# Evolutionary Game Theory based Cooperation Algorithm in Multi-agent System

Yuehai Wang
*North China University of Technology*
*China*

## 1. Introduction

A multi-agent system (MAS) that composed of multiple interacting intelligent agents can be used to solve problems which are difficult or impossible for an individual agent or monolithic system to solve. Since the agent is autonomous and intelligent, it is reasonable to assume that it choice the behavior to bring itself the maximal benefit. Thus, the cooperation and coordination can be achieved successfully if we can wisely design the utility function for every agent so that every agent can get the maximal reward from the cooperation to accomplish a given task.

However, the utility function of one agent usually involves those of others for most "real-world" cooperation needed tasks. Moreover, it is not uncommon that the conflicts between the gains of these agents arise. In other words, the individual optimality is not always consistent with collective optimality in MAS. These conflicts will reduce the collective utility if there is no coordination among these decentralized, autonomous agents.

This paper addresses the essential that in MAS the action of one agent may influence the action of others and there usually be conflicts among the payoff of one another. We investigated the optimal coordination approach for multi-agent foraging, a typical MAS task, from the point view of game theory. After introduced several concepts, we built the equivalence between the optimal solution of MAS and the equilibrium of the game corresponding to that situation, and then we introduced evolutionarily stable strategy into the approach hope that it maybe be of service in addressing the equilibrium selection problem of traditional game theory.

Finally, based on the hawk-dove game model, an evolutionarily cooperation foraging algorithm (ECFA) is proposed to evolve a stable evolutionarily stable strategy (ESS) and bring the maximal reward for the group. If there be some change in the configuration of the environment, ECFA can, then, evolve to the new ESS automatically. And we also proposed a reinforcement factor to accelerate the convergence process of ECFA and thus make a new algorithm Accelerated ECFA (AECFA). These techniques were shown to be successful by the multi-agent foraging simulations.

## 2. Rationality

### 2.1 The concept of rationality

Rationality is an important property we imposed upon the players of a game. It is a central principle for agent to respond optimally by selecting its action based on the beliefs he might

have about the strategies of his opponents and the structure of game, i.e. payoff matrix of the game. Sometimes rationality, also called "hyper-rationality", implies having complete knowledge about all the details of a given situation. Under this concept, a player can calculate its best action of the current situation, and, furthermore, it can also calculate the best response of its opponents' to his action on the flawless premise that no one will make a mistake. However, perfectly rational decisions are not feasible in practice due to the finite computational resources. In fact, if an agent uses finite computational resources to deduce, we say it is bounded rational.

Of course, we assume all the players are honest and flawless weather he is rational or bounded rational when he selects his action. In other words, he never makes mistakes by choosing sub-optimal action intentionally to confuse his opponents.

### 2.2 Autonomous agent and rational player

Autonomous agent is the description of the player in a MAS. Autonomous means it can sense the environment and act on it over time in pursuit of its own goal. If the agent was equipped with learning ability, it can find the optimal way in accomplishing the same or similar job by machine learning techniques such as try-and-error, neutral network, and so on. Agent is egocentric during the selection and improvement of its action.

A rational agent is specifically defined as an agent who always chooses the action which maximizes its expected performance, given all of the knowledge it currently possesses, and this may involve "helping" or "hurting" the other players. This time, the agent is game-centric and the action selection is after a careful consideration about the payoff function of other players as well as game structure.

### 2.3 Rational and selfish

A rational agent always maximizes its payoff function based on the game structure and the common knowledge "other players are rational". But it is not always selfish although it may choice selfish action more often than not. If the game structure shows that cooperation with other player can obtain more benefits for all, it has the incentive to choice this action since both of them are rational and, therefore, they all know these win-win actions. Another exception is repeated game since the Nobel Prize winner Robert Aumann had already shown rational players repeatedly interacting for indefinitely long games can sustain the cooperative outcome in his 1959 paper (Aumann R.J. 1959).

## 3. Individual rationality and collective rationality

Individual rationality indicates that the choices made by individuals are to maximize their benefits and minimize their costs. In other words, agents make decisions about how they should act by comparing the costs and benefits of different courses of action (Sen, A. 1987). And the collective rationality stand for the group as a whole, to maximize the utility of the entire group which is composed every single agent.

As been stated before, usually there exist conflicts between actions that can make individual benefit or collective gains. Let's take the famous classical prisoner's dilemma as an example. In this game, as in all game theory, the only concern of each individual player ("prisoner") is maximizing his/her own payoff, without any concern for the other player's payoff. The unique equilibrium for this game is a Pareto-suboptimal solution—that is, individual

rational choice leads the two players to both play defectly even though each player's individual reward would be greater if they both played cooperately, which is collective rational and Pareto optimal (Poundstone, W. 1992).

## 4. Game theory based cooperation approach for multi-agent system

### 4.1 The relationship between the optimal cooperation solution of MAS and Nash equilibrium of the corresponding game

To accomplish a mission is only the preliminary requirement of a MAS. In fact, the MAS are required to complete the given task efficiently, and finally, optimally. It needs all the actions selected by the agent during every step of the procedure should be optimal. Of course, this is a very hard, if not impossible, problem.

But if we regard the procedure of accomplishing the given task as a Markov game composed by multiple stage games which corresponding to every step that constitute the cooperating work, we can find a optimal solution given that we find the best equilibrium of every stage game of the Markov game. Game theory provides several feasible approaches to find an equilibrium, the most popular one among which is Nash equilibrium.

Nash equilibrium is proven to exist for any game and it also is the only "consistent" prediction of how the game will be played in the sense that if all players predict that a particular Nash equilibrium will occur then no player has an incentive to play differently. Thus, a Nash equilibrium, and only a Nash equilibrium, can has the property that the players can predict it, predict that their opponents predict it, and so on (Fudenberg,D. & Tirole,J. 1991). Therefore, it is reasonable for us to choice Nash equilibrium as the optimal solution for each stage game although a Nash equilibrium can not always be Pareto-optimal.

