IZA

DISCUSSION PAPER SERIES

IZA DP No. 3627

Information and Beliefs in a Repeated Normal-Form Game

Dietmar Fehr
Dorothea Kübler
David Danz

August 2008

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Information and Beliefs in a Repeated Normal-Form Game

Dietmar Fehr
Technical University Berlin

Dorothea Kübler
Technical University Berlin
and IZA

David Danz
Technical University Berlin

Discussion Paper No. 3627
August 2008

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 3627
August 2008

ABSTRACT

Information and Beliefs in a Repeated Normal-Form Game*

We study beliefs and choices in a repeated normal-form game. In addition to a baseline treatment with common knowledge of the game structure and feedback about choices in the previous period, we run treatments (i) without feedback about previous play, (ii) with no information about the opponent's payoffs and (iii) with random matching. Using Stahl and Wilson's (1995) model of limited strategic reasoning, we classify behavior with regard to its strategic sophistication and consider its development over time. We use belief statements to check for the consistency of subjects' actions with the stated beliefs as well as for the accuracy of their beliefs (relative to the opponent's true choice). In the baseline treatment we observe more sophisticated play as well as more accurate beliefs and more best responses to beliefs over time. We isolate feedback as the main driving force of learning to play strategically and to form beliefs that accurately predict the behavior of the opponent.

JEL Classification: C72, C92, D84

Keywords: experiments, beliefs, strategic uncertainty, learning

Corresponding author:

Dorothea Kübler
Faculty of Economics and Management
Technical University Berlin
Straße des 17. Juni 135
10623 Berlin
Germany
E-mail: d.kuebler@ww.tu-berlin.de

* For valuable comments, we thank Harald Uhlig, Georg Weizsäcker, Axel Werwatz and seminar participants at Humboldt-Universität zu Berlin, the European University Institute Florence, the SFB 649 Workshop 2007, the ESA World Meeting 2007 (Rome), the IMEBE 2008 (Alicante) and the Econometric Society NASM 2008 (Pittsburgh). We are indebted to Jana Stöver and Susanne Thiel for research assistance. Financial support from the Deutsche Forschungsgemeinschaft (DFG) through SFB 649 "Economic Risk" is gratefully acknowledged.

1 Introduction

The literature on learning has opened the black box of how an equilibrium is reached. Numerous theoretical and experimental papers have studied learning over a large number of periods and have focused either on the convergence properties of the learning algorithms or on the evolution of observed behavior in experimental data. Here, we focus on the development of strategic behavior in relatively few periods of play. The idea is to take a microscopic view of how beliefs and choices change over time, controlling for the role of information in this process.

We use a repeated two-person normal-form game with a unique Nash equilibrium of the stage game. In this relatively simple setup, we observe whether subjects learn to play the game in the sense that the Nash-equilibrium strategy is chosen more often in later than in earlier periods. A novel feature of the experiment is that we elicit the beliefs of a player about the action of the other player in every period. Thus, we can observe the joint development of beliefs and actions over time. This allows us to answer a number of questions in a dynamic setting that up to now have only been studied in one-shot games.

A widely used classification of behavior is the level-of-reasoning model of Stahl and Wilson (1995). In this model players can be distinguished by their levels of strategic thinking. Players with no sophistication randomize uniformly over their strategy space (level-0 type), whereas a player with one step of thinking best responds to $L0$-types, etc. The model also incorporates other types to capture perfectly rational behavior like the Nash or rational expectation types. We use the most prominent rules ($L1, L2$ and Nash) to classify the actions of our game according to their strategic sophistication and to study how the sophistication of players change over time.

This categorization of choices according to their strategic sophistication is complemented by the elicited beliefs. We show that the beliefs stated by the participants are better predictors of the actual choices than the beliefs estimated from choices with belief-learning models. Assuming that both stated beliefs and estimated beliefs are only a proxy of the true underlying beliefs, we can conclude that stated beliefs are the better proxy. We then use the stated beliefs to analyze whether players' actions are best responses to their beliefs more frequently in later than in earlier periods of the experiment. In addition, we study whether beliefs become more accurate in predicting the opponents' actual behavior in later periods. The (in)accuracy of beliefs can be interpreted as a measure of strategic uncertainty.

In order to better understand the reasons for the development of actions and beliefs over time, we vary the information that is available to the players. Learning theories typically make use only of a limited amount of information. To be able to separate between different forms of learning, we run a baseline treatment with full information about the game and with feedback about one's own payoff (and thereby the other's payoff and action) in the previous period. In addition, we employ a treatment where subjects do not get any feedback about the outcome of play in the previous period and a treatment where subjects do not know the payoffs of the other player in the game, only their own payoffs. As we change only one aspect at a time, we can observe which kind of information is important for the learning process. Finally, we control for repeated game affects by running a treatment with random matching in every period.

Two extreme learning patterns can be distinguished with our experiment. First, subjects can learn inductively, based on the history of play. Players look back to determine which strategy to choose in the next period. For example, belief learning and reinforcement learning fall into this category. Second, deductive reasoning implies that players analyze the game in order to understand its strategic properties and thereby form beliefs about the opponent's choice. This learning without feedback requires more sophistication of the players than most inductive learning algorithms. While both forms of learning have already been studied in different experiments, we provide a unified framework to compare no-feedback learning with inductive learning. The treatment without feedback information and the treatment without information about the opponent's payoffs allow us to separate the two forms of learning. Using the level-of-reasoning model, we characterize behavior as strategic or non-strategic and can then evaluate under which information conditions subjects learn faster to play strategically than in others.

Concerning the results, we find an initially high level of non-strategic behavior in all treatments, i.e., subjects tend to neglect the incentives of their opponents. In the baseline treatment with full information about the game and feedback about past outcomes, this non-strategic behavior decreases in later periods. Experience also has a moderate positive impact on the accuracy of beliefs and on the best-response rates in the baseline treatment.

The control treatments show that the learning path crucially depends on the information available. Information about the other player's payoffs is important for initial play, but not as much as to be expected from rational players. Thus, subjects seem to have only a limited understanding of the strategic properties of the game initially, even if they have full information about the game. Also, behavior over time is very similar in treatments with and without information about the opponent's payoff function. However, our results indicate the importance of feedback. In the treatment without feedback about past outcomes, there is virtually no change in behavior over time. Thus, independent of whether subjects know the complete game or only their own payoffs, it is the experience through feedback which reduces non-strategic behavior.

Both in standard Nash equilibrium and in the level-of-reasoning model, players are assumed to best respond to their beliefs. However, best-response rates are initially only between 50% and 60% in the baseline treatment. We observe an increase in best responses over time in the baseline treatment, but not in any other treatment. Thus, receiving information about the past play of one's opponent and about his incentives in the game allows subjects to learn to best respond. Regarding the accuracy of belief statements in predicting the opponent's behavior, repeated interaction with the same opponent and information about his past choices are the main determinants of success in this task.

The literature related to this study can be summarized as follows. First, the level-of-reasoning model by Stahl and Wilson (1995) has been applied to a number of data sets based on 3x3 one-shot normal-form games. Costa-Gomes, Crawford and Broseta (2001) study decision rules and use the mouselab technique to record how subjects use payoff information. Costa-Gomes and Weizsäcker (2008) elicit subjects' beliefs about the other player's choice and find that subjects perceive the game differently when asked for beliefs than when playing it themselves. Rey-Biel (forthcoming) focuses on constant-sum games to analyze the dependency of equilibrium predictions on the game characteristics. Finally, Ivanov (2006) combines the level-of-reasoning approach with risk aversion to explain observed behavior.

Repeated normal-form games with belief elicitation have been studied in two other papers. Nyarko and Schotter (2002) focus on the matching-pennies game to compare stated beliefs with Cournot and fictitious-play beliefs. Ehrblatt, Hyndman, Özbay and Schotter (2008) use two different normal-form games with a unique Pareto-efficient Nash equilibrium in pure strategies to study convergence to the Nash equilibrium. They focus on the mechanisms underlying the convergence process and on strategic teaching. Our experimental design is closest to the last paper. However, the Nash equilibrium in our game is not Pareto-efficient, leading to less convergence. We focus more broadly on learning how to play strategically and pay close attention to the development and nature of non-strategic play.

Another strand of the literature studies learning in normal-form games under different information conditions. Oechssler and Schipper (2003) and Gerber (2006) use normal-form games with incomplete information about the opponents' payoffs in order to study whether players can

	Left	**Center**	**Right**
Top	78, 68	72, 23	12, 20
Middle	67, 52	59, 63	78, 49
Bottom	21, 11	62, 89	89, 78

Table 1: Game

figure out which game they are playing. Subjects receive feedback about the strategy chosen by the other player and can thereby form a "subjective game" (Kalai and Lehrer, 1993). In contrast, Weber (2003) studies a repeated beauty-contest game without feedback and Weber and Rick (2008) focus on repeated normal-form games without feedback. Both studies observe some amount of no-feedback learning.

The paper is organized as follows. The next section introduces the design of the experiment and provides a description of the level-of-reasoning model applied to the normal-form game we used. In Section 4, we present the results, focusing first on choices and then on belief statements. Section 5 contains a discussion and the conclusions.

2 Experimental design

2.1 Procedures

In all treatments of the experiment, we used the asymmetric normal-form game presented in Table 1. The game has a unique Nash equilibrium in pure strategies in which the row player chooses Top and the column player chooses Left. This equilibrium can be found by applying iterative elimination of dominated strategies. Note that the Nash equilibrium of the stage game is not Pareto efficient. The strategy combination of Bottom and Right leads to higher payoffs for both players. This outcome maximizes the payoff of the player that is least well off, and it also maximizes the sum of payoffs. The unique Nash equilibrium of the stage game is also the unique subgame perfect equilibrium of the repeated game.[1] Finally, note that for the column player choosing Right is strictly dominated by Center.

[1] However, there is a Nash equilibrium of the finitely repeated game in which the players play the Pareto-efficient strategy combination (Bottom, Right) for a number of periods and then switch to the Nash Equilibrium (Top, Left). In case a player deviates in this equilibrium, she is minmaxed by the other player choosing Middle or Center, respectively, for the rest of the game.

