Experimental Social Sciences

Toward Experimentally-based New Social Sciences for the 21st Century

ExpSS21

Working Paper

ExpSS-E-7

Minimax Play by Teams

Yoshitaka Okano
Institute of Social and Economic Research, Osaka University

July, 2008

Grant-in-Aid for Scientific Research on Priority Areas

Minimax Play by Teams*

Yoshitaka Okano†

Institute of Social and Economic Research, Osaka University

First version: November 13, 2006
Current version: June 12, 2008

Abstract

The behaviors of two-person teams and individuals who repeatedly play O'Neill's (1987) game in the laboratory are analyzed. Teams are statistically more rational than individuals in the sense that teams' behaviors are more consistent with the implications of the minimax play than individuals' behaviors are. Furthermore, we investigate the source of difference in rationality between teams and individuals. Individuals seem to have the limitation of cognitive ability to recognize and to avoid their own unprofitable behavior. On the other hand, teams can improve such limitation, and behave more rationally.

Key Words: minimax, team versus individual decision making, experiment
JEL Classifications: C72, C92

1 Introduction

In many real-life situations, decisions are made by groups or teams with two or more individuals, such as families, boards of directors, legislatures or committees. Similarly, political decisions, monetary policy decisions and some business decisions are often taken by teams. Also, there are many economic and strategic situations where individual's decisions are reached through the reference to the expertise of, for instance, lawyers or consultants. These decisions are also considered to be made

*I am grateful to Michihiro Kandori for his helpful discussions and comments. I also thank Hitoshi Matsushima, Akihiko Matsui, Tatsuyoshi Saijo and the participants at Design Science Young Workshop 2007 in Osaka and the 10th Experimental Economics Conference in Tokyo for helpful comments. Financial support from Center for Advanced Research in Finance at the University of Tokyo is gratefully acknowledged. Author is responsible for any errors that may occur.

†email: y-okano@iser.osaka-u.ac.jp

by teams. Thus, team's decision making plays an important role in many economic and game theoretic environments.

This paper reports the results of experiment in which common-purpose freely-discussing minimum-size (two-person) teams and individuals repeatedly play the two-person zero-sum game developed by O'Neill (1987). The results of experiments are examined not only in terms of O'Neill's original experiment but also of many other experiments concerning the games with a unique mixed strategy equilibrium, for example, those in Binmore, Swierzbinski and Proulx (2001), Shachat (2002), Palacios-Huerta and Volij (2008) and so on.

In much of economic theory, game theory, and most experimental investigations of these theories, there is no distinction between decisions by teams and individuals. Only very recently, the importance of whether the decision maker is a team or an individual caught attention. All studies in the past are done by the framework of experimental economics with a variety of decision tasks and games such as dictator game in Cason and Mui (1997), ultimatum game in Bornstein and Yaniv (1998), investment game in Cox (2002), centipede game in Bornstein, Kugler and Ziegelmeyer (2004), monetary policy in Blinder and Morgan (2005), limit pricing game in Cooper and Kagel (2005), beauty contest game in Kocher and Sutter (2005), common value auction in Cox and Hayne (2006) and so on. In general, teams do the same or somewhat better (more rational) than individuals in experimental economics with the possible exceptions of Cason and Mui (1997) and Cox and Hayne (2006)[1].

One of the interests for experimental economics researchers is whether the subjects in the experiment behave rationally or how close their behaviors are to the equilibrium predictions. When we think the "rationality" deeply, in particular from the viewpoint of subjects, required abilities to behave "rationally", that is, to behave in line with game-theoretic predictions differ depending on the game played. For example, in a beauty contest game, players are required to apply greater depth of reasoning. In a dictator game and/or an ultimatum game, there is a conflict between the self-interest and fairness. This indicates that even though we could find the difference in rationality between teams and individuals in one experiment, that result may not be applicable to other experiments.

In the games with a unique mixed strategy equilibrium, players are required to choose their actions that are i.i.d. drawings from a certain stationary multinomial distribution over actions in order to behave rationally. In the psycological literatures, it is known that human beings are not good at the i.i.d. drawings from a stationaly distribution (see, for example, Wagenaar (1972) and Bar-Hillel and Wagenaar (1991) for the survey of subjective randomization in psychology). Human

[1]In Cason and Mui (1997), team choices tended to be dominated by the more other-regarding member. In Cox and Hayne (2006), one treatment revealed that teams tend to be less rational than individuals in the sense that their bids fell prey to the winner's curse. Cooper and Kagel (2005) provides the detailed review.

produced sequences have too few symmetries and long runs, too many alternations among events, and too much balancing of event frequencies over relatively short regions (Lopes and Oden, 1987). Therefore, randomizing the own actions is a difficult and specialized ability for human beings. Our experiment can provide the new evidence of the difference in rationality between teams and individuals in such environment.

Existing laboratory experiments of the games with a unique mixed strategy equilibrium found the significant evidence that the behaviors of subjects are quite inconsistent with the minimax theory, especially at each player level (O'Neill, 1997; Brown and Rosenthal, 1990; Rapoport and Boebel, 1992; Rapoport and Budescu, 1992; Mookherjee and Sopher, 1994; Binmore, Swierzbinski and Proulx, 2001; Walker and Wooders ,2001; Shachat, 2002; Rosenthal, Shachat and Walker, 2003; Palacios-Huerta and Volij, 2008). Main findings in these literatures are: 1) the aggregated data from the experiment was close to some of implications of minimax solution, 2) there is substantially more variation in choice frequencies across subjects than the theory predicts, 3) subjects do not equate the winning rates across strategies and 4) there exists a significant amount of serial correlation in the players' choices across time.

Our first research question is to investigate whether teams are more rational than individuals under the game with a unique mixed strategy equilibrium. There are two types of treatments in our experiment. The individual treatment is that individuals interact with individuals. The team treatment is that teams of two subjects each interact with other teams in which each subject are arrowed to discuss freely with the teammate. In both treatments, subjects were asked to play O'Neill's (1987) game repeatedly. Our experimental results find that teams are more rational than individuals in the sense that the more hypotheses under the minimax play are not rejected in the team treatment compared with the case of the individual treatment: 1) the aggregated data in both team and individual treatments are close to implications of the minimax solution though some statistical tests reject the minimax hypothesis, 2) in the team treatment, the choice frequencies for each decision-maker level are consistently close to the prediction of the minimax theory, while in the individual treatment they are deviated further from the prediction, 3) the winning rates across strategies are identical in the team treatment, which is consistent with the implication of the minimax play, but not in the individual treatment, and 4) the subjects' choices are serially correlated in both treatments.

These findings may give one explanation for the source of a part of rationality of professional sports players empirically observed in the field. The decisions of professional sports players are indirectly, but highly influenced or sometimes determined by analyses, judgements or decisions of teams. Tennis players receive advises from their coaches in their training and reflect those advises on their behaviors in matches, incorporating their own judgments. Team staffs of soccer clubs collect the data of opponents' players and analyse them, so soccer players can utilize the results of those analysis in their matches. Hence, a part of the skills and experiences of professional sports players would be achieved through analyses, judgements or decisions of teams.

Walker and Wooders (2001) found that the winning rates of professional tennis players are the same across directions of serve, though their choices are serially correlated. Hsu, Huang and Tang (2007) extended Walker and Wooders (2001) by analyzing a broader tennis data set and report that not only are the winning rates are identical across strategies, but also the choices are serially independent. Palacios-Huerta (2003) found that the winning rates of professional soccer players in penalty kicks are the same across directions of kicks for each kicker and direction of jumps for each goalkeeper, and player's choices are serially independent. Furthermore, Palacios-Huerta and Volij (2008) found that the professional soccer players behave consistently with the equilibrium play in the laboratory, indicating that field skills and years of experience can be transferred into unfamiliar laboratory setting.

Our analysis actually found that teams can equate their winning rates across strategies, as observed in the field. This evidence is consistent with the idea that team's decision makings affect the skills and experiences of professional sports players. However, the choices of teams are serially correlated, which is different from the ones observed in the field[2]. This indicates that the source of rationality of professional sports players cannot perfectly explained by the team's decision making. The skills of professional players for randomizing the own action over time may be achieved by the other sources.

Given the finding that teams are more rational than individuals, our second research question is to investigate the source of difference in behavior between teams and individuals, which leads to the difference in rationality between teams and individuals. First, examining the serial correlation in the outcomes of the game reveals that a particular outcome (the outcome of (J, J)) tends to continue successively in the individual treatment, while the team treatment does not have such tendency. This indicates that a particular player role (player B) in the individual treatment loses successively by playing a particular strategy (*Joker*), which leads to the bias of the winning rates across strategies. Furthermore, this bias seems to affect the relative frequencies of choice. Second, there are more individuals who adopt the non-equilibrium equiprobable play than teams. This leads to the bias of winning rates across strategies of their opponents in the individual treatment.

These results indicate that individuals have the limitation of cognitive ability to detect more correct or profitable actions. Psychological experiments found that teams generally perform better than individuals on the intellective tasks which have a correct action (or actions) and ex-post evaluation criterion for the quality of per-

[2]Palacios-Huerta and Volij (2008) found that the choices of professional soccer player are also serially independent in the laboratory setting.

formance (see, for example, Davis (1992) and Kerr, MacCoun and Kramer (1996) for a survey of general decision tasks). Our findings are consistent with these existing psychological literatures. Individuals have the limitation of cognitive ability to recognize and avoid their own incorrect (unprofitable) behavior. On the other hand, teams, which is only an aggregation of individuals with communication, can improve the behavior than the case where each member works independently. This indicates that teams can overcome some of the limitations of cognitive ability of individuals. Also, the results indicate that teams behavior is consistent with the underling logic of the minimax argument such that one needs to randomize strategies in order to prevent exploitation by one's opponent. Our experimental results suggest that the theoretical concept of equilibrium may have greater predictive power than previously considered.

The paper is organized as follows. Section 2 describes our experimental design, which allows us to analyse the behaviors of teams and individuals. Section 3 provides the results of a variety of statistical tests for the implications of minimax model. There, we will find that teams are more rational than individuals in the sense that the more implications of minimax hypothesis are not rejected in the team treatment than in the individual treatment. Section 4 investigates the source of the general difference in behaviors between teams and individuals, which explains the difference in rationality between teams and individuals. Furthermore, we argue how the found limitation of the cognitive ability to detect the more correct or profitable action for individuals relates to the existing psychological literatures on the intellective tasks. Section 5 concludes.

2 Experimental Design

There are two treatments in our experiment: one is to have individuals interact with individuals (individual treatment), and another is to have teams of two subjects each interact with teams (team treatment). In both treatments, subjects were asked to play repeatedly O'Neill's (1987) game. Each decision makers (teams/individuals) were randomly assigned one of two player roles, player A and B. Players A and B choose one of four actions $\{Red, Brown, Purple, Green\}$ in each round[3]. The winner is determined according to the following rules:

> Player A wins if there is a match of *Green*s (two *Green*s played) or a mismatch of other choices (*Red-Brown* for example); hence, player B wins if there is a match of choices other than *Green* (*Purple-Purple* for example) or a mismatch of a *Green* (one *Green*, one other choice).

