The One Player Guessing Game: A diagnosis on the relationship between equilibrium play, beliefs, and best responses.*

Ciril Bosch-Rosa†
Thomas Meissner‡

December 10, 2019

Abstract

Experiments involving games have two dimensions of difficulty for subjects in the laboratory. One is understanding the rules and structure of the game and the other is forming beliefs about the behavior of other players. Typically, these two dimensions cannot be disentangled as belief formation crucially depends on the understanding of the game. We present the one-player guessing game, a variation of the two-player guessing game (Grosskopf and Nagel, 2008), which turns an otherwise strategic game into an individual decision-making task. The results show that a majority of subjects fail to understand the structure of the game. Moreover, subjects with a better understanding of the structure of the game form more accurate beliefs of other player's choices, and also better-respond to these beliefs.

Keywords Guessing Game · Strategic Thinking · Cognitive Sophistication

JEL Classification C91 · D12 · D84 · G11

*Corresponding author: cirilbosch@gmail.com.

The authors would like to thank David Freeman, Dan Friedman, Yoram Halevy, Frank Heinemann, Rosemarie Nagel, Luba Petersen, Pedro Rey Biel, Christian Seel, and especially Marina Agranov and Yves Breitmoser for their comments. Additionally, the authors acknowledge financial support from the Deutsche Forschungsgemeinschaft (DFG) through CRC 649 "Economic Risk". The first author thanks the Deutsche Forschungsgemeinschaft (DFG) for funding through the CRC TRR 190 "Rationality and Competition."

†Chair of Macroeconomics, Technische Universität Berlin.

‡Maastricht University

1 Introduction

Subjects in laboratory experiments consistently deviate from equilibrium behavior (Camerer, 2003). Many models of bounded rationality try to explain these deviations through errors in belief formation (e.g., Nagel (1995); Ho et al. (1998); Weizsäcker (2003)). Another explanation is that subjects fail to fully understand the structure of the game (Chou et al. (2009) refer to this as an absence of "game form recognition"). Generally, when analyzing deviations from equilibrium behavior, one would expect both of these effects to play a role. However it is typically hard (if not impossible) to distinguish between the two, as correct belief formation crucially depends on a correct understanding of the structure of the game. With the help of a novel one-player guessing game experiment, we are able to disentangle these two effects, thus improving the understanding of why subjects deviate from equilibrium behavior.

An extensive literature has attempted to analyze both belief formation and understanding the structure of the game. Costa-Gomes and Crawford (2006) present subjects with a series of two-player dominance-solvable games and conclude that most subjects understand the games, but play non-equilibrium strategies due to their "simplified models of others' decisions." In Costa-Gomes and Weizsäcker (2008) the authors look at subject's actions and their stated beliefs, and find that subjects rarely best respond to their stated beliefs. However, Rey-Biel (2009) observes that in simplified versions of the games studied in Costa-Gomes and Weizsäcker (2008), Nash Equilibrium is a better predictor of subject behavior than any other model based on level-K reasoning.

Another strand of the literature focuses on whether subjects understand the structure of the game. Using two-player guessing games, Chou et al. (2009) find that subjects are surprisingly unable to understand the experimental setup they are participating in. By using different sets of instructions for the same game, and by introducing hints, they show that subjects do not deviate from equilibrium because of cognitive biases, but rather because of a lack of game form recognition, which they define as the relationships between possible choices, outcomes, and payoffs. Fragiadakis et al. (2016) let subjects play a two-player guessing game repeatedly against random opponents, and subsequently ask subjects to replicate or best respond to their previous choices. They find that while behavior of only 30% of subjects is consistent with a set of commonly used models (including equilibrium play and level-K), they also identify subjects who play strategically but are not identified by commonly used models. Finally, Agranov et al. (2015) develop an experimental protocol that allows them to track the decision-making process of subjects in a beauty contest game. The results show that around 45% of subjects consider playing weakly dominated strategies at some point in their decision-making process.

In this experiment, we use a one-player guessing game which allows to measure how well subjects understand the structure of the two-player guessing game (Grosskopf and Nagel, 2008). In this "game" subjects play the role of both players in a two-player guessing game. That is, they are asked to pick not one but two numbers between 0 and 100, and are paid according to the proximity of each of their choices to two thirds of the average of both choices. This setup switches off the belief channel, but still demands subjects to understand the structure of the game.[1] By comparing their actions in a two-player guessing game to the choices made in the one-player guessing game we can disentangle the effects of beliefs from "game form recognition," and analyze to what extent understanding the structure of the game determines their belief formation and their best-responses.[2]

Our experimental results show that a majority of subjects fails to fully solve the one-player guessing game, and that subjects with a better understanding of the structure of the one-player guessing game play values closer the Nash Equilibrium in the two-player guessing game. This implies that an important part of non-equilibrium play is likely due to the inability of subjects to fully understand the structure of the game. Additionally, we observe that subjects with a better understanding of the one-player guessing game form more accurate beliefs, are better at best-responding to their own beliefs, and tend to better adjust their beliefs according to the population they face. These results confirm the intuition that understanding the structure of the game is crucial for belief formation.

2 Experimental Design

The experiment consists of four different parts: Subjects first play the one-player guessing game (1PG), followed by the two-player guessing game (2PG). After this, we elicit subjects' beliefs about other subjects' two-player guessing game choices. A subset of subjects

[1] The latter feature would be missing when simply telling subjects what the other player of a two-player guessing game will play.

[2] For convenience we will henceforth refer to the one-player guessing game as a game, even though strictly speaking it is not one.

then participated in an additonal belief elicitaton task ("What-if" belief elicitation). At the end of the experiment, all subjects are asked to answer a battery of cognitive ability tests. In the following we describe each part of the experiment in more detail.

2.1 The One-Player Guessing Game (1PG)

The one-player guessing game, first introduced in Bosch-Rosa et al. (2018), allows to test whether subjects can solve the two-player guessing game introduced by Grosskopf and Nagel (2008) *free of any strategic concerns.*[3]

In essence, subjects play the role of both players in a two-player guessing game, i.e. they play the two player guessing game "against themselves." Accordingly, each subject (i) picks two numbers $x_i \in [0, 100]$ and $y_i \in [0, 100]$ and is paid depending on the absolute distance of each chosen number to the "target value" which is two thirds of the average of both numbers. The further away each chosen number is from this target value, the lower is the payoff. Formally the experimental payoff for choosing number x_i and y_i is:

$$\pi_i^{1PG}(x_i) = 1€ - 0.05€ \left| x_i - \frac{2}{3}\frac{x_i + y_i}{2} \right| \tag{1}$$

$$\pi_i^{1PG}(y_i) = 1€ - 0.05€ \left| y_i - \frac{2}{3}\frac{x_i + y_i}{2} \right| \tag{2}$$

Subjects are paid for both choices, so their combined payoff is:

$$\Pi_i^{1PG} = \max \left[\pi_i^{1PG}(x_i) + \pi_i^{1PG}(y_i), 0 \right]. \tag{3}$$

The payoff function is maximized at ($y_i = 0, x_i = 0$). This solution can be found through logical induction by starting with a random value $x_{0,i}$, and then calculating the "best response" which is $y'_{1,i} = \frac{1}{2}x_{0,i}$. Following this, a "best response to the best response" can be calculated ($x'_{1,i} = \frac{1}{2}y'_{1,i}$) and so on until reaching the fixed point ($x'_{\infty,i} = 0$, $y'_{\infty,i} = 0$).

[3] The idea of letting people play games with themselves, in order to test understanding of the structure of a game is not entirely new. See for instance Blume and Gneezy (2010) or Petersen and Winn (2014).

