Chapter 4

Equilibrium Play and Best Response to (Stated) Beliefs in Constant Sum Games

4.1 Introduction

A substantial portion of the experimental literature shows that game-theoretical predictions do not work well in the laboratory, even when the games played are very simple.[1] This is particularly true if subjects play games for the first time without previous experience. However, first time behaviour is crucial to model a vast number of economic situations which are not repeated, and it helps to understand what people bring into strategic settings. First time behaviour is also important to whatever learning occurs in repeated games. A natural question is to identify the class of games for which game theory predicts well when games are played for the first time and the reasons why it might fail in other games.

We aim to contribute to this question by looking at play and beliefs about opponents' play in simple but non-trivial games with similarities to others in which current experimental evidence shows that game theory predictions do not work so well. In particular, we study two-player 3x3 constant sum normal form games with unique equilibria in pure strategies and with different number of rounds of iterated deletion of (strictly) dominated strategies necessary to reach the Nash equilibrium. We obtain that in this class of games game theory predicts subjects' behaviour better than in previous experiments and we discuss the relation of our results with previous literature

[1] For example, see Güth, Schmittberger and Schwarze (1982), Camerer and Weigelt (1988), Och and Roth (1989), Cooper et al (1990), Brandts and Holt (1992). For an extensive survey, see Kagel and Roth (1995).

in which the theory predictions are not so successful.

For simple games with unique pure strategy equilibria, experimental evidence is not conclusive. While in 2x2 repeated games equilibrium play has found substantial support (McCabe et al. (1994), Mookherjee and Sopher (1994)), in games with more than two strategies for each subject and no possibility of learning equilibrium predictions start to fail. Stahl and Wilson (1995) found equilibrium compliance rates of 68% in 3x3 games with three rounds of dominance solvability. However, Broseta, Costa-Gomes and Crawford (2001) obtain in 2x3 games with three rounds of deletion of dominated strategies to reach equilibrium or with no dominated strategies equilibrium compliance rates ranging from 11% to 28%. For 4x4, 5x5 and 6x6 repeated games, the evidence is even more negative (Brown and Rosenthal (1990), Rapappport and Boebel (1992), Mookherjee and Sopher (1997)). Thus, choosing 3x3 games with different numbers of rounds of iterated deletion of dominated strategies may be a good starting point to disentangle the reasons why game theory loses its predictive power in games with more strategies available to players.

Our experiment is closely related to a previous experiment by Costa-Gomes and Weizsäcker (2004), who found low rates of compliance with equilibrium predictions, low frequency assigned to the belief that opponents would play equilibrium actions and low consistency between actions and belief statements, in the sense that the percentage of actions that were best responses to stated beliefs was low. Our results clearly differ from theirs, which may be caused by our games being constant sum and also to some procedural changes we made with respect to their experiment: payoffs were represented by single-digit numbers, there was no conversion rate between experimental currency and monetary payoffs, and the procedure to elicit beliefs was different. Below we explain why these changes may have made a difference.

First, as we have seen in chapters 2 and 3, one of the causes why game theory predictions may not work well in the laboratory may be that subjects have distributional and/or efficiency concerns. A possible reason for such concerns is that even if laboratory play may be completely anonymous, when games are played for the first time subjects may bring preconceptions on how to behave in strategic situations from previous real life experience that may cause divergences between equilibrium predictions and observed behaviour in experiments.[2] We choose to study constant sum games because, on a theoretical level, behaviour in them should not be affected by distributional and efficiency concerns. Efficiency concerns should not be an issue since subjects' payoffs always add up to the same amount, no matter which actions are chosen. Theoretically, distributional concerns should not affect behaviour as long as

[2] See Binmore (1998).

subjects care more for their own payoffs than for those of others.[3] This is because subjects with distributional concerns would have to give up the same units of payoffs that would go to their opponent in order to increase their opponents' payoffs. On the other hand, this theoretical result seems counter-intuitive.[4] In constant sum games all strategic behaviour refers to how to distribute a pie of a given size and thus, how fair the distribution is should matter to subjects with distributional concerns.[5] Of these preconceptions, a natural one is that, everything else equal, subjects should get equal shares.[6] Therefore, whether it is feasible to equally split payoffs or not, may have an influence on how subjects play constant sum games. We investigate whether distributional concerns influence subjects' choices in constant sum games and whether results are affected by equal splits being feasible. We find that subjects' behaviour is quite close to equilibrium and that it does not matter whether equal splits are feasible or not.[7]

A second issue related to games being played for the first time is whether subjects have formed meaningful beliefs about how their opponents will behave.[8] Although it is not necessary for subjects to hold beliefs about opponents' play for observed behaviour to coincide with equilibrium predictions,[9] it may help us to disentangle the reasons why subjects play or do not play according to game theory predictions in the laboratory. Therefore, we study if subjects are able to correctly predict their opponents' play and if their actions are consistent with their beliefs, in the sense that actions chosen are best replies to beliefs held. Experiments allow us to study this question by eliciting beliefs. We can distinguish between choice-based methods of eliciting beliefs (in which experiments are designed such that actions taken by subjects reveal information about beliefs)[10] and direct elicitation procedures (in which subjects are directly asked how

[3] Camerer et al. (1998).

[4] And in particular there is ample evidence that this result is not satified by Dictator Game data.

[5] There is at least one type of constant sum games where social preferences seem to affect how subjects play games: Dictator games. Becasue dictator games are sequential and more similar to the games in the following chapter, we discuss them further in Chapter 5.

[6] This has been observed in several experiments, for example in ultimatum games (Güth et al, (1982)).

[7] Previous research on contsant sum games has focused on whether subjects' frequencies of play in repeated constant sum games coincide with the probabilities with which subjects should play the one-shot mixed equilibria. Most results have been negative (Rapapport and Boebel (1992), McCabe et al. (1994), Mookherjee and Sopher (1997) and Walker and Wooders (2001)) although O'Neill (1987) and Binmore et al. (2001) are more positive. The discussion has focused on how data for mixed strategy equilibria should look like.

[8] Savage (1954) and Anscombe and Aumann (1963) define "beliefs" as subjective probabilities about uncertain events.

[9] And in particular, in constant sum games the Nash Equilibrium strategies coincide with Minimax and Maximin strategies, for which subjects only need to calculate their "safe" strategy, with no need to predict their opponents' strategy.

[10] See for example, Stahl and Wilson (1994), Nagel (1995), McKelvey and Palfrey (1995) or, more recently, Brañas and Morales (2003).

they think their opponents will play games).[11] We combine both methods and we actually check for consistency between actions chosen and beliefs stated.[12] Our results show that in fact subjects were reasonably accurate at predicting opponents' actions and that most choices were best responses to stated beliefs.

When beliefs are elicited directly, it is crucial to elicit them in a meaningful manner that subjects can understand. First, some authors, starting with Kahneman and Tversky (1973), express doubts on whether subjects can quantify their beliefs. Second, even if subjects may be able to quantify their beliefs, they might find some form of processing quantitative beliefs more meaningful than others. In this sense, following Gigerenzer (2000, 2002), we elicited beliefs by asking about frequencies of play by a pool of subjects instead of asking about probabilities of a single action chosen by a single opponent as it is frequently done.[13] This difference may be important when subjects only choose once in each game.

Having Elicited beliefs allows us to compare our results with previous research in which beliefs are used to study the degrees of complexity with which individuals are able to play games. We explicitly designed our games to be able to discriminate equilibrium behaviour from behaviour predicted by models varying in the degree of cognitive complexity that subjects are assumed to be able to process. Although depth of reasoning is a complex issue, these models approximate subjects' sophistication to whether they are able to best response to their beliefs about opponents' play and whether they form those beliefs anticipating that opponents may also behave strategically. Thus, they define the first degree of depth of reasoning as best responding to believing opponents choose their actions according to a uniform distribution and from them onwards higher degrees of depth are defined as best responding to the assumption that the opponent has the immediately lower degree of depth than one-self. In our games, we find that the equilibrium prediction clearly outperforms these models of depths of reasoning, both for the aggregate of subjects over all games and for a wide majority of individuals.

An alternative way to study cognitive complexity is to associate it with the number of rounds of iterated elimination of strictly dominated strategies subjects are able to perform. Thus, we designed our games such that they differed in the number of rounds of iterated elimination of dominated strategies that were necessary to reach the equilibrium outcome. We compare subjects' behaviour across games differing in the necessary number of rounds and we conclude that this was not a straightforward measure of how complex games were for subjects.

[11] See McKelvey and Page (1990), Offerman et al. (1996) and Nyarko and Schotter (2002).

[12] Costa-Gomes and Weizsäcker (2004) previously combined both methods of belief elicitation in normal form games.

[13] McKelvey and Page (1990), Offerman et al (1996), Costa-Gomes and Weizsäcker (2004).

The remainder of the chapter is organized as follows. Section 4.2 presents the experimental design and procedures. Section 4.3 contains the results and the main descriptive statistics. Section 4.4 concludes. The Appendices contain the instructions and we also show the games.

4.2 Experimental Design and Procedures

4.2.1 Experimental Design

Subjects were presented with a series of ten 3x3 Constant Sum Normal Form Games with Unique Equilibrium in Pure Strategies. For each of the ten games, they were asked to perform two tasks: they had to choose an action (between “UP”, “MIDDLE” or “DOWN”) and they had to report how many of the players on the other subjects' role they thought would play each of the three actions available (“LEFT”, “CENTRE” and “RIGHT”).

We constructed a 2x2 design according to two criteria. The first criterion was the order in which subjects had to perform the two tasks. In treatments BABAF and BABAU (to which we will generically refer as BABA treatments), subjects were asked for each game, first to state their Beliefs (B) and then to chose an Action (A), after which, they moved on to the next game. In treatments ABF and ABU (to which we will generically refer as AB treatments) subjects first chose an action in the ten games, without knowing what the second task would consist of, and then, after answers for all actions were collected, they were presented again with the ten same games and asked to state their beliefs about opponents' play. Comparing the BABA and AB treatments allows us to study whether eliciting beliefs before playing the games influences behaviour.