[3]In O'Neill's (1987) original design, subjects held four cards, *Ace*, *Deuce*, *Trey* and *Joker*, and selected one of those. This design produced *Ace* bias. We follow the suggestion of Shachat (2002) and use color labels in order to avoid this bias.

This game rule was presented on the screen of a computer terminal, and hence we did not use the payoff matrix framing. Subjects learned the game rules by practice. Table 1 describes the payoff structure of the game.

[Table 1 here]

'W' and 'L' denote a win and a loss for player A respectively. The stage and the repeated games have a unique equilibrium which requires both players to randomize their actions *Red*, *Brown*, *Purple* and *Green* with probabilities 0.2, 0.2, 0.2 and 0.4 respectively in each round. The value of the game that corresponds to the probability of a win for player A in equilibrium is equal to 0.4.

Experimental sessions lasted about two hours and proceeded as follows. At the beginning of each session, all participants were gathered in one room. Participants received written instructions, which were read aloud, offering participants the opportunity to ask private questions. After having finished reading instructions, participants were separated according to player roles (player A or B) and moved to two distinct rooms. Hence, the opponent of each subject was not in the same room.

Subjects in the individual treatment were isolated from each other and were not allowed to communicate, whereas each subject in the team treatment was seated with his or her teammate at one computer terminal, was allowed to discuss face-to-face and each team was demanded to reach a single decision in each round. The minimum distance to the next team (computer terminal) was about 3 meters. Team members were requested to speak with a low voice and were strictly forbidden to speak to members of other teams. Assignment to an experimental treatment (individual or team) and to a specific team in the team treatment was random.

Experiment was conducted on computer terminals. Decision makers were anonymously matched with a fixed opponent. They played 15 rounds for practice and then 150 rounds for real money. The winners of each round in the individual treatment were paid 50 yen that is about 43 cents using the exchange rate at the time of experiment. Winning teams were paid 100 yen, which was divided equally across team members. Hence, we kept the per-subject monetary incentives constant across the individual and team treatments. The loser of each round got nothing in both treatments. Subjects were offered 1000 yen as a show-up fee.

The experiment was programmed and conducted with software z-Tree (Fischbacher, 1999). The screen display of the decision making stage was as follows. At the top, the number of the current round and the remaining time of the current round were shown. Decision makers were requested to come up with a decision within 30 seconds. In case they had not entered their decision after 30 seconds, a red sign would pop up on the screen which asked them to reach a decision soon. At the middle-left, game rules described earlier were displayed. At the middle-right of the screen, there were four radio buttons labeled *Red*, *Brown*, *Purple* and *Green* and decision makers were to choose their action by clicking one of buttons and then

‘OK’ button. At the bottom, history information was shown. They included the round number, the decision maker’s own choice, the opponent’s choice and the outcome of the stage game (‘win’ or ‘loss’). After the decision maker chose the action, the outcome was displayed, which included the decision maker’s own choice, the opponent’s choice and the outcome of the game in that round.

There were two experimental sessions in each treatment. All experimental sessions were conducted in August 2006 at the University of Tokyo. In total, 112 undergraduate and graduate non-economics students participated in this experiment. Of 112 students, 36 students were in the individual treatment and 76 students (38 teams) were in the team treatment. Hence we have 18 and 19 play pairs in the individual and team treatments, respectively. None of the subjects participated in more than one experimental session.

In one experimental session of the team treatment, one subjects had left the lab in midstream of the experiment for personal reason. After that, only teammate of this subject played against team. In data analysis, I decided to omit this play pair in order to investigate pure team-versus-team behavior. Hence, we also have 18 play pairs in the team treatment in data analysis.

Furthermore, In one experimental session, when subjects were making a decision at the 133rd round, there was a computer trouble and the experiment stopped for about 20 minutes. When recovered, history information was not correctly displayed on the screen, which confused the subjects. So, I decided to use the data of the first 132 rounds in data analysis which were not affected by the computer trouble. Since data analysis for each decision maker level is one of the main concerns in experiments with a mixed strategy equilibrium, I also use the data of the first 132 rounds in other experimental sessions in data analysis so as to keep the consistency in experiments. Hence, we have in total 4752 choices (132 rounds times 36 decision makers) in each treatment.

3 Data Analysis

This section provides the results of data analysis which test whether the behaviors of team and individual are consistent with the implications of the minimax model. These analyses allow us to evaluate how different, in the sense of rationality, the behaviors of team and individual are. In what follows, actions will be referred to by the names used in O’Neill’s experiment in order to avoid confusion and to facilitate comparison with the literature: J (*Joker*) for *Green*, 1 for *Red*, 2 for *Brown* and 3 for *Purple*.

3.1 Aggregated Data

Tables 2 and 3 present aggregate statistics describing the outcomes of the experiment.

[Tables 2 and 3 here]

Each interior cell in the top panel (panel A) reports the relative frequency with which the outcome of the game corresponding to that cell occurred. To the right and at the bottom, marginal frequencies with which players were observed to play their choices were displayed. Each number in parenthesis is the expected relative frequency of the corresponding cell predicted under the minimax hypothesis. In bracket of each cell, standard deviation for the observed relative frequency under the minimax model is shown. The bottom panel (panel B) reports the observed and the minimax win percentages for player A and standard deviation for the observed win percentage for player A under the minimax model. Since our experimental data for analysis in each treatment involve 132 joint moves for each of 18 pairs of decision makers, there were 2376 joint moves altogether.

One can see the rough adherence of the experimental data to the equilibrium predictions in both treatments. Indeed, all of the joint moves are within 1 to 2 percentage points from the predicted frequencies. Nevertheless, a chi-square goodness-of-fit test of the experimental data to the joint probability distribution implied by the minimax solution can reject the null hypothesis of minimax play at the 10 percent significant level in the team treatment. Test statistic is calculated to be 24.110, has 15 degrees of freedom, and has a p-value of 0.063. Hence, while close to the equilibrium predictions, joint moves in the team treatment are statistically different from predictions. The same test using data in the individual treatment, on the other hand, cannot reject the minimax hypothesis at the conventional significance level. Test statistic is 16.072 whose associated p-value is 0.377. However, we cannot reject the null hypothesis that the data from each treatment is generated by the same probability distribution at the conventional significance level. A chi-square test for homogeneity of two distributions produces a test statistic of 18.086 whose p-value is 0.258. Therefore, although the joint moves are statistically different from the minimax prediction in the team treatment but not in the individual treatment, the difference between them is statistically slight.

The marginal frequencies for each player are also impressively close to the equilibrium predictions in both treatments. For more detail analysis, table 4 reports the results of chi-square goodness-of-fit tests of marginal frequencies of each player and both players to equilibrium ones.

[Table 4 here]

The choice mixtures of player A show a statistical consistency with the minimax predictions in both treatments. P-values of the chi-square tests are 0.306 for the team treatment and 0.419 for the individual treatment. Player B, on the other hand, chooses their frequencies which, while close to the minimax prediction, are statistically different from it. P-values are 0.051 for the team treatment and 0.041 for the individual treatment, which reject the minimax hypothesis at the 10 percent level in both treatments. This leads to the inconsistency of pair-level choice mixtures with equilibrium predictions. P-values of the chi-square tests for both players are 0.077 for the team treatment and 0.086 for the individual treatment, which again reject the minimax hypothesis at the 10 percent level in both treatments.

Another feature of the aggregate data that is strikingly close to the implications of minimax play is the win percentages of player A. In both treatments, the aggregate win percentage of player A is less than one standard deviation away from the theoretically expected value, which is 0.4.

Summarizing, the aggregated data do not show the clear difference in the sense of rationality between team and individual. They are roughly close to the minimax predictions though some of them are statistically different from the minimax predictions. This result is also observed in the existing literatures: that is, although the aggregated data seem to be close to equilibrium predictions, the statistical tests reject the minimax hypothesis[4]. Thus, our two subject pools replicated the results of the existing experiments at the aggregate level.

3.2 Each Decision Maker Level Data

We next address how the choice mixtures of each team and individual conform well to the equilibrium proportions. Tables 5 and 6 report the relative frequencies of choices for each of 18 pairs and the results of the chi-square goodness-of-fit tests of the minimax models conducted for each team and individual.

[Tables 5 and 6 here]

Last three columns of these tables report the p-values of the chi-square goodness-of-fit tests of the minimax multinomial model for each player role and for both players: that is, the null hypothesis is that each decision maker chooses (or both decision makers in the pair choose) J, 1, 2 and 3 with probabilities of 0.4, 0.2, 0.2 and 0.2, respectively. ** and * to the right of the relative frequency for a given choice denote the rejections of the chi-square goodness-of-fit tests of the minimax binomial models for the corresponding choice at the 5 and 10 percent significance level, respectively: that is, the null hypothesis for the test of *Joker* choice is that the decision maker chooses *Joker* and non *Joker* with probabilities of 0.4 and 0.6, and for the test of the

[4]See, for example, Brown and Rosenthal (1990) and Shachat(2002).

number choice, he/she chooses that number and the other choices with probabilities of 0.2 and 0.8.

In the team treatment (table 5), chi-square goodness-of-fit tests of minimax multinomial model for player A, B and both players obtain 1, 2 and 1 rejections at the 5 percent level and 2, 4 and 2 rejections at the 10 percent level, respectively. Note that with 18 decision makers of each player role and hence 18 pairs, minimax play would expect 0.9 and 1.8 rejections at the 5 percent and 10 percent significance level in the tests for each player role and pair. Hence, the numbers of rejections in the team treatment are almost precisely the ones to be expected according to the theory. We reject the minimax binomial model for a given choice for 7 players at the 5 percent level and 16 players at the 10 percent level, respectively. Again note that, with 18 decision makers of each player role with 4 choices, the minimax binomial model indicates that we should expect 7.2 and 14.4 rejections at the 5 and 10 percent significance level. So the numbers of the rejections are again almost precisely the ones to be expected according to the theory. Therefore, teams' behavior is quite consistent with the minimax play.

In the individual treatment (table 6), on the other hand, chi-square goodness-of-fit tests of minimax multinomial model for player A, B and both players obtain 5, 3 and 4 rejections at the 5 percent level and 6, 5 and 6 rejections at the 10 percent level, respectively. Chi-square goodness-of-fit tests of the minimax binomial model for a given choice obtain rejections for 18 players at the 5 percent level and 25 players at the 10 percent level, respectively. These are larger number of rejections than the ones expected under the minimax model. Therefore, in contrast to team treatment, individuals' behavior is inconsistent with the minimax play.