### 4.2 Fundamental equilibria of the game and their relationship

From different viewpoints and based on different solution approaches, a game have multiple kinds of solution equilibria, among which Nash equilibrium, Iterative deletion of strictly dominated strategies, strictly dominance strategies, risk-dominant equilibrium and Pareto-optimal equilibrium are commonly used for static game of complete information. Here gives a very short description of and the relationship among these equilibria, please refer *game theory* (Fudenberg,D. & Tirole,J. 1991) for the details.

Informally, a set of strategies is a Nash equilibrium if no player can do better by unilaterally changing his or her strategy. Thus, Nash equilibrium is a profile of strategies such that each player's strategy is a best response to the other player's strategies. By best-response, we mean that no individual can improve her payoff by switching strategies unless at least one other individual switches strategies as well. There are two kinds of Nash equilibrium: mixed-strategy Nash equilibrium and pure-strategy Nash equilibrium.

Dominance occurs when one strategy is better than another strategy for one player, no matter how that player's opponents may play. The iterated deletion of dominated strategies is one common technique for solving games that involves iteratively removing dominated strategies. Eventually all dominated strategies of the game will be eliminated. Iterative deletion of strictly dominated strategies are those strategies survived.

Strictly dominance strategies are those strategies that can never be dominated by any strategy. They are the subset of iterative deletion of strictly dominated strategies since it also include the weakly dominated strategies. The idea of a dominant strategy is that it is always your best move regardless of what the other guys do. Note that this is a stronger

requirement than the idea of Nash equilibrium, which only says that you have made your best move given what the other guys have done.

Risk-dominant equilibrium (Harsanyi, J.C. & Selten, R. 1988): In a symmetric 2×2 game — that is, a symmetric two-player game with two strategies per player—if both players strictly prefer the same action when their prediction is that the opponent randomizes 1/2-1/2, then the profile where both player play that action is the risk-dominant equilibrium.

Pareto-optimal equilibrium is the equilibrium that has the property that can bring the maximum utilities for all players of the game.

The relationship between these equilibria is depicted in the fig. 1(Li, G.J. 2005).Note that risk-dominant equilibrium may be, or may not be a Nash equilibrium. And also note that a Pareto-optimal equilibrium may be, or may not be a Nash equilibrium.
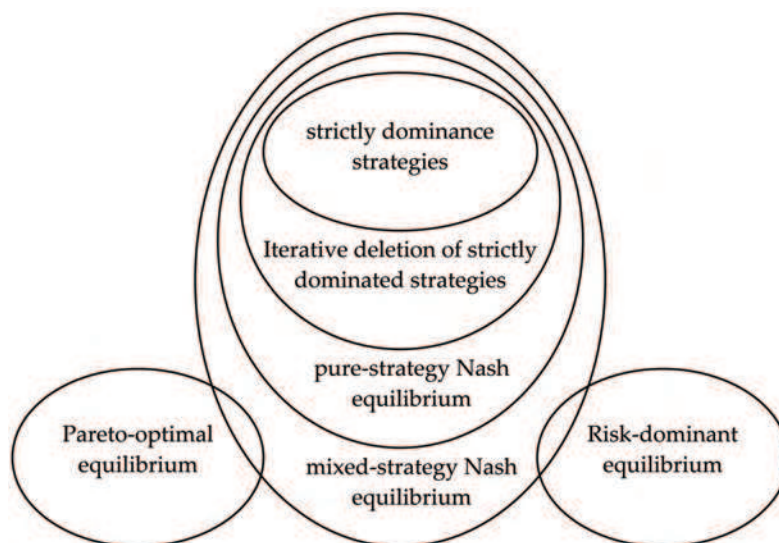


Fig. 1. The relationship between some equilibria

## 4.3 The type of the non-cooperative game and its equilibrium

A non-cooperative game is a one in which players can cooperate, but any cooperation must be self-enforcing, i.e. without the help through third parties by binding commitments or enforcing contracts. According to different standards, there are many categories of games. Fudenberg and Tirole (Fudenberg,D. & Tirole,J. 1991) use complete information and sequence of the players' move as the category standards. Complete information requires that every player knows the structure of the game, the strategies and payoffs of the other players. Static games (or simultaneous games) are games where both players move simultaneously, or if they do not move simultaneously, the later players are unaware of the earlier players' actions (making them effectively simultaneous), whereas the games where later players have some knowledge about earlier actions are called dynamic games (or sequential games). Therefore, there are four category of games: static games of complete information whose equilibrium is Nash equilibrium, dynamic games of complete information whose equilibrium is subgame prefect equilibrium, static games of incomplete

information whose equilibrium is Bayesian equilibrium and the last, dynamic game of incomplete information whose equilibrium is perfect Bayesian equilibrium. Please refer corresponding text for the details.

## 4.4 Equilibrium selection problem in game theory based cooperation approach

Equilibrium is a profile of strategies such that each player's strategy is an optimal response to the other player's strategies. Nash equilibrium is a most frequently used equilibrium among all kinds of equilibria. The fact that a game may exists several, even infinite, Nash equilibria bring about the trouble for the players to predict the outcome of the game. When this is the case, the assumption that one specific Nash equilibrium is played relies on there being some mechanism or process that leads all the players to expect the same equilibrium. However, game theory lacks a general and convincing argument that a Nash equilibrium outcome will occur (Fieser, J. & Dowden, B. 2008). As a result, it is not surprise that different player predict different equilibrium and so as to lead a non-Nash equilibrium come into exists since there is no common acknowledged doctrines for the player to predict and select. This is the equilibrium selection problem that addresses the difficulty for players to select certain equilibrium over another.

