[145] P. J. Darwen, and X. Yao, "Why More Choices Cause Less Cooperation in Iterated Prisoner's Dilemma," in *Proceedings of the 2001 Congress on Evolutionary Computation*, pp. 987-994, IEEE Press, Piscataway, NJ, USA, May 2001.

[146] P. Darwen, and X. Yao, "Speciation as automatic categorical modularization," *IEEE Transactions on Evolutionary Computation,* vol. 1, no. 2, pp. 101-108, 1997.

[147] B. Salles, "Constructing progressive learning routes through qualitative simulation models in ecology," in *Proceedings of the International Workshop on Qualitative Reasoning, QR'01*, May 2001, pp. 82-89.

[148] L. J. Lin, "Self-improving reactive agents based on reinforcement learning, planning and teaching," *Machine Learning*, vol. 8, issue 3-4, pp. 293-321, May 1992.

[149] D. Kraines, and V. Kraines, "Pavlov and the prisoner's dilemma," *Theory and Decision*, vol. 26, pp. 47-79, 1989.

[150] M. A. Maloof, and R. S. Michalski, "Incremental learning with partial instance memory," in *Proceedings of the 13th International Symposium on Foundations of Intelligent Systems,* 2002, pp. 16-27.

[151] P. Moscato, "On evolution, search, optimization, genetic algorithms and martial arts: Towards memetic algorithms," California Institute of Technology, Pasadena, California, USA, Tech. Rep. Caltech Concurrent Computation Program, Report. 826, 1989.

[152] R. Hightower, S. Forrest, and A. Perelson, "The Baldwin effect in the immune system: learning by somatic hypermutation," in Belew, R.K. and Mitchell, M. (eds) *Adaptive Individuals in Evolving Populations: Models and Algorithms*, Chapter 11, Addison-Wesley, 1996.

[153] J. M. Baldwin, "A new factor in evolution," *American Naturalist*, vol. 30, pp. 441-451,536-553, 1896.

[154] T. Deacon, *The Symbolic Species: the Coevolution of language and human brain*. London: Penguin, 1997.

[155] T. Sasaki, and M. Tokoro, "Adaptation towards changing environments: Why Darwinian in nature?" in Husbands, P., & Harvey, I. (Eds.), *Proceedings of the Fourth European Conference on Artificial Life (ECAL'97), Brighton, UK, 28–31 July*, 1997, pp. 145–153 Cambridge, MA. MIT Press / Bradford Books, Cambridge, MA.

[156] R. Suzuki, and T. Arita, "Interactions between Learning and Evolution: Outstanding Strategy generated by the Baldwin Effect," *Biosystems*, vol. 77, issue 1-3, pp. 57-71, 2004.

[157] K. H. Liang, X. Yao, and C. Newton, "Lamarckian evolution in global optimization," in *Proceedings of the IEEE 26th Annual Conference of Industrial Electronics 2000, (IECON 2000)*, vol. 4, October 2000, pp. 2975-2980.

[158] D. McLellan, *The Thought of Karl Marx: An Introduction*. New York: Harper & Row, 1971.

[159] H. Situngkir, "On massive conflict: Macro-micro link," *Journal of Social Complexity*, vol. 1, no.4, pp. 1-12, 2004.

[160] H. Grossman, "Kleptocracy and revolutions," *Oxford Economic Papers*, vol. 51, no.2, pp. 267-283, 1999.

[161] T. R. Gurr, *Why Men Rebel*. New Jersey: Princeton University Press, 1970.

[162] T. R. Gurr, *People Versus States: Minorities at Risk in the New Century*. Washington, DC: United States Institute of Peace Press, 2000.

[163] P. Collier, and A. Hoeffler, "On the economic causes of civil war," *Oxford Economic Papers*, vol. 50, no.4, pp. 563-573, 1998.

[164] P. M. Regan, and D. A. Norton, "Greed, Grievance, and Mobilization: The Onset of Protest, Rebellion, and Civil War," *Journal of Conflict Resolution*, vol. 49, no. 3, pp. 319-336, 2005.

[165] M. W. Doyle, and N. Sambanis, "International peacebuilding: A theoretical and quantitative analysis," *American Political Science Review*, vol. 94, no.4, pp. 779-801, 2000.

[166] R. Licklider, "The consequences of negotiated settlements in civil wars, 1945-1993," *American Political Science Review*, vol. 89, no.3, pp. 681-690, 1995.