Our next task is to test the joint hypothesis that all 36 decision makers in each treatment plays equilibrium play. At this level, we test two kinds of the null hypotheses. The first null hypothesis is that the choices of all 36 decision makers adhere to the equilibrium proportions. We can test this hypothesis by the chi-square joint test using each of 36 chi-square goodness-of-fit tests for the minimax multinomial model. The test statistic is simply the sum of the test statistics of 36 chi-square goodness-of-fit tests for the multinomial model, which is approximately distributed as chi-square with 108 degrees of freedom under the minimax hypothesis. Note that, if the choice mixtures of all decision makers are precisely equal to the equilibrium mixtures, p-value of the chi-square joint test is equal to 1. We have the test statistics of 116.731 and 163.625 with associated p-values of 0.266 and 0.0004 in the team and the individual treatment. This indicates that actions of all teams are statistically close to equilibrium proportions while those of all individuals are not.

The second null hypothesis is that each of all choices is adequately scattered around the equilibrium proportions. Minimax play indicates that each action is a *random draw* from the multinomial distribution equal to mixed strategy equilibrium. Random draws mean that the realized actions of the minimax play are not always exactly equal to equilibrium proportions, though they are close to it. In other words, we must reject the minimax hypothesis, for example, for the experimental data in which the relative frequencies of *all* decision makers are exactly equal to the equilibrium proportions. Speaking on the basis of the chi-square goodness-of-fit test of the minimax multinomial model, minimax plays with 36 decision makers imply that 36 test statistics are realizations of 36 random draws from the chi-square distribution with 3 degrees of freedom. Equivalently, the p-values associated with the realized test statistics should be 36 draws from the uniform distribution $U[0,1]$ by the *probability-integral transformation*[5].

Figure 1 provides the simple visual comparisons of the empirical cumulative distribution function of 36 observed p-values with the uniform distribution for which the cumulative distribution function is the 45° line.

[Figure 1 here]

It is obvious that the empirical and the theoretically predicted distributions are strikingly close to one another in the team treatment. In the individual treatment, on the other hand, the empirical c.d.f. is skewed upward than the theoretically predicted distribution.

We formalize these comparisons of the two distributions by the Kolmogorov-Smirnov one-sample test, which allows us to test the hypothesis that an empirical c.d.f. of observed values is generated from a specified distribution. Test statistic of Kolmogorov-Smirnov test is $K_n = \sqrt{n}\sup_x |S_n(x) - F(x)|$ which has a known distribution, where $S_n(x)$ is the empirical c.d.f. with the sample size of n and $F(x)$ is a hypothesized distribution[6]. In performing the Kolmogorov-Smirnov test for our experimental data, $n = 36$ and the hypothesized c.d.f. is the uniform distribution, $F(x) = x$ for $x \in [0,1]$ and hence $K_{36} = \sqrt{36}\sup_{x\in[0,1]} |S_{36}(x) - x|$. This value is calculated as 0.754 and 1.437 with p-values of 0.621 and 0.032 in the team and the individual treatment, respectively. Therefore, we cannot reject the null hypothesis that the realized p-values are generated from the uniform distribution on $[0,1]$ in the team treatment, which is consistent with the minimax play. On the other hand, we can clearly reject this hypothesis in the individual treatment, because we have many rejections from the chi-square goodness-of-fit test of the minimax multinomial model at the low significance level, which leads to the distribution of the observed p-values being skewed upward than the uniform distribution.

[5]The probability-integral transformation is the following property. Let X be a random variable with c.d.f. F_X. If F_X is continuous, the random variable Y produced by the transformation $Y = F_X(X)$ has the continuous uniform probability distribution over the interval [0,1].

[6]The empirical c.d.f. with the sample size of n, $S_n(x)$ is defined as

$$S_n(x) = \frac{\text{number of sample values } \leqslant x}{n}.$$

Summing up, teams are more rational than individuals based on the relative frequencies for each decision maker level.

Furthermore, it is worth noting that teams' behavior is quite consistent with the theory. Existing laboratory experiments rarely supported the minimax hypothesis for each decision maker level data. As in the individual treatment, there was substantially more variation in choice frequencies across subjects than the theory predicted (see O'Neill, 1987; Brown and Rosenthal, 1990; Rapoport and Boebel, 1992; Binmore, Swierzbinski and Proulx, 2001; Rosenthal, Shachat and Walker, 2003). In particular, of the experiments with many repetition of games (say, more than 100), hence higher power of the statistical tests, the only exception that I am aware of is Palacios-Huerta and Volij (2008) with the professional soccer players who are considered to be familiar with the situations where they are predicted to choose probabilistic mixtures. In our experiment, however, each team is an aggregation of two college students who are unfamiliar with such situations. Therefore, our finding has a clear distinction from the previously found evidences.

3.3 Testing for Equality of Winning Rates across Strategies

One of the implications of the minimax play is that the expected payoffs (win percentages in our case) are identical across strategies. Walker and Wooders (2001) tested this hypothesis for the data from O'Neill's (1987) experiment by the chi-square test for homogeneity of two distributions. We follow their analysis here.

Tables 7 and 8 report the results of the chi-square tests.

[Tables 7 and 8 here]

As in Walker and Wooders (2001), we aggregate actions 1, 2 and 3 into a single non-*Joker* action. In each of 36 chi-square tests, the null hypothesis is that *Joker*'s and non *Joker*'s winning probabilities are the same. For each of 36 decision makers, the two columns labeled "Test Statistic" and "p-value" at the right-hand side of the tables report the value of the test statistic along with its associated p-value.

The team treatment (table 7) obtains 3 rejections at the 5 percent level and 3 rejections at the 10 percent level, respectively. Note that with 36 teams, the expected number of rejections according to the theory is 1.8 rejections at the 5 percent level and 3.6 rejections at the 10 percent level. Hence, the numbers of rejections at the 5 percent and 10 percent level are almost equal to the expected numbers of rejections under the equilibrium play.

In the individual treatment (table 8), on the other hand, the hypothesis is rejected for 8 individuals at the 5 percent level and for 12 individuals at the 10 percent level, respectively. Thus, we have a larger number of rejections than expected under the minimax model.

We turn to the joint hypothesis that all observations are generated by equilibrium play. As in the previous subsection, we perform two kinds of tests. The first null hypothesis is that all 36 decision makers' winning probabilities are simultaneously the same across *Joker* and non-*Joker*. We use again the chi-square joint test. A test statistic is the sum of the test statistics of the chi-square test for the equality of winning rates, which is distributed as chi-square with 36 degrees of freedom under the null hypothesis. Those values are 36.095 and 78.288 with those associated p-values of 0.464 and 0.00006 for the team and the individual treatment, respectively. This indicates that, while the winning rates of all teams are simultaneously identical across *Joker* and non *Joker*, those of all individuals are not.

The second null hypothesis is that the differences between *Joker*'s and non-*Joker*'s winning probabilities for all decision makers are adequately scattered around zero[7]. In other words, it is the null hypothesis that the 36 realized p-values from the tests of equality of winning rates are generated from the uniform distribution on $[0, 1]$.

Figure 2 depicts the empirical and the hypothesized (i.e. uniform) distribution of the p-values for the tests on the equality of winning rates.

[Figure 2 here]

The empirical c.d.f. in the team treatment conform well with the uniform c.d.f., while that in the individual treatment is skewed upward than the uniform c.d.f.. In fact, test statistics of Kolmogorov-Smirnov one sample test for the empirical c.d.f. of p-values to the uniform c.d.f. are 0.857 and 1.715 with p-values of 0.455 and 0.006 for the team and the individual treatment. Therefore, the behaviors of all teams are consistently generated from the minimax play, while those of all individuals are not.

In summary, teams are again more rational than individuals based on the winning rates across strategies.

Furthermore, teams' behavior is quite consistent with the minimax theory. In the existing laboratory experiments with college students, the hypothesis that the winning rates are the same across strategies is often rejected for many subjects as in our individual treatment (see Walker and Wooders, 2001; Palacios-Huerta and Volij, 2008 with college students)[8]. Hence, teams behavior is quite different from the results of those experiments. On the other hand, professional sports players can equate the winning rates across strategies (see Walker and Wooders, 2001; Palacios-Huerta, 2003; Hsu, Huang and Tang, 2007; Palacios-Huerta and Volij, 2008 with professional soccer players). Our result is consistent with the idea that team's decision makings affect the skills and experiences of professional sports players.

[7]We must reject the minimax hypothesis, for example, for the experimental data in which winning rates of all decision makers are exactly the same across strategies.

[8]Palacios-Huerta and Volij (2008) found that professional soccer players consistently equate their winning rates.

3.4 Serial Independence Hypothesis

Another aspect of the implications of minimax play is the serial independence hypothesis. Equilibrium play indicates that a player should randomize his or her action using the same distribution at each stage of the game. This implies that player's choices should be serially independent. To address this question, we perform the runs test as in Walker and Wooders (2001). Tables 9 and 10 report the results of runs tests for our experimental data.

[Tables 9 and 10 here]

As before, we pool three number choices into a single non-*Joker* choice. For each decision maker, the column labeled "Runs r" reports the total number of runs in the action sequence of *Joker* and non-*Joker* in the order of their occurrence. A run is defined as a succession of one or more identical actions, which are preceded and followed by a different action or no action at all. Two columns labeled $F(r-1)$ and $F(r)$ report the conditional probability of obtaining $r-1$ and r or fewer runs under the null hypothesis of serial independence given the number of occurrence of *Joker* and *non-Joker*[9]. At the 5 percent level, for example, the null hypothesis of serial independence is rejected if either $F(r) < 0.025$ or $1 - F(r-1) < 0.025$.

In the team treatment, we obtain 5 rejections at the 5 percent level and 7 rejections at the 10 percent level, respectively. These numbers are larger number of rejections than those expected under the null hypothesis of serial independence. In the individual treatment, on the other hand, we have 3 and 5 rejections at the 5 and 10 percent level, which are slightly larger than expected.

For more detail analysis, we next address the joint hypothesis that the choices are serially independent for each of all 36 decision makers. Walker and Wooders (2001) point out that Kolmogorov-Smirnov test cannot be applied directly to the values in either column $F(r-1)$ or $F(r)$ since these values are neither identically distributed (the distribution of r depends on the numbers of occurrence of *Joker* and non-*Joker*) nor continuously distributed. They suggest that Kolmogorov-Smirnov test can be developed by constructing a random draw d from the uniform distribution $U[F(r-1), F(r)]$ for each decision maker, which is distributed as a $U[0,1]$ under the null hypothesis of serial independence. Figure 3 shows the empirical c.d.f. of a particular realization of 36 d statistics together with the hypothesized (i.e. uniform) distribution.

[Figure 3 here]

In both treatments, the empirical cumulative distribution functions are skewed upwards than the uniform distribution, though the degree of skewness is slightly higher

[9] For more detail functional form of $F(r)$, see, for example, Gibbons and Chakraborti (2003).

in the team treatment than in the individual treatment. After performing ten thousand trials with such random draws for each decision maker, the average p-values of Kolmogorov-Smirnov test for the empirical c.d.f. against the uniform distribution are 0.012 and 0.092 with a standard deviations of 0.005 and 0.027 in the team and the individual treatments, respectively. Thus, the joint hypothesis that each of 36 decision makers is serially independent can be rejected, on average, at the 10 percent level in both treatments.