The second criterion was whether an equal split of payoffs was feasible in each of the games. As the games were constant sum, the sum of payoffs both subjects could earn was always the same and equal to £12, no matter the strategies chosen by both players. In treatments BABAF and ABF (to which we will generically refer as F treatments) an equal split of payoffs was feasible in one of the cells of all the games subjects played. Payoffs were designed such that both subjects would get £6 if they both took the strategy leading to this cell being chosen in each particular game. In treatments BABAU and ABU (to which we will generically refer as U treatments), payoffs in all games were substituted in this cell by a more unequal split, such that one subject would get a payoff of £7 and the other a payoff of £5. For example, in Game 4R below, payoffs when Row subjects chose MIDDLE and Columns subjects chose LEFT were £6 for both subjects in the F treatments, while they were £5 for

Row subjects and £7 for Column subjects in the U treatments. The location of the cell and the changes in payoffs from the F to the U treatments were designed such that it did not affect neither the predictions of the six behavioural models we study nor the degrees of strict dominance solvability, such that in some games it was the Row player who got a better than equal split of payoffs in the U treatments while in other games it was the Column player and such that subjects would get a higher payoff in this cell in some games (lower in other games) than in the Nash equilibrium outcome. Notice also that although the cell in which the equal split was feasible sometimes was in one of the subjects' equilibrium strategies it was never included in both subjects' equilibrium strategies, and therefore such cell never coincided with the Nash equilibrium outcome. Comparing the F and U treatments allows us to study whether the feasibility of an exact equal split influenced behaviour, which may be an indication of subjects' concerns for fairness in constant sum games.

Game 4R (U Treatment)

Row \ Column	Left	Centre	Right
Up	8 4	10 2	11 1
Middle	7 5	1 11	8 4
Down	5 7	4 8	2 10

Game 4R (F Treatment)

Row \ Column	Left	Centre	Right
Up	8 4	10 2	11 1
Middle	6 6	1 11	8 4
Down	5 7	4 8	2 10

4.2.2 Experimental Procedures

The experiment was carried out with pen and paper in the ELSE laboratory during April 2004. Subjects were recruited by e-mail using the ELSE database, which consists of UCL undergraduate and graduate students. As we are interested in behaviour played without previous experience, we only targeted subjects who had not participated in previous game experiments and whose field of study indicated that they would not be familiar with Game Theory and Economics.[14]

Our experiment consisted of four sessions with twenty subjects per session. In each session, ten subjects were randomly assigned "Row" roles in all ten games, while the

[14] Although we targeted unexperienced students, it turned out ex-post that 3 of our 80 subjects had taken introductory courses in Game Theory. However, neither their behavior in the experiment nor their explanations of their behavior in a post-experiment questionnaire were more "game theoretic" than their peers' so we did not exclude them from the sample.

other ten subjects were assigned "Column" roles. However, no subject was aware of their role (nor other subjects' roles) as games were presented to all players from the point of view of row players. Neutral language was used by calling subjects "You" and their opponents "Participants in the other group".

Upon arrival, subjects were randomly assigned seats and were asked to read some preliminary instructions, which described a strategic decision situation and the 3x3 payoff matrix associated with its normal form representation.[15] Then subjects were required to pass an Understanding Test where they had to demonstrate that they knew how to map players' actions in a game to outcomes, and outcomes to players' payoffs. Subjects were told that those who failed the test would act as "assistants" in the experiment. However, no subject failed the test in any treatment and so the over-recruited subjects were asked to assist the experimenter.[16]

The experiment consisted of ten games. In the BABA treatments subjects first read the instructions on stating first order beliefs and choosing actions and how they would be rewarded for these two tasks. Then subjects stated beliefs and chose actions for all ten games with no feedback. Subjects stated beliefs by writing down how many of the 10 subjects in the opponents' role they believed would chose each of their three possible actions in each game. In the AB treatments, subjects first read the instructions about how to choose their actions, and then played those games (Part I). After Part I, answer sheets were collected and subjects read the instructions on beliefs. Next, they stated their beliefs for all 10 games (Part II).[17] This procedure guaranteed that in the AB treatments, when subjects played the games, beliefs had not been mentioned. Finally, all answer sheets were collected. This procedure made sure that all subjects played all games before any feedback had been given. While payments were calculated subjects were asked to fill in an anonymous questionnaire, then subjects received their payments in private and left.[18] The Appendix reproduces the instructions for the BABA treatments.[19]

For each game subjects played they were randomly and anonymously paired with a different participant from the other group. Subjects never learned who their matched

[15] The strategic situations were called "Tables" in the instructions.

[16] All subjects were informed of this.

[17] Games were presented in random and different order to each subject to control for (possible) non-feedback learning. We varied the games and the order in which they were presented to prevent subjects from developing preconceptions about games' strategic structures and to help us discriminate between the different models of behavior we study.

[18] The purpose of the questionnaire was two-fold: it gave us time to calculate payments (with the help of an Excel sheet pre-programmed with the matching of subjects) and it provided information about subjects' fields of study, previous experience in experiments and comments on how they took their decisions.

[19] The text in the instructions for the AB treatments was practically the same. The only difference was in the order in which instructions were received.

participant in each game was, neither the action which was taken by their matched participant or any other participant in any game. Games were also presented in random and different order to all subjects. To ensure that subjects were motivated both to choose preferred actions and to state true beliefs, they were paid according to their answers in both tasks as follows. At the end of each session, a number from 1 to 10 was selected from a bingo urn. This number indicated for which of the 10 games all subjects would be paid for both tasks.[20] To reward actions, subjects were paid exactly the amount of pounds indicated by the number in the lower left corner of the cell chosen as a result of their action and the action chosen by their matched opponent in the particular game selected with the bingo urn.[21]

With respect to payments for stated beliefs, subjects were paid according to a Quadratic Scoring Rule (QSR) which rewarded accuracy between predicted frequencies of play of each action and the frequencies with which each of the three actions available were actually played by the 10 opponents in the game selected.[22] The QSR was designed such that subjects could earn comparatively less money with their belief statements than with their action choices (Maximum of £2 and £11 respectively). Had payoffs for both tasks been similar, risk averse subjects would have found incentives to take actions that were not best responses to their stated beliefs in the aim to average payoffs.

Subjects were paid the sum of a £5 fixed fee, plus their earnings for choosing actions and stating beliefs. Average payments were £12.78 (around $20 at the time). Each session lasted one hour and subjects were allocated forty minutes to perform both tasks in the ten games.[23]

[20] We paid subjects for one random game instead of for an aggregated measure of their answers in all 10 games to be able to maintain the one to one relationship between outcomes and payoffs. Avoiding conversion rates may help clarifying incentives, which may be particularly important in experiments in which beliefs on other subjects' behavior are elicited. As we discuss below, risk aversion did not seem to have been a problem.

[21] A British pound corresponded to 1.85 American dollars at the time of the experiment. Our design allowed us to provide reasonably high incentives while keeping one or two digit numbers to represent payoffs and avoiding conversion rates from experimental currency to monetary currency.

[22] Notice that when subjects are asked to predict the frequencies of play of a finite population of subjects, QSRs are not necessarily incentive compatible as subjects' average expectation of play of each action might not necessarily be equal to one of the possible empirical distributions over the finite set of opponents' actions. In any case, expected payoff maximizers can do no better by stating different beliefs than their true beliefs and given our results we think the problem is minor. For a discussion on QSRs see Offerman, Sonnemans and Schram (1996), Offerman and Sonnemans (2001) and Selten (1998). The particular QSR we used, along with an intuitive explanation for subjects highlighting that understanding the maths of the rule was not essential, can be found in the Instructions (Appendix A).

[23] In the AB treatments, 20 minutes were allocated for each of the two parts.

4.2.3 The Games

We classify our games according to whether they are dominance solvable or not. Eight of our games are dominance solvable. Of these, we classify them according to the number of consecutive rounds of iterated deletion of strictly dominated strategies needed to reach the unique Nash equilibrium. Games 1R and 1C are dominance solvable with one round of dominance to reach the equilibrium for one of the players (Row in 1R, Column in 1C) and two rounds of dominance for the other player. Games 2R and 2C are solvable with two rounds for one player (Row in 2R, Column in 2C) and three rounds for the other. Games 3R and 3C are solvable with three rounds of dominance for one player (Row in 3R, Column in 3C) and two for the other, although the first deletion of strictly dominated strategies is simultaneous for both players. Games 4R and 4C are solvable with four rounds for one player (Row in 4R, Column in 4C) and three rounds for the other. Finally, Games NR and NC are not dominance solvable and have no strictly dominated actions.[24] In the U treatments, Games 1R, 2R, 2C and 3R had additional weakly dominated strategies, apart from the strictly dominated ones.

We chose one-digit numbers to represent payoffs.[25] The sum of Row and Column players' payments in all cells of all games was 12.[26] The ten games were designed such that the equilibrium did not correspond to the same combination of actions by two players in more than two games.

We selected 3x3 games in which the prediction of how subjects would play would not be trivial. Accordingly, we designed the games such that we were able to discriminate Nash Equilibrium choices[27] from the choices predicted by five other models that have proven to be at least partially successful in previous studies on depths of reasoning.[28] These models are named L1, L2, L3, D1 and Maximax. L1 predicts that each subjects' action is a best response against the belief that the opponent is playing each action with equal probability. L2 predicts a best response against the belief that the opponent is playing according to L1 and L3 predicts a best response to the believing the opponent plays according to L2. D1 predicts a best response against a uniform

[24] Out of the six possible types of 3x3 constant sum games with unique pure strategy equilibria, we covered all but one possible case according to their degree of strict dominance solvability. The remaining case has a dominated strategy for one of the subjects and it is not dominance solvable.

[25] We did so because if subjects really chose their actions as a best response to their beliefs, calculating such best response in terms of expected payoffs may have been more difficult if numbers representing payoffs were large, and we did not want to discourage such type of behaviour.

[26] Numbers 10 and 11 were used in a few games to make it possible to discriminate models of behavior. Number 0 was not used to avoid behavior being possibly caused by aversion to getting no payoff.

[27] Simply referred as "Equilibrium", from here onwards.

[28] Stahl and Wilson (1994, 1995), McKelvey and Palfrey (1995), Broseta, Costa-Gomes and Crawford (2001), Costa-Gomes and Weizsäcker (2004), Weizsäcker (2003) and Goeree and Holt (2004).

belief over the opponents' undominated actions. Maximax predicts the action that is part of the action profile leading to the player's highest possible payoff in the game.[29] The Appendix contains the games, indicating the predictions of each of the six models, the round in which a dominated strategy is deleted and the payoffs that were changed to create the F and U treatments.