The researchers had already proposed several approaches and advices to make a reasonable selection for the player. Next list the some most frequently used approaches. The "focal points" theory of Schelling (Schelling, T. C. 1960) suggests that in some "real-life" situations players may be able to coordinate on a particular equilibrium by using information that is abstracted away by the strategic form of the game that may depend on players' culture background, past experiences, and so forth. This focal-point effect opens the door for cultural and environmental factors to influence rational behavior. Correlated equilibrium (Aumann R. 1974) between two players and coalition-proof equilibrium in games with more than two players (Bermheim, B.D., Peleg,B.& Whinstion,M.D. 1987a,1987b) that engage in a preplay discussion and then act independently is another approach. Risk-dominant principle first introduced by Harsanyi and Selten (Harsanyi, J.C. & Selten, R. 1988) is still another. However, please note that the selected Nash equilibrium is not necessarily Pareto-optimal equilibrium.

## 5. Evolutionary game theory approach

### 5.1 Introduction and advantages

Till now, we have motivated the solution concept of Nash equilibrium by supposing that players make their predictions of their opponents' play by introspection and deduction, using their knowledge of the opponents' payoffs, the knowledge that the opponents are rational, the knowledge that each player knows that the others know these things, and so on through the infinite regress implied by "common knowledge".

An alternative approach to introspection for explaining how players predict the behavior of their opponents is to suppose that players extrapolate from their past observation of play in "similar games," either with their current opponents or with "similar" ones. The idea of using learning-type adjustment process to explain equilibrium goes back to Cournot, who proposed a process that might lead the player to play the Cournot-Nash equilibrium outputs(Fudenberg,D. & Tirole,J. 1991).

If players observe their opponents' strategies at the end of each round, and players eventually receive a great many observations, the one natural specification is that each

player's expectations about the play of his opponents converge to the probability distribution corresponding to the sample average of play he has observed in the past. In this case, if the system converges to a steady state, the steady state must be a Nash equilibrium (Weibull, J.W. 1995).

We can use this large-population model of adjustment to Nash equilibrium to discuss the adjustment of population fractions by evolution as opposed to learning. In theoretical biology, Maynard Smith and Price (Smith, J.M. & Price, G. 1973) pioneered the idea that the genes whose strategies are more successful will have higher reproductive fitness. Thus, the population fractions of strategies whose payoff against the current distribution of opponents' play is relatively high will tend go grow at a faster rate, and, any stable steady state must be a Nash equilibrium.

To conclude this section, we know that we can use evolutionary game theory and evolution stable strategies to explain the Nash equilibrium. The advantages of this explanation are if the players play one another repeatedly, then, even if players do not know their opponents' payoffs, they will eventually learn that the opponents do not play certain strategies, and the dynamic of the learning system will replicate the iterative deletion process. And for an extrapolative justification of Nash equilibrium, it suffices that players know their own payoffs, that play eventually converges to a steady state, and that if play does converge all players eventually learn their opponents' steady state strategies. Players need not have any information about the payoff functions or information of their opponents.

## 5.2 Evolutionarily stable strategies and evolutionary game theory

In game theory and behavioral ecology, an evolutionarily stable strategy (ESS) is a strategy which once adopted by an entire population is resistant to invasion by any mutant strategy that is initially rare. ESS was defined and introduced by Maynard Smith and Price (Smith, J.M. & Price, G. 1973) which is presumed that the players are individuals with biologically encoded, heritable strategies who have no control over the strategy they play and need not even be capable of being aware of the game. The individuals reproduce and are subject to the forces of natural selection (with the payoffs of the game representing biological fitness).

Evolutionary game theory (EGT) is the application of population genetics-inspired models of change in gene frequency in populations to game theory. Now it is one of the most active and rapidly growing areas of research. It assumes that agents choose their strategies through a trial-and-error learning process in which they gradually discover that some strategies work better than others. In games that are repeated many times, low-payoff strategies tend to be weeded out, and equilibrium may emerge (Smith, J. M. 1982).

## 5.3 Evolution stable strategies and Nash equilibrium

As we already known, Nash equilibrium is a profile of strategies such that each player's strategy is an optimal response to the other player's strategies as a result of the rational agent's introspection and deduction based on the "common knowledge", such as the opponents' payoffs, while ESSes are only evolutionarily stable result of the simple genetic operation among those agents who even not knows any information about the payoff functions or information of their opponents. Given the radically different motivating assumptions, it may come as a surprise that ESSes and Nash equilibria often coincide. In fact, every ESS corresponds to a Nash equilibrium, but there are some Nash equilibria that are not ESSes. That is to say, an ESS is an equilibrium refinement of the Nash equilibrium --

it is a Nash equilibrium which is "evolutionarily" stable meaning that once it is fixed in a population, natural selection alone is sufficient to prevent alternative (mutant) strategies from successfully invading.

In most simple games, the ESSes and Nash equilibria coincide perfectly. For instance, in the Prisoner's Dilemma the only Nash equilibrium and the strategy which composes it (Defect) is also an ESS. Since ESS is more restrict Nash equilibrium, there may be Nash equilibria that are not ESSes. The important difference between Nash equilibria and ESSes is Nash equilibria are defined on strategy sets (a specification of a strategy for each player) while ESSes are defined in terms of strategies themselves.

Usually the game have more than one ESS, we have to choose one as the solution. To most game, the ESS is not necessary Pareto optimal. But for some specific game, there is only one ESS, and it is the only equilibrium whose utility is maximal for all the players.

## 5.4 Symmetric game and uncorrelated asymmetry

A symmetric game is a game where the payoffs for playing a particular strategy depend only on the other strategies employed, not on who is playing them (Smith, J. M. 1982). If one can change the identities of the players without changing the payoff to the strategies, then a game is symmetric. Symmetries here refer to symmetries in payoffs.

Biologists often refer to asymmetries in payoffs between players in a game as correlated asymmetries. These are in contrast to uncorrelated asymmetries which are purely informational and have no effect on payoffs. Thus, uncorrelated asymmetry only means "informational asymmetry", not "payoff asymmetry".

If uncorrelated asymmetry exists, then the players know which role they have been assigned. i.e. the players in a game know whether they are Player 1, Player 2, etc. If the players do not know which player they are then no uncorrelated asymmetry exists. The information asymmetry is that one player believes he is player 1 and the other believes he is player 2. Let's take the Hawk-Dove game (HDG hereafter), which will be presented in the next section, as an example. If player 1 believes he will play hawk and the other believes he will player dove, then uncorrelated asymmetry exists.