Comparing these results, the choices are serially correlated in both the team and the individual treatments, though the correlation is slightly less in the individual treatment.

There are two ways that the number of runs can indicate the serial correlation, one is too few runs and another is too many runs. In the runs test, rejection by that $F(r)$ is less than the half of the significance level indicates too few runs, and hence positive serial correlation in choices. Rejection by that $1 - F(r-1)$ is less than the half of the significance level indicates too many runs, and hence negative serial correlation in choices. Tables 9 and 10 show that our two subject pools generally reveal positive serial correlation in choices. At the 10 percent level, 6 out of 7 rejections in the team treatment and all 5 rejections in the individual treatment owned to positive serial correlation.

Overwhelming cognitive psychological and experimental economics literatures found that, when people try to generate random sequence of actions, they generally switch the choices too often to be consistent with randomly generated choices, which is called the *negative recency effect* (see, for example, Wagenaar (1972) and Bar-Hillel and Wagenaar (1991) for the survey of literatures in psychology). This indicates that their sequences often have negative serial correlation. Hence, our experimental results are contrast to these existing literatures. However, some experimental economics literatures such as Rosenthal, Shachat and Walker (2003) and Slonim, Erev and Roth (2003) reported evidence of positive serial correlation[10]. Slonim, Erev and Roth (2003) speculates that the anonymity and computer interface in their design may lend itself to less alternation in choices while the face-to-face flipping of cards like O'Neill's (1987) design may lend itself to greater alternation. Our experimental design of the anonymity and computer interface may also accelerate the players' tendencies to repeat the same action.

In the last subsection (testing equal winning rates across strategies), we argue that a part of the skills and experiences of professional sports player might be achieved through the team's decision making. The result of testing serial independence hypothesis in the present subsection, however, is different from the ones observed in the field. This indicates that the source of rationality of professional sports players cannot be perfectly explained by the team's decision making. The

[10]One of the treatments in Mookherjee and Sopher (1994) also reveals positive serial correlation in subjects' choices.

professional players' skills for randomizing the own action over time may be achieved by other sources.

4 Source of Difference in Rationality between Teams and Individuals

Our experimental results have found that: 1) the aggregated data in both the team and the individual treatment are basically close to the ones predicted the minimax hypothesis, though some statistical tests reject it; 2) the choice frequencies for each decision maker level are consistently close to the prediction of the minimax theory in the team treatment, while those are far from it in the individual treatment; 3) the winning rates across strategies are identical in the team treatment, which is consistent with the implication of the minimax play, but it is not the case for the individual treatment; and 4) the subjects' choices are serially correlated in both treatments. These findings suggest that teams are more rational than individuals in the sense that the more implications are not rejected in the team treatment than in the individual treatment. This section attempts to detect the source of the general difference in behaviors between teams and individuals, and argues how this difference yields the rationality of teams and the irrationality of individuals.

4.1 Serial Correlation in the Outcomes of the Game

In order to detect the general difference between team and individual, we pool the data for each player role in each treatment. We focus on the serial correlation in the outcome (J, J). Investigating the serial correlation in the outcome (J, J) can let us detect the nature of bias in the winning rate of *Joker*. *Joker* is a special choice for both players in O'Neill's game, . Player A wins by choosing *Joker* only when player B chooses *Joker*, or equivalently, player B wins by choosing *Joker* only when player A does not choose *Joker*. Therefore, serial correlation of (J, J) causes the bias in the winning rate of *Joker*.

More formally, we estimate the following logit model for each treatment,

$$\begin{aligned}(JJ)_t &= G[\alpha + \beta(JJ)_{t-1} + \gamma(J1)_{t-1} + \delta(J2)_{t-1} + \epsilon(J3)_{t-1} + \\ &\qquad +\zeta(1J)_{t-1} + \eta(11)_{t-1} + \theta(12)_{t-1} + \iota(13)_{t-1} + \\ &\qquad +\kappa(2J)_{t-1} + \lambda(21)_{t-1} + \mu(22)_{t-1} + \nu(23)_{t-1} + \\ &\qquad +\xi(3J)_{t-1} + o(31)_{t-1} + \pi(32)_{t-1}],\end{aligned}$$

where the variable name (ik) denotes the dichotomous indicator of the outcome of the stage game $(i, k \in \{J, 1, 2, 3\})$ where i and k are player A's and B's choice, respectively, and the subscripts t and $t-1$ refer to the current and one-period previous outcomes. The function $G[x]$ denotes the function $e^x/(1+e^x)$. We omit the outcome $(33)_{t-1}$ in order to avoid the multicollinearity.

Table 11 reports the results of the estimation and the significance tests.

[Table 11 here]

In the team treatment, any coefficients are not rejected at the 10 percent or lower significance level. From this, we can conclude that the (J,J) occurs independently to any outcome in the previous period. In the individual treatment, on the other hand, (J,J) in the previous period has statistically significant effect with positive sign on occurring (J,J) in the current period. This indicates that the outcome (J,J) tends to occur successively. In other words, since player A wins and player B loses when the outcome is (J,J), player B tends to lose successively by choosing *Joker*[11].

This finding lead us to infer that the adjustment process of pairs in the individual treatment produces the bias in the winning rate of *Joker*. The direction of the bias is that

$$\text{win rate of } Joker \ > \ \text{win rate of non-}Joker \text{ for player } A, \text{ and} \tag{1}$$

$$\text{win rate of } Joker \ < \ \text{win rate of non-}Joker \text{ for player } B. \tag{2}$$

Let us go back to the results of testing for the equality of winning rates across strategies in the individual treatment as shown in table 8. Most rejections at the 5 and 10 percent levels were rejected by either (1) or (2); 6 out of 8 rejections at the 5 percent level (B of pair 2, A and B of pair 4, A of pair 7, B of pair 8 and A of pair 11) and 8 out of 12 rejections at the 10 percent level (in addition to the subjects listed earlier, B of pair 1 and B of pair 6). This is consistent with the inference from the estimation result.

Continuation of (J,J) indicates that when player B has lost by choosing *Joker*, he/she again tends to choose *Joker*, and his/her opponent (player A) detected that behavior. This is explained by the both players' belief about the excess confidence of player B on *Joker*. In O'Neill's game, *Joker* of player B and only this choice wins against the three choices of player A. So, player B feels that his/her *Joker* looks to be stronger. When he/she has lost the game by choosing *Joker*, he/she may believe that he/she will win the next period by choosing *Joker* because *Joker* is intrinsically stronger and player A will not choose *Joker* successively twice. However, player A

[11] We omitted the outcome $(33)_{t-1}$ in this estimation in order to avoid multicollinearity. We also performed the estimation with omitting one of the outcomes except $(JJ)_{t-1}$. In the team treatment, a few cases (4 out of 14 estimations) reveals that $(JJ)_{t-1}$ has an significant effect on $(JJ)_t$ with positive sign. In the individual treatment, on the other hand, such cases were 10 out of 14 estimations. Hence, the effect of $(JJ)_{t-1}$ on $(JJ)_t$ is a little in the team treatment while it is robust in the individual treatment.

also knows that player B thinks that his/her *Joker* is stronger. So, player A may also choose *Joker* for the next period in order to defeat his/her opponent.

If the argument above is true, the choice frequencies of such pairs may also be affected by the belief in the excess confidence of player B on *Joker*. In fact, comparing the table 8 with table 6, either a player or his/her opponent or both who are rejected at the 10 percent level in testing for equality of winning rates across strategies by either (1) or (2) are also necessarily rejected in the chi-square test of the minimax multinomial or binomial model at the 10 percent level (A of pair 1, A of pair 2, A and B of pair 4, A of pair 6 (rejected only in the binomial model of *Joker*), B of pair 7, A and B of pair 8 and B of pair 11). Furthermore, these subjects choose *Joker* relatively more frequently, such as more than 0.5 in some cases. This is consistent with the idea that these pairs have the belief that player B has excess confidence on *Joker*.

Summing up, there is a significant difference in the adjustment processes of pairs between team and individual: In the adjustment processes of pairs in the individual treatment, there is a tendency that subjects of player B lose successively by playing *Joker*, while the adjustment processes of pairs in the team treatment do not have such tendency. This produces the bias of winning rate of *Joker* in the individual treatment, while not in the team treatment. Most rejections in testing for equality of winning rates in the individual treatment can be explained by this bias, which may be produced by both players' belief that player B has excess confidence on own *Joker*. This bias also produces the inconsistency in the relative frequencies of either the subjects' or the opponents' or both players' choices.

4.2 Another Source of Difference: Equiprobable Play of the Opponent

The cause of most rejections in testing for equality of winning rates across strategies in the individual treatment can be explained by the finding that the outcome (J, J) tends to occur successively. However, this explanation cannot be applied to all rejections because a few (B of pair 12, B of pair 16 and A and B of pair 18) are rejected at the 5 or the 10 percent level by the bias that

$$\text{win rate of } Joker \quad < \quad \text{win rate of non-}Joker \text{ for player } A, \text{ and} \tag{3}$$

$$\text{win rate of } Joker \quad > \quad \text{win rate of non-}Joker \text{ for player } B. \tag{4}$$

Such subjects and only such subjects have in common that the opponents seem to choose each action probabilistically equally, that is, they choose each action withe probability of 0.25. Table 12 re-describes the relative frequencies of players' choices in the individual treatment, and at this time, reports the results of chi-square goodness-of-fit test of the multinomial and binomial equiprobable play.

[Table 12 here]

Last two columns report the p-values of the chi-square goodness-of-fit tests of the multinomial equiprobable play for each player role: that is, the null hypothesis is that the subject chooses J, 1, 2 and 3 with probability 0.25 each, respectively. ** and * to the right of the relative frequency for a given choice denote the rejections of the chi-square goodness-of-fit tests of the binomial equiprobable play at the 5 and 10 percent significance level, respectively: that is, the null hypothesis is that the subject chooses the given choice and the other choices with probabilities 0.25 and 0.75.

While the null hypotheses are clearly rejected for most subjects, player A of pair 12, player A of pair 16 and player A and B of pair 18 are not rejected neither in the test for the multinomial equiprobable play nor in the test for the binomial equiprobable play at the 10 percent level. Hence, the behaviors of these subjects are consistent with the equiprobable play. Furthermore, these subjects are perfectly equivalent to the opponents of the subjects who are rejected in testing for equality of winning rates across strategies at the 5 or 10 percent level by either (3) or (4).