Treatment	Payoff	Feedback	Matching	Periods	Sessions	# of subjects
BASE	own+opponent	own payoff	fixed	20	4	54
PI	own	own payoff	fixed	20	4	48
NF	own+opponent	none	fixed	20	4	50
RM	own+opponent	own payoff	random	20	3	40

Table 2: Treatments

To study the impact of information on choices and belief statements we implemented four treatments, the details of which are given in Table 2. Our main interest is in the baseline treatment, denoted by BASE. In this treatment subjects had all relevant information about the game, i.e. the set of players, the set of strategies and the payoff function of each player. In addition, after each period they received feedback about the payoff earned in this period. Every other treatment differs from BASE only in one respect. In the treatments NF (no feedback) and RM (random matching) subjects had common knowledge of the payoff structure of the game, but we varied either the available feedback after each period or the matching protocol. In treatment NF, subjects received no feedback at all. In treatment RM subjects received feedback about their payoff, but were randomly matched with another participant in each period. In treatment PI (partial information), subjects had incomplete information about some elements of the game. They only knew their own payoff function, but not the payoff function of their opponent. However, they received feedback after each period, just as in treatments BASE and RM, such that they could infer the choice of their opponent. In all treatments subjects did not receive any feedback about their payoffs from the belief elicitation task.[2]

In the beginning of all treatments, subjects were randomly assigned a player role (row player or column player), which they kept during the whole experiment. However, they made all their decisions from the perspective of the row player, i.e. for column players we used a transformation of the matrix game in Table 1. Before choosing an action (choice task), we asked subjects to indicate their beliefs regarding the behavior of their opponent (belief task). In particular, we asked subjects to state the expected frequencies of play, i.e., they had to specify in how many out of 100 times they expect the column player to choose Left, Center and Right in the current period.[3] After the

[2]Nevertheless, they could infer their payoff from this task in treatments BASE, PI and RM. The main reason for not showing the payoffs from the belief elicitation task was to change as few parameters as possible when going from BASE, PI and RM to NF.

[3]For simplicity we restricted the expected frequencies of play to integers. Therefore, we count any belief statement

belief task, subjects had to make their choice by selecting one of the three possible actions (mixing was not possible).[4]

Subjects were paid for both tasks. For the choice task we paid subjects according to the numbers in the payoff matrix, which were exchanged at the commonly known rate of 1 point = € 0.15. To reward the belief task we used a quadratic scoring rule (QSR) which is incentive compatible given that subjects are risk-neutral money maximizers. The QSR we used is defined as follows. The payoff Π_{it}^{QSR} for player i in period t for a given action a_{jt}^k with $k \in \{L, C, R\}$ of player j in period t and belief vector $b_{it} = \left(b_{it}^L, b_{it}^C, b_{it}^R\right) \in \Delta^2$ such that $\Delta^2 = \left\{b_{it} \in \mathbb{R}^3 | \sum_{k\in\{L,C,R\}} b_{it}^k = 1\right\}$ is

$$\prod\nolimits_{it}^{QSR} (b_{it}, a_{jt}) = A - B * \left(\sum\nolimits_{k\in\{L,C,R\}} \left(b_{it}^k - 1_{[a_{jt}^k]}\right)^2\right) \tag{1}$$

where $1_{[a_{jt}^k]}$ is an indicator function equal to 1 if a_{jt}^k is chosen in period t and 0 otherwise. While paying subjects for the choice and the belief task is necessary to ensure incentive compatibility, it allows subjects to engage in hedging. Subjects can for example coordinate on a cell of the payoff matrix that is not an equilibrium and become unwilling to move away from it in order to avoid losses in the belief task. To eliminate such behavior, we decided to determine the final payoffs as follows. First, at the end of the experiment we selected one period randomly and independently to determine the payoffs for each of the two tasks. Second, we used parameters $A = 1.5$ and $B = 0.75$ in the QSR. Thus, the maximum payoff from the belief task (€ 1.50) is relatively low compared to payoffs from choice task. For instance, the Nash equilibrium [Top, Left] would lead to payoffs of € 11.7 and € 10.2 for the two player roles.[5]

The experiments were conducted in the computer lab at Technical University Berlin using the software tool kit *z-Tree*, developed by Fischbacher (2007). Subjects were recruited via a mailing list through which they could voluntarily register to participate in decision experiments (Greiner, 2004). Upon entering the lab, subjects received written instructions and were asked to read them carefully.[6] After everybody had finished reading the instructions, we distributed an understanding

assigning a weight of 34 percent to one action and 33 percent to each of the remaining actions as a uniform belief statement.

[4] We employ belief elicitation in all four treatments to analyze the impact of information on beliefs and choices. For a recent study of the impact of belief elicitation on choices see Rutström and Wilcox (2006).

[5] Note that subjects could guarantee themselves a payoff of € 1 by stating uniform beliefs. Although this would be an attractive choice for a risk-averse subject, we find no evidence of such behavior in our treatments. Only 7.5 percent of belief statements assign no less than 30 and no more than 35 percent to all three of the opponent's actions. (BASE 5.8%, PI 5.9%, NF 12.1% and RM 6.3%)

[6] For the instructions of the baseline treatment see the Appendix A.4.

test that covered both the game and the QSR. Only after all subjects had answered the questions correctly, we proceeded with the experiment. In total 192 students (106 males and 86 females) from various disciplines participated in the four treatments. Sessions lasted about one hour. Subjects' average earnings were about € 12.80, including a show-up fee of € 3 for arriving at the laboratory on time.

2.2 Strategies

Stahl and Wilson (1995) proposed a theory of boundedly rational types, based on a hierarchical model by Nagel (1993). Stahl and Wilson assume that players differ in their level of strategic sophistication. Their model classifies players into types according to their level of reasoning. A level-0 type randomizes uniformly over his strategy space, whereas a level-k type best responds to level-$(k-1)$ behavior for $k \in \{1, 2, .., \infty\}$, hence the term level-$k$ model.[7]

The level-k model is a useful approach to track off-equilibrium behavior. It has been tested and extended by various other studies mainly in the context of normal-form games (e.g. Costa-Gomes et al., 2001, Costa-Gomes and Weizsäcker, 2008, Rey-Biel, forthcoming or Camerer et al., 2004). It is also successful in organizing data from other games such as auctions, as recently shown by Crawford and Iriberri (2007a, 2007b) as well as Gneezy (2005). The most common types found in normal-form games are level-1 ($L1$), level-2 ($L2$) and Nash types, but their distribution crucially depends on the set of games investigated.

All above mentioned studies on normal-form games focus on one-shot interactions. In a repeated setting, additional strategic considerations come into play, and learning becomes possible. The level-k model can accommodate learning by inducing subjects to play higher-level strategies. Suppose a subject starts out by playing the $L1$ action, but then learns to best respond to $L1$ by playing $L2$ and so forth. Thus, a subject can learn by updating his beliefs in the course of the game, and we will investigate this on the basis of our data. In particular, we will test whether the subjects' beliefs become more accurate in predicting the opponents' behavior over time.

We use the level-k model to classify the available strategies in our game (see Table 3). The most important types of the level-k model ($L1$, $L2$ and Nash) can be grouped into two broad

[7] The model contains also other types to capture behavior eventually more in line with traditional game theory. These are the naive Nash type who chooses the Nash equilibrium strategy, the wordly type who plays a best response to a subjective distribution of all other types and the rational expectation type who correctly anticipates the distribution of boundedly rational types and best responds to this distribution.

	Row player		Column player
Top	Nash(L2)	**Left**	Nash
Middle	L1	**Center**	L1(L2)
Bottom	Rawls	**Right**	Rawls

Table 3: Decision rules

categories, namely strategic and non-strategic types. Strategic types form beliefs based on an analysis of what others do and best respond to these beliefs, whereas non-strategic types do not take into account the incentives of others. Given this definition, strategic types are $L2$ and Nash and the non-strategic type is $L1$.

We also introduce a Rawlsian decision rule, defined as choosing the action that maximizes the payoff of the player with the lower payoff, given that the other player has the same objective and chooses accordingly. Remember that in the game we use, the Rawls strategy is the same as the Utilitarian strategy which maximizes the sum of payoffs. With our definition of strategic behavior, the Rawls action is strategic because it requires the belief that the other player has the same preferences and acts accordingly (the same reasoning holds for its interpretation as a Utilitarian rule). Previous studies did not explicitly explore Rawlsian or Utilitarian strategies, but some of them found behavior pointing in this direction (e.g. Costa-Gomes and Weizsäcker, 2008). The game we used, depicted in Figure 1, allows us to separate between Nash play and play of the most efficient and/or fair outcome.[8]

The main focus of this study is on the development of strategic and non-strategic behavior over time. We therefore chose a game that allows us to identify strategic and non-strategic behavior as clearly as possible. In particular our interest was to achieve the best possible separation of the four rules of behavior ($L1$, $L2$, Rawls and Nash). We chose an asymmetric game for which the different rules overlap differently for the two player roles (see Table 3). Only the $L2$ rule cannot be identified clearly for any of the two player roles. For the row player, it prescribes the same action as Nash and for the column player it is the same as $L1$. Assuming that there is a considerable proportion of $L2$ play, which is suggested by previous studies, we will overestimate the proportion of Nash play of the row player and the proportion of $L1$ play of the column player. We will keep this in mind when interpreting the findings. However, our focus is on subjects learning to

[8] In contrast, Ehrblatt et al. (2008) run a similar experiment based on a game where the Nash equilibrium coincides with the Rawlsian/ Utilitarian outcome.

play strategically, and the $L2$ rule represents an intermediate level of strategic reasoning. We are mainly interested in the comparison between $L1$ and Nash behavior as the two extreme ends of the spectrum of strategic play.

Notice that we use the names $L1$, Nash and Rawls also for the three strategies in treatment PI even though a priori the subjects cannot reason about the other player's incentives and consequently cannot identify the Nash and the Rawls strategy in this treatment.[9]

3 Results

In the first part of the analysis, we examine the choices made by the experimental subjects. We begin this analysis with a focus on first period behavior and a comparison of these results to previous experiments. Afterwards we extend our analysis to all periods and focus on the development of behavior over time, considering the impact of the information available. In the second part of the data analysis, we make use of the elicited beliefs. After confirming that the stated beliefs outperform beliefs constructed with standard models of belief formation, we examine the frequency of best responses to the stated beliefs. Furthermore we check the accuracy of the stated beliefs in predicting the opponent's choice as well as the role of feedback and payoff information for the formation of beliefs.

Note that unlike in most other studies on asymmetric one-shot games (e.g. Costa-Gomes and Weizsäcker, 2008), we do not pool the data over player roles. As we study only one specific game, we are able to consider the exact strategic situation of each player role. This differentiation would be lost by pooling the data. Thus, we run all statistical tests separately for row and for column players. All results reported as significant in the paper are based on a 5%-level of significance.

3.1 Choices

3.1.1 First-period choices

In this section, we look at behavior in the first period only. This is of some stand-alone interest, since many experiments on behavior in one-shot 3x3 normal-form games have used similar games, and we can compare our results to them. First-period behavior in the four treatments is presented

[9] However, if the subjective game constructed by the participants happens to be equivalent to the true game, the names of the strategies can be interpreted as decision rules. See Kalai and Lehrer (1993) for the theory of subjective games.