[167] N. Sambanis, "Do ethnic and non-ethnic civil wars have the same causes? A theoretical and empirical inquiry (Part 1)," *Journal of Conflict Resolution*, vol. 45, no.3, pp. 259-82, 2001.

[168] T. Schelhorn, D. O'Sullivan, M. Haklay, and M. Thurstain-Goodwin, STREETS: An Agent-Based Pedestrian Model. London, U.K.: Univ. College London, Center Adv. Spatial Anal., Working Paper no. 9, 1999.

[169] S. Parikh, and C. Cameron, "Riot games: A theory of riots and mass political violence," presented at *7th Wallis Institute Conference on Political Economy*, University of Rochester, New York, October, 2000.

[170] R. B. Myerson, *Game Theory: Analysis of Conflict*. Harvard University Press, 1997.

[171] J. Ginkel, and A. Smith, "So you say you want a revolution: A game theoretic explanation of revolution in repressive regimes," *Journal of Conflict Resolution*, vol. 43, no.3, pp. 291-316, 1999.

[172] T. R. Gulden, "Spatial and temporal patterns in civil violence Guatemala 1977-1986," The Brookings Institution, Washington, DC, Center on Social and Economic Dynamics, Working Paper no.26, 2002.

[173] J. M. Epstein, J. D. Steinbruner, and M. T. Parker, "Modeling civil violence: An agent based computational approach," The Brookings Institution, Washington, DC, Center on Social and Economic Dynamics, Working Paper no.20, 2001.

[174] W. Jager, R. Popping, and H. v. d. Sande, "Clustering and fighting in two-party crowds: Simulating the approach-avoidance conflict," *Journal of Artificial Societies and Social Simulations*, vol. 4, no.3, 2001.

[175] A. Srbljinovic, D. Penzar, P. Rodik, and K. Kardov, "An agent based model of ethnic mobilization," *Journal of Artificial Societies and Social Simulations*, vol. 6, no.1, pp. 1-14, 2003.

[176] S. Y. Yiu, A. Gill, and P. Shi, "Investigating strategies for managing, civil violence using the MANA agent based distillation," in *Proc. Land Warfare Conf.*, Brisbane, Australia, 2002, pp. 475 – 484.

[177] R. Axelrod, "Agent-based modeling as a bridge between disciplines," in *Handbook of Computational Economics, vol. 2: Agent-Based Computational Economics*, L. Tesfatsion and K. Judd (eds.), New York: North-Holland, 2006, pp. 1565-1584.

[178] D. Cliff, and G. F. Miller, "Protean behavior in dynamic games: Arguments for the co-evolution of pursuit-evasion tactics," in *From Animals to Animats III: Proceedings of the Third International Conference on Simulation of Adaptive Behavior*, Cliff, D., Husbands, P., Meyer, J. A. and Wilson, W. S., Eds. Cambridge, USA: Bradford Books, 1994, pp. 411-420.

[179] C. W. Reynolds, "Competition, Co-evolution and the Game of Tag," in *Proceedings of the Fourth International Workshop on Artificial Life*, Brooks, R. and Maes, P., Eds. Cambridge, USA: MIT Press, 1994, pp. 59-69.

[180] R. L. Goldstone, and M. A. Janssen, "Computational models of collective behavior," *Trends in Cognitive Sciences*, vol. 9, no. 9, pp. 424-430, 2005.

[181] T. Schelling, *Micromotives and Macrobehavior*, pp. 137-57, Norton, New York, 1978.

[182] H. Situngkir, and Y. Surya, "Agent-based Model Construction in Financial Economic System," *Complexity Digest*, vol. 13, no.2, pp. 1-10, 2004.

[183] L. Parrott and R. Kok, "Use of an object-based model to represent complex features of ecosystems," in Proc. Presented *3rd Int. Conf. Complex Syst. (ICCS)*, Nashua, NH: New Hampshire, 2000, pp. 169-179.

[184] G. P. Richardson, *Feedback Thought in Social Science and Systems Theory*. Philadelphia, PA: University of Pennsylvania Press, 1991.

[185] L. L. Patrick, "Complexity: The science of change," in *Context: J. Hope Sustainability Change*, vol. 40, Washington, D.C.: Context Inst., 1995, pp. 51–53.

