Is Good Algorithm for Computer Players also Good for Human Players?

Naoki Konno[1], Kyoichi Kijima[2*]

[1,2]Graduate School of Decision Science and Technology Tokyo Institute of Technology, 2-12-1, Ookayama Meguro-ku, Tokyo, Japan, [1]nkonno [2]kijima@valdes.titech.ac.jp

[*]Correspondence can be directed to this author, as the primary contact.

Abstract

This paper aims to examine effectiveness of rational strategies for rough reasoning human players. Nowadays, computer players beat human champion players in many games (ex. Chess, Reversi, etc.) Actually, since computational power of computers transcends the human players, accuracy and volume of the search ability of computer players are superior to the human champion players in the end game phase. Then, the problem is that these computer algorithms are also effective for human players? The algorithms are basically composed by backward induction that is equilibrium concept for rational players. However, human players sometimes make wrong reasoning unlike computer players. In order to investigate the problem, we first propose a rough reasoning model that describes human imperfect reasoning abilities. This model is characterized by following two assumptions. The first is that as the payoff difference decrease, reasoning accuracy tends to decrease. The second is that as length of the tree increase, reasoning accuracy tends to decrease. We then make some examples of games and play them by some kinds of rough reasoning players with various algorithms. In the real game situations, accepted theories sometimes contradict to the rational strategies. We try to reveal the validity and effectiveness of the theories.

Keywords: bounded rationality; algorithm, game theory, rough reasoning.

1. Introduction:

This paper investigates implications of rough reasoning of players in perfect information games. We first develop a rough reasoning process model so as to focus on the payoff difference and depth of the tree. We then apply it to the centipede game for illustrating discrepancy between equilibrium obtained by backward induction and actual experimental results. Finally, we examine a property of rough reasoning by applying it to zero-sum tournament game and centipede game. The first result is that in the zero-sum tournament game, even if our reasoning is not complete, rational approach seems to be better approach than some other simple approaches in the sense of expected payoffs. The second results are in the centipede game case, moderate reasoning ability may lead irrational action with the highest frequency. However, this extent of rationality may lead to socially the most desirable welfare. These results show that in the non-zero-sum game situations, not only rational approach may not be the best one. but also rational reasoning ability may decrease their payoffs.

In May 1997, IBM's Deep Blue Supercomputer played a fascinating match with the reigning World Chess Champion, Garry Kasparov. In such a kind of complicated game, strong computer algorithms usually consisted of three parts. In the opening part of the games, they choose a move from standard move archives. Then, in the middle part, they evaluate alternative moves by their evaluation functions. For example in the case of chess, a king piece has infinity point and a queen piece has 9 point and a pawn piece has 1 point. Finally in the end game part, they first try to calculate all of the nodes correspond to the rational decision making. We then turn to a problem that Is a good algorithm for computer players also a good for human players?

In the traditional game theory, it is usually assumed that players completely recognize the game situation so as to compare all the results without error and to choose rational strategies. So this assumption means the game theory is not for actual incomplete human players but for complete rational players. Therefore even in the extensive form games with perfect information, as Selten (1975) pointed out by using chain store paradox, not only Nash equilibrium but also subgame perfect Nash equilibrium may lead to strange outcomes. Indeed, in perfect information games, though it is theoretically able to calculate reasonable equilibria by backward induction, it is practically difficult to realize them due to various complexity and the limitation of abilities.

In order to describe such kinds of bounded rationality, Selten(1975) proposed a concept of trembling hand equilibrium in normal form games. He refined equilibrium from a viewpoint that players make infinitesimal errors and small errors are more likely to occur than big errors. Noninfifitesimal errors has been studied by Mckelvey and Palfrey(1992). They examined developed the quantal response equilibria by substituting quantal response for best response in the sense that the players are more likely to choose better strategies than worse strategies but do not play a best response with probability *1*. He also examined the further property by using quantal logit functions. On the other hand, in extensive form games, Mckelvey and Palfrey (1998) transformed it into normal form games, and examined quantal response equilibria.

These assumptions about error can be considered as properties of action errors. However, human errors are observed not only in action phase but also in reasoning phase. Our main interest is a kind of player's rough reasoning. Though it may be difficult to describe complete reasoning process itself in normal form games, we deal hear with extensive form games with perfect information situation so as to examine the complete reasoning structure.

We should notice that in these previous papers, investigate action errors such as typing errors. On the other hand, we also make mistake in our reasoning. It is represented by miscalculation in chess. Though it may be difficult to formulate reasoning structure in normal form games, Heifetz and Pauzner(2005) proposed a model with possibility of wrong reasoning in binary action games. Their model is also based on agent normal forms and error rate is given by exogenous manner.

It is also important to emphasize that players do not always try to implement equilibrium strategies therefore we replace the equilibrium approach with heuristics approach. In our model, it is assumed that players try to evaluate each action by rough reasoning, then choose the best action, while they make wrong evaluation with certain probabilities. We look objectively at these game situations, then describe the result of rough reasoning decision making.