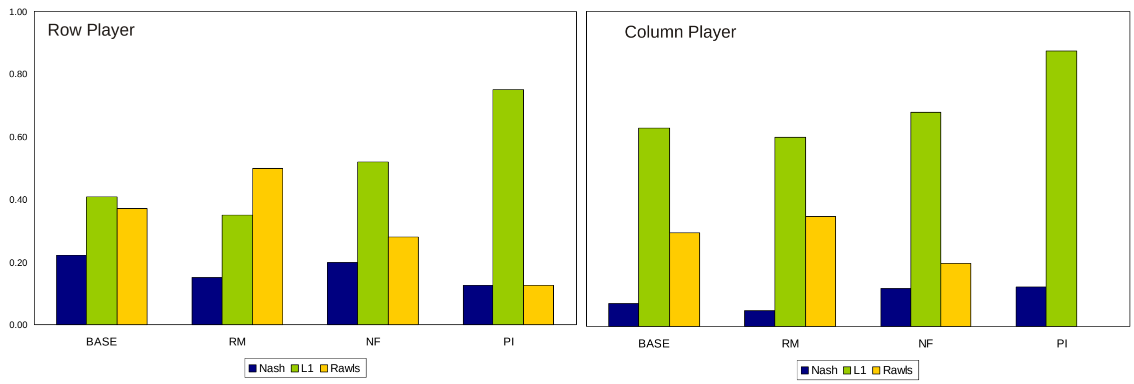


Figure 1: First period choices

in Figure 1. The figure shows the fraction of each action in a given treatment for row players and column players, respectively.

In the first period, subjects in treatments BASE, RM and NF all face the same strategic situation. Therefore we should not observe any differences in behavior in the first period. This is clearly the case, as can be taken from Figure 1. The frequency of chosen strategies of the row players (column players) in all three treatments is 19 (8) percent Nash, 43 (64) percent $L1$ and 38 (28) percent Rawls. We cannot reject the hypothesis that the frequency of strategies is the same in these three treatments using a χ^2-Test.[10] Our finding of 53% $L1$ behavior in the first period in BASE, RM and NF is in line with previous studies.[11] For instance, Costa-Gomes et al. (2001) estimated a $L1$ rate of about 45%, Rey-Biel (forthcoming) found 48% $L1$ behavior in his constant-sum games, whereas Costa-Gomes and Weizsäcker (2008) found slightly higher rates of about 60%.

Now, consider the decision situation in the first period of treatment PI. Subjects only know their own payoffs in the game and therefore cannot base their decisions on strategic considerations. Hence, it is no surprise to see 39 out of 48 subjects (81%) choosing the $L1$ rule in period 1 in PI, which not only maximizes the minimum payoff, but also the expected payoff assuming that the opponent randomizes uniformly over all possible actions. Concerning the column player's choice of the dominated action Right (Rawls), violations of dominance only occur in treatments BASE, RM and NF. It is remarkable that no column player in PI chooses Rawls in the first period, indicating that the choice of dominated actions in the other treatments is due to the payoff structure of the

[10] For both player roles we perform a pairwise comparison of BASE with NF and RM, respectively. The test yields no p-value smaller than 0.64 ($\chi^2_{(2)}$).

[11] For ease of comparison with other studies, we pool $L1$ behavior over player roles.

other player and not only to mistakes. The frequency of the three strategies in PI is significantly different from BASE in the first period for both player roles ($\chi^2_{(2)}$, $p = 0.043$ for row players and $p = 0.014$ for column players). We summarize the findings on choices in the first period in the following result.

Result 1 *(i) First-period behavior in BASE, RM and NF is statistically indistinguishable from each other and comparable to findings from one-shot experiments. (ii) $L1$ is the most frequently chosen strategy in the first period in all treatments and for both player roles. (iii) First-period play in treatment PI is significantly different from BASE.*

3.1.2 Choices over all periods

To give a first impression of how subjects play the game in the different treatments, Figure 2 presents the proportion of the behavioral rules over time for each treatment. The figure shows averages over three periods in a given treatment for row players in the left panel and for column players in the right panel. It emerges from the graphs that the frequency of the behavioral rules differs in the various treatments and for the two player roles.[12] To study these differences, we run a number of regressions, summarized in Table 4.

First, we consider average behavior over all 20 periods in the different treatments. For this purpose, we perform a separate regression for each strategy and player role combination. We regress the strategies on treatment dummies without controlling for time effects, which gives us a first indication of the influence of the different information conditions. To model the repeated decisions of the same subject in each treatment, we use random-effects panel regressions. Since subjects had to choose one out of three possible strategies, a probit model is employed where the dependent variable reflects the inclination to choose one strategy over the other two.

The results of the regressions without time trends, shown in the odd-numbered columns in Table 4, reveal the importance of information about the opponent's payoff and of feedback about past choices. The coefficients of PI are significantly different from BASE for all strategies except for the Nash strategy of the row player. In particular, there is significantly more $L1$ play and less Rawls play in PI than in BASE. In addition, the lack of feedback in NF results in more $L1$ play than in BASE for both player roles and less Rawls play for column players. Similar but weaker

[12] As stated above, we cannot clearly identify $L2$ behavior, since this rule overlaps with Nash for row players and with $L1$ for column players (see also Table 3). A certain proportion of $L2$ play may therefore be the reason why we observe on average less Nash and more $L1$ play of column players compared to row players over all treatments.

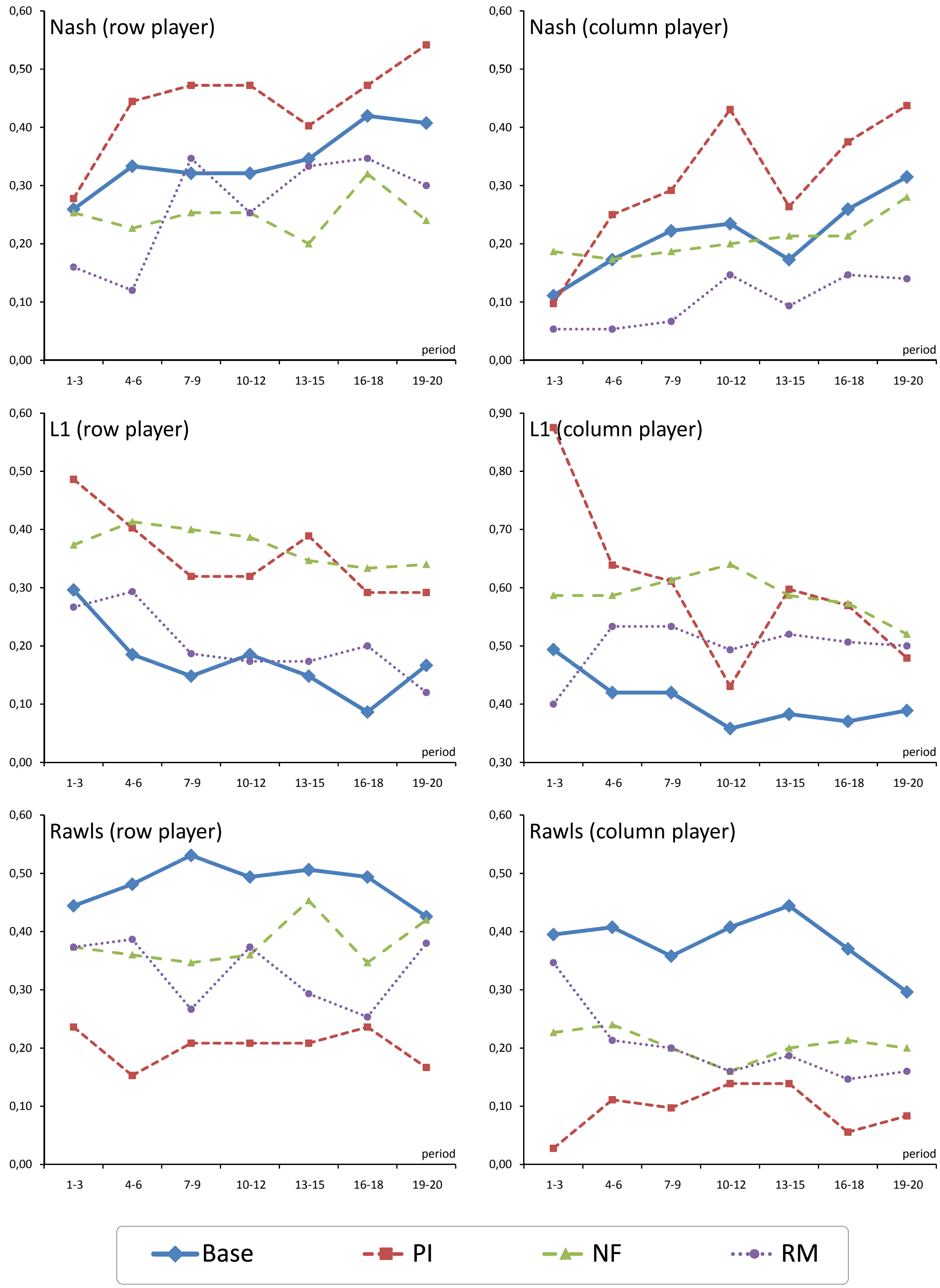


Figure 2: Decision rules over time

	Row Player						Column Player					
	Nash		L1		Rawls		Nash		L1		Rawls	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PI	0.31	0.31	0.82***	0.81**	-1.04***	-1.01***	0.56**	0.44	0.65***	0.99***	-1.2***	-1.41***
	(0.29)	(0.34)	(0.29)	(0.34)	(0.31)	(0.35)	(0.27)	(0.34)	(0.22)	(0.28)	(0.31)	(0.38)
NF	-0.38	-0.08	0.85***	0.56*	-0.42	-0.48	0.10	0.35	0.64***	0.53*	-0.66**	-0.68*
	(0.28)	(0.35)	(0.29)	(0.34)	(0.30)	(0.34)	(0.27)	(0.35)	(0.22)	(0.27)	(0.30)	(0.35)
RM	-0.01	-0.21	0.43	0.41	-0.29	-0.08	-0.24	-0.20	0.69***	0.37	-0.46	-0.11
	(0.31)	(0.36)	(0.31)	(0.36)	(0.32)	(0.36)	(0.30)	(0.38)	(0.24)	(0.29)	(0.32)	(0.37)
Period		0.03***		-0.04***		0.00		0.04***		-0.02*		-0.01
		(0.01)		(0.01)		(0.01)		(0.01)		(0.01)		(0.01)
PI*Per		0.00		0.00		-0.00		0.01		-0.03**		0.02
		(0.02)		(0.02)		(0.02)		(0.02)		(0.01)		(0.02)
NF*Per		-0.02		0.03*		0.00		-0.02		0.01		0.00
		(0.02)		(0.02)		(0.02)		(0.02)		(0.01)		(0.02)
RM*Per		0.02		0.00		-0.02		-0.00		0.03**		-0.04**
		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)
Const	-0.59***	-0.93***	-1.25***	-0.83***	-0.02	-0.03	-1.21***	-1.67***	-0.32**	-0.13	-0.44**	-0.31
	(0.20)	(0.24)	(0.20)	(0.24)	(0.21)	(0.24)	(0.20)	(0.25)	(0.15)	(0.19)	(0.21)	(0.24)
log $\mathcal{L}$	-993.8	-977.7	-925.5	-908.3	-978.0	-978.6	-821.5	-801.9	-1160.6	-1147.7	-803.7	-795.9
$\chi^2_{(3)}/\chi^2_{(7)}$	6.0	36.2***	11.4***	43.7***	12.0***	14.7***	8.1**	45.7***	13.2***	38.0***	15.2***	29.8***
N	1920						1920					

Notes: Random-effects probit regression, standard errors in parentheses,

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 4: Regressions: Decision rules

effects can be observed in RM where the only significant difference to BASE is that $L1$ choices of the column player are more frequent.