## 5.5 Hawk-Dove Game (HDG)

The game of Hawk-Dove, a terminology most commonly used in evolutionary game theory, also known as the Chicken game, is an influential model of conflict for two players in game theory. The principle of the game is that while each player prefers not to yield to the other, the outcome where neither player yields is the worst possible one for both players. The name "Hawk-Dove" refers to a situation in which there is a competition for a shared resource and the contestants can choose either conciliation or conflict.

The earliest presentation of a form of the HDG was by Smith and Price (Smith, J.M. & Price, G. 1973) but the traditional HDG payoff matrix for the HDG, given as Fig. 2, is given in his another book, where $v$ is the value of the contested resource, and $c$ is the cost of an escalated fight. It is (almost always) assumed that the value of the resource is less than the cost of a fight is, i.e., $c > v > 0$. If $c <= v$, the resulting game is not a HDG (Smith, J. M. 1982).

The exact value of the Dove vs. Dove playoff varies between model formulations. Sometimes the players are assumed to split the payoff equally ($v/2$ each), other times the payoff is assumed to be zero (since this is the expected payoff to wait, which is the presumed models for a contest decided by display duration).

While the HDG is typically taught and discussed with the payoffs in terms of $v$ and $c$, the solutions hold true for any matrix with the payoffs in Fig. 3, where W > T > L > X (Smith, J. M. 1982).

|        | Hawk               | Dove       |
|--------|--------------------|------------|
| Hawk   | $(v-c)/2, (v-c)/2$ | $v, 0$     |
| Dove   | $0, v$             | $v/2, v/2$ |

Fig. 2. Payoff matrix of traditional Hawk-Dove game

|        | Hawk | Dove |
|--------|------|------|
| Hawk   | X, X | W, L |
| Dove   | L, W | T, T |

Fig. 3. Payoff matrix of a general Hawk-Dove game

## 5.6 Using Hawk-dove game to model multi-agent foraging

Foraging is a popular, typical, as well as complex, multi-agent cooperation task which can be described as, plainly, a search for provisions (food). How to forage food in an unforeseen environment and evolve coordination mechanisms to make the process effectively and intelligently in itself spans a number of sub tasks. Equipping agents with learning capabilities is a crucial factor to improve individual performance, precision (or quality) and efficiency (or speed) and to adapt the agent to the evolution of the environment.

Generally, there are two kinds of food sources. One type is lightweight and can be carried by a single agent alone which is a metaphor for simple task that can be achieved by single robot, the other is heavy and need multiple agents to work simultaneously to carry it. This heavy food is a metaphor for complex task that must be accomplished by the cooperation of multiple robots (Hayat, S.A. & Niazi, M. 2005). Although coordination of multiple robots are not essential in collecting the lightweight food, the utilities can be increased when coordination indeed appear. Only lightweight foods are considered in this paper to simplify the complexity. In this case, the key to improve the collective utilities lies in how to make a feasible assignment of the food source to every agent so as to the goal for every agent is different since the same food source means there are conflicts between individual optimal assignment and collective optimal assignment in the MAS.

But it is nearly impossible to make an optimal assignment under any situation where there exits lots of agents and foods which scattered randomly. Let's start from an extremely simple situation to illustrate the difficulty. As depicted in Fig. 4, there are two agents A and B (red circle) pursuit two static foods F1 and F2 (two black dots) in a one-dimension world which only permit agent to move left or right and the food will be eaten whenever the agent occupy the same grid as a food. It is obvious that the optimal food for both A and B is F2 since it is nearer than to F1. It is also obvious that if both A and B select F2 as their pursuit target, then utilities of A was sacrificed since it can not capture F2. Thus, it will cause low efficiency as far as the collective utility is considered. In this case, the optimal assignment is B pursuits for F2 while A trying to capture F1. This assignment can be regarded as agent A and B select different policy when confront same food, one is to initiate an aggressive behavior (B), just like hawk in HDG, the other is to retreat immediately (A), like a dove in HDG.

Fig. 4. Simple foraging task in one-dimensional world

And this is only a extremely simple case, if we extend it to two-dimension where the move also extend to {up, down, left, right}, to large number of agents and foods scattered randomly, it will be very hard to make a wise assignment. If we use HDG to model the agent, then we can let the agent select a food by certain doctrine, such as nearest first, and then revise it if the target of multiple agents is the same. In this case, we can let those agents play a HDG to decide who will give up.

As a conclusion, we can abstract the strategies of agents to two categories: one is always aggressive to the food, the other is always yield. The yield agent is dove, and the aggressive one is hawk. In this paper, this HDG model was used to model the strategy of pursuit agents to give the multi-agent foraging a feasible approach.

## 5.7 Evolution dynamics – replicator dynamics

Replicator dynamics is a simple model of strategy change in evolutionary game theory. Shown in equation (1), it describes how the population with strategy *i* will evolve.

$$\dot{x}_i = [u(i,x) - u(x,x)]x_i \tag{1}$$

In the symmetric 2×2 hawk-dove game, a strategy which does better than the average increases in frequency at the expense of strategies that do worse than the average. There are two versions of the replicator dynamics. In one version, there is a single population which plays against itself. In another, there are two population models where each population only plays against the other population (and not against itself).

In the one population model, the only stable state is the mixed strategy Nash equilibrium. Every initial population proportion (except all Hawk and all Dove) converge to the mixed strategy Nash Equilibrium where part of the population plays Hawk and part of the population plays Dove. (This occurs because the only ESS is the mixed strategy equilibrium.) This dynamics of the single population model is illustrated by the vector field pictured in Fig. 5 (Cressman, R. 1995).
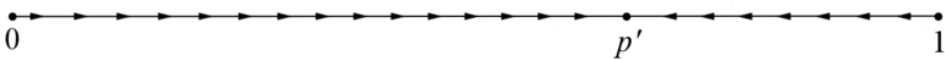


Fig. 5. Vector field for single population replicator dynamics

In the two population model, this mixed point becomes unstable. In fact, the only stable states in the two population model correspond to the pure strategy equilibria, where one population is composed of all Hawks and the other of all Doves. In this model one population becomes the aggressive population while the other becomes passive.