We characterize player's rough reasoning by following two sides. One is the payoff, while the other is the depth of the tree. First, as Mckelvey and Palfrey argued, we assume reasoning accuracy depends on the difference of the payoffs in such a way that error rate is a decreasing function of the difference of payoffs. In addition, we also suppose that as the depth of decision tree increases, reasoning accuracy tends to decrease. This property describes why it is difficult to compare actions in the far future. We call it a general rough reasoning model.

We then assume two more properties about reasoning error rate by studying characteristics of rough reasoning deeply. The first is that as payoff difference decreases, error rate increases exponentially. The second is as payoff difference decreases, error rate increases exponentially. We call it a rough reasoning model with a logit function.

Furthermore, to examine validity of the model, we apply it to zero-sum games and centipede game. First, zero-sum game is a represented by reversi or game of go. In these kinds of games, final results are given by points. both players try to maximize the difference of the points to the opponent. However, especially for human players, complexity can be also considered as a quite important factor. Concretely speaking, if you have an advantage enough, you should prioritize simplification of the situation over maximize evaluation values. We try to validate the principle from the reasoning error viewpoints. Second, the centipede game(Rosenthal, 1981) is known for the discrepancy between equilibrium obtained by backward induction and actual experimental results. According to Mckelvey and Palfrey(1992), the unique subgame perfect Nash equilibrium outcome is not so often observed. They tried to rationalize the results by mathematical model in which some of the players have altruistic preferences. Aumann(1995,1998) insisted that incompleteness of common knowledge assumption causes cooperative behaviors. Although these factors may work in the centipede game, we claim rough reasoning is also an essential factor leading to cooperative behaviors.

This paper is organized as follows. Section *2* presents a general rough reasoning heuristics. In Section *3*, we propose a specific reasoning model in order to examine more precise properties of rough reasoning. We then apply it to the zero-sum game situations and the centipede game and examine influences of the rough reasoning to final outcomes by numerical simulations in Section *4*. Finally some conclusions and remarks are given in Section *5*.

2. General rough reasoning model:

In the traditional game theory, it is usually assumed that all players perceive situation precisely, and essentially compare all the strategies without error. However, such perfect reasoning is quite difficult in most actual decision situations due to the players' reasoning abilities. Reasoning process is quite complication in usual, we deal here with only perfect information games. It is because that reasoning process in perfect information games is easy to study. We first define the true objective game of a perfect information game.

Definition 1

True objective game is a finite perfect information extensive form game given by

$G=(I,N,A,h,(N_i)_I,P,(r_i\}))$

where I is the set of players, while N is the set of nodes. N_T and N_D are partitions of N, where N_T is the set of terminal nodes and N_D is the set of decision nodes. A is the set of actions h: N-$\{n_1\}$: $\rightarrow N_D$ is the function from nodes except initial node n_1 to the prior nodes. P: $N_D \rightarrow I$ is the player function that determines the player who chooses an action at the node. r_i: $N_T \rightarrow \mathbf{R}^I$ is the payoff function that determines the payoffs of each agent.

Since G is a perfect information game, subgame perfect equilibria are obtained by backward induction. However, since the players can not compare all the result without error in the actual situations, we assume that players choose actions by the following heuristics.

To implement it, we need some notations:

N_2: The set of attainable nodes from n_1 i.e. $N_2 =\{n|\ n \in N,\ h(n)= n_1\}$.
$N^1{}_m$: The set of the last decision nodes of G. i.e. $N^1{}_m =\{n|\ n$ N_{D}$,\ \exists n_t \in N_T,\ s.t.\ h(n_t)=n,$ and $\neg(\ \exists n_d \in N_D,\ s.t.\ h(n_d)=n)$.
n^*: A Best node at $n \in N_D$ for $P(n)$. i.e. $n^* \in argmax_{\{n\}},r^{P(n)}(n')|\ h(n')=n$.

Denote by $r(n_d)=(r_1(n_d),...,r_j(n_d),...,r_I(n_d))$ a payoff vector that the player reasons to achieve if the optimal choices are taken at every stage after
$n_d \in N_d$-$\{n_1\}$. Then the heuristics are as follows. (Refer to Fig.1).