By turning the two-player guessing game into an algebraic problem with no strategic uncertainty, we can separate those subjects who can solve the mathematical problem associated with the guessing game from those who cannot.[4]

2.2 The Two-Player Guessing Game (2PG)

The two-player guessing game that we use is an adaptation of the one presented in Grosskopf and Nagel (2008) and Nagel et al. (2016). Subjects are matched in pairs and asked to simultaneously pick a number $z_i \in [0, 100]$. In Grosskopf and Nagel (2008) the winner is whoever picks the number closer to 2/3 of the average of both numbers, so unlike the games with $N > 2$ subjects, now $z_i = 0$ is a (unique) weakly dominant strategy. In our version of the 2PG, the payments are based on the (absolute) distance of each individual pick to 2/3 of the average of both numbers. Formally, the payment for player i depends on the choices of player j and her own in the following way:[5]

$$\Pi_i^{2PG} = \max \left[2€ - 0.10 \left| \left(X_i - \frac{2}{3}\frac{z_i + z_j}{2} \right) \right|, 0 \right]. \tag{4}$$

This small change in payoffs dramatically changes the game as now the equilibrium is reached through iterated deletion of *strictly dominated* strategies, and zero is no longer a weakly dominant strategy. Now the best response is to choose 1/2 of the number a player believes the other player chooses.[6]

We opted for this modification of the original game for two reasons. First, it allows us to *de facto* ask subjects for a point estimate of their belief about the other subject's choice, and secondly, and more important, it makes the game comparable to the 1PG. Note that while certainly not standard, distance-based payoff structures are widely used in the literature. Güth et al. (2002) first utilized such a payoff structure, arguing that it

[4]A penalty which increases in the distance to the target guess is crucial here, as with the more commonly utilized tournament incentives one of the two selves would always win, and the game would become trivial. Additionally, this payoff structure allows us to study the best response of subjects to their beliefs in the 2PG as 0 is not anymore a unique weakly dominant strategy. See Sections 2.2 and 3.4 for more details.

[5]Note that we limited the minimum payoff to zero in order to avoid potential losses for the subject.

[6]See Nagel et al. (2016) for a lengthier discussion on the implications that the different payoff structures have in the 2PG.

more closely resembles the financial decision-making situations that beauty contests are often intended to emulate. Since then a number of experiments have used distance-based beauty contests.[7] Most relevant for our experiment is Nagel et al. (2016), who directly compare distance-based and tournament incentives in two player guessing games and find no significant differences across the choices of subjects.

2.3 Belief Elicitation

After subjects had played the 1PG and the 2PG (with no feedback in both cases) we elicited their beliefs about the other players' decisions in the 2PG. Similar to Lahav (2015), subjects were asked to distribute a total of 19 "tokens" into 20 "bins".

Each token represented a subject in the session (each session consisted of 20 subjects), and each bin had a range of 4 integers that players could play in the 2PG (i.e. the first bin had the range [0,4], the second [5,9], and so on). See Figure 11 in Appendix C for a screen-shot of the experimental interface.

To incentivize subjects, we used a linear scoring rule that paid €0.10 for each token that overlapped with the choice of any other subject in the 2PG. For instance, if a subject put three tokens in the bin "5-9" and in her session only 2 subjects had actually played *any* value within this range, then she would receive a total of 20 cents for the tokens allocated in that bin. If, on the other hand, she placed 5 token in the bin "0-4" and 10 subjects had played a value in this range, then she would be paid 50 cents for the tokens allocated in that bin.

Formally, define b_{ij} as the number of tokens that subject i deposited in bin j, and p_{-ij} as the number of subjects other than player i that chose a value that falls within bin j in the 2PG. Then the payoff for belief formation for subject i is:

$$\Pi_i^B = \sum_{j=1}^{20} 0.10€ * \min\left[b_{ij}, p_{-ij}\right] \quad (5)$$

[7] See for instance Hommes et al. (2004), Costa-Gomes and Crawford (2006), Mauersberger and Nagel (2018), or Duffy (2016) who reviews some of these experiments among other types of macroeconomic experiments.

The resulting distribution of beliefs provides an estimate of what subjects think about other subjects' choices, and allows us to analyze how subject best-respond to their own beliefs.[8]

2.3.1 "What if" Belief Elicitation

Since playing the 1PG could have an influence on the beliefs subjects form in the 2PG, we asked a subset of 40 subjects to additionally guess the choices of players in a 2PG *who had not previously taken part in the 1PG*.[9] The incentives for this elicitation task are the same as the ones described above, and the data came from a random pick of 19 subjects from a sample of 80 subjects who we had invited two weeks earlier to participate in a 2PG without previously taking part in the 1PG.

2.4 Cognitive Ability

Gill and Prowse (2016) show that subjects who score higher in a Raven Test (Raven, 1960) choose numbers closer to equilibrium, earn more, and converge quicker to equilibrium in a three-player guessing game.[10] Since we are interested in studying the ability of subjects to solve the guessing game, we also tested the cognitive ability of our subjects. In particular, all subjects answered a Raven Test and played "Race-to-60," a variant of the Race game (see e.g. Gneezy et al. (2010), Levitt et al. (2011)).[11] The Raven Test is a multiple choice test in which subjects must pick an element that best completes a missing element in a

[8]There is some discussion about how to best incentivize subjects to state their true beliefs. In particular, there is mixed evidence on whether incentive compatibility matters or not (Schotter and Trevino, 2014). Methods to elicit beliefs beyond first moments, such as ours, are typically difficult for subjects to understand. Using a non-linear scoring rule would introduce an additional level of complexity. Hence, while not incentive compatible for risk neutral subjects, we opted for this approach because we believe it provides the best compromise between tractability for subjects and incentivization efficacy.

[9]In the instructions we told subjects they had to guess the choices made by 19 subjects who had played the 2PG a couple of weeks ago, *without having previously played the 1PG*. See Appendix D for the instructions read to subjects.

[10]Note however that Georganas et al. (2015) find only limited evidence of correlations between a variety of cognitive ability tasks and level-k reasoning ability.

[11]While in Gill and Prowse (2016) subjects go through all 60 matrices of the original Raven Test, in our case subjects just took part in three of the hardest blocks of 12 matrices.

matrix of geometrical shapes (see an example in Figure 12 of Appendix C). The score of this test has been found to correlate with measures of strategic sophistication and the ability of subjects to solve novel problems (Carpenter et al., 1990). It is increasingly used in economic research due to its simplicity and the lack of required technical skills.

Since logical induction is a central element of the guessing games, we test this ability with the "Race-to-60" game. In this game, each participant and a computerized player sequentially choose numbers between 1 and 10, which are added up. Whoever is first to push the sum to or above 60 wins the game. The game is solvable by backward induction, and the first mover can always win by picking numbers such that the common pool adds up to the sequence : [5; 16; 27; 38; 49; 60]. Subjects always move first and therefore, independent of the computer's backward induction ability, can always win the game.[12]

3 Results

A total of 80 subjects participated in this experiment. All subjects were recruited through ORSEE (Greiner, 2015) and were mostly undergraduate students with a variety of backgrounds, ranging from anthropology to electrical engineering or architecture. Sessions lasted one and a half hours and were run at the Experimental Economics Laboratory of the Technische Universität Berlin. Subjects who had previously participated in guessing game experiments were not invited. The experiment was programmed and conducted using z-Tree (Fischbacher, 2007). For detailed results on the cognitive ability tests, see Appendix A.

3.1 The One Player Guessing Game

In Figure 1 we present the results of the 1PG in a scatter plot. Recall that in this case subjects have to pick two numbers, (x_i, y_i); the first number is depicted on the horizontal axis, the second on the vertical axis. The diagonal dashed line marks the points where a subject picked the same number for x_i and y_i. The solid circle indicates subjects who fully solved the game (0,0).

[12]As in Bosch-Rosa et al. (2018) the backward induction ability of the computer increased with the rounds of the game. In the first round the computer could do only one step of backward induction, in the second it could do two, in the third three, and so on.