If player B chooses the actions with equal probabilities, the winning probabilities of *Joker* and non-*Joker* for player A are 0.25 and 0.5, respectively. If player A chooses the actions with equal probabilities, the winning probabilities of *Joker* and non-*Joker* for player B are 0.75 and 0.5, respectively. Hence, (3) and (4) hold. In fact, the null hypothesis that the winning rate of *Joker* is equal to 0.25 and that of non-*Joker* is equal to 0.5 for player A or the null hypothesis that the winning rate of *Joker* is equal to 0.75 and that of non-*Joker* is equal to 0.5 for player B is not rejected at the 10 percent level for these subjects. P-values are 0.377 for player B of pair 12, 0.346 for player B of pair 16, 0.289 for player A of pair 18 and 0.289 for player B of pair 18, respectively[12]. Hence, we can conclude that the equiprobable play of the opponent causes the rejections in testing for equality of winning rates across strategies by either (3) or (4).

In the team treatment, only player B of pair 11 is rejected in testing for equality of winning rates across strategies by (4)[13]. This team is also affected by the equiprobable play of the opponent. Table 13 also reports the relative frequencies of teams' choices and the results of the chi-square goodness-of-fit test of the equiprobable play.

[Table 13 here]

Player A of pair 11 and only this player is not rejected neither in the test for the multinomial equiprobable play nor in the test for the binomial equiprobable play

[12]Note that test statistic is approximately distributed as chi-square with two degree of freedom under the null hypothesis.

[13]No one is rejected by (3) in the team treatment.

at the 10 percent level. Hence, the behavior of this player is consistent with the equiprobable play. Furthermore, the null hypothesis that the winning rate of *Joker* is equal to 0.75 and that of non-*Joker* is equal to 0.5 is not rejected at the 10 percent level for player B of pair 11 (p-value is 0.459). Hence, we can again conclude that the equiprobable play of player A of pair 11 causes the rejections for player B in testing for equality of winning rates across strategies by (4).

Most players who adopt the equiprobable play in our experiment are player A (4 out of 5 players). In O'Neill's game, *Joker* of player A wins only against *Joker* of player B. This gives player A the impression that his/her *Joker* is weaker, and so player A may hesitate to choose *Joker*. This consideration may reduce the frequency of player A choosing *Joker*, which, as a result, would lead to the equiprobable play.

These findings indicate another source of the difference between teams and individuals: that is, there are relatively larger number of decision makers who choose according to the equiprobable play in the individual treatment than in the team treatment (4 players in the individual treatment against 1 player in the team treatment). This fact leads to a larger number of rejections in testing for the equality of winning rates across strategies by either (3) or (4) in the individual treatment than in the team treatment.

4.3 Discussion: Teams' Superiority on Intellective Tasks in terms of Psychology

We have found the sources of the differences in behaviors between teams and individuals. It is that pairs of individuals follow the behavioral pattern by which player B has tendency to lose successively by playing *Joker*, while teams do not have such tendency. Another is that a larger number of individuals (mostly player A role) adopt the equiprobable play than teams. Individuals seem to have the limitation of cognitive ability to recognize and to avoid their own unprofitable behavior. This limitation leads individuals to the successive loss of a particular player and non-equilibrium equiprobable play. On the other hand, teams seem to be able to improve some of such limitations, and behave more rationally.

Why could not individuals of player B role perceive that they are detected of tendency to choose *Joker* just after they have lost by playing *Joker*? Why did individuals choose the irrational equiprobable play, which gives their opponent the opportunity to win the game more (relative to the equilibrium prediction)? Or conversely, why could teams, which are only an aggregation of two individuals with communication, avoid such behaviors? These questions may be answered by the findings from the psychological literatures that compare the decision makings between groups and individuals.

The conventional wisdom of group superiority has been challenged by numerous psychological experiment[14]. When we separate the types of decision tasks into judgmental tasks and intellective tasks, we can find some specific differences between groups and individuals. Judgmental tasks are settings where there is no correct action and so lacks a clear ex-post evaluation criterion for the quality of performance. They involve, for example, studies comparing the attitudes of groups and individuals toward risk. On judgmental tasks, there does not seem to be a systematic difference between groups and individuals in terms of performance. On the other hand, intellective tasks have a correct action (or actions) and clear ex-post evaluation criterion for the quality of performance. On intellective tasks, we can find that groups typically perform better than individuals, meaning that groups more often guess correctly than individuals, or groups are, on average, closer to the correct solution than individuals[15].

In our experiment, subjects are involved in the settings where they try to seek the own and the opponent's behavioral pattern, which allow them to defeat the opponent or make harder for opponent to predict their behavior. They can evaluate their performance from the stream of the outcome (either win or loss) as the experiment proceeds. Hence, we can interpret that our subjects work on the intellective tasks which endogenously occur through the interaction. Therefore, teams' superiority on the intellective tasks might work in such situation and lead teams to rational behaviors: that is, many individuals might not be able to perceive that a particular outcome continues or to detect that there are other profitable plays than the equiprobable play, while many teams could find and avoid those. It also indicates that teams behavior is consistent with the underlying logic of the minimax argument such that one needs to randomize strategies in order to prevent exploitation by one's opponent.

5 Concluding Remarks

This paper has compared the behaviors between teams and individuals who play the game with a unique mixed strategy equilibrium. The focus of this paper is on differences in rationality and its source between two-person teams versus individual subjects. Teams are statistically more rational than individuals in the sense that the more implications of the minimax play are not rejected in the team treatment than in the individual treatment. This finding may give one explanation for the source of a part of rationality for professional sports players empirically observed in the field.

Furthermore, we have found the source of the difference in rationality between teams and individuals. Individuals seem to have the limitation of cognitive ability to

[14] See Davis (1992) and Kerr, MacCoun and Kramer (1996) for the survey.

[15] In particular, on intellective tasks whose correctness of action are highly demonstrable to others, this is clearly the case, because in such situations, the group members can easily confirm the correctness of the action and are most likely to adopt the correct solution.

recognize and to avoid their own unprofitable behavior. This limitation lead individuals to the successive loss of a particular player and non-equilibrium equiprobable play. On the other hand, teams, each of which is only an aggregation of two individuals with communication, can improve such limitation, and behave more rationally.

However, team formation also have bound on their improving of team behavior. There was a significant serial correlation in teams' choice. The choices of professional sports player in the field are observed to be serially independent. Hence, achievement of such skill of randomization of professional sports players cannot be explained by team's decision making.

From the viewpoint of theory, the broader range of team's behaviors are very close to the equilibrium one, so the theoretical concept of equilibrium may have greater predictive power than previously considered.

As an implication of our experimental results to reality, there is a significant advantage in forming a team when players confront with the strictly competitive situations. Also, our research provides the reasonable reason for the fact that the decisions are made by teams in many real-life social and economic environments.

Our research should be extended in a variety of directions. In particular:

1. *Are teams more rational than individuals in other strategic environments?* Although, there is growing number of literatures comparing the difference in behavior between teams and individuals in experimental economics, economists still know too little on under which strategic environments teams can behave more rationally than individuals. As noted in the introduction, required abilities are different depending on the game played in order to behave rationally. Teams might not perform better than individuals in other games.

2. *What is the optimal size of teams in order to behave rationally?* We have examined the behavior of teams which is an aggregation of two individuals. It would be interesting to investigate whether increasing number of teammates promote the rationality of teams. The effects may include not only the capability of more information processing, but also increasing costs for reaching a decision.

3. *How do teams reach the decision?* In our experiment, the contents of the conversations within a team are not clear. It would be interesting to know what kind of conversations is exchanged within a team until they reach the decision. In particular, one of the interests in this context is whether there is a leadership in a team or not. It would be also interesting to investigate whether there is the difference in rationality between teams with and without leadership. Such investigation may clarify which configuration of a team is more appropriate.

References

[1] Bar-Hillel, Maya, and Willem A. Wagenaar. 1991. "The Perception of Randomness." *Advances in Applied Mathematics*, 12: 428-454.

[2] Binmore, Ken, Joe Swierzbinski, and Chris Proulx. 2001. "Does Minimax Work? An Experimental Study." *Economic Journal*, 111: 445-464.

[3] Blinder, Alan S., and John Morgan. 2005. "Are Two Heads Better Than One?: Monetary Policy by Committee." *Journal of Money, Credit and Banking*, 37(5): 789-811.

[4] Bornstein, Gary, Tamar Kugler, and Anthony Ziegelmeyer. 2004. "Individual and Group Decisions in the Centipede Game: Are Groups More "Rational" Players?" *Journal of Experimental Social Psychology*, 40: 599-605.

[5] Bornstein, Gary, and Ilan Yaniv. 1998. "Individual and Group Behavior in the Ultimatum Game: Are Groups More "Rational" Players?" *Experimental Economics*, 1: 101-108.

[6] Brown, James N., and Robert W. Rosenthal. 1990. "Testing the Minimax Hypothesis: A Re-examination of O'Neill's Game Experiment." *Econometrica*, 58(5): 1065-1081.

[7] Cason, Timothy N., and Vai-Lam Mui. 1997. "A Laboratory Study of Group Polarisation in the Team Dictator Game." *Economic Journal*, 107: 1465-1483.

[8] Chiappori, Pierre-Andre, Steven D. Levitt, and Timothy Groseclose. 2002. "Testing Mixed-Strategy Equilibria When Players Are Heterogeneous: The Case of Penalty Kicks in Soccer." *American Economic Review*, 92(4): 1138-1151.

[9] Cooper, David J., and John H. Kagel. 2005. "Are Two Heads Better than One? Team versus Individual Play in Signaling Games," *American Economic Review*, 95(3): 477-509.

[10] Cox, James C., and Stephen C. Hayne. 2006. "Barking Up the Right Tree: Are Small Groups Rational Agents?" *Experimental Economics*, 9: 209-222.

[11] Davis, James H. 1992. "Some Compelling Intuitions about Group Consensus Decisions, Theoretical and Empirical Research, and Interpersonal Aggregation Phenomena: Selected Examples, 1950-1990." *Organizational Behavior and Human Decision Processes*, 52: 3-38.

[12] Fischbacher, Urs. 1999. z-Tree. Toolbox for Readymade Economic Experiments. IEW Working paper 21, University of Zurich.

[13] Gibbons, Jean D., and Subhabrata Chakraborti. 2003. *Nonparametric Statistical Inference*. New York: Marcel Dekker.

[14] Hsu, Shih-Hsun, Chen-Ying Huang, and Cheng-Tao Tang. 2007. "Minimax Play at Wimbledon: Comment." *American Economic Review*, 97(1) 517-523.

[15] Kerr, Norbert L., Robert J. MacCoun, and Geoffrey P. Kramer. 1996. "Bias in Judgment: Comparing Individuals and Groups." *Psychological Review*, 103(4): 687-719.

[16] Kocher, Martin G., and Matthias Sutter. 2005. "The Decision Maker Matters: Individual Versus Group Behavior in Experimental Beauty-Contest Games." *Economic Journal*, 115: 200-223.

[17] Lopes, Lola L., and Gregg C. Oden. 1987. "Distinguishing Between Random and Nonrandom Events" *Journal of Experimental Psychology: Learning Memory and Cognition*, 13(3): 392-400.

[18] Mookherjee, Dilip, and Barry Sopher. 1994. "Learning Behavior in an Experimental Matching Pennies Game." *Games and Economic Behavior*, 7: 62-91.