The single population model presents a situation where no uncorrelated asymmetries exist, and so the best players can do is randomize their strategies. The two population models

provide such an asymmetry and the members of each population will then use that to correlate their strategies, and thus, one population gains at the expense of another.

Note that the only ESS in the uncorrelated asymmetric single population hawk-dove model is the mixed strategy equilibrium, and it is also a Pareto optimal equilibrium (Smith, J. M. 1982). If some problem can be solved by this model, including our HDG modeled multi-agent foraging, and then the evolutionarily stable strategy is the only Pareto-optimal Nash equilibrium of the system.

## 6. Evolutionarily cooperation foraging algorithm for MAS

Multi-agent foraging is popular to verify the effectiveness of different cooperation algorithms. In evolving game theory, equilibrium is the result of long process in which the bounded-rational players are trying to optimize their payoff by a natural-selection like mechanism. From the learning process based on replicator dynamic, every player can obtain enough information of personalized equilibrium selection pattern of other agents, and then attain an optimal unanimous equilibrium for the whole MAS. For HDG, the sole evolutionarily stable strategy is also the sole Pareto-optimal Nash equilibrium and thus give a solution to the equilibrium selection of the traditional game theory.

Using evolutionarily stable strategy as optimal solution, we built a HDG model to simulate the interaction between agents, and then proposed a evolutionarily coordinating foraging algorithm (ECFA) to find certain consistent maximal reward equilibrium for the group. Finally, we also add an accelerating factor to make ECFA converge faster, and thus make a new Accelerated ECFA (AECFA). The simulation verified the efficiency of the proposed algorithm.

### 6.1 Description of problem

Suppose a group of agent ($n$ agents) were to capture as much as possible random moving preys ($m$ preys) in a bounded rectangle field during a fixed period of time. The agents, having same bounded visual field, start at WANDER state to find a prey. Once it found the food, the agent change its state to GETIT to capture till it eat the food and change its state back to WANDER.

If the agent is the sole pursuer for its target food, it just eats it by moving near to it. Eating occurs when the distance between the food and agent is less than a threshold distance. Another food will be generated at a random position right after to mimic a food abundant environment.

But if the agent find another agent who pursuit the same food (suppose all agent know the goal of other agents), these two agents will play a HDG to determine the rewards they can get. As described in the previous part, two hawks compete for the food with sufficient large cost, while two doves both give up the food and get nothing. If a hawk meet a dove, the hawk eat the food and the dove give up.

Agent can change its strategy to be hawk or dove. As stated in the replicator dynamics, a strategy which does better than the average increases in frequency at the expense of strategies that do worse than the average. Thus, the average reward of the whole system produced by the replicator dynamic is monotonically increasing with time for the symmetric HDG (Losert, V. & Akin, E. 1983). And as a result, the agent with worse strategy would change his strategy to better one and thus lead the whole system to a dynamic stable state with best reward for the agent group (Smith, J. M. 1982).

## 6.2 Introduction of evolutionarily coordinating foraging algorithm-ECFA

This part describes how the replicator dynamics works so that the system evolves to the sole ESS. In replicator dynamics, the increasing quota of certain strategy is in proportion to the ratio of its average payoff to the average payoff of the population (Weibull, J.W. 1995). Therefore, a strategy which does better than the average increases in frequency at the expense of strategies that do worse than the average. The agent select its strategy based on the accumulated experience or on the observation and imitation of the strategies adopted by opponents. The more popular of a strategy, the more possibility it would be imitated.

During the learning process, agent makes introspection to its strategy from time to time and this gives the possibility that it may change its strategy. Suppose those agents using less successful strategy are more likely to introspection and let $r_i(x)$ be the average rate of introspection of agent using strategy $i \in K$, where $K$ is the strategy set and $e^i$ are strategies of agent $i$ in $K$.

$$r_i(x) = 1 - \frac{u(e^i, x)}{\sum_{h \in K} u(e^h, x)} \qquad (2)$$

Next, we use $p_i^j(x)$ denote the probability that agent $i$ will change to use strategy $j$ and let $p_i^j(x)$ is proportion to the popularity of strategy $j$ and its rewards.

$$p_i^j(x) = \frac{\omega[u(e^j, x), x]x_j}{\sum_{h \in K} \omega_i[u(e^h, x), x]x_h} \qquad (3)$$

here $\omega$ is continuous Lipschitz function that non-decrease in its first independent variables.

Also, to show the agents using less successful strategy are more likely to introspection we suppose $r_i(x) = \phi[u(e^i, x), x]$ and $\phi$ is continuous Lipschitz function that strictly decreases in its first independent variables.

And at last we get the replicator dynamics of this symmetric revised Hark-Dove game as

$$\dot{x} = \sum_{j \in K} x_j r_j(x) p_j^i(x) - r_i(x) x_i \qquad (4)$$

which will lead to the average fitness of the whole system increase monotonically with the time until the system evolve to an Pareto-optimal ESS, the sole evolutionarily stable state(Wang,Y.H., Liu, J., & Meng, W. 2007).

## 6.3 Description of ECFA

*Initialization:*

　　Generate all preys and agents.

　　Assign random strategy (hawk or dove) to each agent.

　　Set the state of agent to WANDER to enable the agent looking for food.

　　Let $RAND \in (0,1)$ is a random generated threshold.

*Main:*

　　for every agent, run Step1 to Step3 infinitely until the MAS converge to ESS.

*Step 1: //Agent pursues food*

```
    if (prey found) {
       Agent change its state to GETIT; goto Step 2;
    }
    else goto Step 1;
```
*Step 2: //Single pursuer*
```
    If (the agent is the only pursuer of the prey) {
       Eat the prey and get reward;
       Generate a new prey at random position;
       goto Step1;
    }
    else goto Step 3.
```
*Step 3://Multiple pursuer executing introspection - imitation*

Play the hawk-dove game and get reward;

Update its environment model $x(x_i,x_j)$, where $x_i$ and $x_j$ are the proportion of encountered hawks and doves.