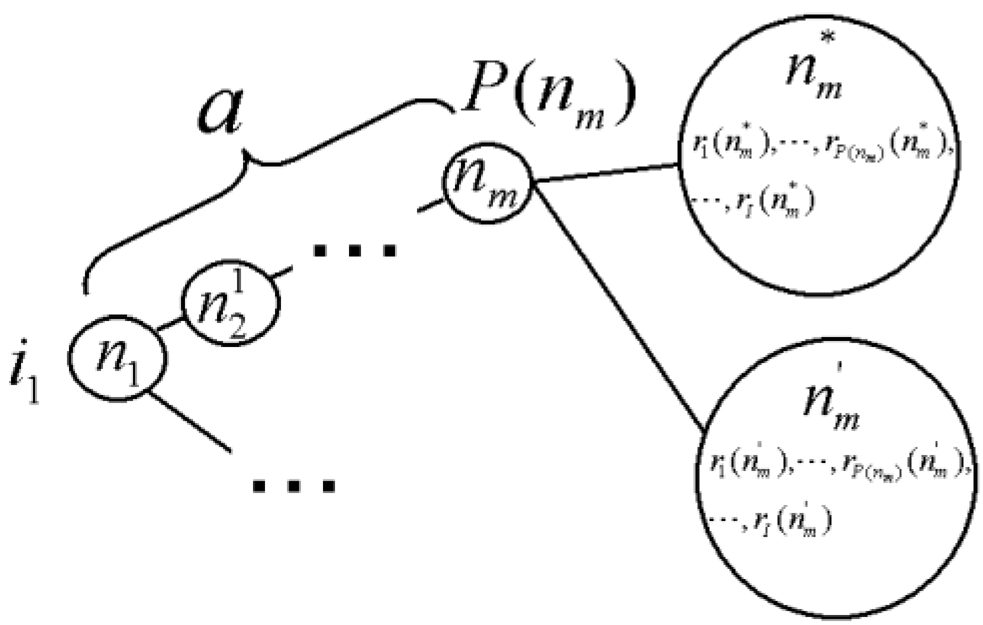

Fig.1. General reasoning rule

[1] Let i_1 be the player that chooses an action at the initial node n_1. i_1 tries to reason the estimated payoff vector at $n_1 \in N_2$ by backward induction.

[2] Indeed, i_1 tries to reason estimated payoff vector at node $n_m \in N_m^1$. Let a be the depth form the initial node to n_m. Let $b(n_m)$ be the difference between $r_{P(nm)\}(nm^*)}$ and $r_{P(nm)\}(nm')}$. i.e. $b(n_m')= r_{P(nm)\}(nm^*)} - r_{P(nm)\}(nm')}$, where $n_m' \in \{n|h(n)=n_m\}$

[3] $i_1$assigns $r_{(nm^*)}$ to estimated payoff vector $r_{(nm)}$, while it may occurs an error with a certain probability. We assume that the error probability is an increasing function of a and a decreasing function of b. If there are some best responses, each best action is taken with same probability.

[4] When the above operations have been finished for every $n_m \in N_m^1$, i_1 identifies every $n_m \in N_m^1$ with terminal nodes. Then i_1 generates N_m^2 as a set of last decision nodes of a new truncated game. Start to reason next reasoning process. This process is iterated until n_2. By this process, i_1 generates a payoff vector at n_2.

[5] Finally, i_1 compares the payoff vector of $n_2 \in N_m$ and chooses a best action. (This heuristics is an kind of backward induction with errors.)

[6] Let i_2 be a next player after i_1. Then i_2 reasons independently of reasoning of i_1 and chooses a best action for i_2.

[7] The players implement these processes until they reach a terminal node.

This process can be considered as a situation that all players are try to implement rational choices as possible as their reasoning abilities. The result is non-deterministic, we only take probability distribution over N_T. It is note that even if players' reasoning ability is not equal, they tries to reason by their reasoning abilities. Furthermore, if one player chooses actions more than once in the true game, reasoning at the subsequent nodes may contradict to that at the prior node. Our model can also describe such situations.

3. Rough reasoning model based on logit function:

First, we need the following notations:
Suppose that player i at node n_k reasons about the decision node n_l.

j : The decision player at node n_l.

N_s: A set of attainable nodes from n_l. i.e. $N_s=\{n| n \in N, h(n)=n_l\}$.

σ : A reasoning ability parameter.

We should notice that σ works as a fitting parameter with respect to the unit. For example, if description about payoffs change from dollar to cent, σ will be 1/100. Furthermore, if unit is fixed, as the rationality of agent is increased, σ will be increased.

Suppose that $n_{sx} \in N_s$, the rough reasoning model based on logit function with parameter σ, as follows:

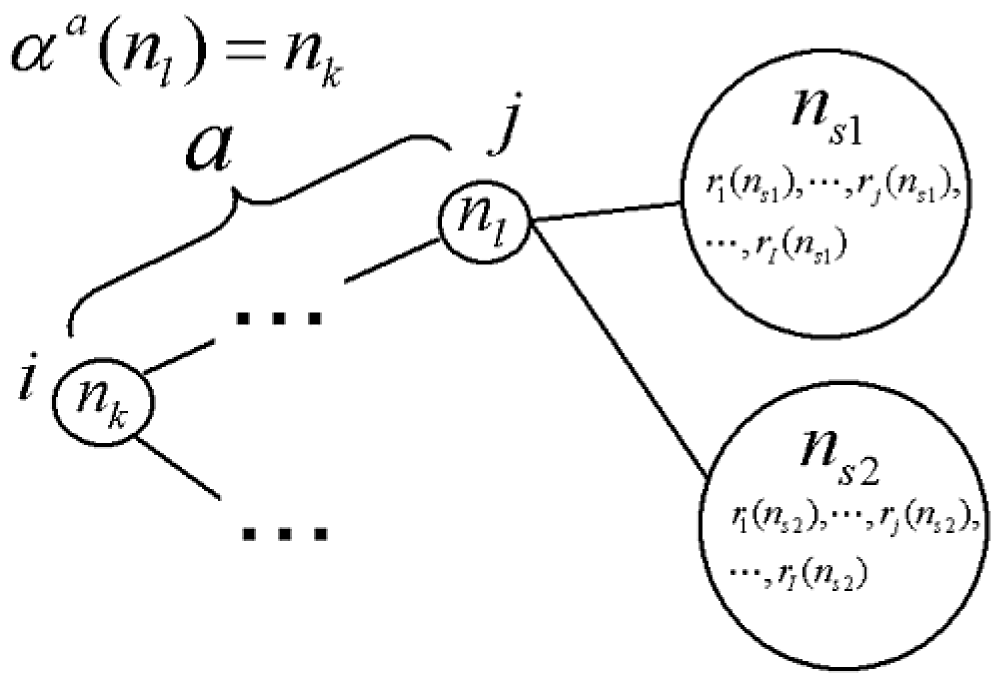

Fig.2 Rough reasoning model based on logit function.