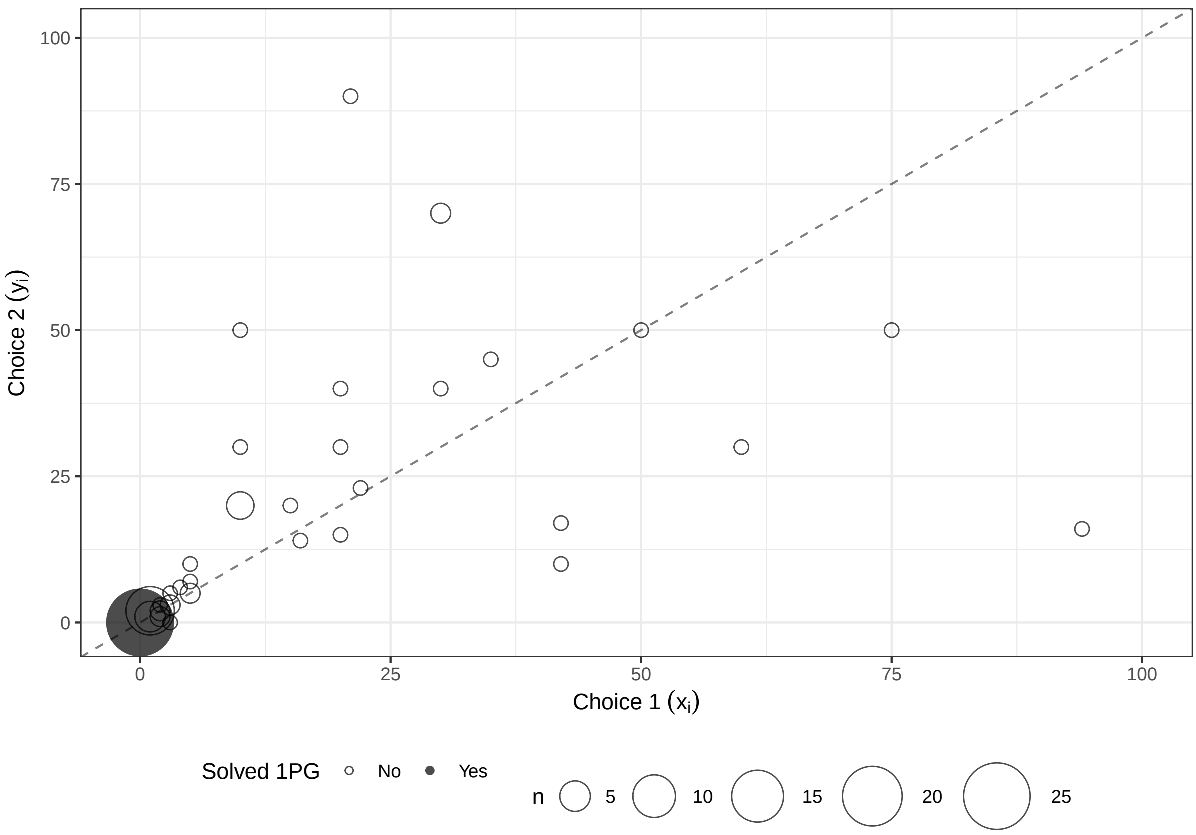

Figure 1: Scatter plot of the choices made by each subject. Darker dots refer to subjects that fully solved the game (0,0). See Figure 13 in Appendix C for a zoom in plot depicting only the choices from [0,50].

As can be seen, only a minority (≈31%) of subjects is able to fully solve the 1PG, i.e., pick zero for both numbers. In the remainder of this paper we will use this ability to fully solve the game as our primary measure of understanding of the structure of the guessing game.

Result 1: *Only 31% of our subjects fully understand the one-player guessing game.*

Another interesting observation in Figure 1 is that subjects who play numbers closer together also play numbers closer to the origin. This is relevant, as in the 1PG there are two ways in which a subject (who has not fully solved the game) can improve her payoffs: by picking numbers closer to zero, and/or by picking numbers that are closer to each other. A Spearman test confirms the correlation between higher average of both choices and the distance between them (Spearman $\rho = 0.83$, p-value < 0.001). As subjects with high payoffs played both numbers that were close to each other, and to zero, one could interpret the payoffs of the 1PG as a measure of (partial) understanding of the structure of the guessing game. Therefore, we will use the payoffs of the 1PG as a secondary measure to complement to our primary measure of understanding, “Solved 1PG”/“Not solved 1PG”.

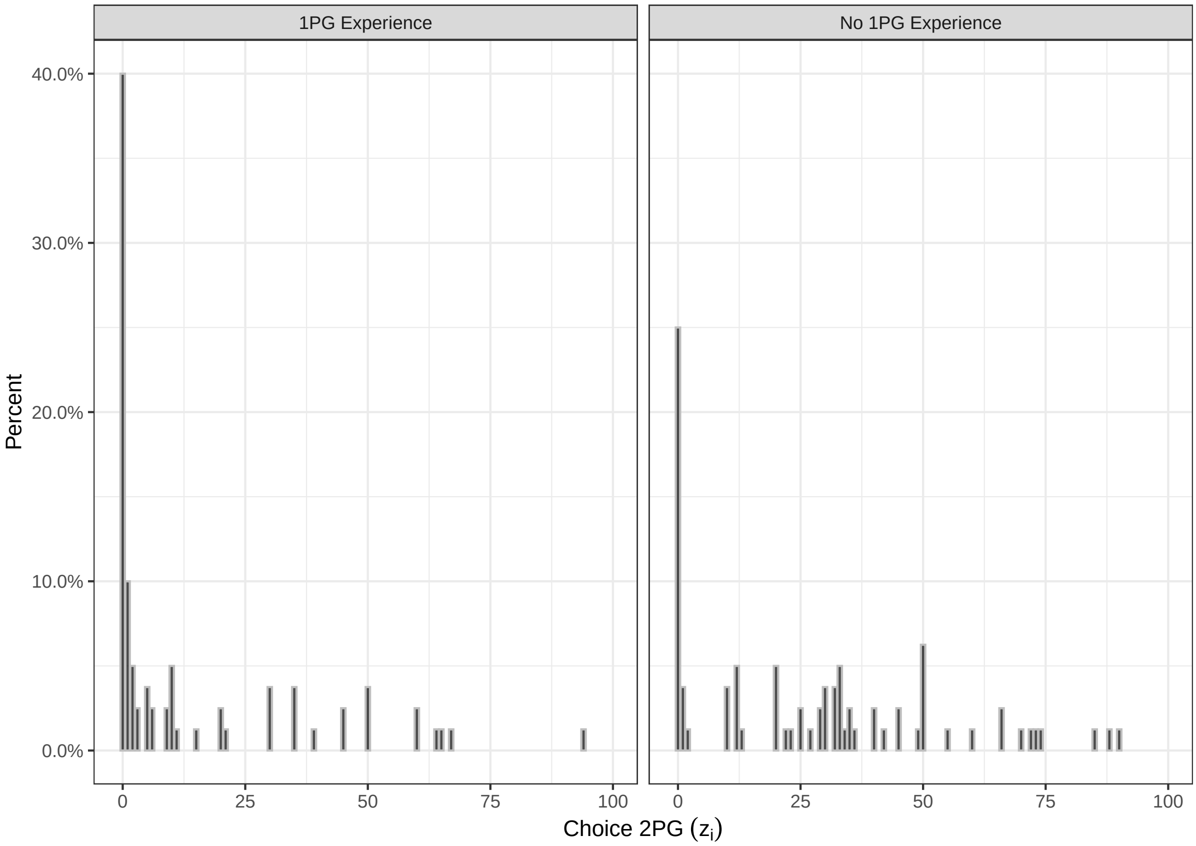

Figure 2: Distribution of choices in the 2PG with and without 1PG experience.

3.2 The Two Player Guessing Game

The left panel of Figure 2 shows the distribution of choices in the 2PG, for subjects who have played the 1PG before. The distribution appears to be quite different from the typical distribution one sees with guessing game “first timers.” The mass of the distribution is close to zero with 50% of subjects playing Nash Equilibrium.[13] The mean

[13]Note that in the 2PG there are two Pareto ranked Nash Equilibria in pure strategies: i) both players choose zero and ii) both players choose one. The latter is because the best response to the counterpart picking 1 would be picking 1/2. Since in our experiment subjects can only chose integers, both 0 and 1 are best responses. Such Pareto-ranked equilibria will *always exist* in laboratory experiments as it is impossible to implement a truly continuous choice.

is 13.47 and the median choice is 2. As mentioned in Section 2.3.1 we also collected data on 80 subjects who played the 2PG *without previously* taking part in the 1PG. Choices of these subjects are presented in the right panel of Figure 2. While a relatively large number of subjects with no prior 1PG experience also play Nash Equilibrium (28.75%), the overall distribution of 2PG choices for those subjects is significantly shifted to the right compared to that of subjects with prior 1PG experience (Kolmogorov Smirnov test, p-value < 0.001, see Figure 14 in Appendix C for cumulative density plots). This shift results in a mean and median choice without 1PG experience of 27.8 and 26 respectively.

The difference in behavior between both groups could be the result of two phenomena: introspective learning from having played the 1PG (Weber, 2003), or a change in the beliefs of subjects that previously played the 1PG given that they are facing a more "experienced" pool of subjects (Agranov et al., 2012).[14] In section 3.4.1 we show that shifts of beliefs are relatively small. Therefore, we attribute most of the difference in behavior to introspective learning. So, while most subjects are not able to fully solve the 1PG, there appears to be some learning that carries over to the 2PG.