To investigate potential learning paths, we extend our regressions by including a time trend and interaction terms for treatment with time. The results of these regressions are presented in the even-numbered columns in Table 4. In these regressions the dummy variables are coded such that the corresponding coefficients represent the intercept and the development over time in each treatment relative to the baseline treatment. In order to assess the absolute time trends in each treatment, we additionally test the hypothesis that the sum of the coefficient for Period and the relevant coefficient for Treatment*Period is equal to zero (see Appendix A.1)

First, let us focus on the development of the three strategies in BASE. The coefficient of Period shows that in BASE subjects tend to choose the $L1$ strategy less often in later periods while Nash play increases and Rawls choices are more or less stable over time. We can compare this learning path to the time trend in treatment PI. The inclusion of time controls reveals that behavior in PI changes in a similar way as in BASE, with an even stronger decrease of $L1$ play for the column player. The average difference in the choices between the BASE and the PI treatment is therefore mainly due to differences in initial play.

Now consider treatment NF. Although the removal of feedback in treatment NF does not produce significant differences in the time trend compared to BASE, the time trends in NF are so small that they are no longer significant when tested directly (see Table A.1). Finally, we compare the effect of random matching compared to fixed matching on the time trend. While we do not find differences between RM and BASE for row players over time, column players in RM choose Rawls less often and $L1$ more often than in BASE. This is consistent with the fact that reputation building is not possible in RM, and a deviation from Rawls to $L1$ which gives a higher payoff cannot be sanctioned effectively by the row player.

The findings based on the various regressions can be summarized as follows.

Result 2 *(i) In treatment BASE there is significantly less* $L1$ *and more Rawls play than in PI and NF. (ii) Over time the proportion of the Nash strategy increases in all treatments and for both player roles except in NF. (iii) The proportion of the* $L1$ *strategy decreases over time in BASE (row player) and PI (both player roles). There is no similar time trend in NF. (iv) The proportion of Rawls choices is almost constant over time for all treatments and player roles (except for the column player in RM).*

Thus, in the sense of Stahl and Wilson we observe a trend towards more strategic play (that is more Nash and less $L1$ play) in all treatments with feedback information. There is an increase in Nash and a decrease in $L1$ play in BASE and PI. In PI, the overall lower proportion of strategic behavior compared to BASE can be ascribed to the lack of information about the opponent's payoffs. However, the fact that players in PI can observe the choices of their opponent over time and react to these observations leads to a development of behavior away from the $L1$ rule, just as in BASE. In treatment NF behavior does not change over time. As the NF treatment is comparable to a repeated one-shot situation, this finding lends support to the frequently applied method of giving no feedback between different tasks in experiments in order to minimize learning effects. Finally, as our control for repeated game effects, treatment RM reveals no differences to BASE for the row player. But we observe that the column player's behavior is affected by the matching protocol in that she chooses on average more non-strategic $L1$ play in RM than in BASE. And over time she is less likely to choose the dominated strategy (Rawls) in RM compared to BASE, probably due to a lack of repeated-game effects.

3.2 Belief formation

In this section, we focus on the relationship between the elicited beliefs and the subjects' own as well as their opponents' actions. In standard equilibrium analysis it is assumed that subjects form beliefs about the behavior of the opponent and then best respond to these beliefs. The level-k model departs from this view by positing that subjects differ in their strategic sophistication when thinking about the behavior of other players, i.e., they differ in their beliefs (Stahl and Wilson, 1995). In particular, level-1 behavior implies that beliefs are naive in that uniform randomization by the opponent is assumed. Level-2 types hold the belief that others best respond to uniform randomization. Thus, we can use belief statements to measure the level of strategic sophistication and to track the development of strategic thinking over time.

There are some caveats concerning the elicitation of beliefs. First, subjects need not hold beliefs about the opponent's play at all. For example, they might choose some non-strategic decision rule in the first period and then condition play on received payoffs (as in reinforcement learning). Forcing them to state beliefs could alter the choices if these subjects move their decisions in the direction of belief-based play.[13] However, our design is based on a comparison between treatments which all use belief elicitation. Unless the effects of belief elicitation interact with our treatment

[13] See Rutström and Wilcox (2006) for an argument along these lines.

variables, our results are immune to such problems. More importantly, the assumption of best-responses to beliefs in decision theory can be understood as an "as if" assumption. With this interpretation, subjects do not necessarily have to best respond to their stated beliefs as these beliefs might be unrelated to the true underlying beliefs. In order to address this concern, we compare the stated beliefs to beliefs constructed from previous play of the opponent. The stated beliefs emerge as a better predictor of actual choices than the constructed beliefs, which lends support to the hypothesis that the elicited beliefs are good approximations of the true underlying beliefs. Also, subjects might make mistakes when stating their beliefs, just as when taking decisions. We therefore propose that the belief statements should only be taken as a proxy of the true underlying beliefs of subjects.[14] Finally, even though we asked explicitly to state myopic beliefs, i.e. beliefs only for the current period, we cannot rule out that subjects follow repeated-game strategies and hold beliefs consistent with this. As the choices that are part of repeated-game strategies are not necessarily best responses to myopic beliefs, we will use treatment RM to check for repeated-game effects.

3.2.1 Stated beliefs vs. models of belief formation

We follow the approach used in Nyarko and Schotter (2002) and compare the explanatory power of elicited beliefs compared to standard belief learning models. The purpose of this comparison is to establish whether stated beliefs are a good measure of strategic uncertainty or whether stated beliefs are inferior to beliefs derived indirectly from the opponents' choices.

Standard belief learning models assume that players update their beliefs based on the opponent's history of play and then best-respond to these beliefs. The two most prominent models based on this assumption are the fictitious-play and the Cournot best-response model. While in the Cournot model subjects best respond to the opponent's play in the very last period, players in a pure fictitious-play model best respond to beliefs based on all previous actions of the opponent. The γ-weighted fictitious-play model introduced by Cheung and Friedman (1997) contains Cournot best response and fictitious-play as special cases. In this model subject i's belief $b_{i,t+1}^k$ that subject j will choose action $a_{jt}^k, k \in \{L, C, R\}$ in period $t+1$ is defined as:

$$b_{i,t+1}^k = \frac{1_{[a_{jt}^k]} + \sum_{u=1}^{t-1} \gamma_i^u 1_{[a_{j,t-u}^k]}}{1 + \sum_{u=1}^{t-1} \gamma_i^u}. \tag{2}$$

The parameter γ_i is the weight player i gives to the past actions of his opponent. It is obvious

[14] See Costa-Gomes and Weizsäcker (2008) for a thorough analysis of belief statements.

from (2) that $\gamma_i = 0$ leads to the Cournot best-response model and $\gamma_i = 1$ yields fictitious-play, respectively. We incorporate this model into a standard logistic choice model to allow subjects to best respond to their beliefs with noise. Subject i chooses action k with probability

$$\Pr\left(a_{it}^{k}\right) = \frac{\exp\left(\lambda\pi[a_{it}^{k}, b_{it}]\right)}{\sum_{l\in\{L,C,R\}}\exp\left(\lambda\pi[a_{it}^{l}, b_{it}]\right)}, \tag{3}$$

where $\pi[a_{it}^{k}, b_{it}]$ is the expected payoff of player i when she chooses an action k given her beliefs b_{it} over the action set of her opponent. The parameter λ determines the impact of this expected payoff on her own choice probability and can be interpreted as a rationality parameter. A player with $\lambda = 0$ chooses all actions with equal probability disregarding the expected payoff of her choice. On the other hand if $\lambda \to \infty$ the player is fully rational, i.e. she always best responds to her beliefs.

We now turn to the estimation and probabilistic comparison of the choice model (3) based on the γ-weighted fictitious-play model (2) on the one hand and on the stated beliefs on the other hand. Since the belief-learning model assumes that subjects process only information about their own payoffs and about the history of their opponent's play, we only use the data of treatments BASE and PI in the following analysis, while we do not consider treatment NF. We also analyze the data from treatment RM, since the process described in (2) can also be interpreted as the formation of beliefs over the average play of the population rather than over individual choices. The estimation results for each treatment and player role are presented in Table 5.[15]

As a first result we observe that the stated beliefs play a significant role in explaining the behavior of our subjects, since appropriate likelihood-ratio tests reject the hypothesis that the rationality parameter λ is equal to zero ($p = 0.00$ for all treatments and player roles).

Using tests for the selection between non-nested models introduced by Vuong (1989) and Clarke (2003), the hypothesis of equal explanatory power of the models can be rejected at all usual significance levels for all treatments and player roles, the only exception being the column player in the random-matching treatment.[16] In our notation the negative signs of the test statistics reveal

[15] For the γ-weighted fictitious-play model we estimated γ and λ simultaneously. All ML-estimations and tests have been conducted with MATLAB and R.

[16] Vuong's test statistic is based on the overall likelihood ratio of two rival models and is asymptotically normally distributed under the null. Clarke's test statistic consists of the number of single likelihood ratios being greater than 1 which is binomially distributed under the null with parameters $\theta = 0.5$ and the number of observations in each subset of the data. Vuong's test is outperformed by Clarke's test when the distribution of the single log-likelihood ratios is highly peaked. Both tests were calculated using corrections for the dimension of the models as proposed by Schwarz (1978) and Clarke (2007) respectively.