Rough reasoning model based on logit function is a reasoning model that assigns $r(n_{s1})$ to $r(n_l)$ with probability

$$\frac{e^{\frac{r_{P(n_m)}(n_{m1})}{a(n_m)}\sigma}}{\sum e^{\frac{r_{p(n_m)}(n_{mj})}{a(n_m)}\sigma}} \quad (1)$$

The probability essentially depends on the ratio of payoff against *a* in such a way that if *a* is sufficiently large, then the choice can be identical with random choice. If *a* is sufficiently small and *b* is sufficiently large, the choice can be seem as by the best response.

Property 1
Rough reasoning model based on logit function satisfies translation invariance of payoff origin. i.e. If a payoff unit changes from x to $x+k$, given result distributions are equal.

Property 2
Rough reasoning model based on logit function satisfies transformation linearity of payoff scale and reasoning parameter.
i.e. If a payoff unit changes from x to kx and reasoning a parameter changes from from σ to σ/k , given result distributions are equal.

These two property shows the rough reasoning model based on logit function holds enough generality.

4. Simulation results and their implications in tournament zero-sum games

We then apply our model to the most basic zero-sum game situations where there are *2* players and each player has *2* alternatives in order to examine the validity of the model.

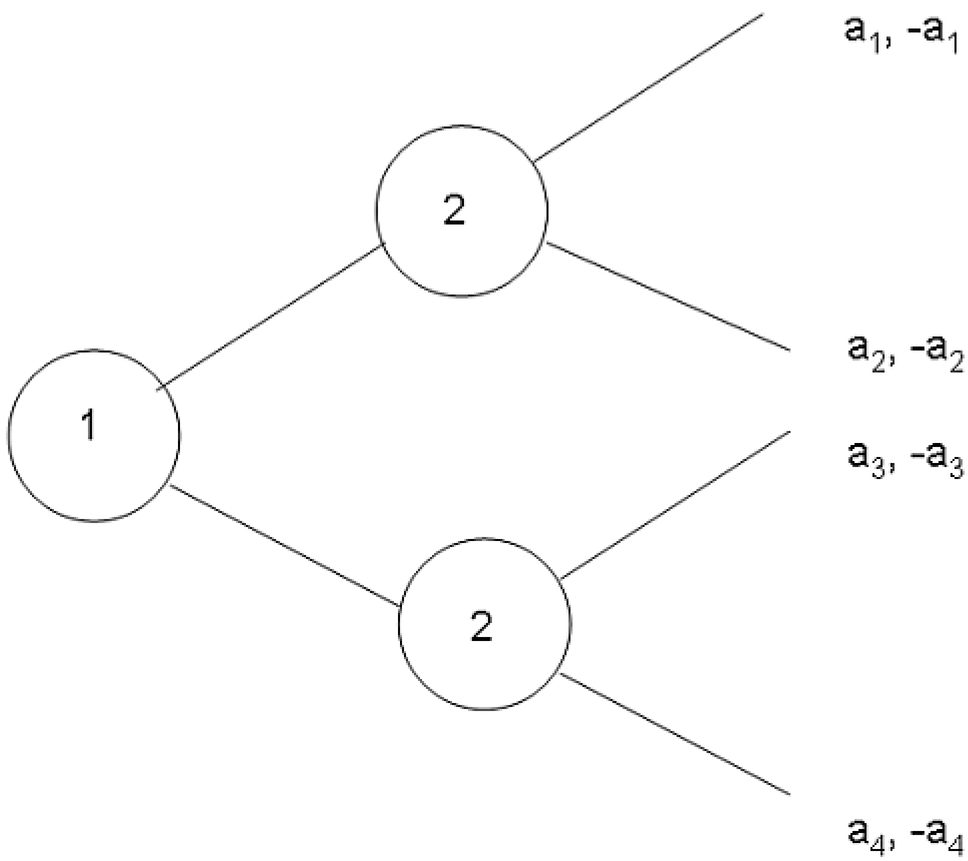

Fig.3 Basic zero-sum game

Fig. 3 is the most popular type of 2×2 extensive form game situations. We now sort these for payoff value and normalize the best one for player *2* is *1* and the worst one for player *2* is *0*. Let us assume that there are no tie and $1>b_1>b_2>0$. Then, we can classify in three cases by focusing magnitude relation.