3.3 Relationship between the 1PG and the 2PG

Figure 3 shows the decisions of subjects in the 2PG on the vertical axis, and their payoffs for the 1PG on the horizontal axis. Subjects who fully solved the 1PG (solid circles) mostly chose zero in the 2PG (24/25, 96%), and picked significantly lower numbers in the 2PG than subjects who did not fully solve the 1PG (Mann-Whitney U Test, p-value < 0.001). In line with this, we also observe that subjects who earn higher payoffs in the 1PG play lower numbers in the 2PG (Spearman ρ = -0.745, p-value < 0.001)

Result 2: *Subjects with a better understanding of the structure of the one-player guessing game play numbers closer to the Nash Equilibrium in the two-player guessing game.*

But, is playing numbers near the Nash Equilibrium the best strategy in the 2PG? To answer this question we construct $\bar{\Pi}_i^{2PG}$. This variable represents the payoff that each

[14] Arguably, some learning could be due to a switch to simple heuristics (e.g. "play zero") rather than real learning about the structure of the game. Using games that reverse the end of the strategy space containing the equilibrium such as Ho et al. (1998) or Rick and Weber (2010), or games with interior equilibria as in Costa-Gomes and Crawford (2006) would allow to distinguish between these different kinds of learning. We thank the editor for this comment.

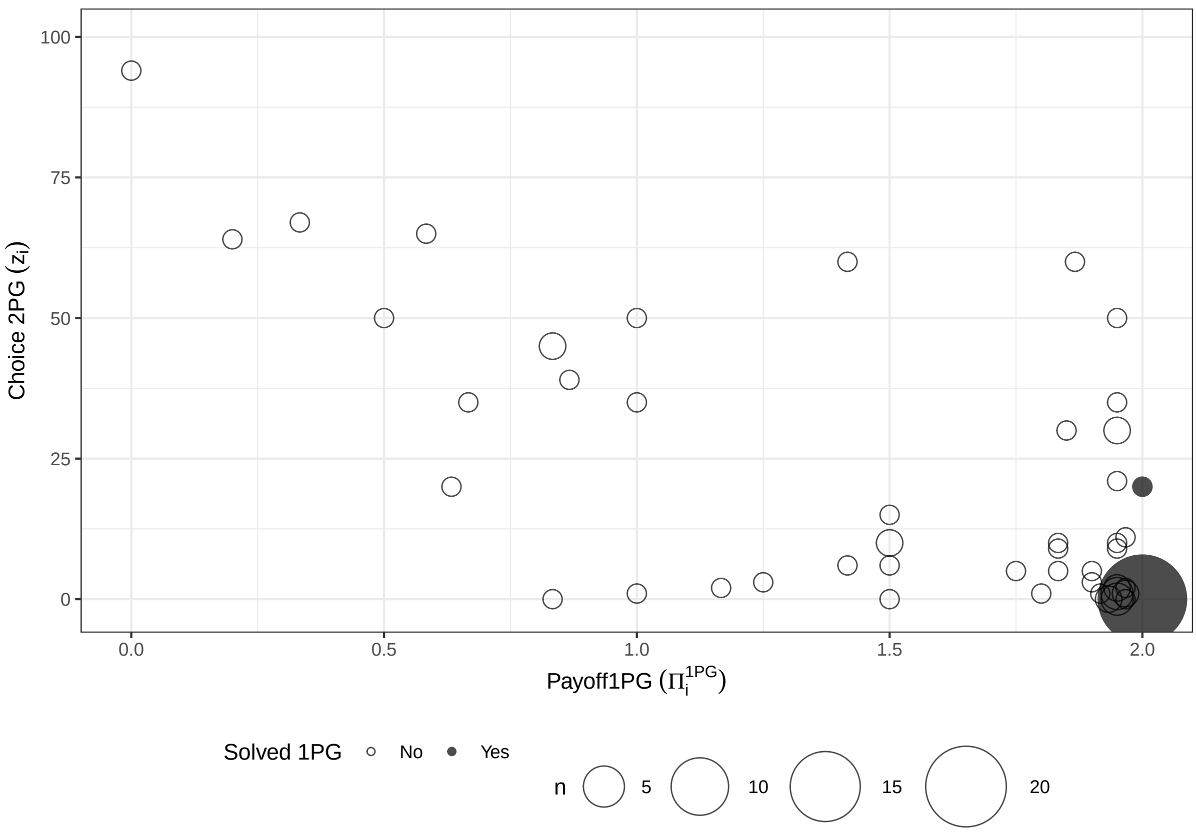

Figure 3: Distribution of choices in the 2PG (vertical axis) and payoff in the 1PG (horizontal axis).

subject i would have gotten had she played against the average choice of all other subjects j except herself (i.e., $j \neq i$). Formally $\bar{\Pi}_i^{2PG}$ is defined as:

$$\bar{\Pi}_i^{2PG} = 2€ - 0.10 \left| z_i - \frac{2}{3} \frac{z_i + \sum_{j \neq i}^{N} z_j/(N-1)}{2} \right| \tag{6}$$

Figure 4 illustrates the relationship of $\bar{\Pi}_i^{2PG}$ with both, the payoffs of the 1PG (Π_i^{1PG}, left panel) and choice in the 2PG (z_i, right panel). Interestingly, subjects who fully solved the 1PG don't have the highest $\bar{\Pi}_i^{2PG}$. This is because they play Nash Equilibrium, when payoffs would have been maximized by playing a number close to 9 as can be seen in the right plot. Overall, subjects who fully solved the game did not earn a significantly different payoff compared to subjects who did not fully solve the game (Mann-Whitney U test p-value = 0.465).

Analyzing our secondary measure of understanding of the structure of the game (Π_i^{1PG}) reveals a more nuanced pattern: it appears that there is a non-monotonic relationship between understanding of the structure of the game and expected payoffs in the 2PG

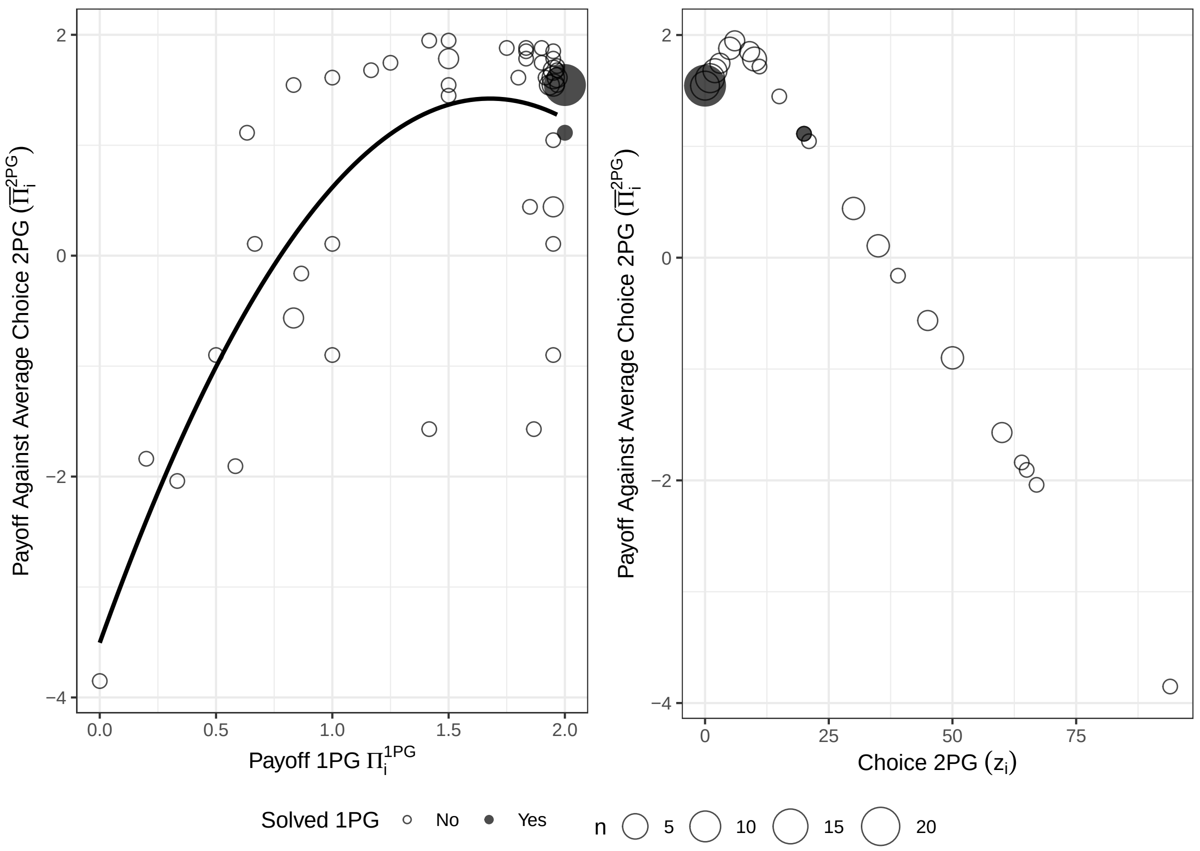

Figure 4: Relationship of payoff in the 1PG (Π_i^{1PG}) and $\bar{\Pi}_i^{2PG}$ (left panel). The line in the left panel is a fitted quadratic function. The right panel shows the relationship of choice in the 2PG (z_i) and $\bar{\Pi}_i^{2PG}$. In both panels the darker dots indicate subjects who fully solved the 1PG.