		ML-estimation of model (3) using					Model selection tests			
		Fictitious play (2)			Stated beliefs		Vuong's test		Clarke's test	
Treatment	**Role**	λ	γ	$\log \mathcal{L}$	λ	$\log \mathcal{L}$	**Z**	**p-value**	**Z**	**p-value**
BASE	row	0.0575	0.7418	-484.01	0.1005	-422.12	-3.46	0.0005	-9.12	0.0000
	column	0.0373	0.6009	-492.96	0.0586	-421.93	-6.88	0.0000	-5.42	0.0000
PI	row	0.0442	0.6488	-487.68	0.0646	-451.16	-3.72	0.0002	-3.47	0.0005
	column	0.0571	0.6220	-413.59	0.1066	-307.98	-5.82	0.0000	-14.97	0.0000
RM	row	0.0233	0.5821	-427.04	0.0825	-372.34	-5.35	0.0000	-5.90	0.0000
	column	0.0729	0.9067	-350.21	0.0604	-334.25	-1.42	0.1548	-1.30	0.1936

Notes: p-values are two-sided. Clarke's corrected B has been approximated by the standard normal distribution.

Table 5: Model Estimation and Selection.

that the stated belief model is closer to the real data generating process than the beliefs generated by the belief-learning models.

To summarize, we extend the finding of Nyarko and Schotter (2002) from a matching-pennies game to our normal-form game with a unique subgame perfect Nash equilibrium in pure strategies. We find that stated beliefs are better at explaining observed choices than beliefs that are implied by the standard models of belief formation. In the following, we therefore use the stated beliefs when analyzing the impact of experience and information on the consistency and accuracy of beliefs.

3.2.2 Consistency of actions and stated beliefs

Both in standard Nash equilibrium and in the level-k model it is assumed that subjects best respond to their beliefs. Using the elicited beliefs, we can investigate the consistency of actions and stated beliefs, i.e. whether subjects best respond to their stated beliefs. This helps us to evaluate the relative descriptive validity of both models in the four different treatments.

In Figure 3 the proportion of players best responding to their stated beliefs is displayed for each player role and treatment separately. The figure shows the average proportion of best responses over three periods. In all treatments, the average best response rates are rather low, ranging from 45% to 75%. In order to compare our results to other studies, it is useful to look at the aggregated best-response behavior of all subjects. Averaging over all treatments and player

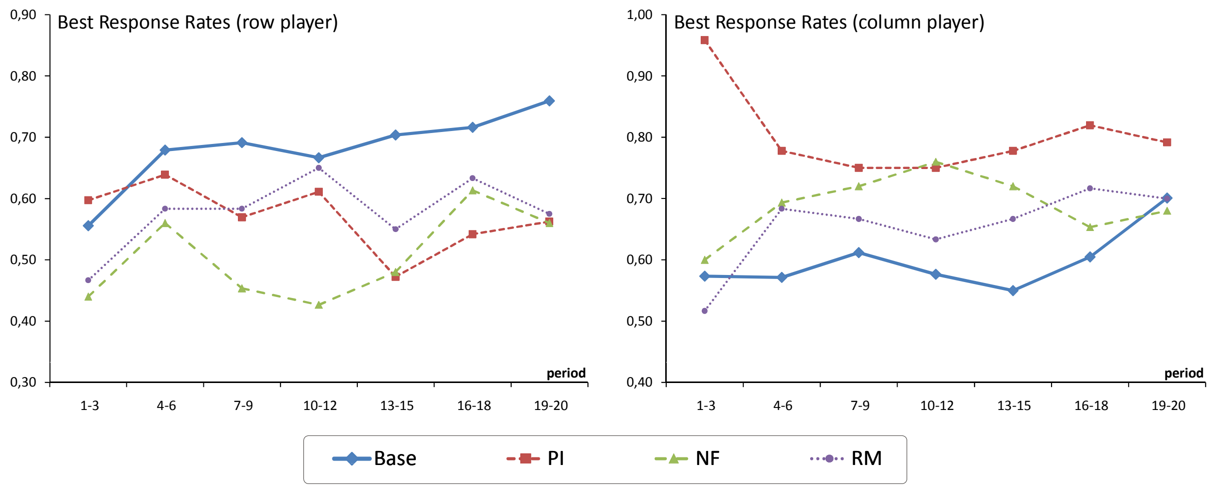


Figure 3: Best-response rates over time

	Costa-Gomes & Weizsäcker (2008)	Rey-Biel (forthcom.)	Ehrblatt et al. (2008)	Nyarko & Schotter (2002)	our study
Games	various 3x3	various 3x3	two 3x3	one 2x2	one 3x3
Interaction	one-shot	one-shot	repeated	repeated	repeated
⌀	54	69	49	75	63

Table 6: Best-response rates (in %) in various studies

roles, subjects best-respond to their stated beliefs in 63% of the cases. The best-response rates found in similar studies are summarized in Table 6. In simple games like 2x2 games (Nyarko and Schotter, 2002) or constant-sum games (Rey-Biel, forthcoming) consistency rates are about 70%, whereas the rates range from 49% to 63% in more complicated games like ours or the games used in Costa-Gomes and Weizsäcker (2008) and Ehrblatt et al.(2008), respectively.

For statistical evidence on differences between the treatments and the development of best-response rates over time, we run random-effects panel regressions. As the dependent variable is either 0 (no best response) or 1 (best response), we use a probit model. Besides the constant, the independent variables are dummies for PI, NF and RM, a linear time trend and interaction dummies for time trend and treatment. The regression results are summarized in Table 7. Again, we run additional direct tests of the absolute time trends in the control treatments (see Appendix A.2).

Due to the asymmetry of the game and in particular due to the fact that only the column

	Best-response rates			
	row player		column player	
	(1)	(2)	(3)	(4)
PI	-0.356*	0.105	0.846***	1.262***
	(0.212)	(0.263)	(0.257)	(0.314)
NF	-0.517**	-0.387	0.499**	0.578*
	(0.209)	(0.260)	(0.253)	(0.302)
RM	0.326	-0.176	0.335	0.244
	(0.222)	(0.275)	(0.269)	(0.320)
Period		0.032***		0.018*
		(0.011)		(0.011)
PI*Period		-0.046***		-0.039**
		(0.015)		(0.017)
NF*Period		-0.014		-0.007
		(0.015)		(0.016)
RM*Period		-0.016		0.009
		(0.016)		(0.017)
Constant	0.557***	0.234	0.176	-0.012
	(0.146)	(0.181)	(0.174)	(0.208)
log $\mathcal{L}$	-1160.74	-1152.70	-1008.56	-1002.94
$\chi^2_{(3)}/\chi^2_{(7)}$	6.48*	22.13***	11.22**	22.09***
N	1920		1920	

Notes: Random-effects probit regressions,

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 7: Regressions: Best-response rates

player has a dominated strategy, it is necessary to differentiate between the row and the column player in this section. First we investigate whether subjects learn to best respond in BASE in the course of the experiment. The significant and positive coefficient of Period reveals that this is the case for the row player. There is a positive trend also for the column player, but it fails to be significant.

In what follows, we will compare each control treatment with BASE separately, starting with PI. For the row player, the average number of best responses in BASE is slightly higher than the number of best responses in PI. The opposite holds for the column player which is due to less Rawls play in PI (as the Rawls strategy cannot be identified in PI), thereby avoiding violation of dominance. Using direct tests for the time trends, there is no significant development over time for both player roles in PI. When comparing BASE to NF the overall level of best responses is again higher in BASE than in NF for the row player and lower for the column player (because the dominated Rawls strategy is played less often in NF than in BASE). As in PI, the time trends in NF are not significant when tested directly. Aside from the row players in BASE, only the column players in RM display higher best-response rates in later periods (the time trend of column players in RM is significant when using a direct test).

These findings raise the question why the best-response rates of the row player are higher in BASE compared to treatments PI and NF. Internal consistency requires best responding to one's beliefs, independent of the information conditions. We can merely offer potential explanations of our observations, but further research is necessary to disentangle the causes of behavior more thoroughly. In treatment NF, subjects might be doubtful about the accuracy of their beliefs, lacking information about the other player's behavior. This might induce them to put less weight on their beliefs when choosing an action. But this reasoning fails to explain the similar result in treatment PI where there is also no discernible increase in best-response behavior. In PI, players have to learn about the structure of the game over time. Two possible explanations come to mind. First, the complexity of learning both the structure of the game and of best responding to one's beliefs at the same time may be too high. Second, in treatment PI many subjects start with uniform beliefs and best respond to them. As the belief set of $L1$ is large and $L1$ is an attractive strategy initially, there is a high rate of consistency at the outset. This effect is absent in BASE and NF.

The focus of the preceding analysis was on myopic beliefs. However, in our repeated-game setting, Folk Theorem results are possible. If subjects aim at a cooperative outcome, column players might choose their dominated action (Rawls) in response to Rawls play of row players. A necessary condition for a repeated-game strategy is the observability of past behavior such that subjects can condition their actions on their opponents' play. To achieve a cooperative outcome a minimum of information is needed to allow for sanctions of deviations.[17] As mentioned before, the fact that the choice of Rawls by the column player can never be a best response to any myopic belief explains why we observe very low best-response rates for column players in BASE compared to NF and PI where less Rawls play is observed. If the low best-response rates are indeed a result of repeated-game strategies, we should observe significantly higher best-response rates in RM. The reason is that the finite time horizon and the random-matching protocol do not allow for cooperation based on the Folk theorem. But we observe a substantial proportion of Rawls play also in RM in both player roles. Moreover, the regressions reveal no significant differences between BASE and RM neither for the overall proportion of Rawls play (see Table 4) nor for the average best response rates (see Table 7).

The insignificant difference of best response rates in BASE and RM could be due to a higher number of failures to best respond to undominated actions in RM, which would push best-response rates down in the direction of BASE. But this is not the case. When considering only the best-response behavior to Nash and $L1$, we find best response rates of about 92% in BASE and 88% in RM. We can further support this finding of equal best-response rates in BASE and RM by a Kolmogorov-Smirnov test which compares the number of best responses to Nash and $L1$ of each subject. The test yields a p-value of $p > 0.88$.[18] For these reasons we consider the evidence for repeated-game strategies as weak.

Result 3 *(i) Row player: The best-response rates in PI and NF are on average significantly lower than in BASE. While the proportion of best responses increases over time in BASE, there is no significant time trend in all other treatments. (ii) Column player: Best-response rates are overall higher in treatments NF and PI than in BASE. This difference disappears when restricting attention to undominated actions. There is no significant time trend for any treatment except RM. (iii) For both player roles, treatments BASE and RM do not significantly differ from each other with respect to overall best-response rates.*

Interpreting the stated beliefs as proxies for the true underlying beliefs, we can conclude that

[17] For instance, Ellison (1994) and Kandori (1992) have shown for infinitely repeated games with random matching that a cooperative outcome is possible through contagious sanctions.

[18] We use each column player as an independent observation and compare the empirical distribution of the number of best responses to Nash and $L1$ between BASE and RM.