[186] J. Weibull, *Evolutionary game theory*. Cambridge, MA: MIT Press, 1992.

[187] C. M. Frayn, A. N. Pryke, and S. Y. Chong, "Exploring the effect of proximity and kinship on mutual cooperation in the iterated prisoner's dilemma," in *Proceedings of the Ninth Conference on Parallel Problem Solving from Nature*, 2006, pp. 701-710.

[188] I. Nishizaki, M. Sakawa, and H. Katagiri, "Influence of environmental changes on cooperative behavior in the Prisoner's Dilemma game on an artificial social model," *Applied Artificial Intelligence*, vol. 18, no.7, pp. 651-671, 2004.

[189] R. K. Sawyer, "Artificial societies: Multiagent systems and micro-macro link in sociological theory," *Sociological Methods & Research*, vol. 31, no.3, pp. 325-363, 2003.

[190] M. Berdal, and D. M. Malone, (eds.), *Greed and Grievance: Economic Agendas in Civil Wars*, Lynne Rienner Publishers, 2000.

[191] P. Collier, and A. Hoeffler, "Greed and grievance in civil war," *Oxford Economic Papers*, vol. 56, no.4, pp. 563-595, 2004.

[192] J. Hugues, and J. B. Pollack, "Coevolving the 'Ideal' trainer: Application to the discovery of cellular automata rules," in *Proceedings of the Third Annual Genetic Programming Conference*, 1998, pp. 519-528.

[193] D. Cliff, and G. F. Miller, "Tracking the Red Queen: Measurements of adaptive progress in co-evolutionary simulations," in *Proceedings of the Third European Conference on Artificial Life*, 1995, pp. 200-218.

[194] J. Hofbauer, and K. Sigmund, *Evolutionary games and population dynamics*. Cambridge University Press, 1998.

[195] R. P. Wiegand, W. C. Liles, and K. A. De Jong, "Analyzing Cooperative Coevolution with Evolutionary Game Theory," in *Proceedings of the 2002 Congress on Evolutionary Computation*, 2002, pp. 1600-1605.

[196] R. P. Wiegand, W. C. Liles, and K. A. De Jong, "Modeling variation in cooperative coevolution using evolutionary game theory," in *Found. Genetic Algorithms VII*, R. Poli, J. Rowe, and K. D. Jong, (eds.), 2002, pp. 231–248, Morgan Kaufmann.

[197] M. D. Schmidt, and H. Lipson, "Coevolution of Fitness Predictors," *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 6, pp. 736 – 749, 2008.

[198] S. Y. Chong, P. Tiño, and X. Yao, "Measuring Generalization Performance in Coevolutionary Learning," *IEEE Transactions on Evolutionary Computation*, vol. 12, no.4, pp. 479 – 505, 2008.

[199] K. C. Tan, T. H. Lee, and E. F. Khor, "Evolutionary algorithm with dynamic population size and local exploration for multiobjective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 5, no.6, pp. 565-588, 2001.

[200] T. Kuran, "Sparks and prairie fires. A theory of unanticipated political revolution," *Public Choice*, vol. 61, no. 1, pp. 41–74, 1989.

[201] T. T. Mao, "A single spark can start a prairie fire," in *Selected Military Writings of Mao Tse-Tung*, Peking: Foreign Languages Press, 1930/1972, pp. 65-76.

[202] J. D. Fearon, "Why do some civil wars last so much longer than others?" *Journal of Peace Research*, vol. 41, no.3, pp. 275–301, 2004.

[203] L. B. Gustave, *The Crowd: A Study of the Popular Mind*. Fraser Publishing Company, 1982.

[204] S. Freud, *Group Psychology and the Analysis of the Ego*. W. W. Norton and Company, 1921/1975.

[205] N. Sambanis, "Partition as a solution to ethnic war: An empirical critique of the theoretical literature," *World Politics*, vol. 52, no.4, pp. 437-483, 2000.

[206] R. Ames, *Sun Tzu: The Art of Warfare*. New York: Ballantine Books, 1993.

[207] M. Lloyd, *The Art of Military Deception*. Pen and Sword Books, 1997.

[208] S. Gerwehr, and R. W. Glenn, *Unweaving the Web: Deception and Adaptation in Future Urban Operations*. Rand Books and Publications, 2003.