(see Figure 15 in Appendix C for a close-up of Figure 4.). Regressing $\bar{\Pi}_i^{2PG}$ on Π_i^{1PG} and $(\Pi_i^{1PG})^2$ yields coefficients that are significantly positive and negative respectively. This gives statistical support to the fitted quadratic function in the left panel of Figure 4, implying that increased understanding leads to increased expected payoffs, but that this relationship reverses for very high levels of understanding.

Result 3: *The relationship between understanding of the structure of the one-player guessing game and payoffs in the two-player guessing game follows a non-monotonic pattern.*

3.4 Subjective Beliefs

On the left panel of Figure 5, we plot the number of tokens subjects have placed correctly in the belief elicitation task against the payoff in the 1PG. Subjects who fully solved the 1PG placed a larger number of tokens correctly (Mann-Whitney U test, p-value =

	$\bar{\Pi}_i^{2PG}$
Π_i^{1PG}	5.519***
	(0.904)
$(\Pi_i^{1PG})^2$	-1.565***
	(0.353)
Constant	-3.389***
	(0.500)
N	80
adj. R^2	0.580
Joint test p-value	0.000

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1: Regression of $\bar{\Pi}_i^{2PG}$ on the payoff in the 1PG (Π_i^{1PG}) and its square value $(\Pi_i^{1PG})^2$.

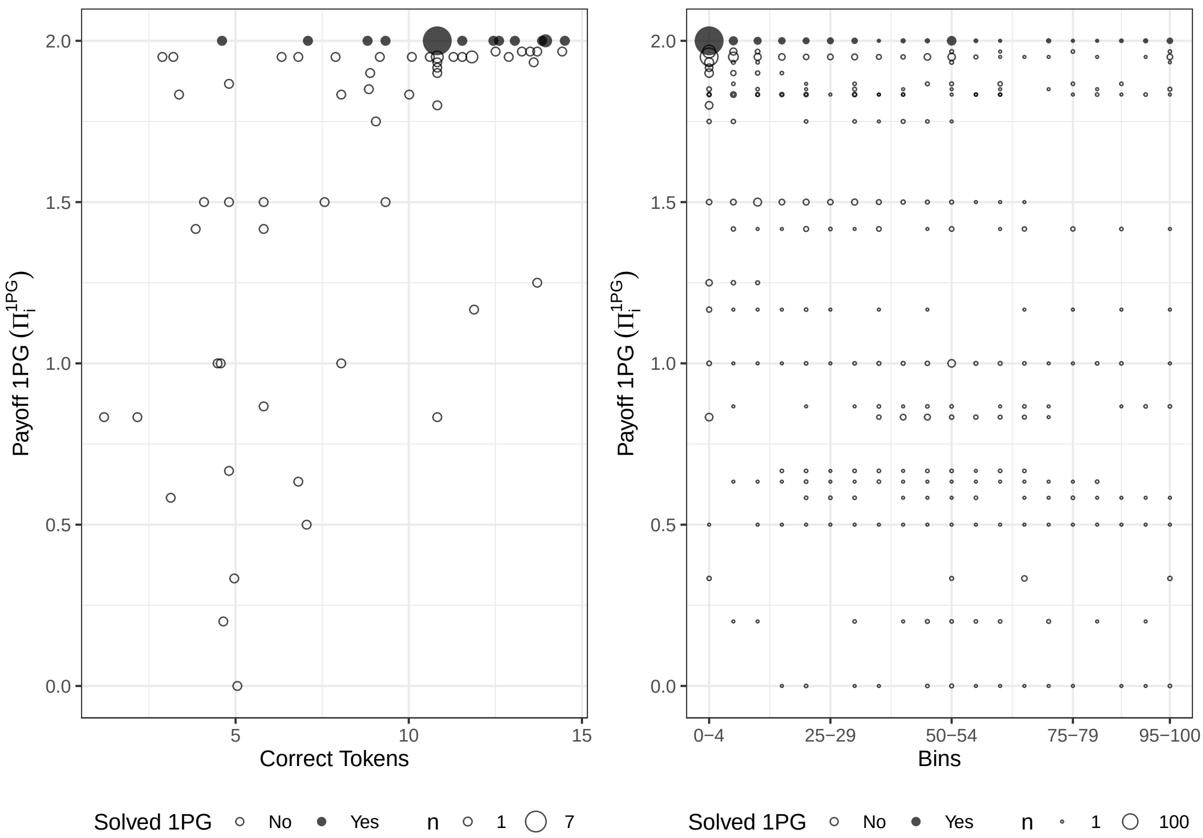

Figure 5: Relationship between the payoff in the 1PG (Π_i^{1PG}) and number of correct tokens deposited in the belief elicitation phase (left panel) and distribution of tokens across bins (right panel). In a darker color those subjects that solved the 1PG.

0.001).[15] This result is confirmed by the strong correlation that we observe between the 1PG payoff and the number tokens placed correctly (Spearman ρ=0.583, p-value < 0.001). On the right panel of Figure 5 we plot the distribution of tokens (horizontal axis) against the payoff in the 1PG (vertical axis). While subjects who did not fully solve the 1PG spread out their tokens across most of the strategy space, subjects who fully solved the 1PG expect their counterparts to play numbers closer to the Nash Equilibrium (Mann-Whitney U test, p-value < 0.001). Again, the correlation between the distance of tokens to NE and payoffs in the 1PG confirms this result (Spearman ρ = -0.359 with p-value = 0.001).

To test how the accuracy of beliefs relates to the understanding of the structure of

[15]Again, in order to avoid noise due to session specific outliers, we compute the number of correct tokens by comparing individual beliefs to the distribution of the 2PG choices we collected across all sessions. For more details see Appendix B.

the game, we plot the mean of the belief distribution of each subject against their payoff in the 1PG (vertical axis) on the left panel of Figure 6.[16] The vertical dotted line marks the mean choice across all subjects in the 2PG (13.63). The right panel of Figure 6 plots the absolute distance of individual mean beliefs to mean 2PG play against earnings in the 1PG. Two things are clear from the graph: First, the mean beliefs of some subjects differ quite a bit from mean actual play in the 2PG. Second, subjects who fully solved the 1PG have a lower absolute difference of their mean beliefs and mean choice of all subjects in the 2PG (Mann-Whitney U test p-value = 0.030). This result is supported using our secondary measure of understanding (Spearman ρ = -0.473 with p-value < 0.001).

Result 4: *Subjects with a better understanding of the structure of the one-player guessing game form more accurate beliefs about their counterparts' choices in the two-player guessing game.*

Additionally, we analyze whether choices in the 2PG are best responses to the stated beliefs (i.e., the token distribution). To do so we compute the choice in the 2PG that would maximize the payoff of a subject conditional on her stated beliefs being correct:

$$z_i^*(B) = \arg\max_{\tilde{z}_i} \sum_{j=1}^{20} \frac{b_{ij}}{19} \left[2 - 0.10 \left| \tilde{z}_i - \frac{2}{3} \frac{\tilde{z}_i + \bar{b}_j}{2} \right| \right], \tag{7}$$

where $z_i^*(B_i)$ is the choice of subject i that maximizes her payoffs given her beliefs $B_i = (b_{i1}, b_{i2}, ..., b_{i20})$, b_{ij} is the number of tokens that subject i put in bin j, and $\bar{b}_j$ is the average value of the bin (so for example, for the first bin [0,4], $\bar{b}_1 = 2$, for the second [5,9], $\bar{b}_2 = 7$, etc.).[17] We then create an individual variable $\Delta z_i^* = |z_i - z_i^*(B_i)|$ which is the absolute difference between actual choice of subject i in the 2PG minus the optimal choice conditional on her stated beliefs. Figure 7 illustrates the relation of Δz_i^* and the payoffs for the 1PG. It appears that subjects who fully solved the 1PG are better at best responding to their own beliefs and therefore have a lower Δz_i^* (Mann-Whitney U test p-value = 0.001). This is confirmed by a significantly negative correlation between 1PG payoffs and Δz_i^* (Spearman ρ =-0.531, p-value < 0.001). These results imply that better

[16] The individual means of the belief distributions are calculated as $m_i = \sum_{j=1}^{20} \frac{b_{ij}}{19} \bar{b}_j$, where b_{ij} is the number of tokens that subject i put in bin j, and $\bar{b}_j$ is the average value of the bin (so for example, for the first bin [0,4], $\bar{b}_1 = 2$, for the second [5,9], $\bar{b}_2 = 7$, etc.).