[19] O'Neill, Barry. 1987. "Nonmetric Test of the Minimax Theory of Two-Person Zerosum Games." *Proceedings of the National Academy of Sciences*, 84: 2106-2109.

[20] Palacios-Huerta, Ignacio. 2003. "Professionals Play Minimax." *Review of Economic Studies*, 70: 395-415.

[21] Palacios-Huerta, Ignacio., and Oscar Volij. 2008. "Experientia Docet: Professionals Play Minimax in Laboratory Experiments." *Econometrica*, 76(1): 71-115.

[22] Rapoport, Amnon, and Richard B. Boebel. 1992. "Mixed Strategies in Strictly Competitive Games: A Further Test of the Minimax Hypothesis." *Games and Economic Behavior*, 4: 261-283.

[23] Rapoport, Amnon, and David V. Budescu. 1992. "Generation of Random Series in Two-Person Strictly Competitive Games." *Journal of Experimental Psychology: General*, 121(3): 352-363.

[24] Rosenthal, Robert W., Jason Shachat, and Mark Walker. 2003. "Hide and Seek in Arizona." *International Journal of Game Theory*, 32: 273-293.

[25] Shachat, Jason M. 2002. "Mixed Strategy Play and the Minimax Hypothesis." *Journal of Economic Theory*, 104: 189-226.

[26] Slonim, Robert L., Ido Erev, Alvin E. Roth. 2003. "Regularities in Play of Zero Sum Games with Unique Mixed Strategy Equilibria." http://wsomfaculty.cwru.edu/slonim/.

[27] Wagenaar, Willem A. 1972. "Generation of Random Sequences by Human Subjects: A Critical Survey of Literature." *Psychological Bulletin*, 77(1): 65-72.

[28] Walker, Mark, and John Wooders. 2001. "Minimax Play at Wimbledon." *American Economic Review*, 91(5): 1521-1538.

Table 1: O'Neill's Game

Player B

		Green	*Red*	*Brown*	*Purple*
Player A	*Green*	W	L	L	L
	Red	L	L	W	W
	Brown	L	W	L	W
	Purple	L	W	W	L

'W' and 'L' denote a win and a loss for player A respectively.

Table 2: Relative Frequencies of Choices and Win Percentages in the Team Treatment

A. Frequencies

		Player B's Choice				Marginal frequencies for player A
		J	1	2	3	
Player A's Choice	J	0.178 (0.160) [0.008]	0.078 (0.080) [0.006]	0.082 (0.080) [0.006]	0.081 (0.080) [0.006]	0.419 (0.400) [0.010]
	1	0.082 (0.080) [0.006]	0.035 (0.040) [0.004]	0.037 (0.040) [0.004]	0.042 (0.040) [0.004]	0.196 (0.200) [0.008]
	2	0.078 (0.080) [0.006]	0.045 (0.040) [0.004]	0.040 (0.040) [0.004]	0.029 (0.040) [0.004]	0.191 (0.200) [0.008]
	3	0.085 (0.080) [0.006]	0.035 (0.040) [0.004]	0.042 (0.040) [0.004]	0.031 (0.040) [0.004]	0.194 (0.200) [0.008]
Marginal frequencies for player B		0.424 (0.400) [0.010]	0.192 (0.200) [0.008]	0.201 (0.200) [0.008]	0.183 (0.200) [0.008]	

B. Win Percentages

Observed player A win percentage:	0.409
Minimax player A win percentage:	0.400
Minimax player A win std. deviation:	0.010

In panel A, the numbers in parentheses represent minimax predicted relative frequencies, whereas those in brackets represent standard deviations for observed relative frequencies under the minimax hypothesis. In panel B, minimax player A win percentage and std. deviation are the mean and the standard deviation of the observed mean percentage win under the minimax hypothesis for player A.

Table 3: Relative Frequencies of Choices and Win Percentages in the Individual Treatment

A. Frequencies

		Player B's Choice				Marginal frequencies for player A
		J	1	2	3	
Player A's Choice	J	0.171 (0.160) [0.008]	0.075 (0.080) [0.006]	0.074 (0.080) [0.006]	0.085 (0.080) [0.006]	0.406 (0.400) [0.010]
	1	0.087 (0.080) [0.006]	0.043 (0.040) [0.004]	0.041 (0.040) [0.004]	0.039 (0.040) [0.004]	0.209 (0.200) [0.008]
	2	0.085 (0.080) [0.006]	0.030 (0.040) [0.004]	0.037 (0.040) [0.004]	0.037 (0.040) [0.004]	0.189 (0.200) [0.008]
	3	0.082 (0.080) [0.006]	0.035 (0.040) [0.004]	0.040 (0.040) [0.004]	0.039 (0.040) [0.004]	0.195 (0.200) [0.008]
Marginal frequencies for player B		0.425 (0.400) [0.010]	0.182 (0.200) [0.008]	0.192 (0.200) [0.008]	0.201 (0.200) [0.008]	

B. Win Percentages

Observed player A win percentage:	0.392
Minimax player A win percentage:	0.400
Minimax player A win std. deviation:	0.010

In panel A, the numbers in parentheses represent minimax predicted relative frequencies, whereas those in brackets represent standard deviations for observed relative frequencies under the minimax hypothesis. In panel B, minimax player A win percentage and std. deviation are the mean and the standard deviation of the observed mean percentage win under the minimax hypothesis for player A.

Table 4: Results of Chi-square Goodness-of-fit Test of Marginal Frequencies of Each Player and Both Players to Equilibrium Mixtures

		Player A	Player B	Both Players
Team	Test Statistic	3.616	7.767	11.383
	p-value	0.306	0.051*	0.077*
Individual	Test Statistic	2.825	8.268	11.092
	p-value	0.419	0.041**	0.086*

** and * denote rejections of the minimax multinomial model at the 5 and 10 percent level. Test statistics for the tests of each player and both players are approximately distributed as $\chi^2(3)$ and $\chi^2(6)$ respectively under the minimax hypothesis.

Table 5: Relative Frequencies of Choices for Each Team and Results of Chi-square Goodness-of-fit Tests for the Minimax Models

Pair #	Player A's Choice				Player B's Choice				p-values of χ^2 tests of minimax multinomial models		
	J	1	2	3	J	1	2	3	Player A	Player B	Both Players
1	0.386	0.227	0.205	0.182	0.402	0.220	0.174	0.205	0.853	0.871	0.960
2	0.379	0.174	0.205	0.242	0.318*	0.227	0.220	0.235	0.618	0.289	0.476
3	0.477*	0.197	0.182	0.144	0.477*	0.174	0.235	0.114**	0.234	0.043‡	0.054†
4	0.477*	0.227	0.136*	0.159	0.379	0.242	0.189	0.189	0.101	0.686	0.259
5	0.576**	0.144	0.136*	0.144	0.492**	0.144	0.144	0.220	0.001‡	0.065†	0.000‡
6	0.455	0.159	0.220	0.167	0.417	0.174	0.189	0.220	0.380	0.835	0.685
7	0.394	0.227	0.212	0.167	0.402	0.258*	0.235	0.106**	0.721	0.032‡	0.119
8	0.470	0.174	0.182	0.174	0.417	0.242	0.174	0.167	0.441	0.484	0.525
9	0.439	0.159	0.189	0.212	0.500**	0.129**	0.197	0.174	0.618	0.069†	0.181
10	0.432	0.197	0.212	0.159	0.462	0.205	0.167	0.167	0.673	0.431	0.637
11	0.295**	0.258*	0.235	0.212	0.439	0.205	0.197	0.159	0.082†	0.651	0.215
12	0.379	0.205	0.197	0.220	0.402	0.167	0.242	0.189	0.935	0.573	0.877
13	0.371	0.205	0.174	0.250	0.386	0.242	0.182	0.189	0.498	0.673	0.688
14	0.379	0.167	0.227	0.227	0.424	0.144	0.212	0.220	0.601	0.454	0.611
15	0.424	0.182	0.197	0.197	0.462	0.189	0.182	0.167	0.935	0.513	0.843
16	0.402	0.212	0.205	0.182	0.447	0.174	0.212	0.167	0.954	0.573	0.887
17	0.364	0.205	0.189	0.242	0.417	0.167	0.205	0.212	0.634	0.817	0.852
18	0.439	0.212	0.144	0.205	0.394	0.159	0.258*	0.189	0.441	0.337	0.415

** and * denote rejections of the minimax binomial model for a given choice at the 5 and 10 percent levels, respectively. Test statistic is approximately distributed as $\chi^2(1)$ under the minimax hypothesis. Similarly, ‡ and † denote rejections of the minimax multinomial model at those levels. Test statistics for the tests of each player and both players are approximately distributed as $\chi^2(3)$ and $\chi^2(6)$ respectively under the minimax hypothesis.

Table 6: Relative Frequencies of Choices for Each Individual and Results of Chi-square Goodness-of-fit Tests for the Minimax Models

Pair #	Player A's Choice				Player B's Choice				p-values of χ^2 tests of minimax multinomial models		
	J	1	2	3	J	1	2	3	Player A	Player B	Both Players
1	0.508**	0.212	0.136*	0.144	0.409	0.159	0.212	0.220	0.034‡	0.686	0.119
2	0.530**	0.152	0.144	0.174	0.470	0.197	0.174	0.159	0.022‡	0.369	0.046‡
3	0.424	0.220	0.182	0.174	0.432	0.197	0.174	0.197	0.776	0.853	0.930
4	0.515**	0.167	0.136*	0.182	0.523**	0.136*	0.182	0.159	0.046‡	0.030‡	0.009‡
5	0.417	0.197	0.167	0.220	0.386	0.258*	0.205	0.152	0.780	0.282	0.557
6	0.492**	0.152	0.182	0.174	0.402	0.182	0.159	0.258*	0.170	0.319	0.201
7	0.432	0.159	0.227	0.182	0.523**	0.159	0.144	0.174	0.542	0.035‡	0.097†
8	0.455	0.265*	0.129**	0.152	0.508**	0.083**	0.227	0.182	0.034‡	0.004‡	0.001‡
9	0.386	0.227	0.220	0.167	0.447	0.197	0.129**	0.227	0.673	0.206	0.411
10	0.356	0.189	0.227	0.227	0.409	0.144	0.250	0.197	0.638	0.289	0.487
11	0.424	0.189	0.174	0.212	0.508**	0.159	0.182	0.152	0.849	0.082†	0.277
12	0.295**	0.250	0.212	0.242	0.341	0.189	0.220	0.250	0.088†	0.372	0.139
13	0.432	0.159	0.197	0.212	0.371	0.182	0.235	0.212	0.673	0.708	0.817
14	0.379	0.220	0.220	0.182	0.364	0.205	0.227	0.205	0.831	0.812	0.934
15	0.341	0.212	0.197	0.250	0.485**	0.159	0.205	0.152	0.406	0.168	0.242
16	0.280**	0.295**	0.205	0.220	0.333	0.242	0.159	0.265*	0.012‡	0.087†	0.007‡
17	0.333	0.227	0.250	0.189	0.432	0.189	0.159	0.220	0.298	0.622	0.487
18	0.311**	0.273**	0.205	0.212	0.311**	0.242	0.212	0.235	0.101	0.193	0.090†

** and * denote rejections of the minimax binomial model for a given choice at the 5 and 10 percent levels, respectively. Test statistic is approximately distributed as $\chi^2(1)$ under the minimax hypothesis. Similarly, ‡ and † denote rejections of the minimax multinomial model at those levels. Test statistics for the tests of each player and both players are approximately distributed as $\chi^2(3)$ and $\chi^2(6)$ respectively under the minimax hypothesis.