Compute the utilities $u(i,x)$; $i \in K = \{Hawk,Dove\}$

Using equation (3) to compute the introspective probability of agent who execute strategy $i$:

If ($r_i(x) > RAND$){

Compute the strategy change probability $p_{Dove}^{Hawk}$ and $p_{Hawk}^{Dove}$; here $p_i^j$ denotes the probability that agent change its strategy from $i$ to $j$.

if( $p_{Dove}^{Hawk}$ > $RAND$)) strategy = Hawk;

if( $p_{Hawk}^{Dove}$ >$RAND$)) strategy = Dove;

}

## 6.4 Simulation results of ECFA

Several simulations had been done to verify the efficiency of ECFA. The following parameters were used for the simulations: agent number $n$ =50, prey number $m$ = 130, the benefits of capture prey $v$ =4, the cost to injury to self $c$ = 6. The environment is defined as an 1150*650 grid. Each grid location represent an $x$ and $y$ location which can be occupied by one or more agents at the same time. Preys were randomly disposed in the field before the simulation start. They can move randomly with a lower velocity than that of agents (70%). Right after a prey was eaten, a new prey would be regenerated at a random chosen grid.

The first group of simulation is to test the validity and efficiency of the ECFA, we compare ECFA with another three algorithms, namely random forage, fixed strategy 1 forage with 30% hawk and 70% dove, fixed strategy 2 forage with 70% hawk and 30% dove. While a random forage agent will try to eat every food it found, a fixed strategy forage agent will play the HDG when two agents compete for the same food, but the number of hawk agent and dove agent remains unchanged. In ECFA, however, the number of hawk and dove agent will evolve until they finally converge to a stable state. The performance index is the average number of the preys captured by the agent group in a given span of time.

In either situation, the four algorithms were tested for 10 times respectively, and fig. 6 gives the graphic depiction of simulation results.
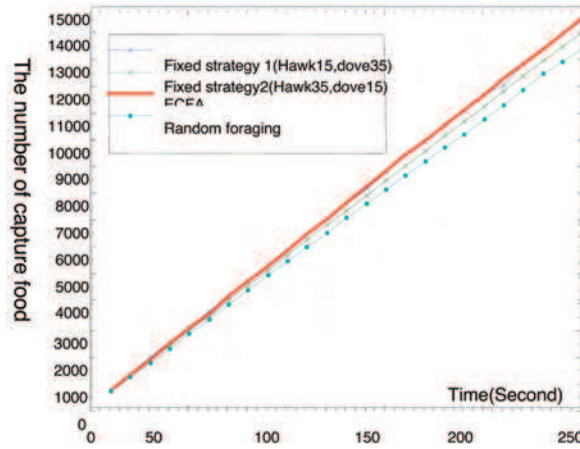
Fig. 6. The average number of preyscaptured by different foraging algorithm

These results show that ECFA captured more foods than other three algorithms averagely. If we investigate the results more carefully, we can see fixed strategy 1 foraging with 30% hawk and 70% dove outperforms fixed strategy 2 that forage with 70% hawk and 30% dove. It is easy to see that the agents with different hawks and doves have different performance. Then, it would be natural to ask how much hawk and dove will be evolved in ESS in various HDG model. That is what we want to show in the second group of simulation.

The second group of simulation is to find how many hawk agents in the evolutionarily stable state for different configuration of the hawk-dove games. Here we suppose $v+c=10$ and we and we test 6 situations from $(c=4,v=6)$, $(c=5,v=5)$,…, to $(c=9,v=1)$. Note that even it is not HDG when $(c=4,v=6)$ and $(c=5,v=5)$ since $c < v$, we are also eager to know the result.

Each situation was test for 10 times and table 1 lists the simulation result as well as the corresponding theoretical result of the average number of hawk agent in the ESS of every situation. Fig. 7 is the corresponding graph.
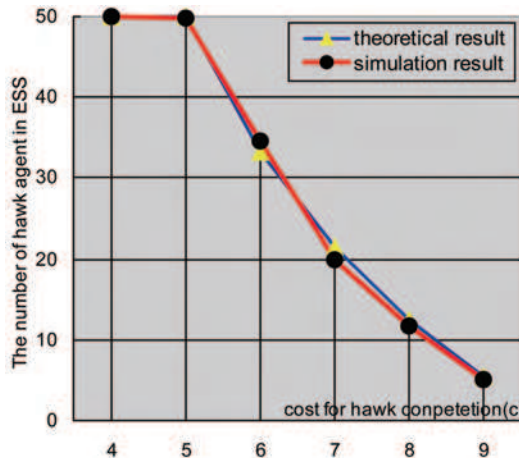


Fig. 7. The average number of hawk agents in convergent ESS of different game model

From these simulations, we can see that the number of hawk agent in the convergent ESS is decreasing with the increasing of the cost for two hawk competition. And it also shows that the simulation results are close to their theoretical values. The error between these two values is probably because the convergent threshold value for our simulation and the theoretical value is limit point which hardly achieved in finite trials.

| *cost* | *4* | *5* | *6* | *7* | *8* | *9* |
|---|---|---|---|---|---|---|
| theoretical result | 50 | 50 | 33.3 | 21.4 | 12.5 | 5.5 |
| simulation result | 50 | 49.8 | 34.6 | 20 | 11.8 | 5 |

Table 1. The number of hawk agent and corresponding theoretical result in different *c,v*

### 6.5 Improving of ECFA

As stated in replicator dynamics of this symmetric HDG, a strategy better than the average increases in frequency at the expense of strategies that worse than the average. And the changing quota of certain strategy is in proportion to the ratio of its average payoff to the average payoff of the population. But because the difference between strategies in the early evolutionarily stage is small, the better strategies or worse strategies can only impose a little impact on the agent group. Thus the evolution to the ESS of agent adopted ECFA is slow. Moreover, just as this evolving is a dynamic process and the strategies adopted by the agents are keep changing, which would make the system is not stable enough.