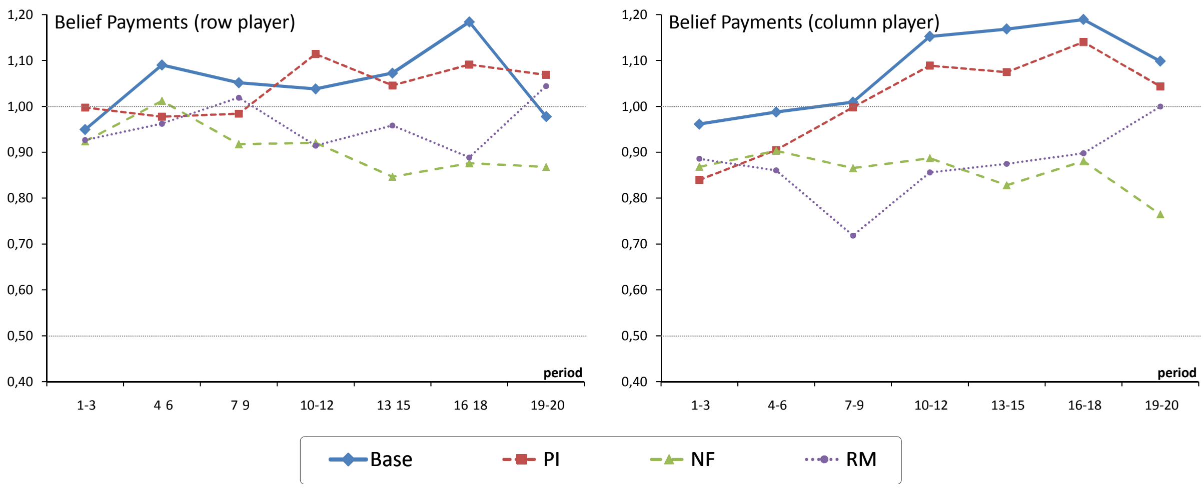


Figure 4: Accuracy of stated beliefs

actors best respond more often to their beliefs in repeated games with feedback information and information about the game structure with some experience of the situation, compared to situations with less information and experience. Best-response rates are low when there is a dominated action leading to the Pareto-efficient outcome.[19]

3.2.3 Accuracy of stated beliefs

We will now focus on whether the elicited beliefs are accurate in predicting the behavior of the opponents. The accuracy of beliefs is also a measure of strategic uncertainty. Subjects who hold accurate beliefs about opponent's behavior do not experience strategic uncertainty. The baseline treatment together with the control treatments allow us to identify the factors enabling subjects to reduce strategic uncertainty and to state accurate beliefs. In addition, the predicted accuracy of beliefs is different in the Nash equilibrium prediction and the level-k model. In the Nash equilibrium of the stage game, subjects hold accurate beliefs about their opponent's choice. In the level-k model, however, this is typically not the case as subjects' beliefs can be at odds with their opponents' behavior. In order to measure how well stated beliefs predict the opponent's play, we use the earnings from the quadratic scoring rule (QSR). Figure 4 shows the average earnings over three periods from the QSR for all treatments and for both player roles.[20]

[19] Of course, the strategy may not be dominated for other specifications of the utility function. However, in this paper we restrict attention to payoffs representing utilities.

[20] In principle, the accuracy of predicting other's behavior should not depend on the player role. Indeed, we only find a weakly significant difference between player roles in RM (Mann-Whitney test, $p = 0.098$). In all other treatments

The average payoff across treatments and player roles is about € 1.[21] This corresponds to the payoff for subjects who state uniform beliefs, which is indicated by the vertical line in Figure 4. The second benchmark to which we can compare the earnings is € 0.5, representing the expected payoff from randomizing uniformly over degenerate beliefs. Although subjects earn hardly more than € 1, their beliefs are much better than in the case where they simply try to predict the choice of their opponent with a probability of one (Wilcoxon signed-rank test, all p-values < 0.01). How can the belief statements be further characterized? We do not observe many uniform belief statements (see also footnote 5). Although 49% of all belief statements assign a positive probability to each action, only a fraction of 7.5% submit uniform beliefs. About 21% of the belief statements assign a positive probability to two of the actions, and 30% of the statements are degenerate. Row players in BASE and PI earn on average more than € 1 and in NF and RM they earn less than € 1. But we cannot reject the hypothesis of equal means at a 5% level of significance for all treatments (Wilcoxon signed-rank test, p-values > 0.085). The same holds for column players in BASE and PI, but column players in NF and RM earn on average significantly less than € 1 (Wilcoxon signed-rank test, for NF and RM p-values < 0.01).

Figure 4 displays improvements over time in predicting the play of the opponent in treatments BASE and PI. Apparently, this is not the case for treatments NF and RM. We run a random-effects panel regression where the dependent variable is the payoff from the belief elicitation task. In addition to the constant and a reference time trend for BASE, the regression includes treatment and time interaction dummies for the controls PI, NF and RM as independent variables in order to measure the corresponding performance relative to BASE. Again, direct tests of the absolute time trends in the control treatments were performed separately (see Appendix A.3).

Averaged over all periods, beliefs are significantly less accurate in NF than in BASE while PI and BASE show the same accuracy of beliefs. When comparing BASE to RM, only the column players differ significantly due to a lower accuracy of beliefs in RM than in BASE. Focusing on the development over time, we observe some learning in BASE since the beliefs become more accurate over time for the column players. We observe the same pattern over time in PI as in BASE. But for treatments NF and RM, tests of the absolute time trends reveal that there is no learning. Finally, the row players show no significant learning path in any treatment.

The findings can be summarized as follows:

the same test yields p-values higher than 0.45.

[21] The average payoff across player roles is in BASE € 1.07, in PI € 1.02, in NF € 0.89 and in RM € 0.91.

	Payment for belief task			
	row player		column player	
	(1)	(2)	(3)	(4)
PI	-0.017	-0.030	-0.069	-0.087
	(0.063)	(0.082)	(0.063)	(0.081)
NF	-0.144**	-0.024	-0.219***	-0.039
	(0.063)	(0.081)	(0.062)	(0.080)
RM	-0.100	-0.053	-0.216***	-0.138
	(0.067)	(0.086)	(0.066)	(0.085)
Period		0.005		0.013***
		(0.003)		(0.003)
PI*Period		0.001		0.002
		(0.005)		(0.005)
NF*Period		-0.011**		-0.017***
		(0.005)		(0.005)
RM*Period		-0.005		-0.007
		(0.005)		(0.005)
Constant	1.056***	0.999***	1.080***	0.941***
	(0.044)	(0.056)	(0.043)	(0.055)
$\chi^2_{(3)}/\chi^2_{(7)}$	6.84*	15.60**	17.61***	55.56***
R^2	0.07	0.07	0.16	0.16
N	1920			

Notes: Random-effects regressions,

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 8: Regressions: Accuracy of stated beliefs

Result 4 *(i) In treatments BASE and PI, behavior is characterized by the same learning path. The column player's beliefs are more accurate in later periods while the row players do not display any learning. (ii) Overall, the beliefs are significantly less accurate in NF than in BASE since there is no learning at all in NF. (iii) While for the row player the accuracy of beliefs is the same in RM as in BASE over all periods, the column players in RM exhibit less accurate beliefs on average.*

The results indicate that feedback about past behavior of one's opponent is more important for reducing strategic uncertainty than information about the strategic incentives of one's opponent. Without feedback or with noisy feedback because the opponent is not the same in each period, it is difficult to predict one's opponent's behavior. Further support for the relatively minor role of information about the structure of the game comes from the fact, displayed in Figure 4, that in the first periods beliefs are not significantly less accurate in PI than in BASE. The observed low accuracy of beliefs is an indicator for the strategic uncertainty in all treatments. However, we see that in the treatments with feedback (BASE and PI) the accuracy of beliefs is not only higher than in NF and RM, but it also increases over time. The non-significance of the time trend of the row players in BASE is mainly due to the large drop of the accuracy of beliefs in the last two periods.[22] This drop is associated with the deviation from Rawls of almost all column players in the last two periods which was not anticipated by the row players.

4 Summary and Conclusions

We have performed an experiment to study the development of strategic reasoning over a limited number of periods. To classify the strategies of the 3x3 normal-form game employed in our study, we used the level-of-reasoning model of Stahl and Wilson (1995). This classification allowed us to track strategic play over time. In order to understand the determinants of strategic play, we varied the information available to the players and elicited their beliefs about opponents' play.

We find that feedback information and information about the payoffs of the opponent have an impact on choices. When either type of information is lacking, this leads to an increase in non-strategic ($L1$) and a decrease in Rawls play on average. However, not revealing the opponent's payoff function has almost no impact on the learning path compared to the baseline treatment. In both treatments (BASE and PI), subjects exhibit less non-strategic and more Nash play over time.

[22] If we exclude the last two periods in regression (2) and (4) in Table 8, the coefficients of period are significantly positive for BASE and significantly negative for NF and RM.

In contrast, in the no-feedback treatment there is no increase in strategic play in the course of the experiment. This fact clearly highlights the importance of feedback and the limits of strategic sophistication of the subjects.

Regarding the analysis of beliefs, we first evaluate whether stated beliefs or beliefs constructed with belief-learning models are a better proxy for the underlying true beliefs of the subjects. We find that the stated beliefs are more consistent with actual choices than beliefs constructed with belief models such as weighted fictitious play or Cournot best response. Given this result, we study the best-response rates to the stated beliefs. In the baseline treatment, actions are consistent with stated beliefs more frequently in later periods (significant for the row players). Missing information about the opponent's payoff function or no feedback destroys this trend towards more best responses in later periods.

The accuracy of the subjects' beliefs with respect to the opponent's choices is increasing over time in the baseline treatment (this is significant for column players). Surprisingly, removing the information about the opponent's payoff function does not decrease the overall accuracy of beliefs nor its development over time in a significant manner. However, without feedback information about the other player's past actions, the overall accuracy of beliefs is significantly lower, and players do not make any improvements in predicting the other player's behavior.

This study should be seen as a first step in understanding the development of strategic thinking with the help of stated beliefs in a game. Many issues remain to be investigated. For example, other games should be used in order to be able abstract from the specifics of our game. Also, the accuracy and consistency of beliefs over time is by now very little understood and in our view deserves thorough empirical scrutiny.

References

[1] Camerer, C., Ho, T., and Chong, J.-K. (2004). A Cognitive Hierarchy Model of Games, *Quarterly Journal of Economics*, 119, 861-898.

[2] Clarke, K. A. (2003). Nonparametric Model Discrimination in International Relations, *Journal of Conflict Resolution*, 47, 72-93.

[3] Clarke, K. A. (2007). A Simple Distribution-Free Test for Nonnested Model Selection, *Political Analysis*, 15, 347-363.

[4] Costa-Gomes, M.A., Crawford, V. and Broseta, B. (2001). Cognition and Behavior in Normal-form Games: An Experimental Study, *Econometrica*, 69, 1193-1235.