[209] I. Faurby, and M. L. Magnusson, "The battle(s) of Grozny," *Baltic Defence Review*, vol. 2, no.7, pp. 75-87, 1999.

[210] S. N. Kalyvas, "The logic of violence in civil war," presented at the Laboratory in Comparative Ethnic Processes, Duke University, USA, 2000.

[211] F. Caselli and J. Coleman, "On the theory of ethnic conflict," National Bureau of Economic Research, Cambridge, MA, Working Paper 12125, 2006.

[212] L. M. Woolf, and M. R. Hulsizer, "Intra- and inter-religious hate and violence: A psychosocial model," *Journal of Hate Studies*, vol. 2, no.1, pp. 5-26, 2003.

[213] J. M. Epstein, and R. L. Axtell, *Growing Artificial Societies: Social Science from the Bottom Up*. The MIT Press, 1996.

[214] G. Hardin, "Tragedy of the Commons," *Science*, vol. 162, pp. 1243-1248. 1968.

[215] H. Goren, R. Kurzban, and A. Rapoport, "Social loafing vs. social enhancement: Public goods provisioning in real-time with irrevocable commitments," *Organizational Behavior and Human Decision Processes*, vol. 90, pp. 277 – 290, 2003.

[216] S. Bikhchandani, "Ex post implementation in environments with private goods," *Theoretical Economics*, vol. 1, no. 3, pp. 369-393, 2006.

[217] E. Ostrom, *Governing the commons: The evolution of institutions for collective action*. Cambridge: Cambridge University Press, 1990.

[218] C. M. Tiebout, "A Pure Theory of Local Expenditures," *Journal of Political Economy*, vol. 64, no. 5, pg. 416-424, 1956.

[219] R. Hardin, *Collective Action*. Baltimore, Md.: John Hopkins University Press, 1982.

[220] P. J. Klenow, and A. Rodriguez-Clare, "Externalities and Growth," in: Aghion, P. and Durlauf, S. Eds., *Handbook of Economic Growth*, *ed. 1*, vol. 1, chapter 11, Elsevier, 2005, pg. 817-861.

[221] M. Olson, *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge: Harvard University Press, 1971.

[222] M. I. Lichbach, *The Cooperator's Dilemma*. Ann Arbor: The University of Michigan Press, 1996.

[223] C. Landesman, "The Voluntary Provision of Public Goods," Ph.D. dissertation, Princeton University, NJ, USA, 1995.

[224] S. Calabrese, D. Epple, T. Romer, and H. Sieg, "Local public good provision: Voting, peer effects, and mobility," *Journal of Public Economics*, vol. 90(6-7), no. 2, pp. 959-981, 2006.

[225] I. Kaul, P. Conceic¸a˜o, K. L. Goulven, and R. U. Mendoza, "How to improve the provision of global public goods," *Providing Global Public Goods*, Oxford Scholarship Online, 2003, pp. 21-59.

[226] R. O. Zerbe Jr., and H. McCurdy, "The End of Market Failure," *Regulation*, vol. 23, no. 2, pp. 10-14, 2000.

[227] B. P. Brownstein, "Pareto Optimality, External Benefits and Public Goods: A Subjectivist Approach," *The Journal of Libertarian Studies*, vol. 4, no. 1, 1980.

[228] J. Akin, P. Hutchinson, and K. Strumpf, "Decentralisation and government provision of public goods: The public health sector in Uganda," *Journal of Development Studies*, vol. 41, no. 8, pp. 1417-1443, 2005.

[229] A. Tabarrok, "The private provision of public goods via dominant assurance contracts," *Public Choice*, vol. 96, no. 3-4, pp. 345-362, 1998.

[230] M. Feldman, and J. Chuang, "Overcoming free-riding behavior in peer-to-peer systems," *ACM Sigecom Exchanges*, vol. 5, no. 4, pp. 41-50, 2005.

[231] D. W. Boyd, "Vertical restraints and the retail free riding problem: An Austrian perspective," *Review of Austrian Economics*, vol. 9-1, no. 6, pp. 119-34, 1996.

[232] J. Venugopal, "Drug imports: the free-rider paradox," *Express Pharma Pulse*, vol. 11, no. 9, pp. 8, 2005.

[233] A. Sen, *Collective Choice and Social Welfare*. San Francisco, Holden-Day, 1970.

[234] O. Kim, and M. Walker, "The free rider problem: Experimental evidence," *Public Choice*, vol. 43, no. 1, pp. 3-24, 1984.