4.3 Experimental Results

4.3.1 Descriptive Statistics

Table 1 below presents the main descriptive statistics for each game when grouping all treatments and subject roles. We report, for each of the ten games, the percentage of equilibrium actions taken, the percentage of frequencies assigned to opponents' choosing equilibrium actions and the percentage of best responses to stated beliefs. A first look at results shows that roughly 80% of actions taken were according to equilibrium, that subjects believed the equilibrium action would be played with highest frequency (although with lower frequency than it was actually played), and that subjects actions were best responses to the distribution of stated beliefs in 73% of the cases. Frequencies were similar across games and the number of rounds of iterated dominance does not seem to affect percentages in a clear cut manner. However, notice that in the two games which are not dominance solvable (NR and NC) the equilibrium and best response frequencies show percentages that were lower than in the other games. This is particularly true for the percentage of best responses.

Game	Equilibrium Actions	Equilibrium Beliefs	Best Response to Stated Beliefs	Nº Rounds Iterated Dominance (Row, Column)
1R	76.25	58.5	80	1,2
1C	75	59.375	75	2,1
2R	82.5	55.875	83.75	2,3
2C	81.25	51.125	71.25	3,2
3R	82.5	64.75	77.5	3,3
3C	86.25	63.125	77.5	3,3
4R	87.5	59.625	82.5	4,3
4C	78.75	59	80	3,4
NR	72.5	52.875	47.5	No
NC	73.75	51.5	55	No
Average	**79.625**	**57.575**	**73**	

Table 1: Descriptive statistics (percentages).

[29] Stahl and Wilson (1994) use a more sophisticated version of these models. According to their definition, L2 is a best response to a belief distribution which assigns positive weights to a portion of the population choosing actions randomly (L0) and the remaining portion to subjects best responding to uniform beliefs (L1). The reason to define the zero-level of rationality as an equal probability to play each possible strategy, and thus define degrees of rationality from there on, remains open.

The results of the informal questionnaire subjects answered after the experiment shows that a high percentage (95%) of the subjects who answered this questionnaire claimed to have taken their actions strategically, i.e., taking into account what their opponents would choose. Also, a high percentage of subjects (92%) claimed that they believed their matched participants would choose strategically, i.e., they would take into account the choice themselves made. Notice that in this questionnaire there were no monetary incentives for truth telling and that as subjects' answers were relatively vague, it was difficult to classify the level of strategic sophistication subjects claimed they had used from their answers. Therefore, in the following we use subjects' actual choices in the experiments and we compare the performance of different models that assume subjects' different level of strategic sophistication are able to predict subjects' choices and beliefs about how opponents would play.

4.3.2 Treatment Effects

In this section we study whether the different treatments in our design had an effect on subjects' choices or beliefs stated. In particular we study two questions: 1) whether eliciting beliefs immediately before actions had any effect on actions played or beliefs stated and 2) whether allowing for equal payoff splits changed behaviour.

We start with the first question. There are several reasons why we may think that belief elicitation prior to choosing actions can affect play. First, eliciting beliefs may make beliefs a more salient aspect of game play than they would otherwise. Second, asking subjects what they believe their opponents will do may cause subjects to predict the behaviour of their opponents more accurately than they otherwise would and thus, they may best respond more frequently to their stated beliefs. Third, when subjects are rewarded both for their stated beliefs and their actions, risk averse subjects could state their beliefs in a way that insures them against ex-post strategic mistakes (and vice-versa), so that strategy choices and belief statements could become mutually endogenous. Finally, rewarding beliefs subsidizes using beliefs to choose actions versus other procedures that might have been used when the rewards are not offered. Most of previous studies have found no effect of eliciting beliefs prior to actual play[30] although a recent article by Ruström and Wilcox (2004) reports some effects. We study this question by comparing the actions chosen and the beliefs stated in the BABA against the AB treatments.

First we look at actions chosen. We use Fisher's Exact Probability Test (FEPT) for count data[31] which tests if differences in observed proportions of actions chosen between two treatments might be expected by chance. The null hypothesis (two-tailed)

[30] See Nyarko and Schotter (2002) and Costa-Gomes and Weizsäcker (2003).
[31] Developed by Fisher (1935), Irwin (1935) and Yates (1934).

is that there is no difference in the probability of playing each strategy generating the observed proportion of play of each strategy in each treatment.[32] As with all statistical tests in this thesis, we used the free software R (2003) to perform FEPTs.

We conduct FEPTs separately for each game. We first compare subjects' aggregate actions for each player role (Row or Column) in each of the ten games between the BABA and the AB treatments (without aggregating the F and U treatments). Out of the 40 possible comparisons, we can never reject the null hypothesis that the underlying probability is the same at the 5% significance level.[33] [34] We then perform a stronger test by pooling the data for the F and U treatments[35] and again compare aggregate actions across players' roles between the BABA and the AB treatments. There is no p-value smaller than 5% so we cannot reject the hypothesis that there is no effect of the order of tasks performed in the aggregate actions.

Our next step is to test if the order of tasks affected subjects' belief statements. We collapse each agents' belief statements into one of four categories: for each of the three actions all the stated beliefs that assigned more than half of the frequency to an action were classified in the same category (thus creating three categories), and the last category comprises all the beliefs that do not assign more than half of the frequency to any of the three actions opponents can take. Again, this allows us to create a contingency table and use FEPTs to test for differences in belief statements between BABA and AB treatments.[36]

When comparing subjects' aggregate belief statements for each player role in each of the ten games between treatments BABA and AB treatments (without aggregating the F and U treatments) we cannot reject the null hypothesis of no difference in all comparisons. When we perform a stronger test by pooling the F and U treatments we can only reject it once (p-value equal to 0.003 for Row subjects in Game NC). Thus, we conclude the following:

> Result 1 *The order in which subjects performed both tasks did not affect behaviour.*

[32] Although less common than the Chi-square test, Fisher's test requires less data in each category to be correctly calculated. Chi-square tests would require at least five subjects playing each action in each treatment which, given that most subjects chose the same actions, was not satisfied in our games. The main assumption required for both of these tests is independence between observations of the games in each treatment.

[33] Although FEPT is specifically designed for small samples it is still not a very powerful test with only ten observations in each treatment. For example using this test, we cannot reject that distribution of answers (3,2,5) in one treatment is the same as the distribution (1,7,2) in another treatment at the 5% significanc level. However, we can reject that it is different than (1,8,1). The power of the test increases with the number of observations.

[34] Qualitative results of all FEPTs in this section are the same at the 10% signicance level.

[35] This is allowed by results below.

[36] This procedure was previously used by Costa-Gomes and Weizsäcker (2004).

We now study whether the feasibility of equal payoff splits had an effect on behaviour. We proceed similarly as before by carrying out FEPTs for both actions and stated beliefs under the null hypothesis that there was no difference across treatments in the probability of playing (or stating) the observed proportions of play (or beliefs stated) of each action.

When comparing aggregate actions between the F and the U treatments for each player role (without aggregating the BABA and the AB treatments), no p-value is smaller than 5% out of 40 comparisons. When we pool the BABA and AB treatments and we compare the F and U treatments across player roles, only one out of the 20 possible p-values is smaller than 5% (p-value equal to 0.006 for Row subjects in Game 4C), which indicates that there is no significant effect. We also performed Mann-Whitney tests under the null hypothesis that the median of the distribution of games in which subjects chose the strategy containing the equal split was not different between the F and U treatments at the 5% significance level. Both when we aggregate the BABA and the AB treatments and when we do not, we could never reject the null hypothesis. Thus, we conclude that actions chosen were not affected by whether equal splits were available or not.[37]

Moving on to beliefs, we created a contingency table using the four categories mentioned before to classify beliefs stated and we performed FEPTs comparing same games under the F and U treatments to test for effects of equal splits on subjects' beliefs. We obtain no p-value smaller than 0.05 for the 40 comparisons when we do not aggregate treatments with respect to the order of tasks. When we do aggregate them, only one of the 20 possible p-values is smaller than 0.05 (p-value of 0.0189 for Column subjects in Game NC), which indicates that there is no effect of the feasibility of equal splits. We also performed Mann-Whitney tests comparing the distribution of average frequencies assigned to the strategy which contained the equal payoff splits between the F and U treatments, again for each game and player role. We could never reject the null hypothesis that the median of the distribution of frequencies assigned to the strategy containing the equal split was not different at the 5% significance level, both when aggregating the BABA and AB treatments and when not. Thus, we conclude the following:

Result 2: *Behaviour was not affected by the feasibility of equal splits.*

Small payoff differences between the equal and unequal split might explain Result 2. It would be worthwhile to study robustness to higher payoff differences. An alternative

[37] Same results were obtained for the null hypothesis that the feasibility of equal splits did not affect the median of the distribution of the number of games in which subjects played the equilibrium action neither of the number of games in which they best responded to their stated beliefs.

explanation is that the equal split was feasible (or not) in *all* the games subjects played. As subjects were only paid for one of the games, our experiment resembles the strategy method, in which a weakening of the “equal split effect” has previously been observed (Güth et al. (2001)). In any case, and admitting these caveats, our results show that there are circumstances in which subjects do not change their behaviour whether equal splits are feasible or not when deciding how to share pies of given sizes.

Given that we have obtained that the different treatments did not have an effect on subjects' choices or beliefs stated, we use results 1 and 2 to pool the data across treatments. For the remainder of the analysis we will report statistics on pooled data, although we will refer to the different treatments when required.

4.3.3 Actions

In this section we study subjects' actions and their compliance with dominance, iterated dominance and equilibrium predictions.

Subjects almost never played strictly dominated strategies. Each subject could have played a strictly dominated strategy in five of the ten games He/She played. However only 21 out of the 800 actions taken were strictly dominated (2.65% of the total actions chosen and 5.25% out of the possible dominated actions). Dominated actions were taken in only a few games and for specific player roles: 8 dominated actions were taken by Row subjects in Game 1R, 4 by Column subjects in Game 1C, 6 by Column subjects in Game 3R and 3 by Column subjects in Game 4C. Only one subject (Column subject 1 in treatment BABAF) chose more than one dominated strategy across the ten games played (she chose dominated strategies in Game 1C and in Game 3R).

Now we look at dominant strategies. Only Games 1R and 1C had a dominant strategy for one of the players (DOWN for Row subjects in 1R, RIGHT for Column subjects in 1C). Out of the 80 subjects who had a dominant strategy in one of these games, 68 (85%) chose the dominant strategy. The 12 subjects who did not choose the dominant strategy all chose the same strategy across player roles (MIDDLE for Row in 1R, LEFT for Column in 1C) which accounts for 57% of the total of dominated actions chosen over all games. Although the action chosen by these 12 subjects coincided in all cases with the one that had the equal split cell in the F treatments, this action was actually chosen more times in U treatments than in F treatments (seven times against five). Also notice that LEFT in Game 1C, was not only strictly dominated by the dominant strategy (RIGHT), but also by the other dominated strategy (CENTRE). Row subjects' dominated choices may have been explained by a desire to have certainty over own payoffs, as when choosing the dominated action, Row subjects would obtain the same payoff no matter what their opponents' choices were. Column subjects' dominated actions are more difficult to explain.