[5] Costa-Gomes, M.A. and Weizsäcker, G. (2008). Stated Beliefs and Play in Normal-form Games, *Review of Economic Studies*, 75, 729-762.

[6] Crawford, V. and Iriberri, N. (2007a). Level-k Auctions: Can a Non-Equilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private Value Auctions, *Econometrica,* 75, 1721-1770.

[7] Crawford, V. and Iriberri, N. (2007b). Fatal Attraction: Salience, Naivete, and Sophistication in Experimental Hide-and-Seek Games, *American Economic Review*, 97, 1731-1750.

[8] Ehrblatt, W. Z., Hyndman, K., Özbay, E.Y. and Schotter, A. (2008). Convergence: An Experimental Study of Teaching and Learning in Repeated Games, Working Paper.

[9] Ellison, G. (1994). Cooperation in the Prisoner's Dilemma with random matching, *Review of Economic Studies*, 61, 567-588.

[10] Fischbacher, Urs (2007). z-Tree: Zurich toolbox for ready-made economic experiments, *Experimental Economics*, 10, 171-178.

[11] Gerber, A. (2006). Learning In and About Games, Working paper.

[12] Gneezy, U. (2005). Step-Level Reasoning and Bidding in Auctions, *Management Science*, 51, 1633-1642.

[13] Greiner, B. (2004). An Online Recruitment System for Economic Experiments, in: Kremer,K., Macho, V. (eds.): *Forschung und wissenschaftliches Rechnen 2003*, GWDG Bericht 63, Göttingen: Ges. für Wiss. Datenverarbeitung, 79-93.

[14] Ivanov, A. (2006). Strategic Play and Risk Aversion in One-Shot Normal-Form Games: An Experimental Study, Working Paper.

[15] Kandori, M. (1992). Social Norms and Community Enforcement, *Review of Economic Studies*, 59, 63-80.

[16] Kalai, E. and Lehrer, E. (1993). Subjective Equilibrium in Repeated Games, *Econometrica*, 61, 1231-1240.

[17] Nyarko, Y. and Schotter, A. (2002). An Experimental Study of Belief Learning Using Elicited Beliefs, *Econometrica*, 70, 971-1005.

[18] Oechssler, J. and Schipper, B. (2003). Can You Guess the Game You Are Playing?, *Games and Economic Behavior*, 43, 137-152.

[19] Rey-Biel, P. (forthcoming). Equilibrium Play and Best Response to (Stated) Beliefs in Constant Sum Games, *Games and Economic Behavior*.

[20] Rutström, E. and Wilcox, N. (2006). Stated Beliefs Versus Empirical Beliefs: A Methodological Inquiry and Experimental Test. Mimeo.

[21] Schwarz, G. (1978). Estimating the Dimension of a Model, *The Annals of Statistics*, 6, 461-464.

[22] Stahl, D.O. and Wilson, P.W. (1995). On Players' Models of Other Players: Theory and Experimental Evidence. *Games and Economic Behavior,*10, 218-254.

[23] Vuong, Q. H. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses, *Econometrica*, 57, 307-333.

[24] Weber, R. (2003). 'Learning' With No Feedback in a Competitive Guessing Game, *Games and Economic Behavior*, 44, 134-144.

[25] Weber, R. and Rick, S. (2008). Meaningful learning and transfer of learning in games played repeatedly without feedback, Working Paper.

A APPENDIX

A.1 Tests for absolute time trends of choices

	Row player			Column player		
	(1) **Nash**	(2) **L1**	(3) **Rawls**	(4) **Nash**	(5) **L1**	(6) **Rawls**
BASE	-0.929***	-0.831***	-0.03	-1.671***	-0.126	-0.31
	(0.24)	(0.239)	(0.236)	(0.251)	(0.187)	(0.242)
PI	-0.622**	-0.023	-1.037***	-1.227***	0.861***	-1.723***
	(0.244)	(0.244)	(0.258)	(0.236)	(0.203)	(0.29)
NF	-1.012***	-0.273	-0.507**	-1.317***	0.406**	-0.993***
	(0.251)	(0.235)	(0.249)	(0.244)	(0.199)	(0.256)
RM	-1.137***	-0.421	-0.112	-1.875***	0.245	-0.417
	(0.274)	(0.266)	(0.274)	(0.296)	(0.22)	(0.28)
Period*BASE	0.032***	-0.045***	0.001	0.042***	-0.018*	-0.013
	(0.012)	(0.013)	(0.011)	(0.013)	(0.01)	(0.012)
Period*PI	0.033***	-0.041***	-0.002	0.053***	-0.05***	0.007
	(0.011)	(0.012)	(0.012)	(0.012)	(0.011)	(0.016)
Period*NF	0.008	-0.013	0.007	0.019	-0.007	-0.012
	(0.012)	(0.011)	(0.012)	(0.013)	(0.011)	(0.013)
Period*RM	0.051***	-0.042***	-0.019	0.038**	0.013	-0.051***
	(0.013)	(0.014)	(0.013)	(0.016)	(0.012)	(0.014)
$\log\mathcal{L}$	-977.66	-908.25	-978.62	-801.85	-1147.71	-795.93
N		1920			1920	

Notes: Random-effects probit regression, standard errors in parentheses,

* significant at 10-percent level; ** significant at 5-percent level; *** significant at 1-percent level.

Table A.1: Regressions: Decision rules with absolute time trends.

A.2 Tests for absolute time trends of best-response rates

	Best-response rates	
	row player	**column player**
BASE	0.234	-0.012
	(0.181)	(0.208)
PI	0.339^{*}	1.250^{***}
	(0.191)	(0.235)
NF	-0.153	0.566^{***}
	(0.186)	(0.218)
RM	0.059	0.232
	(0.207)	(0.243)
Period*BASE	0.032^{***}	0.018^{*}
	(0.011)	(0.011)
Period*PI	-0.013	-0.021^{*}
	(0.011)	(0.013)
Period*NF	0.019^{*}	0.011
	(0.01)	(0.011)
Period*RM	0.017	0.027^{**}
	(0.012)	(0.013)
log $\mathcal{L}$	-1152.70	-1002.94
N	1920	1920

Notes: Random-effects probit regression,

* significant at 10%; ** significant at 5%; *** significant at 1%

Table A.2: Regressions: Best-response rates with absolute time trends.

A.3 Tests for absolute time trends of the accuracy of beliefs

	Payment in belief task	
	row player	column player
BASE	0.999***	0.941***
	(0.056)	(0.055)
PI	0.970***	0.854***
	(0.059)	(0.058)
NF	0.975***	0.902***
	(0.058)	(0.057)
RM	0.947***	0.803***
	(0.065)	(0.063)
Period*BASE	0.005	0.013***
	(0.003)	(0.003)
Period*PI	0.007*	0.015***
	(0.004)	(0.004)
Period*NF	-0.006*	-0.004
	(0.004)	(0.003)
Period*RM	0.001	0.006
	(0.004)	(0.004)
log $\mathcal{L}$	-1290.10	-1236.34
N	1920	1920

Notes: Random-effects regressions,

* significant at 10%; ** significant at 5%; *** significant at 1%

Table A.3: Regressions: Accuracy of stated beliefs with absolute time trends.

A.4 Instructions (for BASE)

The experiment you are about to participate in is part of a project financed by the German Research Foundation (DFG). Its aim is to analyze economic decision-making behavior. You can earn a considerable amount of money in this experiment, dependent on your decisions and the decisions of the other participants. Consequently, it is extremely important that you read these instructions very carefully.

Please note: these instructions are for your eyes only, and it is not permitted to hand on any information whatsoever to other participants. Similarly, you are not allowed to speak to the other participants throughout the whole experiment. Should you have a question, please raise your hand and we will come to you and answer your question individually. Please do not ask your question(s) aloud. If you break these rules, we will unfortunately be compelled to discontinue the experiment.

General information The experiment is made up of several rounds where decisions must be made and questions answered. You can win points with your decisions. These points represent your earnings and will be converted into euros at the end of the game and paid out in cash. The exact procedure of the experiment, the various decisions and the method of payment are clearly explained in the next section.

The decision-making situation At the beginning of the experiment, you will be assigned by draw to another participant, randomly and anonymously. This allocation is maintained throughout the whole of the remaining experiment. The participant who has been assigned to you will be called "the other one" from now on.

In each round, you and the other one will be confronted with the same decision-making situation. Each time, you must choose between the three alternatives: "top", "middle", and "bottom".

Each of these three alternatives has been given three possible payoffs (as points). The other one must also decide between three alternatives ("left", "center" or "right"), and each of these alternatives has also three possible payoffs, as above. You will see the following input screen on the computer:

Your three alternatives, “top”, “middle”, and “bottom”, are listed in the first column of the table. Next to your alternatives, you can see three boxes, each with two numbers. The subscript (lower) number is always your possible payoff. On the input screen illustrated above, the alternative “top” has been allocated the payoff of 78, 72 and 12, the alternative “middle” the payoff of 67, 59 and 78, and the alternative “bottom” the payoff of 21, 62 and 89. This means that should you decide on “top”, for example, then your payoff is 78, 72 or 12 points. The payoff you actually receive depends on whether the other one selects “left”, “center” or “right”. Thus your payoff depends on your own decision as well as that of the other one. The superscript (raised) number in any box is always the possible payoff of the other one. For example, if the other one decides on “left”, then his/her possible payoff points are 68, 52 and 11. This means, for example, that if you decide on “middle” and the other one decides on “right”, your payoff is 78 points. The payoff for the other one is 49 points in this case.

The possible payoff points on the input screen above are therefore as follows:

You choose “top”; the other one chooses “left”:

Your payoff is:	78 points
The payoff for the other one is:	68 points

You choose “top”; the other one chooses “center”

Your payoff is:	72 points
The payoff for the other one is:	23 points

You choose “top”; the other one chooses “right”:

Your payoff is:	12 points
The payoff for the other one is:	20 points

You choose “middle”; the other one chooses “left”:

Your payoff is:	67 points
The payoff for the other one is:	52 points

You choose “middle”; the other one chooses “center”:

Your payoff is:	59 points
The payoff for the other one is:	63 points

You choose “middle”; the other one chooses “right”:

Your payoff is:	78 points
The payoff for the other one is:	49 points

You choose “bottom”; the other one chooses “left”:

Your payoff is:	21 points
The payoff for the other one is:	11 points

You choose “bottom”; the other one chooses “center”:

Your payoff is:	62 points
The payoff for the other one is:	89 points

You choose “bottom”; the other one chooses “right”:

Your payoff is:	89 points
The payoff for the other one is:	78 points

Please note that the possible payoff points for you and the other one remain the same in every round.