[17] We pick this instead of the lowest value of the bin, because it is a more stringent test to our "high ability" subjects.

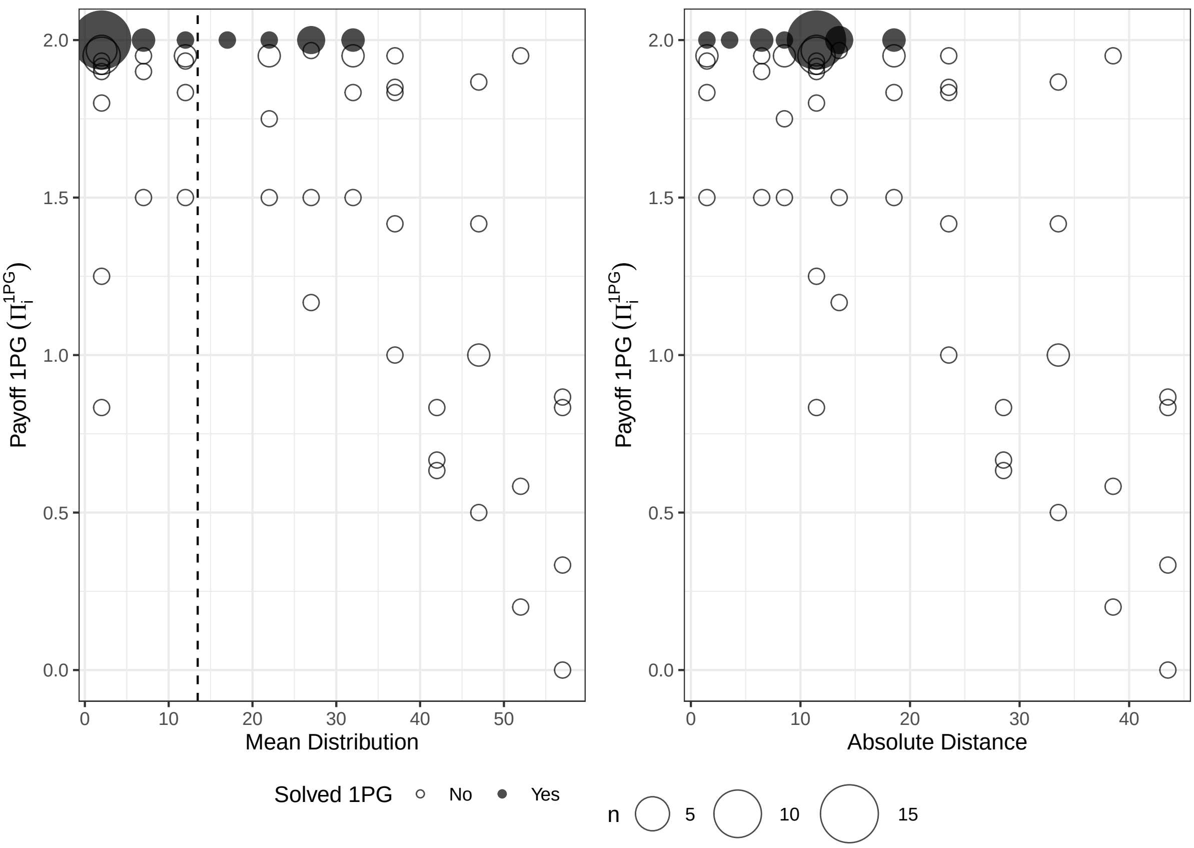

Figure 6: In the left panel we present the relationship between the payoff in the 1PG (Π_i^{1PG}) and the mean value of the distributed tokens (horizontal axis). The vertical dotted line marks the mean of all choices in the 2PG (which is 13.63). The right panel illustrates the relationship between the payoff in the 1PG (Π_i^{1PG}, vertical axis) and the absolute distance between mean choice of subjects in the 2PG and the mean value of the distributed tokens.

understanding of the structure of the guessing game improves the ability to best respond to own beliefs.

Result 5: *Subjects with a better understanding of the structure of the one-player guessing game choose numbers closer to the best response of their beliefs in the two-player guessing game.*

3.4.1 "What-if" Beliefs

As there could be some influence of having played the 1PG on the beliefs in the 2PG, we asked 40 subjects to use 19 tokens to guess the choices of 19 subjects that had played the 2PG "a couple of weeks ago, *without having previously played the 1PG*". We will refer to these distributions as "what-if" distributions, as opposed to the elicited distribu-

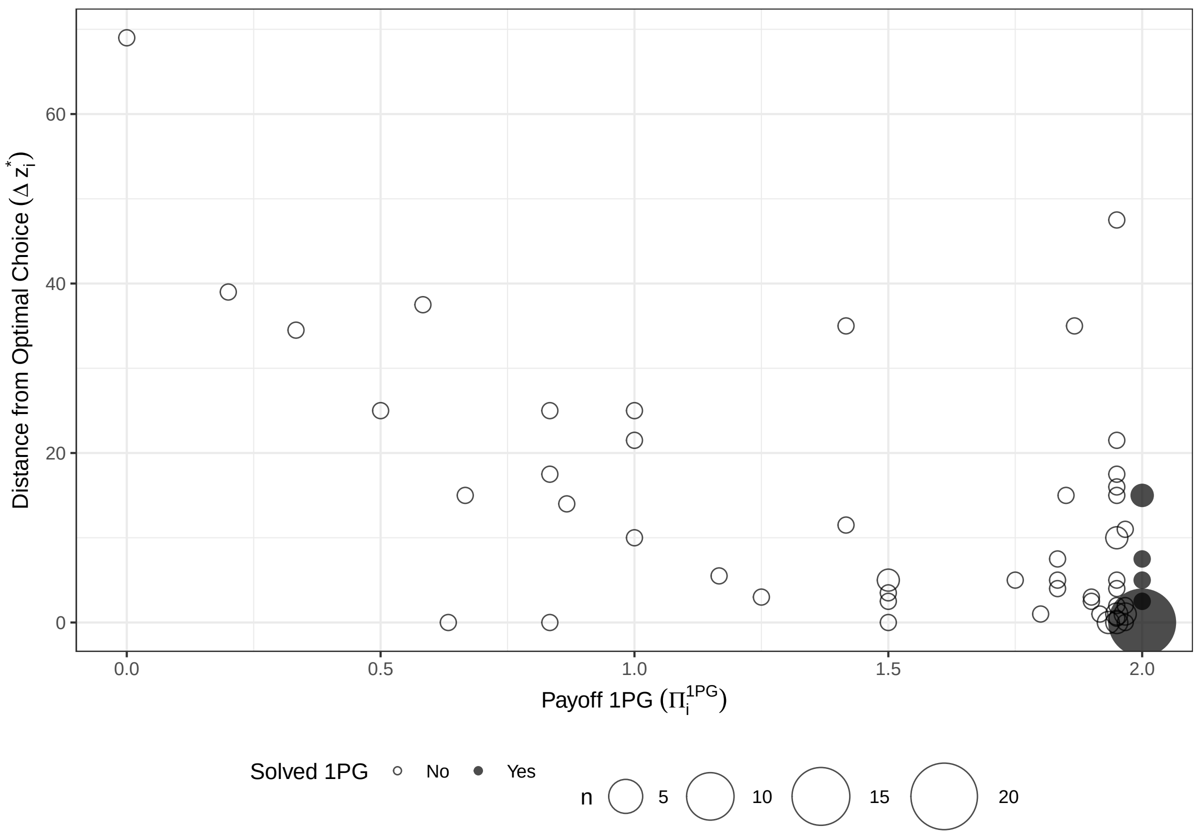

Figure 7: Difference between actual choice and optimal choice conditional on beliefs (Δz_i^*) vs. payoff in the 1PG (Π_i^{1PG}).

tions in the belief elicitation part of the experiment which we will refer to as "original" distributions.