We now look at whether the number of rounds of iterated deletion of dominated strategies in each game affected the percentage of actions according to equilibrium taken. Table 2 shows the percentage of compliance with equilibrium predictions for each game by subject role.

Game	Row Subjects	Column Subjects	All Subjects	N° Rounds Iterated Dominance (Row, Column)
1R	80	72.5	76.25	1, 2
1C	60	90	75	2, 1
2R	95	70	82.5	2, 3
2C	75	87.5	81.25	3, 2
3R	92.5	72.5	82.5	3, 3
3C	87.5	87.5	86.25	3, 3
4R	87.5	87.5	87.5	4, 3
4C	67.5	90	78.75	3, 4
NR	92.5	52.5	72.5	No
NC	72.5	75	73.75	No
Average	**81**	**78.5**	**79.625**	

Table 2: Percentage of equilibrium actions by game and subject role.

On average, subjects played equilibrium actions in 79.625% of the cases. Notice that there is no clear pattern between the number of rounds of iterated deletion of dominated strategies required to reach the equilibrium and the percentage of equilibrium actions played. For example, games 1R and 1C show a lower percentage of equilibrium actions than games 3C or 4R. We also noticed that the lowest percentage of equilibrium play occurred in the non-dominance solvable games (NR and NC). We created contingency tables with the number of subjects who played equilibrium actions in each of the games (aggregating both subject roles) and performed McNemar's tests[38] under the null hypothesis that there was no statistically significant difference in the proportion of compliance with equilibrium between each pair of games. We do not find statistically significant differences between games at the 5% level, when we group both subject roles.[39] When we do not, some differences are significant, for example between Row subjects in game 2R and NC, but no clear pattern emerges.

[38] In the following, we use McNemar's to exploit the statistical power derived from having the same subjects playing across different games. When this is not fulfilled, we use Chi-square tests.

[39] This creates seven categories: subjects who reach the equilibrium strategy in 1 round of iterated deletion, 2 rounds, 2 rounds with simultaneous deletion in the first round, 3 rounds, 3 rounds with simultaneous deletion in the first round, 4 rounds and non dominance solvable. Notice that not all these categories have the same number of subjects, but that the Chi-square test allows us to do this comparison.

Thus the degree of iterated dominance needed to reach equilibrium is not a straightforward measure of the proportion of equilibrium play and thus, this may indicate that if subjects really reasoned in game theoretic terms, deleting more rounds to reach the unique equilibrium is not a good indicator of how complex these games were for subjects.

Overall we conclude:

> Result 3: *Subjects almost never played strictly dominated strategies and played equilibrium strategies in 80% of the cases. The number of rounds of necessary deletion of strictly dominated strategies to reach the Nash equilibrium was not a clear indicator of the percentage with which the equilibrium strategies were played.*

We now compare how well the equilibrium model predicted actions taken in comparison to other models. Table 3 shows the percentage of actions taken that were predicted by the standard equilibrium model, together with the percentage rates predicted by each the other five models described in section 4.2.3.

Game	Equilibrium	L1	L2	L3	D1	Maximax
1R	76.25	51.25	76.25	76.25	76.25	51.25
1C	75	62.5	75	75	75	47.5
2R	82.5	17.5	62.5	82.5	62.5	17.5
2C	81.25	56.25	81.25	81.25	56.25	16.25
3R	82.5	38.75	82.5	82.5	82.5	38.75
3C	86.25	48.75	51.25	86.25	86.25	13.75
4R	87.5	50	87.5	87.5	87.5	12.5
4C	78.75	61.25	78.75	78.75	61.25	17.5
NR	72.5	66.25	22.5	62.5	66.25	21.25
NC	73.75	50	46.25	11.25	50	15
Average	**79.625**	**50.25**	**66.37**	**66.75**	**70.375**	**25.125**

Table 3: Percentage of equilibrium actions predicted by each model.

Equilibrium outperforms the predictions of the other models in all games.[40] Although the games were intentionally constructed to highlight differences between models' predictions, it is noticeable that the two models that have been most successful in previous research on depths of reasoning perform clearly worse across all games than the standard equilibrium (L1 predicts 50.25% of the actions, while L2 predicts

[40] Equilibrium also outperforms each of the other models in all games when subject roles are not pooled.

66.375%).[41] Of the models analyzed, the one that comes second in predicting the aggregate of actions is D1, with a percentage of 70.375%. D1 predicts the same action as Equilibrium for five of the ten games. In the five games were the predictions of both models are different, Equilibrium outperforms D1 in all games, with an overall success rate of 77.75% against 49.35%.

Notice however, that L1 has strong predictive value for non Equilibrium actions. The action that coincided with the L1 prediction was the one which was chosen with at least second highest frequency in all 10 games for both subject roles. Additionally, out of the 163 actions that were not taken according to Equilibrium, 98 (60.12%) were taken according to L1. As a reference, only 43 (26.38%) of the non Equilibrium actions taken coincided with L2, with lower percentages for the other models. Thus, most subjects when they did not choose the Equilibrium action, they chose the action that gave them the highest expected payoff against a uniform distribution of play by their opponent (L1). This gives some support to L1 as a decision model when subjects do not know what to choose and do not have any particular beliefs on how their opponents will choose.

We now look at individual behaviour. First, table 4 shows the cumulative distribution function (CDF) of the percentage of subjects who played at least a certain number of games according to each models' predictions. We observe that while 20% of the subjects played according to the Equilibrium prediction in all ten games, at most only 1.25% of the subjects played in all ten games according to any of the other models here studied. Also notice that 70% of the subjects chose at least 8 actions according to the Equilibrium model.

N° Predictions	Equilibrium	L1	L2	L3	D1	Maximax
10	20	1.25	1.25	0	1.25	0
9	43.75	3.75	2.5	3.75	16.25	0
8	70	10	27.5	31.25	37.5	1.25
7	81.25	16.25	61.25	55	70	3.75
6	87.5	36.25	78.75	76.25	85	6.25
5	96.25	65	87.5	91.25	95	8.75
4	98.75	87.5	93.75	98.75	98.75	15
3	98.75	95	98.75	100	100	28.75
2	100	100	100	100	100	43.75
1	100	100	100	100	100	75
0	100	100	100	100	100	100

Table 4: CDF Of the percentage of subjects who play at least a number of times according to models' predictions.

[41] Notice that L2 predicts the same outcome as Equilibrium in six games, while L1 does not predict the same outcome as Equilibrium in any game. Thus, we should not infer that L2 captures behavior better than L1. L3 coincides with Equilibrium in all but Games NR and NC, where it performs significanty worse.

Second, we classified subjects according to the model whose predicted action subjects chose in the highest number of games. Table 5 shows the percentage of subjects who could be classified according to each model category. First, there were 56 out of the 80 subjects that could be clearly classified to a model according to this criterium. i.e., who responded the highest number of times according to only one model. Of these, 69.6% of subjects were classified as "Equilibrium". There are 24 subjects who could not be classified in this manner, as there were ties between various models. This is the reason why the sum of percentages of columns "Ties" and "Overall" adds up to more than a hundred percent. In any case, 87.5% of the subjects who tied between two models, chose the highest number of actions according to "Equilibrium" and some other model, while only 50% did it according to "L1" and another model. In the column "Overall", we add up both the clear cases and the ties to conclude that 75% of the 80 subjects can be classified as "Equilibrium", while only 26.25% of subjects can be classified as D1. Other models show lower percentages. Finally, we show in parenthesis the average number of games in which subjects classified in each model category chose actions according to each model. Notice that this average measures the intensity with which subjects were classified with respect to each model and thus, it shows that subjects classified in each category were quite consistent with the model in which they were classified.[42]

Classification	**Clear Cases**	**Ties**	**Overall**
Equilibrium	69.6 (9.05)	87.5 (7.86)	75 (8.63)
L1	5.36 (9.33)	8.33 (6.5)	6.25 (8.2)
L2	5.36 (8.66)	41.66 (7.2)	16.25 (7.53)
L3	1.78 (9)	25 (7.33)	8.75 (7.33)
D1	16.07 (8.66)	50 (8.08)	26.25 (8.34)
Maximax	1.78 (8)	8.33 (6.5)	3.75 (7)

Table 5: Classification in models to which subjects respond most times.

Thus, we conclude:

Result 4: *Equilibrium captures actions played by subjects better than the alternative models, both at the individual and aggregate levels.*

Although we cannot discard that there may be other models that capture behaviour better than those studied here or that, as we have seen, a small percentage of players'

[42] Had subjects chosen actions randomly they would have answered on average in 3.3 games according to each model and, given the structure of the games, the average intensity of subjects classified in each category would have been 5.1.

behaviour might be better captured by one of the other models presented here, it is clear that Equilibrium is a good predictor of actions taken for the particular class of games we study in this experiment.

Below we check if subjects also believed that their opponents would play according to Equilibrium.

4.3.4 Stated Beliefs

In this section we study subjects' stated beliefs about opponents' frequencies of play. We study whether subjects expected their opponents to play dominated strategies, whether they expected them to comply with the equilibrium prediction and we also check the accuracy of the frequency predictions with respect to the frequency of actions chosen.

Subjects believed their opponents would play dominated actions with higher frequency than they actually did. Over all games, subjects assigned 6.575% of frequency to dominated actions (13.15% of the possible frequency that could be assigned in the five games with dominated actions for each role subject) while dominated actions were played in 2.625% of the cases (5.25%).

On average, subjects expected their opponents to play the Equilibrium action with the highest frequency in each game, although the frequency with which Equilibrium was played was higher than the frequency with which subjects believed it was going to be played. We will refer to these as "conservative beliefs", following Huck and Weizsäcker (2001).

Overall, subjects assigned on average 57.6% of frequency to their opponents playing the equilibrium action. Frequencies assigned to equilibrium play were disperse. The lowest average frequency assigned to equilibrium play is 35.2% by Row subjects in Game 1R to Column players' action. The highest, 81.7% by Column subjects in the same game. Notice that Row subjects have a dominant strategy in this game, so these results indicated that most Column subjects did expect Row subjects to play the dominant strategy, while Row subjects where uncertain of how Column subjects would choose, as Column subjects did not have dominated strategies. Results in game 1C extend this intuition, although percentages are less clear cut.