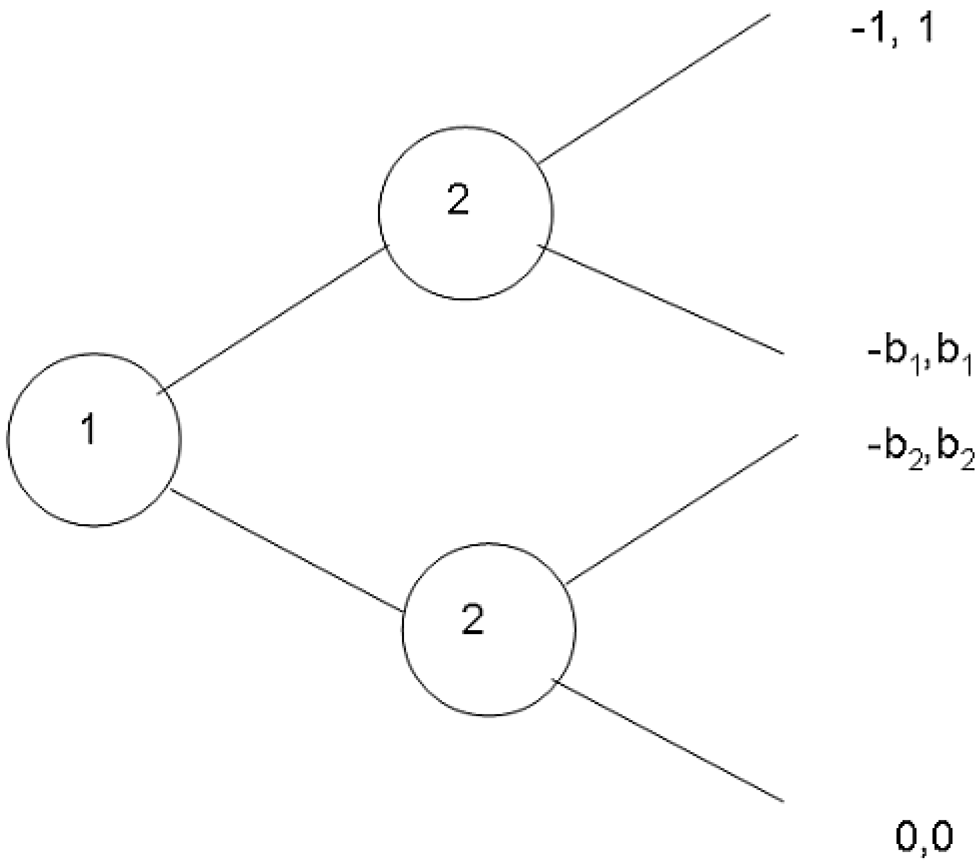

Fig.4 Basic zero-sum game case 1

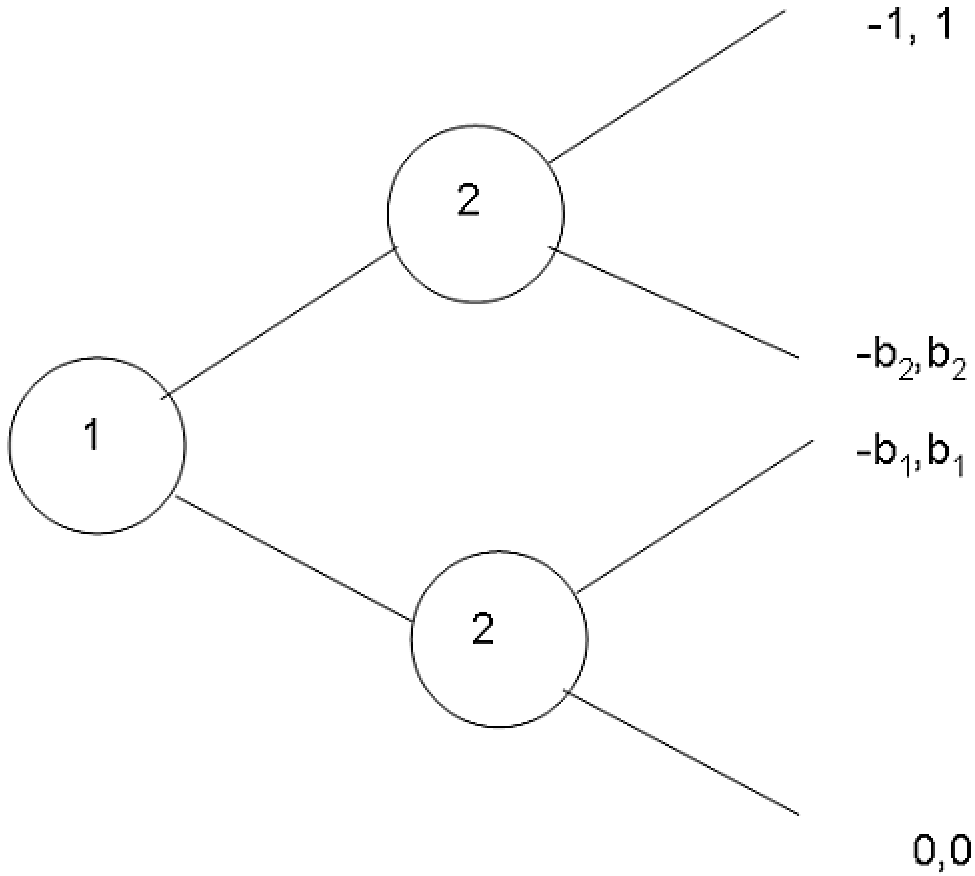

Fig.5 Basic zero-sum game case 2

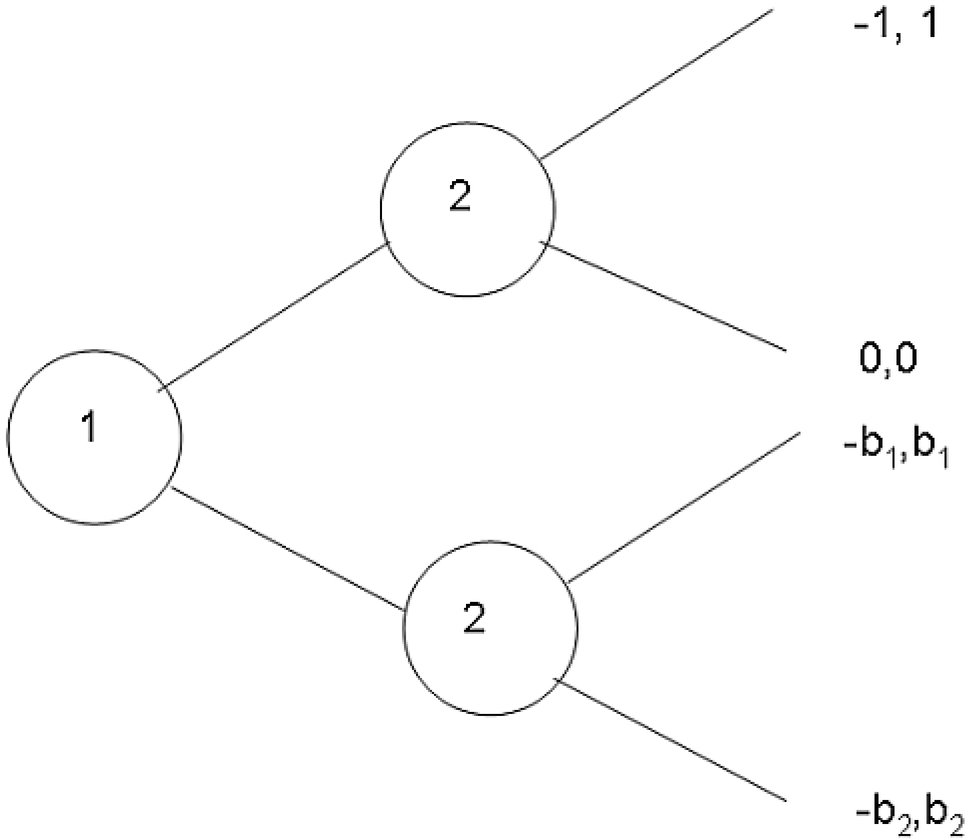

Fig.6 Basic zero-sum game case 3

In the first case, even if there are wrong reasoning, player *1* will highly evaluate the lower node than upper node. In this case, every kind of player 1's wrong reasoning does not change the results. In the second case, if player *1* perceives b_2 better than *1*(wrong reasoning) and b_1 better than 0 (correct reasoning), then player *1* chooses the lower altenatives. The probability is given by

$$\frac{e^{(b-1)\sigma}}{e^{(b-1)\sigma}+1} \bullet \frac{e^{a\sigma}}{e^{a\sigma}+1}$$

This is an increasing function of *a* and b.
Finally, in the third case, if player *1* percieves *0* better than *1*(wrong reasoning), player *1* chooses the lower altenatives.The probability is given by
1/(1+e), This is irrelevant to *a* and *b*.

These three analysis and 2 properties in previous section show that two statements. The first is as difference between the best and the worst payoff increase, the expected payoff of player 1 tend to decrease.The second is if the difference is constant, the best outcome is outstanding tend to trigger wrong choices.

5. Simulation results and their implications in centipede game

To examine systematic deviation from Nash equilibrium, we first focus on the Rosenthal's (Rosenthal, 1981) centipede game by using the model. Centipede game is well known as an example illustrating differences between results by backward induction and those by actual experiments.

The centipede game is two-person finite perfect information game. We call player *1* is "she", and player *2* is "he". Each player alternately chooses Pass *(P)* or Take *(T)* in each decision node. If she chooses action *P*, her payoff decreases while his payoff increases by more than his decrease. If she chooses action *T,* the game is over and they receive payoffs at that node. Symmetrically if he chooses action *P*, his payoff decreases while her payoff increases by more than his decreases. If the game has n decision nodes, we call the *n-move* centipede game.