[235] H. Goren, A. Rapoport, and R. Kurzban, "Revocable Commitments to Public Goods Provision under the Real-Time Protocol of Play," *The Journal of Behavioral Decision Making*, vol. 17, pp. 17-37, 2004.

[236] D. Fudenberg, and J. Tirole, *Game Theory*, MIT Press, 1991.

[237] J. Sonnemans, A. Schram, and T. Offerman, "Strategic behavior in public good games: when partners drift apart," *Economics Letters*, vol. 62, pp. 35-41, 1999.

[238] T. Decker, A. Stiehler, and M. Strobel, "A comparison of punishment rules in repeated public good games: An experimental study," *Journal of Conflict Resolution*, vol. 47, no. 6, pp. 751-772, 2003.

[239] P. M. Todd, "The causes and effects of evolutionary simulation in the behavioral sciences," in: Belew, R. and Mitchell, M. Eds., *Adaptive Individuals in Evolving Populations: Models and algorithms*, MA: Addison-Wesley, 1996, pp. 211—224.

[240] M. Spence, "Informational aspects of Market Structure: An Introduction," *The Quarterly Journal of Economics*, vol. 90, no. 4, pp. 591-597, 1976.

[241] J. E. Stiglitz, and A. M. Weiss, "Asymmetric Information in Credit Markets: Implications for Macro-Economics" *Oxford Economic Papers*, vol. 44, no. 4, pp. 694-724, 1992.

[242] P. Milgrom, and J. Roberts, "Relying on the information of interested parties," *Rand Journal of Economics*, vol. 17, no. 1, 1986.

[243] B. Hillier, *The Economics of Asymmetric Information*. London, Macmillan Press, 1997.

[244] X. Yao, and P. Darwen, "An experimental study of N-person iterated prisoner's dilemma games," *Informatica*, vol. 18, no. 4, pp. 435-450, 1994.

[245] V. L. Smith, "Experimental methods in economics," *The New Palgrave: A Dictionary of Economics vol. 2*, pp. 241-49, 1987.

[246] C. Hauert, and G. Szabó, "Prisoner's dilemma and public goods games in different geometries: compulsory versus voluntary participation," *Complexity*, vol. 8, pp. 31-38, 2003.

[247] C. Hauert, S. De Monte, J. Hofbauer, and K. Sigmund, "Replicator dynamics for optional public good games," *Journal of Theoretical Biology*, vol. 218, no. 2, pp. 187-194, 2002.

[248] G. Szabó, and G. Fáth, "Evolutionary games on graphs," *Physics Reports*, vol. 446, no. 4-6, pp. 97-216, 2007.

[249] R. Kurzban, K. McCabe, V. L. Smith, and B. J. Wilson, "Incremental commitment and reciprocity in a real time public goods game," *Personality and Social Psychology Bulletin*, vol. 27, no. 12, pp. 1662-1673, 2001.

[250] H. Brandt, C. Hauert, and K. Sigmund, "Punishment and reputation in spatial public goods games," in *Proceedings of the Royal Society of London Biology*, vol. 270, pp. 1099-1104, 2003.

[251] R. K. Wilson, and J. Sell, "'Liar, Liar ...' Cheap Talk and Reputation in Repeated Public Goods Settings," *Journal of Conflict Resolution*, vol. 41, no. 5, pp. 695-717, 1997.

[252] M. F. Shakun, "Unbounded Rationality," *Group Decision and Negotiation*, vol. 10, no. 2, pp. 97-118, 2001.

[253] J. H. Miller, and J. Andreoni, "Can evolutionary dynamics explain free riding in experiments?" *Economics Letters*, vol. 36, pp. 9-15, 1991.

[254] J. Andreoni, "Why free ride? Strategies and learning in public goods experiments," *Journal of Public Economics*, vol. 37, pp. 291-304, 1988.

[255] L. M. Marx, and S. A. Matthews, "Dynamic voluntary contribution to a public project," *Review of Public Economics*, vol. 15, pp. 111-194, 2000.

[256] R. Kurzban, and P. DeScioli, "Reciprocal cooperation in groups: Information-seeking in a public goods game," *European Journal of Social Psychology*, (in press)

[257] E. H. Hagen, and P. Hammerstein, "Game theory and human evolution: A critique of some recent interpretations of experimental games," *Theoretical Population Biology*, vol. 69, no. 3, pp. 339-348, 2006.