Table 6 shows the average frequency assigned to Equilibrium actions by game and subject role. We do not observe any straightforward pattern between the number of rounds of iterated elimination of dominated strategies required to solve for the Equilibrium and the frequency of beliefs of Equilibrium play stated. However, it is still the case that when aggregating subject roles, in games NR and NC the frequency assigned to equilibrium actions is lower than in most of the other games (although not lower than in 2C). Performing Wilcoxon tests under the null hypothesis that there are no differences in the median of the distribution of frequencies of beliefs assigned to equilibrium actions between pairs of games, we obtain statistically significant differences at the 5% level between on the one hand games 1C, 1R, 3R, 3C, 4R and 4C and on the other games 2C, NC and NC.[43] When we performed Wilcoxon tests for each player role, we obtained several statistically significant differences, although there was not a clear pattern between the number of rounds of iterated dominance and the frequency assigned to equilibrium actions by opponents.

Game	Row Subjects	Column Subjects	All Subjects	N° Rounds Iterated Dominance (Row, Column)
1R	35.25	81.75	58.5	1, 2
1C	68.75	50	59.375	2, 1
2R	45.25	66.5	55.875	2, 3
2C	57.25	45	51.125	3, 2
3R	63.25	66.25	64.75	3, 3
3C	53	73.25	63.125	3, 3
4R	53	66.25	59.625	4, 3
4C	67.75	50.25	59	3, 4
NR	45.5	60.25	52.875	No
NC	49.5	53.5	51.5	No
Average	53.8	61.3	57.575	

Table 6: Average frequency assigned to equilibrium actions.

We now look at how the Equilibrium model for beliefs performs in the aggregate with respect to the other five models we are considering. Table 7 shows the frequency assigned to the predictions of the six models for each of the games. We observe a clear pattern: although in all the games the highest frequency was assigned according to the Equilibrium model (but in the already mentioned Game 1R where the L1 model performs slightly better), the percentage of predictions captured by each model are much closer when we look at belief statements than when we look at actions. In particular, contrary to what happens with actions, the average percentage of frequencies matched with Equilibrium model predictions (57.6%) and the average percentage matched with the D1 model predictions (55.4%) are very close. Notice also that the order in which each of the models is successful is practically the same with beliefs stated as it happened with actions (although with beliefs L3 outperforms L2).

[43] The distribution of the median of game 2R was also statistically different than the distribution of the median of games 3R and 3C.

Game	Equilibrium	L1	L2	L3	D1	Maximax
1R	58.50	61.25	58.50	58.50	58.50	61.25
1C	59.38	51.38.	59.38	59.38	59.38	42.38
2R	55.88	33.00	56.13	55.88	56.13	33.00
2C	51.13	49.75	51.13	51.13	49.75	36.63
3R	64.75	44.38	64.75	64.75	64.75	44.38
3C	63.13	49.75	36.75	63.13	63.13	23.38
4R	59.63	49.63	59.63	59.63	59.63	30.63
4C	59.00	54.25	59.00	59.00	54.25	31.25
NR	52.88	49.00	29.63	19.13	49.00	28.00
NC	51.50	39.88	40.75	22.13	39.88	26.38
Average	**57. 58**	**48. 23**	**51. 56**	**51. 26**	**55. 44**	**35. 73**

Table 7: Frequencies stated matched by models' predictions.

We thus conclude:

Result 5: *While Equilibrium still captures belief statements better than the other models studied here, differences with other models are smaller than with actions.*

The lower predictive value of the Equilibrium model and the small differences between the predictive value of competing models in belief statements seems to be caused by a tendency to conservatism in belief statements already observed in previous experiments (Huck and Weizsäcker (2001), Costa-Gomes and Weizsäcker (2004)). This tendency is also reflected in the low percentage of belief statements that assigned frequency one to all ten opponents playing one particular strategy (11.125%). However, tendency to conservatism does not mean that subjects assigned equal frequency to their opponents playing each of their three available actions. The percentage of uniform belief statements[44] is only 5.875%, much lower, in fact, than the percentage of belief statements that assigned zero frequency to at least one of the opponents' actions (42%). Costa-Gomes and Weizsäcker (2004) argue that the higher percentage of zero-belief statements than uniform beliefs is a reason to discard the hypothesis that the tendency to conservative beliefs might be caused by risk aversion. They argue that since QSRs punish large mispredictions, risk averse subjects would avoid losses by making roughly uniform belief statements, which subjects did not make in most of the cases. However, notice that even a highly risk averse subject would state zero beliefs to two of his opponents' actions if he was sufficiently certain about the actions that all opponents would take in a particular game. The reason for conservatism seems to be different. Given the high percentage of Equilibrium actions played, and the lower expectations

[44] Defined as statements that assigned frequency of 3 to two actions and 4 to the other one.

of opponents playing Equilibrium, it seems that subjects genuinely believed that their opponents would play Equilibrium less frequently than they did.

We now study the accuracy of belief statements by comparing stated frequencies with the frequencies with which each action was actually played in each game. As belief statements were conservative we should not expect them to be very precise. Before looking at precision we assess accuracy of belief statements in the aggregate by looking at whether subjects predicted the "structure of frequencies" correctly. We define correct structure of beliefs as subjects assigning highest frequency to the actions which were played with highest frequency and assigning lowest average frequency to the actions which were played with lowest frequency. Table 8 compares, for each game and subject role, the average frequency with which each of the three actions was played by subjects, with the average percentage of stated frequency assigned by the opponents to those same actions. It is noticeable that for all but three comparisons, aggregate average beliefs get the "structure of frequencies" played correctly.[45]

Game	Row Actions			Column Beliefs		
	UP action	MIDDLE action	DOWN action	UP Belief	MIDDLE Belief	DOWN Belief
Game 1R	0	20	80	3	15.25	81.75
Game 1C	5	35	60	16	34	50
Game 2R	95	5	0	66.5	20.25	13.25
Game 2C	25	0	75	42.25	12.75	45
Game 3R	92.5	2.5	5	66.25	8.25	25.5
Game 3C	0	85	15	6.25	73.25	20.5
Game 4R	0	12.5	87.5	5.5	28.25	66.25
Game 4C	67.5	32.5	0	50.25	40.75	9
Game NR	2.5	92.5	5	18.25	60.25	21.5
Game NC	2.5	72.5	25	16.25	53.5	30.25
	Column Actions			Row Beliefs		
Game 1R	22.5	5	72.5	40.75	24	35.25*
Game 1C	2.5	7.5	90	11.5	19.75	68.75
Game 2R	30	0	70	45.75	9	45.25*
Game 2C	5	87.5	7.5	11.75	57.25	31
Game 3R	15	72.5	12.5	27.25	63.25	9.5
Game 3C	0	12.5	87.5	20.25	26.25	53
Game 4R	87.5	0	12.5	53	14	33
Game 4C	90	2.5	7.5	67.75	21.75	10.5**
Game NR	7.5	52.5	40	16.75	45.5	37.75
Game NC	5	20	75	22.5	28	49.5

Table 8: Comparison of the percentage with which actions were played with percentages of belief frequencies assigned.

[45] The difference between frequencies assigned in those three games was, however, very small. These games are indicated in Table 7 with a star (*). The double star (**) in game 4C indicates that the order of beliefs with which the second and the third actions were played was inverted.

However, when looking at each subject individually, we observe that the patterns of aggregate behaviour do not translate well into individual behaviour across games. Table 9 shows the cumulative distribution function of the percentage of subjects who assigned the highest frequency to the correct action, as well as the lowest frequency at least in a number of games. It also reports the percentage of subjects who predicted the correct and the opposite structure of frequencies of actions in at least a number of games. Table 9 shows that subjects were good at assigning highest frequency to the actions which was played with highest frequency, but they were not so good in ranking the two other actions.

Number of Games	Highest Frequency	Lowest Frequency	Same Structure	Opposite Structure
10	10	0	0	0
9	25	1.25	0	0
8	36.25	5	1.25	1.25
7	50	15	5	2.5
6	75	31.25	8.75	3.75
5	87.5	53.75	18.75	6.25
4	92.5	70	35	12.5
3	95	88.75	50	23.75
2	96.25	98.75	76.25	32.5
1	100	100	90	71.25
0	100	100	100	100

Table 9: CDF of the percentage of subjects who predicted the structure of actions by their oponents at least in a number of games.

While 75% of the subjects assigned highest frequency in six or more games to the action that was played with highest frequency, only 8.75% of the subjects did, at the same time, assigned the lowest frequency to the action that was played with lowest frequency in those six or more games, and thus, answered with the same structure of frequencies of beliefs as their opponents played. Very few subjects assigned the ranking of frequencies in the opposite order as they were played by their opponents (32.5% of the subjects made this mistake two or more times). Also, 28.25% of belief statements assigned the same frequency to the two actions that were not believed to be played with highest frequency.[46] Figure 1 shows the scatter plot for each game of the frequency of stated beliefs for each action against the frequency with which those actions were played. The tendency is clearly increasing, supporting the evidence that, on average, subjects assigned higher frequency to actions that were played with higher frequency.

[46] These cases do not qualify for either the "Same structure" or the "Opposite structure" categories.

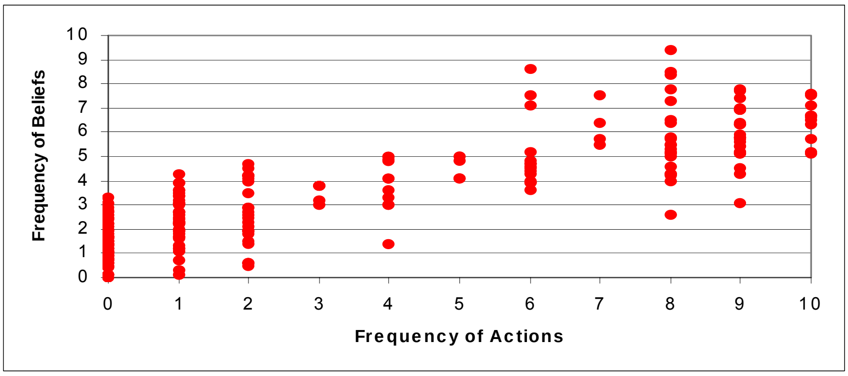

Figure 1: Average Frequency of Actions and Average Frequency assigned.