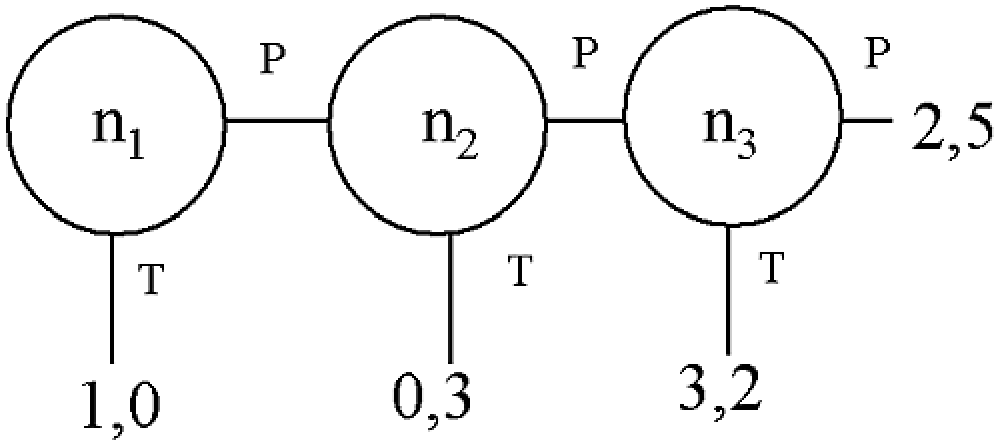

Fig.: 3-move centipede game

The pair of strategies that both the players choose *T* at every decision node is only subgame perfect equilibrium because the centipede game is finite. This equilibrium leads to the result that the game is over at the first period.

The centipede game has many variants about payoff structures. However we adopt the original Rosenthal's structure, where if she chooses *P*, her payoff is reduced by *1* and his payoff is increased by *3*.

This model represents a situation that players reasoning accuracy is determined by relative manner.

In order to examine frequency of cooperative behavior *P* with relation with *FCPF*, we calculated several simulations, where *FCPF* denotes frequencies of choice *P* at first period. We focus on the choice at the first period, because if *P* is chosen at the first period, the remained subgame can be considered as the *(n-1)-move* centipede game. Figures *4* and shows the simulation results of *FCPF* on the model.

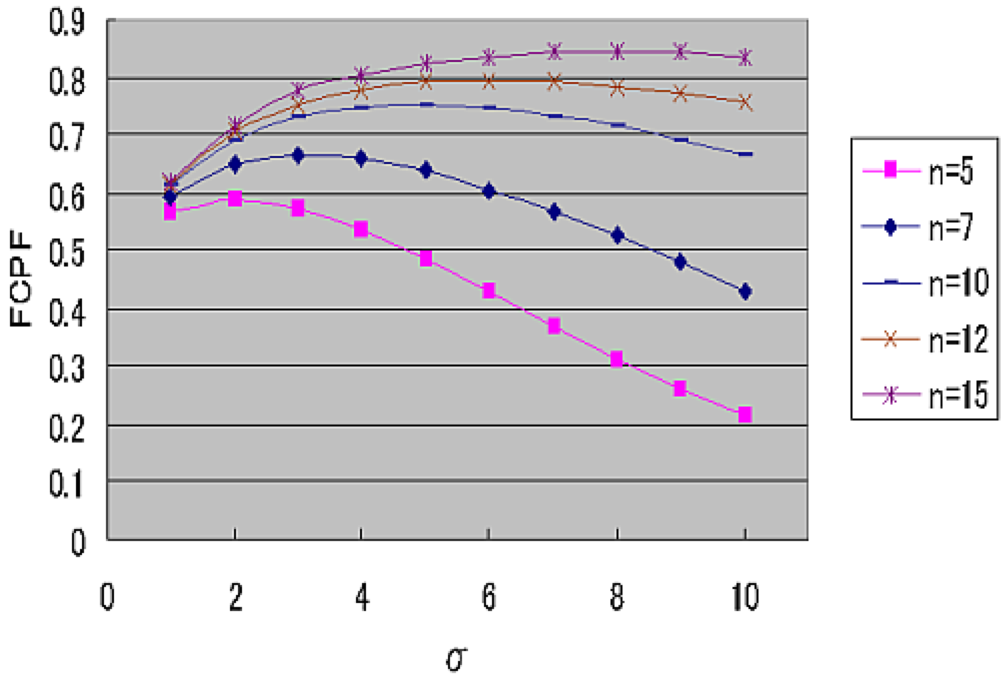

Fig.4 Relation between FCPF and reasoning ability

Note that in the Figure *4*, larger σ means more rational.

One implication is about relation between FCPF and the reasoning ability. For every n in the both models, it is observed that there is a turning point. Until the turning point, as the rationality is increased, cooperative behavior P tends to increase. However if reasoning ability exceeds the turning point, as the rationality is increased, cooperative behavior P tends to decrease.

Figures *4* gives following implications about the relation between FCPF and the reasoning ability: Moderate rationality may maximize the frequency of irrational actions. This property can be showed , so that, we think that this property holds that if human reasoning abilities proportional to the exponential rate of the depth of the tree.