For these deficiencies, we added a reinforcement factor to ECFA to make an Accelerated ECFA (AECFA) to strengthen the outstanding strategies and weaken inferior strategies. The process of convergence will be accelerated and the convergent ESE will be more stable for the impact of the worse mutation strategies is weakened.

### 6.6 Reinforcement factor and the description of AECFA

Let $\theta_{i,t}^{e^j}$ be the reinforcement factor of agent $i$ with respect to strategy $e^k \in K$ at time $t$ and $\Delta\theta_{i,t}^{e^k}$ be a degree that the reinforcement factor is changed at time $t$. When $t=0$, $\theta_{i,t}^{e^j}=0$ and $\sum_{e^k \in K} \theta_{i,t}^{e^j} = 0$. Then the reinforcement factor can be defined as

$$\theta_{i,t+1}^{e^k} = \theta_{i,t}^{e^k} + \Delta\theta_{i,t}^{e^k} \tag{5}$$

which means the better the strategy does, the more positively it is reinforced. And vice versa, the worse the strategy does, the more negatively it is reinforced.

Now let $q_{i,t}^{e^k}$ denote the probability that agent $i$ execute strategy $e^k$ at time $t$ and let

$$\Delta\theta_{i,t}^{e^k} = \frac{q_{i,t}^{e^k} - \frac{1}{n}}{\frac{1}{n}} = n \cdot q_{i,t}^{e^k} - 1 \tag{6}$$

where $n$ denote the number of the set of strategies $K$. Then, there is positive correlation between $q_{i,t}^{e^k}$ and the utility of $e^k$. And we let

$$q_{i,t}^{e^k} = \frac{u(e^k, x)}{\sum_{h \in K} u(e^h, x)} \tag{7}$$

At time $t$, let the agent $i$ executes the strategy whose reinforcement factor is maximal. If the number of maximal reinforcement factor is more than one, the agent executes one of them according to some probability. At any time, each strategy in the set of agent-strategies is reinforced positively or negatively with respect to its current utility (Wang,Y.H., Liu, J. 2008). The algorithm description of this accelerated ECFA is given in Fig. 8.

```
Step 1:
    Compute the utilities u(e^k, x), e^k ∈ K = {Hawk, Dove}
    Using (2), Computing the reflecting probability of agent i
Step 2:
    IF ( r_i(x) > RAND ){
        Compute θ_{i,t}^{hawk} and θ_{i,t}^{dove} ;
        IF ( θ_{i,t}^{hawk} > θ_{i,t}^{dove} )
        {
            STRATEGY=Hawk;
            Update its environment model;
        }
        ELSE IF ( θ_{i,t}^{hawk} < θ_{i,t}^{dove} )
        {
            STRATEGY=Dove;
            Update its environment model;
        }
        ELSE
            Keep the strategy as it is;
}
```

Fig. 8. The algorithm description Accelerated ECFA

## 6.7 Simulation results of AECFA

To verify effectiveness of the reinforcement factor for the algorithm, we use multi-agent foraging task to test the difference of the stability and the time of convergence between AECFA and ECFA.

The parameters: agent number $n$ =64, prey number $m$ =30, $c$=8, $v$=2. Theoretically, the number of hawk agent should be 16 in ESE on the condition of this simulation; the simulation sampled the number of hawk once 500 seconds (Liu, J. 2008).

The performance index: the number difference between the number of hawk and the number of hawk in the equilibrium. Here is an example to make it clear. Suppose at certain time, the sampled hawk agent is 18, then the number difference is |18-16|=2 and the

performance ratio is 2/64=3.1%. Here 2 and 64 denote the number difference and total number of the agents respectively.

Considering the size of the MAS, we take the following difference ratio as the criterion to decide weather the system is in the convergent state or not. This threshold is determined to be 5% according to several simulations. If the performance ratio is less than the difference ratio 5%, the system would be considered in the ESE. In this case all these two algorithms had been tested for 10 times respectively; the results are given in Fig.9 and table 2. Fig.9 plots the number difference between the average hawk number and the hawk number of theoretical equilibrium in convergent ESE.

As shown in the Fig.9, in convergent ESE, the stability of the AECFA is less than 2 %( between 15 and 18), yet the stability of the ECFA is only close to 5 %( between 12 and 20). The AECFA evolved into ESE at average 100 seconds and the ECFA at average 56590 seconds. Thus, AECFA gives a much faster convergence process than the ECFA does.
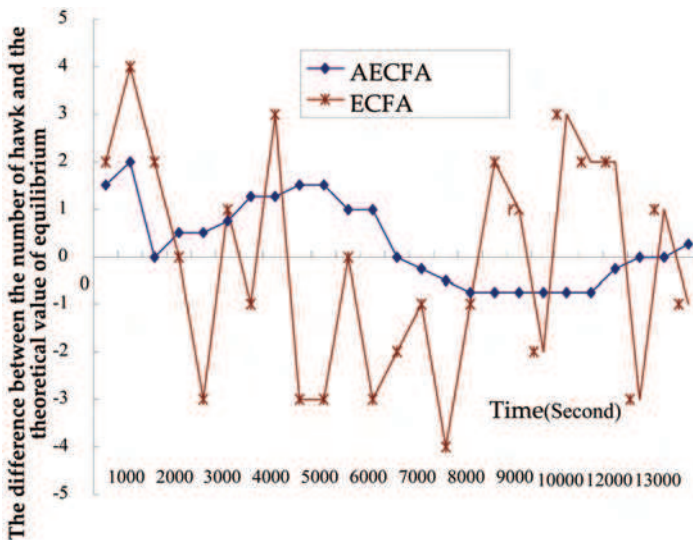