[258] M. M. Pillutla, and X. P. Chen, "Social norms and cooperation in social dilemmas: the effects of context and feedback," *Organizational Behavior and Human Decision Processes*, vol. 78, pp. 81–103, 1999.

[259] D. Kahneman, and A. Tversky, "Choices, Values and Frames," *American Psychologist*, vol. 39, no. 4, pp. 341-350.

[260] R. M. Isaac, and J. M. Walker, "Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism," *The Quarterly Journal of Economics*, vol. 103, no. 1, pp. 179-199, 1988.

[261] G. Gigerenzer, P. M. Todd, and The ABC Research Group, *Simple Heuristics that make us smart*. New York: Oxford University Press, 1999.

[262] P. Darwen, and X. Yao, "Co-Evolution in Iterated Prisoner's Dilemma with Intermediate Levels of Cooperation: Application to Missile Defense," *International Journal of Computational Intelligence and Applications*, vol. 2, no. 1, pp. 83-107, 2002.

[263] B. Hayes, "Unions and Strikes with Asymmetric Information," *Journal of Labor Economics*, vol. 2, no. 1, pp. 57-83, 1984.

[264] D. Aboody, and B. Lev, "Information Asymmetry, R&D, and Insider Gains," *Journal of Finance*, vol. 55, no. 6, pp. 2747–2766, 2000.

[265] Á. Kun, G. Boza, and I. Scheuring, "Asynchronous snowdrift game with synergistic effect as a model of cooperation," *Behavioural Ecology*, vol. 17, pp. 633-641, 2006.

[266] R. Boyd, and P. J. Richerson, *Culture and the evolutionary process*. Chicago IL: University of Chicago Press, 1985.

[267] M. A. Janssen, and R. L. Goldstone, "Dynamic-persistence of cooperation in public good games when group size is dynamic," *Journal of Theoretical Biology*, vol. 234, pp. 134-142, 2006.

[268] H. Y. Quek, and A. Tay, "A evolutionary, game theoretic approach to the modeling, simulation and analysis of public goods provisioning under asymmetric information," in *Proceedings of the 2007 Congress on Evolutionary Computation*, pp. 4735-4742, Singapore, Sep 2007.

[269] T. Killingback, J. Bieri, and T. Flatt, "Evolution in group-structured populations can resolve the tragedy of the commons," in *Proceedings of the Royal Society of London B*, vol. 273, pp. 1477–1481, 2006.

[270] D. P. Myatt, H. S. Shin, and C. Wallace, "The Assessment: Games and Coordination," *Oxford Review of Economic Policy*, vol. 18, no. 4, pp. 397-417, 2002.

[271] R. Kurzban, and D. Houser, "Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations," in *Proceedings of the National Academy of Sciences*, vol. 102, no. 5, pp. 1803-1807, 2005.

[272] C. Baray, "Effects of population size upon emergent group behavior," *Complexity International*, vol. 6, 1999.

[273] M. Haag, and R. Lagunoff, "On the size and structure of group cooperation," *Journal of Economic Theory*, vol. 135, no. 1, pp. 68-89, 2007.

[274] M. A. Nowak, and K. Sigmund, "Evolution of indirect reciprocity by image scoring," *Nature*, vol. 393, no. 6685, pp. 573-557, 1998.

[275] M. A. Nowak, and K. Sigmund, "Evolution of indirect reciprocity," *Nature*, vol. 437, no. 7063, pp. 1291-1298, 2005.

[276] P. Barclay, "Trustworthiness and competitive altruism can also solve the "tragedy of the commons," *Evolution and Human Behavior*, vol. 25, no. 4, pp. 209-220, 2004.

[277] R. Mashima, and N. Takahashi, "The emergence of indirect reciprocity: Theoretical and empirical approaches toward indirect reciprocity," presented at the 7th 21st century Cultural and Ecological Foundations of the Mind International Workshop, Hokkaido University, Japan, 2005.

[278] S. Suzuki, and E. Akiyama, "Evolution of indirect reciprocity in groups of various sizes and comparison with direct reciprocity," *Journal of Theoretical Biology*, vol. 245, pp. 539–552, 2007.