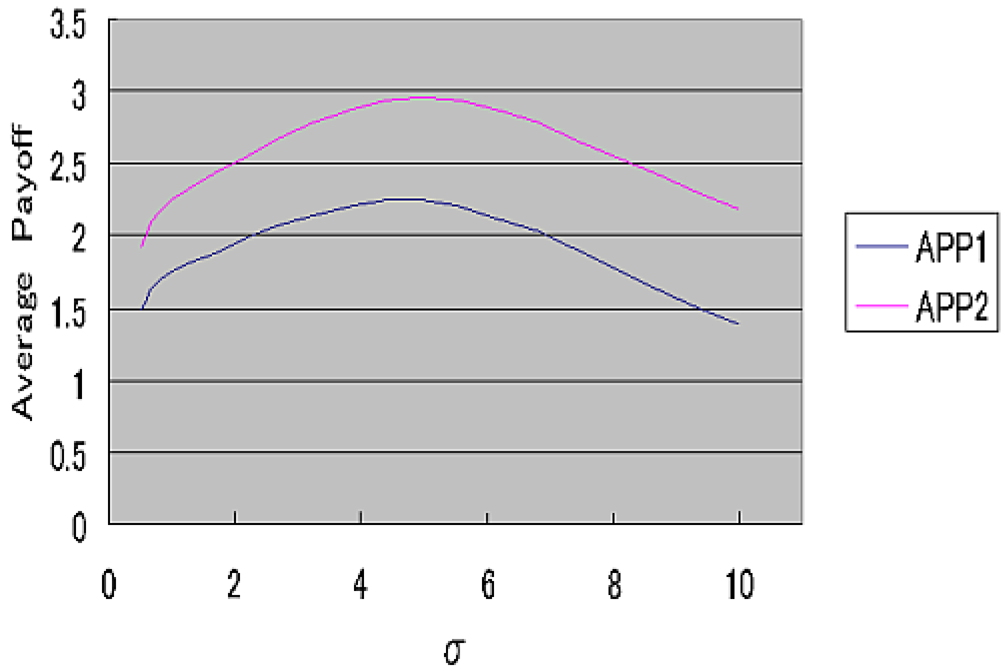

Fig.5 APP1 in the logit function model.

The other implication is concerning relation between social welfare and the reasoning ability. Figure *5* shows that the relation among the *APP1* and the *APP2* and reasoning ability, where *APP1* is the average payoff of player *1* and *APP2* is that of player *2*. Figure *5* is the result of *10-move* centipede game with logit function model, The case of σ=5 in the logit model, in other words, moderate rational society may maximize the social welfare.

This result implies that the centipede game can be considered as a kind of situation that cooperation is desired. Since cooperative behavior is not always increase their payoffs, Pareto efficiency is not guaranteed. To implement Pareto optimal results with certainly, we need to introduce a certain penalty system. However, since introduction of such a penalty system inevitably requires social cost, it does not always increase social welfare in the real world. These arguments indicate severe penalty system may not required to implement cooperative state in the real situations. In addition, repetition of cooperative actions may generate a kind of moral so that the players may perceives the centipede game as if it were a game which the cooperative actions are equilibrium strategies.

6.Conclusions and further remarks

The main contributions of this paper are as follows: First, we proposed a dynamic mathematical models expressing rough reasoning. Reasoning ability is defined as dependent not only on the payoff but also on the depth of decision tree. Second, we first examine the most basic extensive form game situation from the rough reasoning viewpoint and give some insights. In this game, good argorithm for computer players also work well for human palyers. Futhermore, improvement of the reasoning accuracy tend to increase their expected payoff. We next apply to the centipede game and give a new interpretation of our intuition in centipede game. We pointed out some implications of rough reasoning from two sides, frequency of rational action and social welfare. This means especially non zero sum game situstions, good argorithm for computer sometimes work worse. Furtheremore, improvement of the reasoning accuracy may decrease their expected payoff.

In this paper, we only discussed cases where each of players is equally rational. It was shown that the increase of agent's rationality is not necessarily connected with the rise of social welfare. It is future task to analyze what strategy is stabilized from an evolutional viewpoint by assuming a social situation is repeated.

References

(1)R., Aumann: Correlated Equilbrium as an Expression of Beysian Rationality. Econometrica **55** (1992) 1-18.
(2)R., Aumann: On the Centipede game: Note. Games and Economic Behavior **23** (1998) 97-105.
(3)R., Myerson: Refinements of the Nash Equilibrium Concept. Int. Journal of Game Theory **7** (1978) 73-80.
(4)R., Mckelvey, T., Palfrey: An Experimental Study of the Centipede Game. Econometrica **60** (1992) 803-836.
(5)R., Mckelvey, T., Palfrey: Quantal Response Equilibria for Normal Form Games. Games and Economic Behavior **10** (1995) 6-38.
(6)R., Mckelvey, T., Palfrey: Quantal Response Equilibria in Extensive Form Games. Experimental Economics **1** (1998) 9-41.
(7)R., Rosenthal: Games of Perfect Information, Predatory Pricing and the Chain-Store Paradox. Journal of Economic Theory **25** (1981) 92-100.
(8)R., Selten: The Chain-Store Paradox. Theory and Decision **9** (1978) 127-159.
(9)A., Heifetz, A., Pauzner: Backward Induction with Players Who Doubt Others' Faultlessness. Mathematical Social Science **50** (2005) 252-267.
(10)D.,Kreps, R.,Wilson, : Sequential Equilibria, Econometrica **50** (1982) 863-894.
(11)D.,Kreps, R.,Wilson, : Reputation and Imperfect Imformation, Journal of economic Theory **27** (1982) 253-79.