Table 7: Testing for Equality of Winning Rates across Strategies in the Team Treatment

		Mixtures		Win Rates		Test	
Pair #	Player	J	Non-J	J	Non-J	Statistic	p-value
1	A	0.386	0.614	0.373	0.296	0.828	0.363
	B	0.402	0.598	0.642	0.696	0.432	0.511
2	A	0.379	0.621	0.380	0.476	1.153	0.283
	B	0.318	0.682	0.548	0.567	0.042	0.837
3	A	0.477	0.523	0.540	0.435	1.451	0.228
	B	0.477	0.523	0.460	0.565	1.451	0.228
4	A	0.477	0.523	0.333	0.435	1.430	0.232
	B	0.379	0.621	0.580	0.634	0.384	0.535
5	A	0.576	0.424	0.474	0.375	1.280	0.258
	B	0.492	0.508	0.446	0.687	7.772	0.005**
6	A	0.455	0.545	0.317	0.306	0.019	0.891
	B	0.417	0.583	0.655	0.714	0.535	0.465
7	A	0.394	0.606	0.500	0.438	0.495	0.482
	B	0.402	0.598	0.509	0.557	0.288	0.591
8	A	0.470	0.530	0.387	0.400	0.023	0.880
	B	0.417	0.583	0.564	0.636	0.711	0.399
9	A	0.439	0.561	0.517	0.324	5.006	0.025**
	B	0.500	0.500	0.545	0.636	1.128	0.288
10	A	0.432	0.568	0.474	0.427	0.290	0.590
	B	0.462	0.538	0.557	0.549	0.009	0.926
11	A	0.295	0.705	0.333	0.344	0.014	0.905
	B	0.439	0.561	0.776	0.568	6.279	0.012**
12	A	0.379	0.621	0.520	0.512	0.008	0.931
	B	0.402	0.598	0.509	0.468	0.214	0.643
13	A	0.371	0.629	0.449	0.434	0.029	0.865
	B	0.386	0.614	0.569	0.556	0.022	0.883
14	A	0.379	0.621	0.500	0.463	0.167	0.683
	B	0.424	0.576	0.554	0.500	0.371	0.543
15	A	0.424	0.576	0.411	0.355	0.421	0.516
	B	0.462	0.538	0.623	0.620	0.001	0.970
16	A	0.402	0.598	0.377	0.367	0.014	0.905
	B	0.447	0.553	0.661	0.603	0.475	0.491
17	A	0.364	0.636	0.375	0.393	0.041	0.839
	B	0.417	0.583	0.673	0.571	1.389	0.239
18	A	0.439	0.561	0.379	0.338	0.244	0.621
	B	0.394	0.606	0.577	0.688	1.681	0.195

** and * denote rejections at the 5 and 10 percent levels respectively.

Table 8: Testing for Equality of Winning Rates across Strategies in the Individual Treatment

Pair #	Player	Mixtures J	Mixtures Non-J	Win Rates J	Win Rates Non-J	Test Statistic	p-value
1	A	0.508	0.492	0.388	0.385	0.002	0.968
	B	0.409	0.591	0.519	0.679	3.487	0.062*
2	A	0.530	0.470	0.371	0.274	1.415	0.234
	B	0.470	0.530	0.581	0.757	4.663	0.031**
3	A	0.424	0.576	0.429	0.342	1.024	0.311
	B	0.432	0.568	0.579	0.653	0.762	0.383
4	A	0.515	0.485	0.500	0.266	7.639	0.006**
	B	0.523	0.477	0.507	0.730	6.902	0.009**
5	A	0.417	0.583	0.491	0.468	0.070	0.791
	B	0.386	0.614	0.471	0.556	0.906	0.341
6	A	0.492	0.508	0.400	0.388	0.020	0.888
	B	0.402	0.598	0.509	0.671	3.463	0.063*
7	A	0.432	0.568	0.544	0.333	5.877	0.015**
	B	0.523	0.477	0.551	0.603	0.371	0.543
8	A	0.455	0.545	0.533	0.333	5.359	0.021**
	B	0.508	0.492	0.522	0.631	1.587	0.208
9	A	0.386	0.614	0.412	0.346	0.586	0.444
	B	0.447	0.553	0.644	0.616	0.107	0.744
10	A	0.356	0.644	0.404	0.424	0.046	0.830
	B	0.409	0.591	0.648	0.538	1.579	0.209
11	A	0.424	0.576	0.518	0.303	6.256	0.012**
	B	0.508	0.492	0.567	0.646	0.862	0.353
12	A	0.295	0.705	0.282	0.398	1.592	0.207
	B	0.341	0.659	0.756	0.575	4.192	0.041**
13	A	0.432	0.568	0.316	0.307	0.013	0.911
	B	0.371	0.629	0.633	0.723	1.172	0.279
14	A	0.379	0.621	0.380	0.439	0.445	0.505
	B	0.364	0.636	0.604	0.571	0.135	0.714
15	A	0.341	0.659	0.444	0.345	1.251	0.263
	B	0.485	0.515	0.688	0.559	2.320	0.128
16	A	0.280	0.720	0.405	0.484	0.665	0.415
	B	0.333	0.667	0.659	0.477	3.901	0.048**
17	A	0.333	0.667	0.409	0.375	0.144	0.705
	B	0.432	0.568	0.684	0.560	2.108	0.147
18	A	0.311	0.689	0.244	0.418	3.684	0.055*
	B	0.311	0.689	0.756	0.582	3.684	0.055*

** and * denote rejections at the 5 and 10 percent levels respectively.

Table 9: Runs Tests in the Team Treatment

		Choices		Runs		
Pair #	Player	J	Non-J	r	$F(r-1)$	$F(r)$
1	A	51	81	53	0.021	0.032*
	B	53	79	59	0.140	0.185
2	A	50	82	62	0.382	0.451
	B	42	90	60	0.599	0.665
3	A	63	69	76	0.935	0.954
	B	63	69	53	0.006	0.009**
4	A	63	69	69	0.613	0.678
	B	50	82	58	0.149	0.194
5	A	76	56	71	0.815	0.859
	B	65	67	71	0.731	0.785
6	A	60	72	69	0.640	0.704
	B	55	77	64	0.382	0.451
7	A	52	80	60	0.204	0.258
	B	53	79	55	0.035	0.052
8	A	62	70	67	0.482	0.552
	B	55	77	53	0.011	0.018**
9	A	58	74	63	0.265	0.327
	B	66	66	81	0.991**	0.994
10	A	57	75	64	0.343	0.410
	B	61	71	50	0.001	0.002**
11	A	39	93	50	0.091	0.124
	B	58	74	60	0.123	0.163
12	A	50	82	68	0.793	0.840
	B	53	79	57	0.074	0.104
13	A	49	83	60	0.280	0.343
	B	51	81	61	0.283	0.350
14	A	50	82	63	0.451	0.528
	B	56	76	64	0.361	0.429
15	A	56	76	56	0.037	0.054
	B	61	71	63	0.235	0.292
16	A	53	79	54	0.024	0.035*
	B	59	73	71	0.773	0.823
17	A	48	84	53	0.035	0.053
	B	55	77	61	0.200	0.255
18	A	58	74	50	0.002	0.003**
	B	52	80	62	0.322	0.388

** and * denote rejections at the 5 and 10 percent levels respectively.

Table 10: Runs Tests in the Individual Treatment

Pair #	Player	Choices: J	Choices: Non-J	Runs r	$F(r-1)$	$F(r)$
1	A	67	65	76	0.932	0.952
	B	54	78	57	0.066	0.093
2	A	70	62	73	0.843	0.882
	B	62	70	59	0.074	0.101
3	A	56	76	73	0.895	0.925
	B	57	75	67	0.551	0.621
4	A	68	64	58	0.049	0.070
	B	69	63	66	0.405	0.475
5	A	55	77	56	0.041	0.060
	B	51	81	62	0.350	0.418
6	A	65	67	75	0.906	0.932
	B	53	79	64	0.432	0.502
7	A	57	75	66	0.481	0.551
	B	69	63	64	0.278	0.340
8	A	60	72	64	0.301	0.365
	B	67	65	64	0.271	0.332
9	A	51	81	67	0.702	0.765
	B	59	73	68	0.587	0.654
10	A	47	85	52	0.029	0.043*
	B	54	78	68	0.686	0.746
11	A	56	76	64	0.361	0.429
	B	67	65	59	0.069	0.095
12	A	39	93	52	0.178	0.228
	B	45	87	58	0.293	0.357
13	A	57	75	69	0.686	0.747
	B	49	83	63	0.487	0.566
14	A	50	82	58	0.149	0.194
	B	48	84	60	0.313	0.378
15	A	45	87	56	0.176	0.226
	B	64	68	54	0.009	0.015**
16	A	37	95	41	0.002	0.004**
	B	44	88	43	0.000	0.001**
17	A	44	88	59	0.403	0.488
	B	57	75	73	0.884	0.916
18	A	41	91	52	0.112	0.150
	B	41	91	48	0.022	0.033*

** and * denote rejections at the 5 and 10 percent levels respectively.

Table 11: Comparison of Serial Correlation in Outcomes between the Team and the Individual Treatment

Variable Name	Team Treatment	Individual Treatment
Constant	-1.529***	-1.804***
	(0.306)	(0.299)
$(JJ)_{t-1}$	0.289	0.719**
	(0.328)	(0.320)
$(J1)_{t-1}$	-0.514	0.039
	(0.384)	(0.367)
$(J2)_{t-1}$	-0.358	-0.141
	(0.374)	(0.376)
$(J3)_{t-1}$	-0.191	-0.194
	(0.366)	(0.370)
$(1J)_{t-1}$	0.150	0.461
	(0.355)	(0.346)
$(11)_{t-1}$	-0.051	-0.108
	(0.424)	(0.422)
$(12)_{t-1}$	-0.196	0.195
	(0.429)	(0.406)
$(13)_{t-1}$	-0.129	0.246
	(0.410)	(0.406)
$(2J)_{t-1}$	0.304	0.532
	(0.353)	(0.344)
$(21)_{t-1}$	-0.262	-0.004
	(0.414)	(0.454)
$(22)_{t-1}$	-0.145	-0.329
	(0.416)	(0.462)
$(23)_{t-1}$	-0.133	0.287
	(0.449)	(0.407)
$(3J)_{t-1}$	0.131	0.025
	(0.353)	(0.362)
$(31)_{t-1}$	-0.168	-0.156
	(0.430)	(0.451)
$(32)_{t-1}$	0.143	-0.117
	(0.395)	(0.430)
Log Likelihood	-1093.737	-1058.997

Variable $(ik)_{t-1}$ denotes the outcome of the stage game in the previous period ($i, k \in \{J, 1, 2, 3\}$) where i and k are player A's and B's choices respectively. Numbers in parenthesis are the standard errors. ***, ** and * denote rejections at the 1, 5 and 10 percent levels respectively.