[279] S. Abele, and G. Stasser, "Continuous versus Step-Level Public Good Games," Erasmus University Rotterdam, Netherlands, Tech. Rep. ERS-2005-015-ORG, 2005.

[280] F. McAndrew, "New Evolutionary Perspectives on Altruism - Multilevel-Selection and Costly-Signaling Theories," *Current directions in Psychological Science*, vol. 11, no. 2, pp. 79-82, 2002.

[281] D. S. Wilson et al, "Hunting, Sharing, and Multilevel Selection: The Tolerated-Theft Model Revisited," *Current Anthropology*, vol. 39, no. 1, pp. 73-97, 1998.

[282] D. S. Wilson, "Evolutionary Biology: Struggling to Escape Exclusively Individual Selection," *The Quarterly Review of Biology*, vol. 76, no. 2, pp. 199-205, 2001.

[283] T. C. Bergstrom, "Evolution of Social Behavior: Individual and Group Selection," *The Journal of Economic Perspectives*, vol. 16, no. 2, pp. 67-88, 2002.

[284] J. A. Fletcher, and M. Zwick, "The evolution of altruism: Game theory in multilevel selection and inclusive fitness," *Journal of Theoretical Biology*, vol, 245, no. 1, pp. 26-36, 2007.

[285] B. Cooper, and C. Wallacey, "Group selection and the evolution of altruism," *Oxford Economic Papers*, vol. 56, pp. 307–330, 2004.

[286] A. Traulsen, and M. A. Nowak, "Evolution of cooperation by multilevel selection," in *Proceedings of the National Academy of Science U.S.A.*, vol. 103, no. 29, pp. 10952-10955, 2006.

[287] K. Lee, "Moral Hazard, Insurance and Public Loss Prevention," *The Journal of Risk and Insurance*, vol. 59, no. 2, pp. 275-283, 1992.

Appendix A

Ranking Poker Combinations

Figure A.1: Name of poker cards combinations

- Each card has a value (A, K, Q, J, 10, 9, 8, 7, 6, 5, 4, 3, 2) and a suit (♠, ♣, ♥, ♦). The values from largest to smallest are: A, K, Q, J, 10, 9, 8, 7, 6, 5, 4, 3, 2. All suits are equal.

- In Texas Hold'em, it is to be noted that each player form the best 5-cards combination from the seven cards they can use. The unused two cards are not used in any way in determining whose combination has a higher ranking.

- The highest ranked combination is the "Royal Flush". It is made up of the cards A, K, Q, J, 10 of any suits. All royal flush are equal.

- The 2nd ranked combination is "Straight Flush" and is made up of any five consecutive cards of the same suit. If there is more than one "Straight Flush", the one that is made up of larger values is higher ranked, otherwise they are equal.

- The 3rd ranked combination is "Four of a Kind", made up of four cards of the same value and 1 any other card. A "Four of a Kind" with larger value for the

four same-valued cards will be higher ranked than one with a smaller value. If there are still ties, the value of the 5th card will determine the better combination. If all the cards are equal in value, then the combinations are also equal.

- The 4th ranked combination is "Full House", made up of three cards of the same value and another two cards of the same value. For more than one "Full House", the one with larger value for three cards wins. If there is still a tie, one with larger value for two cards wins.

- The 5th ranked combination is "Flush", which is made of all five cards of the same suit. If there is more than one "Flush", the one with the higher highest value wins. If the highest values are equal, then the next highest value is compared and so on.

- The 6th ranked combination is "Straight", consisting of five cards of consecutive values. A "Straight" made up of larger values will be bigger than one with smaller values.

- The 7th ranked combination is "Three of a Kind". The "Three of a Kind" with larger value for the three same-valued cards will be ranked higher. Otherwise the larger of the last two cards will be compared, finally followed by the last card.

- The 8th ranked combination is "Two pairs". If there are more than one "Two pairs", the larger pair of all combinations will be compared. The largest of them will be ranked the highest. If the larger pairs are all equal, the smaller pairs will be compared. If there is still a tie, the last card with the highest value will be highest ranked, otherwise all are equal.

- The 9th ranked combination is the "Pair". A "Pair" with higher valued pair will be larger than one with the smaller value. If the "Pairs" are the same, then each remaining card will be compared staring with the largest one.

- The smallest combination is the "High Card". If there is more than one "High Card", the largest card of each player will be compared first. If it is still tied, then the next largest card will be compared, and so on.