Finally, we look at the precision of belief statements. The average mean square error for the predictions of Row subjects about the frequencies of play of Column subjects was 2.49, while for Column subjects was 2.16. Random belief statements would have generated average mean square errors of 3.3 for Row Subjects and 4.1 for Column Subjects. We will use the average mean square errors in the next section to associate accuracy of predictions and best response behaviour. We conclude:

> Result 6: *Subjects were good at predicting the actions that were played with highest frequency by their opponents, although stated beliefs tended to be "conservative".*

4.3.5 Best Response of Actions to Stated Beliefs

We finally look at the consistency of stated beliefs and actions at the individual level. We check for consistency by checking whether actions chosen by each subject were best replies to the same subject's stated beliefs (BR). We define best replying behaviour as choosing the action that gives the highest expected payoff given the distribution of beliefs stated. According to this definition, best replying implies that subjects' utilities only depend on own monetary payoffs and that subjects are risk neutral. Results below show that a majority of subjects satisfied this definition. Given that subjects were better at identifying the action which was played with highest frequency than the frequencies of the other two actions, we also check whether actions taken were a best response only to the action assigned the highest frequency (BR Max F).[47]

First, as it would be obvious from previous results, subjects clearly best responded to their stated beliefs more often than they would have had they chosen their actions randomly. Kolmogorov-Smirnoff Goodness of Fit Tests comparing the empirical CDFs

[47] As there are only three actions available for each subject and only ten units of frequency to be assigned, in many cases both models of best response behavior predict the same action.

to the CDF implied by random behaviour gives p-values of virtually zero. Table 10 shows the percentage of best responses by game and player role. Overall, subjects best responded to their stated beliefs in 73.375% of the cases (75.625% for best response to the highest frequency belief). This percentage is much higher than the observed in the only other study with elicited beliefs we are aware of on one-shot behaviour in a similar setting (Costa-Gomes and Weizsäcker (2004) with 50% of best responses). Furthermore, even if in our experiment there is no chance for learning, the percentage of best response behaviour is as high as the one observed in experiments that allowed for learning (Nyarko and Schotter (2002), with 75% of best responses in their 2x2 games).

Thus, we conclude:

Result 7: *Subjects best responded to their stated beliefs a high number of times (73% of the cases).*

By comparing the percentage of best responses across games for all subjects using McNemar's test (5% significance level), we again observe the familiar pattern that the number of rounds of iterated dominance does not seem to affect in a clear way the percentage of best replies. However, it is true that the percentage of best responses was significantly lower in the two non-dominance solvable games (NR and NC) than in some of the other games. Both models of best response (BR and BR Max F) perform similarly, which may be due to stated beliefs not being too extreme.

	Row Subjects		Colum Subjects		All Subjects	
Game	BR	BR Max F	BR	BR Max F	BR	BR Max F
1R	80	80	77.5	75	78.75	77.5
1C	60	57.5	90	90	75	73.75
2R	90	92.5	77.5	75	83.75	83.75
2C	80	72.5	80	80	80	76.25
3R	80	90	72.5	72.5	76.25	81.25
3C	77.5	65	80	87.5	78.75	76.25
4R	82.5	82.5	85	75	83.75	78.75
4C	57.5	57.5	87.5	87.5	72.5	72.5
NR	60	52.5	40	80	50	66.25
NC	60	75	50	65	55	70
Average	**72.75**	**72.5**	**74**	**78.75**	**73.375**	**75.625**

BR: Best Response. BR Max F: Best Reponse to the action assigned highest frequency

Table 10: Average percentage of best responses.

Looking at the individual level, Figure 2 draws the empirical probability density function (PDF) of the number of games for which each subject best responded to their stated beliefs, overall and for each player role. Although only 3.75% of subjects best responded to their stated beliefs in all ten games, 70% of subjects best responded to their stated beliefs in seven or more games.

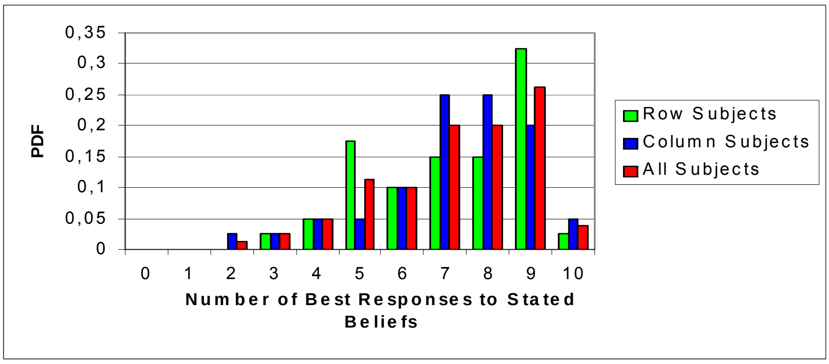

Figure 2: Empirical PDF of the number of times subjects best responded to their stated beliefs.

Although the proportion of non-best response behaviour is not insignificant, it is small. We look into the nature of non-best response behaviour by calculating how much subjects lost for not best responding to their stated beliefs. We use the monetary losses subjects made when non-best responding to their stated beliefs as a proxy for how important it was for subjects to best respond in each of the games.

We proceed by calculating, for each subject, the sum of his expected loss when not best responding to their stated beliefs averaged over the ten games each subject played. We find that Row subjects lost on average £0.3037 per game and Column subjects lost on average £0.3205 per game. Given that subjects were only paid for their actions in one game, these were the average losses per subject. Next, we calculate the average maximum feasible loss had subjects have played, in all games, the action that gave them the lowest possible expected payoff, given their stated beliefs. On average, Row subjects could have lost £3.05 per game while Column Subjects could have lost £2.69 per game. Finally, we divide both numbers to calculate for each subject in each game, the percentage of the maximum loss they incurred by not best responding. Averaging over all games for each subject role we obtain that Row subjects lost on average 10.97% of the maximum losses they could have made, while Column subjects lost 15.96% of the maximum possible losses. To put things in perspective, Row subjects would have lost 40.21% of the maximum possible losses they could have made had they chosen the action that neither was a best response nor the worst response to their stated beliefs in all ten games. Column subjects would have lost 55.24% of the maximum possible losses had their chosen this action in all 10 games. Therefore, we conclude:

Result 8: *As subjects best responded in most of the games, they did not lose much with respect to the maximum losses they could have made.*

Table 11 shows the percentage of average losses in each game for each subject role. Notice that these calculated losses are only hypothetical, as we obtain them using stated beliefs, not the actual matching of subjects in each game.[48] Differences between average losses for Row and Column subjects were probably caused by a variety of factors, including the beliefs stated but also the design of the games, which created higher payoff differences for Column subjects than for Row subjects.

Game	Row Subjects	Column Subjects
1R	9.56	14.61
1C	22.16	5.49
2R	8.73	3.74
2C	5.17	13.99
3R	6.12	21.60
3C	12.01	13.86
4R	6.72	10.33
4C	6.11	7.79
NR	16.81	46.64
NC	16.32	21.52
Average	**10.97**	**15.96**

Table 11: Percentage of average loss per game and subject role.

Not best responding is not the only kind of mistake subjects could have made. Subjects could also err in the accuracy of their predictions of opponents' play. Although the monetary loss derived from this mistake would be minimal, as payments for stated beliefs have an upper bound of £2, a bad prediction of how opponents play, even if it was a best response to stated beliefs, could result in taking a non-optimal action, given the frequencies with which opponents really played. We address whether both types of mistakes (bad predictions and non-best response behaviour) are related, by calculating the correlation between each subjects' average mean square error of his predictions and the average percentage of maximum loss for not best responding each subject makes. We find that there is positive significant correlation between both series (Pearson's coefficient of 0.559 with a p-value of 6.8e-08).[49] This high correlation means that subjects who chose equilibrium actions, also expected a high proportion

[48] An alternative way of calculating the hypothetical loses is to use the real frequencies of play by the opponents instead of the stated beliefs. Given that the percentage of best response to stated beliefs is similar to the percentage of best response to "real" play by the opponents, overall percentages by game differ only slightly.

[49] We also calculated the correlation between each subject's number of best responses with the mean square error of predictions and Pearson's coefficient was, as expected, negative and significant (Pearson's coefficient: 0.55, p-value 8.6e-08).

of their opponents to choose equilibrium actions, and that this prediction was right. This suggests that subjects may have believed that their opponents would choose their actions in a similar way as they did. We thus conclude:

> Result 9: *Subjects who are better at predicting the frequencies of play of their opponents are also the ones who lost, on average, less for not best responding.*

Figure 3 draws the correlation between average losses and errors of predictions.

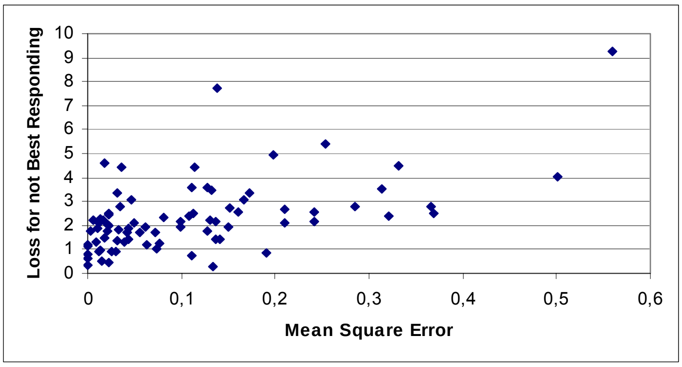

Figure 3: Correlation Between Average Losses for Not Best Responding and Mean Squared Errors of Predictions

4.4 Discussion

We have identified a class of non-trivial games for which game-theoretical predictions work reasonably well, even when games are played for the first time by subjects with no previous experience in laboratory games or knowledge of game theory. These games are constant sum games with unique equilibria in pure strategies. Our results imply that most subjects not only played according to the Equilibrium prediction but that they were reasonably good at predicting the actions that would be played with highest frequency by their opponents and they best responded to their beliefs on opponents play.

When surveying the experimental evidence in dominance solvable games, Camerer (2003, Chapter 5), claims that the joint hypothesis of game theoretic behaviour and social preferences that value only one's own payments is easily rejected. He then claims that the interesting question is whether the rejection is due to the pure self-interest part of the joint hypothesis or to the game theoretic reasoning part or even to both. We have here designed a simple experiment in which by using a theoretically useful control for social preferences, we check if subjects play according to the game theoretic prediction, and thus, this may indicate whether subjects are able to reason in game theoretic terms. Notice that this procedure does not allow us to answer whether individuals have social preferences or not, but only helps us to identify a class of games in which whether they have social preferences or not, the equilibrium prediction is reasonably accurate. Therefore, for our simple but non trivial games, the game theoretic part of the hypothesis is not rejected in a context in which we would not expect social preference to influence behaviour.