Table 12: Relative Frequencies of Choice and Results of Chi-square Goodness-of-fit Tests for the Equiprobable Play in the Individual Treatment

Pair #	Player A's Choice				Player B's Choice				p-values of χ^2 tests of equiprobable play for all choices	
	J	1	2	3	J	1	2	3	Player A	Player B
1	0.508**	0.212	0.136**	0.144**	0.409**	0.159**	0.212	0.220	0.000‡	0.000‡
2	0.530**	0.152**	0.144**	0.174**	0.470**	0.197	0.174**	0.159**	0.000‡	0.000‡
3	0.424**	0.220	0.182*	0.174**	0.432**	0.197	0.174**	0.197	0.000‡	0.000‡
4	0.515**	0.167**	0.136**	0.182*	0.523**	0.136**	0.182*	0.159**	0.000‡	0.000‡
5	0.417**	0.197	0.167**	0.220	0.386**	0.258	0.205	0.152**	0.000‡	0.001‡
6	0.492**	0.152**	0.182*	0.174**	0.402**	0.182*	0.159**	0.258	0.000‡	0.000‡
7	0.432**	0.159**	0.227	0.182*	0.523**	0.159**	0.144**	0.174**	0.000‡	0.000‡
8	0.455**	0.265	0.129**	0.152**	0.508**	0.083**	0.227	0.182*	0.000‡	0.000‡
9	0.386**	0.227	0.220	0.167**	0.447**	0.197	0.129**	0.227	0.003‡	0.000‡
10	0.356**	0.189	0.227	0.227	0.409**	0.144**	0.250	0.197	0.038‡	0.000‡
11	0.424**	0.189	0.174**	0.212	0.508**	0.159**	0.182*	0.152**	0.000‡	0.000‡
12	0.295	0.250	0.212	0.242	0.341**	0.189	0.220	0.250	0.598	0.079†
13	0.432**	0.159**	0.197	0.212	0.371**	0.182*	0.235	0.212	0.000‡	0.011‡
14	0.379**	0.220	0.220	0.182*	0.364**	0.205	0.227	0.205	0.007‡	0.026‡
15	0.341**	0.212	0.197	0.250	0.485**	0.159**	0.205	0.152**	0.086†	0.000‡
16	0.280	0.295	0.205	0.220	0.333**	0.242	0.159**	0.265	0.369	0.042‡
17	0.333**	0.227	0.250	0.189	0.432**	0.189	0.159**	0.220	0.118	0.000‡
18	0.311	0.273	0.205	0.212	0.311	0.242	0.212	0.235	0.255	0.416

** and * denote rejections of tests of equiprobable play for a given choice at the 5 and 10 percent levels, respectively. Test statistic is approximately distributed as $\chi^2(1)$ under the equiprobable hypothesis. Similarly, ‡ and † denote rejections of tests of equiprobable play for all choices at those levels. Test statistic is approximately distributed as $\chi^2(3)$ under the equiprobable hypothesis.

Table 13: Relative Frequencies of Choice and Results of Chi-square Goodness-of-fit Tests for the Equiprobable Play in the Team Treatment

Pair #	Player A's Choice				Player B's Choice				p-values of χ^2 tests of equiprobable play for all choices	
	J	1	2	3	J	1	2	3	Player A	Player B
1	0.386**	0.227	0.205	0.182*	0.402**	0.220	0.174**	0.205	0.003‡	0.001‡
2	0.379**	0.174**	0.205	0.242	0.318*	0.227	0.220	0.235	0.005‡	0.343
3	0.477**	0.197	0.182*	0.144**	0.477**	0.174**	0.235	0.114**	0.000‡	0.000‡
4	0.477**	0.227	0.136**	0.159**	0.379**	0.242	0.189	0.189	0.000‡	0.005‡
5	0.576**	0.144**	0.136**	0.144**	0.492**	0.144**	0.144**	0.220	0.000‡	0.000‡
6	0.455**	0.159**	0.220	0.167**	0.417**	0.174**	0.189	0.220	0.000‡	0.000‡
7	0.394**	0.227	0.212	0.167**	0.402**	0.258	0.235	0.106**	0.001‡	0.000‡
8	0.470**	0.174**	0.182*	0.174**	0.417**	0.242	0.174**	0.167**	0.000‡	0.000‡
9	0.439**	0.159**	0.189	0.212	0.500**	0.129**	0.197	0.174**	0.000‡	0.000‡
10	0.432**	0.197	0.212	0.159**	0.462**	0.205	0.167**	0.167**	0.000‡	0.000‡
11	0.295	0.258	0.235	0.212	0.439**	0.205	0.197	0.159**	0.572	0.000‡
12	0.379**	0.205	0.197	0.220	0.402**	0.167**	0.242	0.189	0.008‡	0.000‡
13	0.371**	0.205	0.174**	0.250	0.386**	0.242	0.182*	0.189	0.008‡	0.003‡
14	0.379**	0.167**	0.227	0.227	0.424**	0.144**	0.212	0.220	0.005‡	0.000‡
15	0.424**	0.182*	0.197	0.197	0.462**	0.189	0.182*	0.167**	0.000‡	0.000‡
16	0.402**	0.212	0.205	0.182*	0.447**	0.174**	0.212	0.167**	0.001‡	0.000‡
17	0.364**	0.205	0.189	0.242	0.417**	0.167**	0.205	0.212	0.020‡	0.000‡
18	0.439**	0.212	0.144**	0.205	0.394**	0.159**	0.258	0.189	0.000‡	0.001‡

** and * denote rejections of tests of equiprobable play for a given choice at the 5 and 10 percent levels, respectively. Test statistic is approximately distributed as $\chi^2(1)$ under the equiprobable hypothesis. Similarly, ‡ and † denote rejections of tests of equiprobable play for all choices at those levels. Test statistic is approximately distributed as $\chi^2(3)$ under the equiprobable hypothesis.

Figure 1: Empirical Cumulative Distribution Function for Observed P-values from Chi-square Goodness-of-fit Tests of Minimax Multinomial Model in the Team Treatment (Left) and the Individual Treatment (Right).

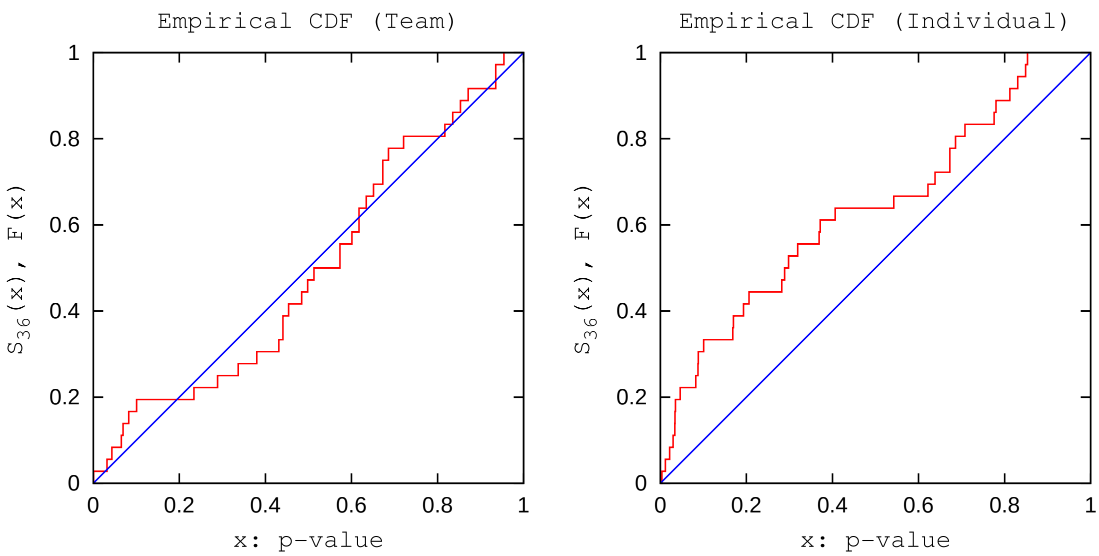

Figure 2: Empirical Cumulative Distribution Function for Observed p-values from Chi-square Tests for Equality of Winning Rates across Strategies in the Team Treatment (Left) and the Individual Treatment (Right).

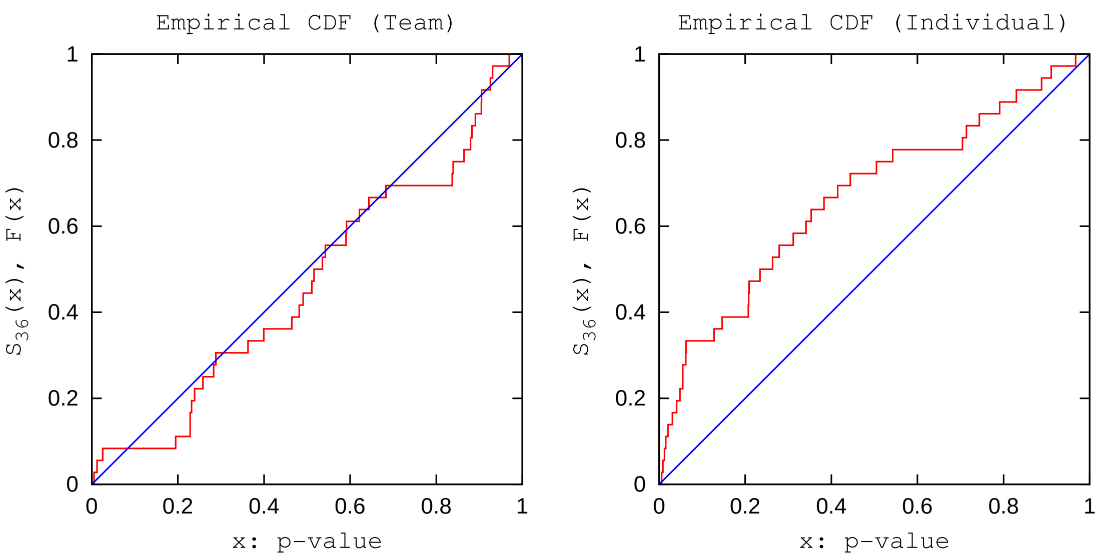

Figure 3: Empirical Cumulative Distribution Function for a Particular Realization of Statistics drawn from $U[F(r-1), F(r)]$ for Each Decision Maker in the Team Treatment (Left) and the Individual Treatment (Right).