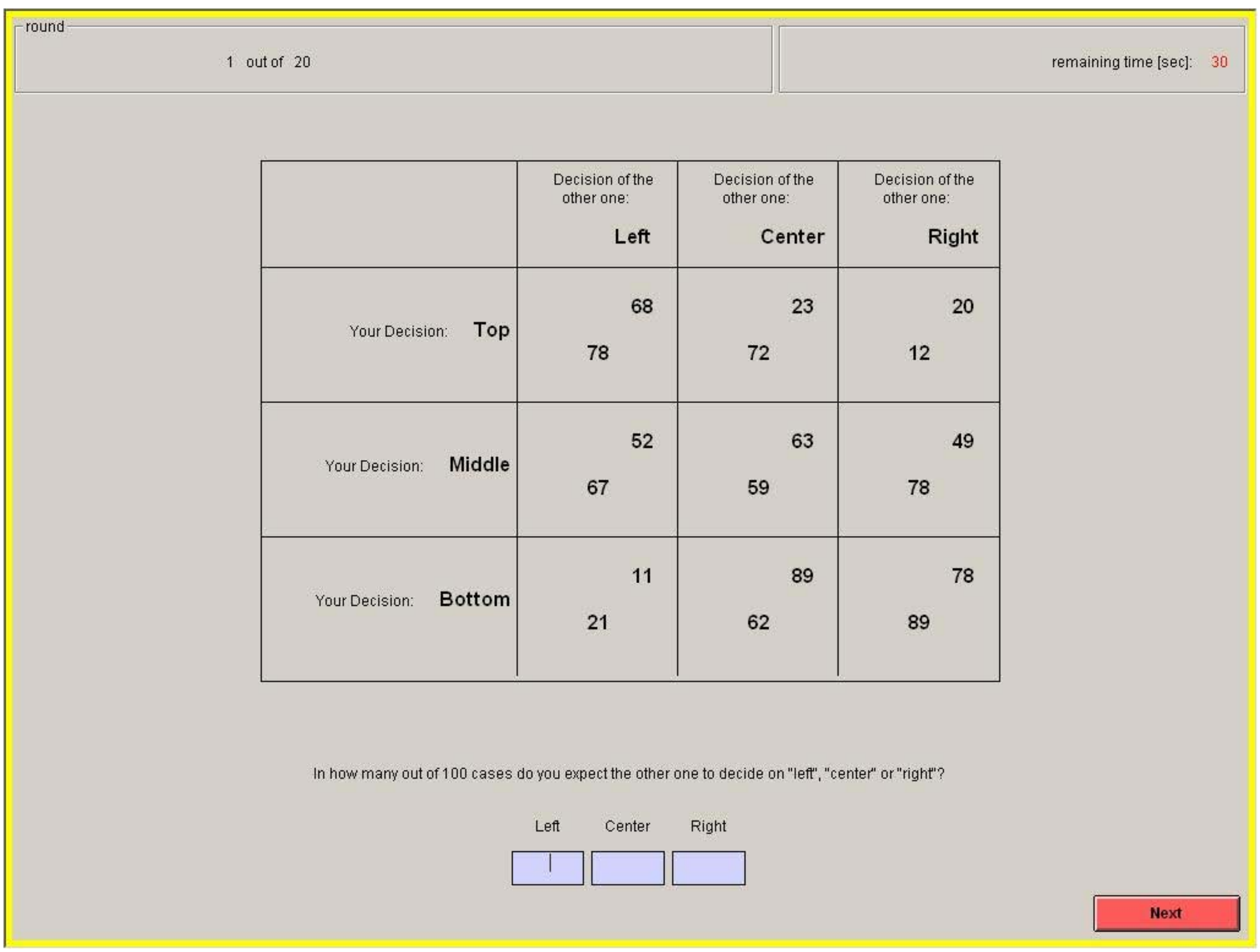
The other one always has exactly the same input screen in front of him/her as you do. After you and the other one have chosen between the three alternatives, you will be informed of your payoff in this round. This is the only information you will be given during the experiment in each round. The next round begins after that.

Statement of expectations

a) How can you state your expectations? Before each decision-making situation, you will be asked how you estimate the decision-making behavior of the other one. This means that at the beginning of each round we will require you to predict how the other one will decide in this round. You will have to answer the following question:

In how many out of 100 cases do you expect the other one to decide on "left", "center" or "right"?

Of course, the other one makes his decision only once in each round. You could also consider the question as asking you to state the likelihood that each of the three alternatives is chosen by the other one. You will see the following input screen on the computer:



Your three alternatives, “top”, “middle” and “bottom”, are listed in the table above, as well as the corresponding possible payoff. Below that, there is the question with the three boxes.

Let us assume that you are sure that the other one will choose “right”, and definitely not “center” or “left”. Then you would respond to our question by entering the number 100 in the box for “right” and the number 0 in the boxes for “center” and “left”. Alternatively, we could assume that you think the other one will probably choose “center”, but there is still a small chance that s/he will choose “right”, and an even smaller chance that s/he will choose “left”. Then, for example, you might respond to our question by entering the number 70 for “center”, 20 for “right” and 10 for “left”.

If you think it is even more unlikely that s/he will choose “center”, then you could enter, for example, 60 for “center”, 24 for “right” and 16 for “left”. Or it is possible that you think it is equally likely that the other one will choose “left”, “center” and “right”. Then you should enter, for example, the numbers 33, 33, 34 in the boxes.

Please note that the three numbers may not be decimal, and that they must always add up to 100.

N.B.: The numbers used in the examples have been chosen arbitrarily. They give you no indication how you and the other one decide.

b) How is the payoff for your stated expectations calculated? Your payoff is calculated after you have guessed how frequently the other one chooses his/her three alternatives. Your payoff depends on the difference between your estimate of the frequency of the decision and the actual decision made. Your payoff is higher when you have guessed that the other one often makes the “true” decision (which s/he really made), and it is lower when you have guessed that the other one will make this decision infrequently. Similarly, your payoff is higher when you have correctly predicted that the other one will not make a particular decision and then s/he in fact does not make the decision.

The exact calculation of the payoff is as follows: We calculate a number for each of the three alternatives. This number reflects how appropriate your estimate of the decision frequency of the corresponding alternative was. We take these three numbers to calculate your payoff.

First, we consider how well you predicted the alternatives which were actually chosen. Let us assume that the other one chose “left”. We then compare your estimate of how often the other one would choose “left” out of 100 cases with the number 100, and calculate the difference between the two. This difference is then multiplied by itself and the resulting number multiplied by the factor 0.0005. Thus, if you expected the other one to choose “left” in many out of 100 cases, then this number will be smaller (since the difference between your estimate and 100 is small) than if you expected that s/he would choose “left” in few out of 100 cases.

Then we consider how well you predicted that the other two alternatives would not be chosen. Let us assume again, for example, that the other one chose “left”, which at the same time means that “center” and “right” were not chosen. Then we take your estimate for the alternative “center” and multiply this by itself. The resulting number is again multiplied by the factor 0.0005. We apply this procedure again to your estimate for the alternative “right”. We then take the three numbers thus calculated and deduct them from the number 10. This determines the number of points you receive for your statement of expectations.

As an illustration of how your payoff might appear, let us consider three examples. Let us assume that the other one chose “left” and that your estimate for “left” was 100 and correspondingly 0 for the other two alternatives. This means that you have stated an estimate that is exactly right. Consequently, you earn the following points:

$$10 - 0.0005 * (100 - 100)^2 - 0.0005 * 0^2 - 0.0005 * 0^2 = 10$$

Let us assume again that the other one chose “left”. Your estimate for “left” was 60, for “center” 20 and for “right” 20, which means that your stated estimate predicted that the other one would choose “left” more frequently than “center” and “right”. Consequently, you earn the following points:

$$10 - 0.0005 * (100 - 60)^2 - 0.0005 * 20^2 - 0.0005 * 20^2 = 8.8$$

If we still assume that the other one chose “left”, but your estimate for “left” was 0, for “center” also 0 and for “right” 100, this means that your stated estimate was exactly wrong. Consequently, you earn the following points:

$$10 - 0.0005 * (100 - 0)^2 - 0.0005 * 0^2 - 0.0005 * 100^2 = 0$$

N.B.: The numbers used in the examples have been chosen arbitrarily. They give no indication how you and the other one decide.

These examples should make it clear that you will always receive a payoff of at least 0 points, and at most 10 points for your stated expectations. And the closer your estimations, the more money you earn. (You may be asking yourself why we have chosen such a payoff ruling as described above. The reason being that with such a payoff ruling, you can expect the highest payment when you state numbers that are closest to your own estimate.)

Procedure and payment The experiment consists of 20 rounds altogether. In each round, you have to first state your estimate of the behavior of the other one, and then make your own decision.

At the end of the experiment, a round each for the decision-making situation and for the statement of expectations will be chosen randomly in order to determine your earnings in the experiment. The choice of both rounds will be made randomly by the experiment leader throwing a dice. The chosen rounds will then be entered onto the input screen by the experiment leader. At the end of the experiment, you will see an overview of your earnings from the decision-making situation and your earnings from the statement of expectation, as well as the total amount. The payoff that you have attained in the corresponding round chosen will be converted at a rate of

1 point = 15 cents

and will be paid out in cash.

Do you have any questions?

Control questions Now you have to answer 7 questions. In this way we are checking whether you have understood the decisions you have to make during the experiment. Should you have any further questions, please raise your hand and one of the experiment leaders will come to you. The experiment will not start until all participants have answered the control questions correctly.

The decision-making situation:

1. If you choose “bottom” and the other one chooses “center”, how many points do you earn?

   ————————

2. If you choose “middle” and the other one chooses “left”, how many points does the other one earn?

   ————————

3. If we assume your payoff amounts to 12, which decision did the other one make?

   ————————

4. If you choose “bottom” and the other one chooses “left”, how much do you earn and how much does the other one earn?

   The other one:____________ You:___________

5. Consider the following two cases:

   You expect the other one to choose “left” in 80 out of 100 cases. The other one actually does choose “left”. You expect the other one to choose “left” in 20 out of 100 cases. The other one actually chooses “right”. In both cases we assume that you expect the other one to choose “center” in 0 out of 100 cases.

   Is your payoff for the statement of expectation in the first case:

   higher the same lower (Please underline your answer!)

   than in the second case?

6. Imagine that Participant 1 states the following expectation: The other one chooses “left” in 50 out of 100 cases, “center” in 20 out of 100 cases, and “right” in 30 out of 100 cases. Participant 2 expects the following: the other one chooses “left” in 60 out of 100 cases, “center” in 20 out of 100 cases, and “right” in 20 out of 100 cases. We will assume that the other one chose “left” by Participant 1 as well as by Participant 2. Who will receive the highest payoff?

   Participant _____

7. If you consider all three alternatives to be equally possible, which numbers should you then enter?

   left:_________ center:__________ right:__________

Thank you for participating in the experiment!