The number of rounds of iterated deletion of strictly dominated strategies to reach the equilibrium is not a straightforward measure of how complex games are for subjects and thus we can not conclude that in games with more rounds of deletion, the percentage of equilibrium played was lower. Whether games are dominance solvable or not seems to have an effect on the predictive power of game theory over these games, specially for the percentage of best responses. However, notice that in those two games at least one of the subjects could obtain the same payoff by choosing some particular strategy no matter the action chosen by their matched opponent. This, together with the impossibility of deleting dominated strategies, may have caused that subjects were worse able to predict how their opponents would play and, as a consequence, the percentage of equilibrium actions may have been lower in these games. We aim to conduct further research on non dominance solvable games in which this unfortunate characteristic is not present.

In any case, constant sum games seem like a good starting point to study how subjects reason in simple games as issues like fairness and efficiency concerns seem not to affect their choices. Equal splits neither influenced actions chosen nor beliefs stated about others' behaviour, although this point requires further investigation as the payoff differences between equal and unequal splits were low in our experiment. Belief elicitation does not affect how subjects play games and how they think other subjects will play them, although comparing our results to previous experiments, the different belief elicitation procedure we used may be important.

Our results may have been influenced by procedural changes with respect to previous experiments and in particular, to Costa-Gomes and Weizsäcker (2004). Apart from focusing on constant sum games, we used one-digit numbers to represent payoffs and there was no conversion rates between experimental payoffs and final monetary payoffs. The procedure for eliciting beliefs may be also of some importance when studying best response to stated beliefs. Given our results, it would be interesting to study how each of these changes affect results. It would also be interesting to estimate a model with noise, as they did, in which to jointly check the consistency between subjects' actions and the beliefs they have about their opponents. However, our results already hint that whether games are constant sum or not and some of the procedural changes we made make a difference in the predictive value of game theory. It seems wrong to generally dismiss Nash equilibrium as a good predictor of behaviour in simple games even if they are played for the first time by subjects with no particular training in Economics. Once we have this evidence, further research should aim to identify reasons for differences with previous evidence and ultimately, identify a possibly larger set of games for which game theory predictions work well.

4 Instructions for Chapter 4 (BABA Treatments)

WELCOME TO OUR EXPERIMENT!

This is a serious scientific experiment and, as such, no talking, looking around or walking around will be permitted. If you have any questions or need any assistance, please raise your hand and an experimenter will come to you. If you talk, exclaim out loud, etc, YOU WILL BE ASKED TO LEAVE AND YOU WILL NOT BE PAID. Thank you.

This is an experiment on individual decision making. The ESRC Centre for Evolutionary Learning and Social Evolution (ELSE) has provided the funds for this experiment. You will be paid £5 (five pounds) for having arrived on time. Additionally, if you follow the instructions and pass an Understanding Test you will be allowed to continue in the experiment. Once in the experiment, depending on your decisions you may earn a considerable additional amount of money. This additional amount will be determined both by your decisions and by those of other participants in the experiment. Before making your decisions, you will be informed about how your earnings and the other participants' earnings depend on your and their decisions. All that you earn is yours to keep, and will be paid to you in private, in cash, after today's session.

We need 20 people for this session. Thus, if more than 20 people pass the Understanding Test, some of you will be asked not to participate in the experiment but to help the experimenter as "assistants". These assistants will check that everything is done as explained in the instructions. The assistants will be paid the average of the payments of the 20 participants in the experiment.

For each decision you take in the experiment, You will be anonymously matched with one of the other participants. We will refer to the other participants as "S / HE". You and s/he will be presented with a TABLE. For this table, You and S/HE separately and independently will make a DECISION. Together, the two decisions determine the number of POUNDS each of you earns, which may be different.

The table in the next page shows an illustrative example. IT IS ONLY AN ILLUSTRATION. The tables you will see during the experiment will be different from this one. AS YOU LOOK AT THIS TABLE, PLEASE CONTINUE READING THIS HANDOUT FOR INSTRUCTIONS ON HOW TO UNDERSTAND THE TABLE:

S / HE

		LEFT	CENTRE	RIGHT
	UP	9 1	8 2	7 3
YOU	MIDDLE	6 4	5 5	4 6
	DOWN	3 7	2 8	1 9

In the actual experiment, you will be shown tables like this one (but with different numbers), and asked to choose one of your decisions ("UP", "MIDDLE" or "DOWN"). The participant to whom you are matched for each table will be asked, independently, to choose one of her/his decisions ("LEFT", "CENTRE" or "RIGHT").

The combination of your decision and her/his decision determines the cell of the table chosen. The number of pounds you and s/he receive for the cell chosen is a whole number ranging from 1 to 11.

The number of pounds you receive appears in the lower left corner of each cell of the table.

The number of pounds s/he receives appears in the upper right corner of each cell of the table.

To interpret the table, consider the results of the possible combinations of decisions.

-If you choose UP and S/HE chooses LEFT, you earn 1 Pound and S/HE earns 9 Pounds.
-If you choose UP and S/HE chooses CENTRE, you earn 2 Pounds and S/HE earns 8 Pounds.
-If you choose UP and S/HE chooses RIGHT, you earn 3 Pounds and S/HE earns 7 Pounds.
-If you choose MIDDLE and S/HE chooses LEFT, you earn 4 Pounds and S/HE earns 6 Pounds.
-If you choose MIDDLE and S/HE chooses CENTRE, you earn 5 Pounds and S/HE earns 5 Pounds.
-If you choose MIDDLE and S/HE chooses RIGHT, you earn 6 Pounds and S/HE earns 4 Pounds.
-If you choose DOWN and S/HE chooses LEFT, you earn 7 Pounds and S/HE earns 3 Pounds.
-If you choose DOWN and S/HE chooses CENTRE, you earn 8 Pounds and S/HE earns 2 Pounds.
-If you choose DOWN and S/HE chooses RIGHT, you earn 9 Pounds and S/HE earns 1 Pound.

Please be sure you understand this table. Raise your hand if you would like further explanation. Otherwise, please start with the Understanding Test in the next page. Please raise your hand once you have finished the Understanding Test.

UDERSTANDING TEST

You will now take a short UNDERSTANDING TEST. After you have finished the TEST, it will be graded and you will ONLY be allowed to continue in the experiment if you have answered ALL the QUESTIONS CORRECTLY. If one or more of your answers is not correct, we will ask you to be our assistant and to check that everything proceeds as explained in the instructions. Notice that even if all your answers are correct, you may be asked to be our assistant.
This test has 5 questions. After you have answered all 5 questions, please re-check your answers. Please raise your hand when you are finished so as we can grade this test.

S / HE

		LEFT	CENTRE	RIGHT
	UP	2, 6	7, 3	5, 4
YOU	MIDDLE	9, 8	6, 9	4, 2
	DOWN	1, 1	3, 5	3, 7

Using the table above, please answer the following questions.

Questions:

1. If you choose MIDDLE and S/HE chooses RIGHT, how many Pounds will you earn?

2. If you choose UP and S/HE chooses LEFT, how many Pounds will S/HE earn?

3. If you choose UP and S/HE chooses RIGHT, how many Pounds will you earn?

4. If you choose DOWN and S/HE chooses CENTRE, how many Pounds will you earn?

5. If you choose DOWN and S/HE chooses LEFT, how many Pounds will S/HE earn?

YOU HAVE JUST COMPLETED THE TEST.
Please re-check your answers and raise your hand when you are done.

INSTRUCTIONS

There are 20 participants in this experiment. We have randomly divided the 20 participants in two groups of 10 participants. Everyone has been recruited for this experiment using the same procedure and everyone is receiving the same instructions.
In this experiment we are going to show you 10 different tables, similar to the one you have already seen.
For each table, you will have to answer two questions. One question asks you to choose a decision and the other question asks you about what you THINK other people's decisions are. Below we explain how to answer these questions and how you will be paid for your answers.

For each table, you have to choose between "UP", "MIDDLE" and "DOWN".
The 10 participants in the other group, choose between "LEFT", "CENTRE" and "RIGHT" in each of the 10 tables.

For the first question, you will have to write down how many of the 10 participants in the other group YOU THINK have chosen each of their 3 options (LEFT, CENTRE and RIGHT) in each of the 10 tables.

For the second question, you have to circle your decision (UP, MIDDLE or DOWN) in each of the 10 tables.
Notice that for each of the 10 tables, you have been anonymously and randomly matched with one of the 10 participants in the other group (who chooses between "LEFT", "CENTRE" and "RIGHT").
YOU HAVE BEEN MATCHED WITH A DIFFERENT PARTICIPANT IN EACH TABLE.
None of the participants will know who they are matched with in each table.

Lets see an example on how to answer the 2 questions:

In the table below I have written down that out of the 10 participants in the other group, I THINK 4 will choose "LEFT", I THINK 1 will choose "CENTRE" and I THINK 5 will choose "RIGHT". (Notice that guesses about how others play must always add up to 10).
Also, I have circled "MIDDLE" to indicate that "MIDDLE" is my decision.

Example:
Out of the 10 participants in the other group I think they will choose:

LEFT	CENTRE	RIGHT	TOTAL
__4___	__1___	__5___	__10__

		S/HE LEFT	S/HE CENTRE	S/HE RIGHT
YOU	UP	5 2	4 8	3 4
YOU	(MIDDLE)	7 5	4 11	3 7
YOU	DOWN	9 3	11 1	8 5

Caution: The numbers used in this example were selected arbitrarily.They are NOT intended to suggest how anyone might respond in any situation.

Below we explain how we will pay you according to your answers and the answers of your matched participant to each of the 2 questions.

PAYMENT FOR YOUR ANSWERS TO QUESTION 1

You will be paid an amount of money according to the difference between the number of participants in the other group you have guessed have chosen each option ("LEFT", "CENTRE" or "RIGHT") and the actual number of participants in the other group who have in fact chosen each decision.

We will pay you according to a formula that we explain below. Do not worry if the formula seems complicated as it is not important that you understand the workings of the formula completely. However, notice that with this formula, the closer the numbers you write down (your "guesses") to the actual number of participants who have chosen each decision in each table, the more money you will get.

For example, if 6 participants have chosen LEFT and you guessed that 6 participants would choose LEFT, you would get more money than if you guessed that 5 or 7 participants would choose LEFT for that table.