We plot the resulting aggregated distributions in Figure 8. At first glance, the differences between what-if and original distributions appear to be small.[18] However, when comparing the means and variances of subjects' individual distributions, we find that both mean and variance are significantly higher in the what-if distributions (Wilcoxon matched-pairs signed-ranks test, p-value $= 0.039$ and p-value $= 0.009$ for the difference in means and variance respectively). This indicates that subjects adjust their beliefs depending on the population they face.[19]

To get a better understanding of how subjects change their beliefs in the 2PG when

[18]The mean bin for the original distributions is 5.18, while the mean of the what-if distributions is 5.77 (bin number 5 contained values 20-24 while bin 6 contained values 25-29.).

[19]See Figure 16 of Appendix C for a graphical representation of these results.

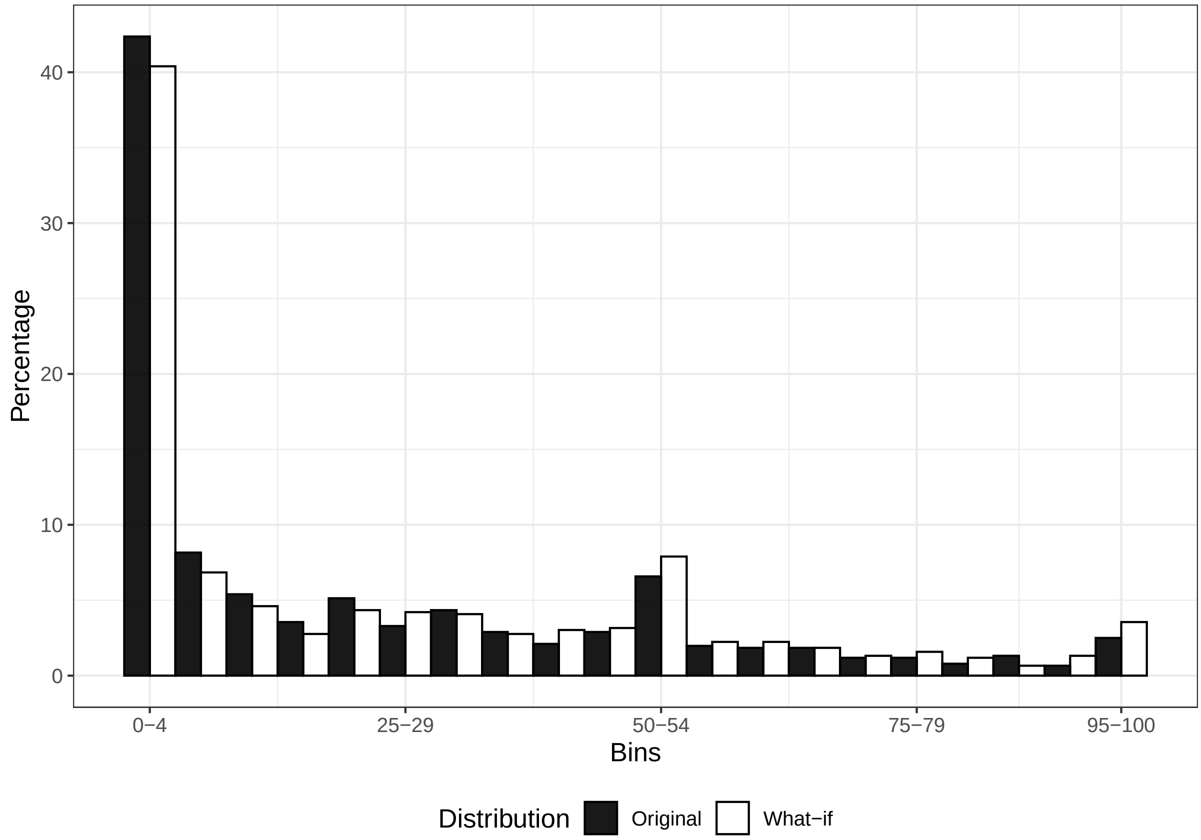

Figure 8: Distribution of “original” and “what-if’ beliefs for the 2PG.

faced with different populations, we plot the difference in means between the what-if and original distributions (ΔB_i) against the individual payoff from the 1PG (Π^{1PG}) in Figure 9. In this figure, any value above the horizontal dotted line indicates a shift of the what-if distribution, with respect to the original one, away from the NE.

As can be seen in Figure 9, whenever subjects who fully solved the 1PG adjust their beliefs, they seem to do so in the right direction (i.e., away from the NE). However, we cannot reject the null hypothesis of no differences in the distribution means (ΔB_i) between subjects who fully solved the 1PG and those who did not (Mann-Whitney U Test, p-value $= 0.683$). Using our secondary measure of understanding, we find a significant correlation between payoffs in the 1PG and ΔB_i (Pearson $\rho = 0.308$, p-value $= 0.053$), but we cannot reject the hypothesis that this relationship is not monotonic (Spearman ρ=0.262, p-value $= 0.101$). Therefore, if we interpret a higher payoff in the 1PG as a better understanding of the structure of the game, then it would appear that a better understanding is associated

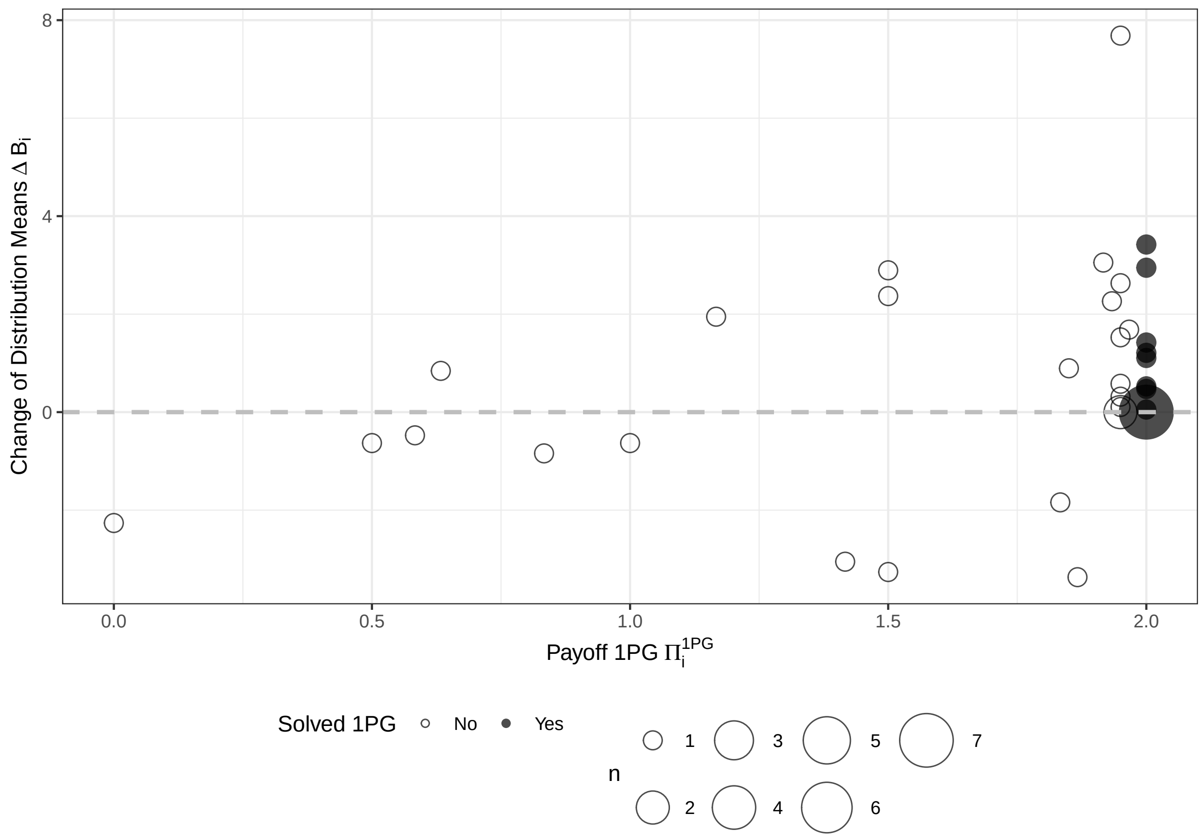

Figure 9: Payoff in the 1PG on the horizontal axis, and change in the mean between belief distributions (ΔB_i) on the vertical axis. Any value of ΔB_i above zero is a change in the mean away from the Nash Equilibrium.

with better adjustment of beliefs in response to facing an inexperienced population.[20]

Result 6: *There is weak evidence suggesting that subjects with a better understanding of the structure of the one-player guessing game are better at adjusting their beliefs in response to facing an inexperienced population in the two-player guessing game.*

[20] An important consideration in within-subject designs are experimenter demand effects. In our setup is not clear in what direction such effects may work. One may conjecture that, if subjects aim to appear rational, the interaction of experimenter demand and the extra opportunity for introspective learning may shift the what-if distribution closer to the Nash Equilibrium. As can be seen in Figures 9 and 16, the shift in distributions seems to be in the opposite direction. This makes us somewhat optimistic that experimenter demand was either small or non-existent.