Fig. 9. The stability of two algorithms in the convergent ESE

| Time hawks | 0 | 500 | 1000 | 1500 | 2000 | 2500 | 3000 | 3500 | 4000 | 4500 | 5000 | 5500 | 6000 | 6500 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AECFA | 18 | 18 | 16 | 17 | 17 | 17 | 17 | 17 | 18 | 18 | 17 | 17 | 16 | 16 |
| ECFA | 18 | 20 | 18 | 16 | 13 | 17 | 15 | 19 | 13 | 13 | 16 | 13 | 14 | 15 |

Table 2. The stability of two algorithms in the convergent ESE (Part I)

| Time hawks | 7000 | 7500 | 8000 | 8500 | 9000 | 9500 | 10000 | 10500 | 11000 | 11500 | 12000 | 12500 | 13000 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AECFA | 16 | 15 | 15 | 15 | 15 | 15 | 15 | 16 | 16 | 16 | 16 | 15 | 15 |
| ECFA | 12 | 15 | 18 | 17 | 14 | 19 | 18 | 18 | 13 | 17 | 15 | 14 | 19 |

Table 2. The stability of two algorithms in the convergent ESE (Part II)

## 7. Conclusion and future works

An evolutionarily stable strategy was adopted to be an optimal solution for the multi-agent foraging task, and hawk-dove game model was used to simulate the interaction between the agents, and finally, an evolutionarily cooperation foraging algorithm (ECFA) that can adapt the environment change, evolve an stable ESS and bring the maximal reward for the group eventually is proposed. Moreover, to the disadvantages such as long convergence process of the ECFA, we imposed a reinforcement factor to accelerate the convergence process.

The technique was shown to successfully apply in the multi-agent foraging task and to increase the efficiency of the agent group. Thus, at least for the domain under study, large population of agents and random moving preys, the technique proposes very fast foraging mechnisms. The simulation shows that the proposed algorithm can convergence to a maximal-reward ESS which in turn give an solution to equilibrium selection problem of game theory based cooperation MAS.

For the second type of preys, a big food that need more than one agent together to capture, and dynamic environments where the value and cost to contest can change due to some reason need to studied in the future work.

## 8. Acknowledgements

## 9. References

Aumann R.J. (1959), Acceptable points in general cooperative n-person games, in: *Contributions to the Theory of Games IV, Annals of Mathematics Study*, Luce, R.D. & Tucker, A.W. (Ed.), pp. 287-324, Princeton University Press, Princeton NJ.

Aumann R. (1974), Subjectivity and correlation in randomized strategies. Journal of Mathematical Economics, Vol. 1, pp. 67–96

Bermheim, B.D., Peleg,B. & Whinstion, M.D. (1987), Coalition-proof Nash equilibria - I. Concepts, Journal of Economic Theory, Vol. 42, pp.1-12

Cressman, R. (1995). Evolutionary stability for two-stage Hawk-Dove games. The Rocky Mountain Journal of Mathematics, Vol. 25, No. 1, Winter 1995. pp.145–155.

Fieser, J. & Dowden, B. (2008), Game theory, in: the Internet encyclopedia of philosophy, http://www.iep.utm.edu/g/game-th.htm, accessed on 8 August 2008

Fudenberg,D. & Tirole,J. (1991), Game Theory, the MIT Press, Cambridge, Massachusetts, London, England

Harsanyi, J. C. & Selten, R. (1988), A General Theory of Equilibrium Selection in Games, the MIT Press, Cambridge, Massachusetts, London, England

Hayat, S. A. & Niazi, M. (2005), Multi-agent foraging – taking a step further Q-learning with search. Proceedings of IEEE international conference on emerging technologies (ICET05), September 2005, Islamabad, pp. 215-220

Li, G.J. (2005), A fundamental course in game theory, the Chemical industry press. Beijing, China, (Feb. 2005), p30-32

Liu, J., The study on multi-agent cooperative algorithm based on game theory. Master thesis, North China University of Technology, Beijing, China, June 2008

Losert, V. & Akin, E. (1983), Dynamics of games and genes: discrete versus continuous time, Journal of Mathematical Biology, 1983

Poundstone, W. (1992), Prisoner's Dilemma: John Von Neumann, Game Theory and the Puzzle of the Bomb, Doubleday, (February 1993). ISBN-13: 978-0385415804

Schelling, T. C. (1960), The Strategy of Conflict, Harvard University Press, Cambridge, Massachusetts.

Sen, A. (1987), Rational behaviour, The New Palgrave: A Dictionary of Economics, Vol. 3,(1987) pp. 68-76.

Smith, J. M. & Price, G. R. (1973), The logic of animal conflict. Nature, Vol. 246. pp.15-18.

Smith, J. M. (1982), Evolution and the theory of games, Cambridge University Press, ISBN 978-0-521-28884-2

Wang,Y.H., Liu, J., & Meng, W. (2007), Cooperative algorithm for multi-agent foraging task based on modified hawk-dove game. Proceeding of International Conference on Computational Intelligence and Security Workshops (CIS07), December 2007, IEEE Computer Society Press, Harbin, China. pp.179-182

Wang, Y.H., Liu, J. (2008), Ponder-reinforcement cooperative algorithm for multi-agent foraging task based on evolutionary stable equilibrium, Proceedings of 7th World Congress on Intelligent Control and Automation (WCICA08), June 2008, Chongqing, China. pp. 4527-4530.

Weibull, J.W. (1995), Evolutionary Game Theory. The MIT Press, Cambridge, Massachusetts. ISBN 0-262-23181-6

**Multiagent Systems**

Edited by Salman Ahmed and Mohd Noh Karsiti

Multi agent systems involve a team of agents working together socially to accomplish a task. An agent can be social in many ways. One is when an agent helps others in solving complex problems. The field of multi agent systems investigates the process underlying distributed problem solving and designs some protocols and mechanisms involved in this process. This book presents an overview of some of the research issues in the field of multi agents. It is a presentation of a combination of different research issues which are pursued by researchers in the domain of multi agent systems as they are one of the best ways to understand and model human societies and behaviours. In fact, such systems are the systems of the future.

**How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

INTECH

open science | open minds