Here is the formula:

$$Payment = 2 - 0.01 * [(a - X)^2 + (b - Y)^2 + (c - Z)^2]$$

Where:

a: Number of participants you think have chosen LEFT	*X: Number of participants who have chosen LEFT*
b: Number of participants you think have chosen CENTRE	*Y: Number of participants who have chosen CENTRE*
c: Number of participants you think have chosen RIGHT	*Z: Number of participants who have chosen RIGHT*

Please follow the next examples to see how the formula works.

Examples:

-In some table, you write that you think 7 participants have chosen LEFT, 0, participants have chosen CENTRE and 2 participants have chosen RIGHT. If, in fact, 7 participants have chosen UP, 0 participants have chosen CENTRE and 2 participants have chosen RIGHT, you get:

$$Payment = 2 - 0.01* [(7-7)^2 + (0-0)^2 + (2-2)^2] = 2- 0.01*[0] = \textbf{2 Pounds}.$$

- In some other table, you write that you think, 5 participants have chosen LEFT, 2 participants have chosen CENTRE and 3 participants have chosen RIGHT. If, in fact, 1 participant has chosen UP, 8 participants have chosen CENTRE and 1 participant has chosen RIGHT, you get:

$$Payment = 2 - 0.01 [(5-1)^2 + (2-8)^2 + (3-1)^2] = 2 - 0.01[4^2 + 6^2 + 2^2] =$$
$$= 2 - 0.01*[56] = \textbf{1.44 Pounds}$$

- Finally, in some other table, you write that you think, 0 participants have chosen LEFT, 10 participants have chosen CENTRE and 0 participants have chosen RIGHT. If, in fact, 10 participants have chosen UP, 0 participants have chosen CENTRE and 0 participants have chosen RIGHT, you get:

$$Payment = 2 - 0.01\,[(0\text{-}10)^{2} + (10\text{-}0)^{2} + (0\text{-}0)^{2}] = 2 - 0.01[0] = \mathbf{0\ Pounds}$$

Caution: The numbers used in this example were selected arbitrarily. They are NOT intended to suggest how anyone might respond in any situation.

These examples should illustrate that with this formula you will always receive a payment of at least £0 and at most £2 for question 1 and that you will earn more money the more accurate your written guesses are.

PAYMENT FOR YOUR ANSWERS TO QUESTION 2

You will be paid a number of pounds equal to the number that appears in the lower left corner of the cell that you and your matched participant in that table have chosen. Your matched participant in the table will be paid the amount of pounds that appears on the upper right corner of the cell that you and her/him have chosen.

FINAL INSTRUCTIONS

We will wait until all participants have finished answering the 2 questions in the 10 tables. Please take some time to think and check your answers. We will allow a maximum of 40 minutes to answer all questions. **Please, if you finish before time raise your hand and we will collect your answers.** However, you are asked to remain in your seat quiet until all participants have finished.
After all participants have finished, we will randomly select ONE table from which all payments to all participants will be done. This table will be selected using a bingo urn with 10 numbered balls. The number on the ball selected determines for which of the 10 tables all participants are paid for both question 1 and question 2.
You will be paid the sum of three things:

- £5 for arriving on time
- The result of applying the formula explained in question 1 to the selected table.
- The amount of pounds indicated in the lower left corner of the cell that you and your matched participant in the selected table have chosen.

You have been given an identification number. Please write this number at the top of each of your answer sheets and keep the number. You will need this number to be paid.
While we calculate the payments you will be asked to fill in an anonymous questionnaire. After we have done the calculations, you will be asked to come with the questionnaire and your identification number to a room where you will be paid your earnings in cash and in private.

PLEASE WAIT UNTIL THE EXPERIMENTER TELLS YOU TO START
(Please raise your hand if there are any doubts with these instructions, and we will answer them privately)

5 The Games in Chapter 4

Below we show each of the ten games subjects played in treatments BABAU and ABU. Below each game we indicate the cells that were changed to create treatments BABAF and ABF. Additionally, we indicate the predictions of each of the six models studied ("Eq" for Equilibrium, "Max" for Maximax, other names are identical), the percentage of subjects who played each action ("Act"), the percentage of beliefs assigned to that action ("Bel") and the round of iterated strict dominance in which a strategy would be deleted ("*Do1*", "*Do2*", "*Do3*" and "*Do4*").

Game 1R

Column

Row	Left *Do2*	Centre *Do2*	Right	
Up *Do1*	3, 9	4, 8	5, 7	Act: 0 Bel: 3
Middle *Do1*	5, 7	7, 5	7, 5	Act: 20 Bel: 15.25
Down	9, 3 L1 Max	9, 3	8, 4 Eq L2 L3 D1	Act: 80 Bel: 81.75
	Act: 22.5 Bel: 40.75	Act: 5 Bel: 24	Act: 72.5 Bel: 35.25	

In treatments BABAF and ABF, the Middle-Left Payoff was changed by (6,6).

Game 1C

Column

Row	Left *Do1*	Centre *Do1*	Right	
Up *Do2*	10, 2	2, 10	1, 11 Max	Act: 5 Bel: 16
Middle *Do2*	9, 3	8, 4	2, 10 L1	Act: 35 Bel: 34
Down	7, 5	4, 8	3, 9 Eq L2 L3 D1	Act: 60 Bel: 50
	Act: 2.5 Bel: 11.5	Act: 7.5 Bel: 19.75	Act: 90 Bel: 68.75	

In treatments BABAF and ABF, the Down-Left Payoff was changed by (6,6).

Game 2R

Row	Column: Left *Do3*	Centre *Do1*	Right	
Up	**5**, **7** (L2, D1)	**5**, **7**	**4**, **8** (Eq, L3)	Act: 95 Bel: 66.5
Middle *Do2*	**2**, **10** (L1, Max)	**11**, **1**	**3**, **9**	Act: 5 Bel: 20.25
Down *Do2*	**1**, **11**	**10**, **2**	**3**, **9**	Act: 0 Bel: 13.25
	Act: 30 Bel: 45.75	Act: 0 Bel: 9	Act: 70 Bel: 45.25	

In treatments BABAF and ABF, the Up-Left Payoff was changed by (6,6).

Game 2C

Row	Column: Left *Do2*	Centre	Right *Do2*	
Up *Do3*	**11**, **1**	**4**, **8** (L1, D1)	**7**, **5** (Max)	Act: 25 Bel: 42.25
Middle *Do1*	**4**, **8**	**4**, **8**	**1**, **11**	Act: 0 Bel: 12.75
Down	**7**, **5**	**5**, **7** (Eq, L2, L3)	**7**, **5**	Act: 75 Bel: 45
	Act: 5 Bel: 11.75	Act: 87.5 Bel: 57.25	Act: 7.5 Bel: 31	

In treatments BABAF and ABF, the Down-Left Payoff was changed by (6,6).

Game 3R

Column

Row	Left *Do3*	Centre	Right *Do1*	
Up	**5**, **7**	**4**, **8** Eq L2 L3 D1	**5**, **7**	Act: 92.5 Bel: 66.25
Middle *Do1*	**3**, **9**	**1**, **11**	**4**, **8**	Act: 2.5 Bel: 8.25
Down *Do2*	**3**, **9**	**3**, **9** L1 Max	**11**, **1**	Act: 5 Bel: 2.55
	Act: 15 Bel: 27.25	Act: 72.5 Bel: 63.25	Act: 12.5 Bel: 9.5	

In treatments BABAF and ABF, the Up-Left Payoff was changed by (6,6).

Game 3C

Column

Row	Left *Do1*	Centre *Do2*	Right	
Up *Do1*	**9**, **3**	**1**, **11**	**8**, **4**	Act: 0 Bel: 6.25
Middle	**10**, **2**	**10**, **2** L1	**9**, **3** Eq L3 D1	Act: 85 Bel: 73.25
Down *Do3*	**8**, **4**	**11**, **1** Max	**7**, **5** L2	Act: 15 Bel: 20.5
	Act: 0 Bel: 20.35	Act: 12.5 Bel: 26.39	Act: 87.5 Bel: 53.27	

In treatments BABAF and ABF, the Down-Right Payoff was changed by (6,6).

Game 4R

Column

Row	Left	Centre *Do2*	Right *Do4*	
Up *Do1*	**8** / **4**	**10** / **2**	**11** / **1**	Act: 0 Bel: 6
Middle *Do3*	**7** / **5**	**1** / **11**	Max **8** / **4**	Act: 12.5 Bel: 28.25
Down	Eq **5** L2 / **7** L3 D1	**4** / **8**	L1 **2** / **10**	Act: 87.5 Bel: 66.25
	Act: 87.5 Bel: 53	Act: 0 Bel: 14	Act: 12.5 Bel: 33	

In treatments BABAF and ABF, the Middle-Left Payoff was changed by (6,6).

Game 4C

Column

Row	Left	Centre *Do3*	Right *Do1*	
Up	Eq **5** L2 / **7** L3	**4** / **8**	**3** / **9**	Act: 67.5 Bel: 50.25
Middle *Do4*	L1 **7** D1 / **5**	Max **1** / **11**	**3** / **9**	Act: 32.5 Bel: 40.75
Down *Do2*	**9** / **3**	**11** / **1**	**2** / **10**	Act: 0 Bel: 9
	Act: 90 Bel: 67.75	Act: 2.5 Bel: 21.75	Act: 7.5 Bel: 10.5	

In treatments BABAF and ABF, the Middle-Left Payoff was changed by (6,6).

Game NR

Column

Row		Left	Centre	Right	
	Up	**4** **8**	**7** **5**	**11** Max **1**	Act: 2.5 Bel: 18.25
Row	Middle	**7** **5**	Eq **7** **5**	L1 **7** D1 **5**	Act: 92.5 Bel: 60.25
	Down	L3 **10** **2**	**7** **5**	L2 **5** **7**	Act: 5 Bel: 21.5
		Act: 7.5 Bel: 16.75	Act: 52.5 Bel: 45.5	Act: 40 Bel: 37.75	

In treatments BABAF and ABF, the Down-Right Payoff was changed by (6,6).

Game NC

Column

Row		Left	Centre	Right	
	Up	**11** **1**	**5** L3 **7**	**9** **3**	Act: 2.5 Bel: 16.25
Row	Middle	**8** **4**	**8** L2 **4**	Eq **8** **4**	Act: 72.5 Bel: 53.5
	Down	Max **4** **8**	**10** **2**	L1 **9** D1 **3**	Act: 25 Bel: 9
		Act: 5 Bel: 22.5	Act: 20 Bel: 28	Act: 75 Bel: 49.5	

In treatments BABAF and ABF, the Up-Centre Payoff was changed by (6,6).