4 Conclusion

In laboratory experiments, subjects often deviate from equilibrium play. These deviations can be the result of either subjects not understanding the structure of the underlying game or from not forming the correct beliefs about the strategies of their counterparts. One strand of the literature has tried to explain these deviations as errors in belief formation (e.g., Costa-Gomes and Crawford (2006); Ho et al. (1998)). Yet, some recent research shows that subjects might not fully understand the experimental environment.

In this paper we use an individual decision-making task that allows us to uncouple subjects' understanding of the game from their belief formation, and thus to establish to what extent understanding of the structure of the game contributes to non-equilibrium play in our experiment.

We find that a majority of subjects fail to fully understand the structure of the game. Moreover, subjects who understand the structure of the game play closer to the Nash Equilibrium, are better at best-responding to their own beliefs, and seem to modify their beliefs (correctly) depending on the population they are facing. This result is inconsistent with models of the Level-K type (e.g., Costa-Gomes and Crawford, 2006) which assume that agents fully understand the structure of the game and only play out of equilibrium due to flaws in belief formation. Our findings suggest, otherwise, that out of equilibrium play is not only the result of a limited ability to form correct beliefs, but that it also results from the inability of subjects to fully understand the game's structure.

In light of these results, we believe the 1PG could be a useful "quick and easy" test for researchers interested in, or aiming to control for, understanding of the structure of guessing games. More generally, we believe that the reduction of strategic games into one-player forms could be a useful tool in the analysis of other games too.[21] Such transformation would allow researchers to study the degree of understanding that subjects have of the structure of the game, and to control for any deviations from Nash equilibrium play independent of errors in belief formation.[22]

Finally, a potential extension of this experiment could be to vary the number of selves subjects play in the 1PG and compare play to standard guessing games populated by

[21] For instance in ring-networks (e.g., Kneeland, 2015) or minimum effort games (e.g., Van Huyck et al., 1990).

[22] We would like to thank an anonymous referee for suggesting this wider interpretation of our results.

the same number of strategic players. On the one hand, increasing the number of selves and strategic players increases the complexity of the game, and may therefore make understanding the structure of the game and belief formation more difficult. On the other hand, increasing the number of selves and strategic players may lead subjects to better understand the unraveling mechanic of the game. We leave it to future extensions of this work to test whether the general findings in this paper would also hold under such conditions.

References

Agranov, M., A. Caplin, and C. Tergiman (2015): "Naive play and the process of choice in guessing games," *Journal of the Economic Science Association*, 1, 146–157. Cited on page 3.

Agranov, M., E. Potamites, A. Schotter, and C. Tergiman (2012): "Beliefs and endogenous cognitive levels: An experimental study," *Games and Economic Behavior*, 75, 449–463. Cited on page 11.

Blume, A. and U. Gneezy (2010): "Cognitive forward induction and coordination without common knowledge: An experimental study," *Games and Economic Behavior*, 68, 488–511. Cited on page 4.

Bosch-Rosa, C., T. Meissner, and A. Bosch-Domènech (2018): "Cognitive bubbles," *Experimental Economics*, 21, 132–153. Cited on pages 4 and 8.

Camerer, C. (2003): *Behavioral game theory: Experiments in strategic interaction*, Princeton University Press. Cited on page 2.

Carpenter, P. A., M. A. Just, and P. Shell (1990): "What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices Test." *Psychological review*, 97, 404. Cited on page 8.

Chou, E., M. McConnell, R. Nagel, and C. R. Plott (2009): "The control of game form recognition in experiments: Understanding dominant strategy failures in a simple two person "guessing" game," *Experimental Economics*, 12, 159–179. Cited on page 2.

Costa-Gomes, M. A. and V. P. Crawford (2006): "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study," *American Economic Review*, 96, 1737–1768. Cited on pages 2, 6, 11, and 21.

Costa-Gomes, M. A. and G. Weizsäcker (2008): "Stated Beliefs and Play in Normal-Form Games," *The Review of Economic Studies*, 75, 729–762. Cited on page 2.

Duffy, J. (2016): "Macroeconomics: a survey of laboratory research," in *The handbook of experimental economics*, ed. by J. H. Kagel and A. E. Roth, Princeton university press, vol. 2. Cited on page 6.

Fischbacher, U. (2007): "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental economics*, 10, 171–178. Cited on page 8.

Fragiadakis, D. E., D. T. Knoepfle, and M. Niederle (2016): "Who is strategic?" *Working Paper*. Cited on page 2.

Georganas, S., P. J. Healy, and R. A. Weber (2015): "On the persistence of strategic sophistication," *Journal of Economic Theory*, 159, 369–400. Cited on page 7.

Gill, D. and V. Prowse (2016): "Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis," *Journal of Political Economy*, 124, 1619–1676. Cited on page 7.

Gneezy, U., A. Rustichini, and A. Vostroknutov (2010): "Experience and insight in the race game," *Journal of Economic Behavior & Organization*, 75, 144–155. Cited on page 7.

Greiner, B. (2015): "Subject pool recruitment procedures: organizing experiments with ORSEE," *Journal of the Economic Science Association*, 1, 114–125. Cited on page 8.

Grosskopf, B. and R. Nagel (2008): "The two-person beauty contest," *Games and Economic Behavior*, 62, 93–99. Cited on pages 1, 3, 4, and 5.

Güth, W., M. Kocher, and M. Sutter (2002): "Experimental "beauty contests" with homogeneous and heterogeneous players and with interior and boundary equilibria," *Economics Letters*, 74, 219–228. Cited on page 5.

Ho, T.-H., C. Camerer, and K. Weigelt (1998): "Iterated dominance and iterated best response in experimental" p-beauty contests"," *The American Economic Review*, 88, 947–969. Cited on pages 2, 11, and 21.

Hommes, C., J. Sonnemans, J. Tuinstra, and H. Van de Velden (2004): "Coordination of expectations in asset pricing experiments," *The Review of Financial Studies*, 18, 955–980. Cited on page 6.

KNEELAND, T. (2015): "Identifying Higher-Order Rationality," *Econometrica*, 83, 2065–2079. Cited on page 21.

LAHAV, Y. (2015): "Eliciting beliefs in beauty contest experiments," *Economics Letters*, 137, 45–49. Cited on page 6.

LEVITT, S. D., J. A. LIST, AND E. SALLY (2011): "Checkmate: Exploring backward induction among chess players," *The American Economic Review*, 101, 975–990. Cited on page 7.

MAUERSBERGER, F. AND R. NAGEL (2018): "Levels of reasoning in Keynesian Beauty Contests: a generative framework," in *Handbook of Computational Economics*, Elsevier, vol. 4, 541–634. Cited on page 6.

NAGEL, R. (1995): "Unraveling in guessing games: An experimental study," *The American Economic Review*, 85, 1313–1326. Cited on page 2.

NAGEL, R., C. BÜHREN, AND B. FRANK (2016): "Inspired and inspiring: Hervé Moulin and the discovery of the beauty contest game," *Mathematical Social Sciences*. Cited on pages 5 and 6.

PETERSEN, L. AND A. WINN (2014): "Does money illusion matter? Comment," *American Economic Review*, 104, 1047–62. Cited on page 4.

RAVEN, J. C. (1960): *Guide to the standard progressive matrices: sets A, B, C, D and E*, HK Lewis. Cited on page 7.

REY-BIEL, P. (2009): "Equilibrium play and best response to (stated) beliefs in normal form games," *Games and Economic Behavior*, 65, 572–585. Cited on page 2.

RICK, S. AND R. A. WEBER (2010): "Meaningful learning and transfer of learning in games played repeatedly without feedback," *Games and Economic Behavior*, 68, 716–730. Cited on page 11.

SCHOTTER, A. AND I. TREVINO (2014): "Belief elicitation in the laboratory," *Annu. Rev. Econ.*, 6, 103–128. Cited on page 7.

Van Huyck, J. B., R. C. Battalio, and R. O. Beil (1990): "Tacit coordination games, strategic uncertainty, and coordination failure," *The American Economic Review*, 80, 234–248. Cited on page 21.

Weber, R. A. (2003): "Learning with no feedback in a competitive guessing game," *Games and Economic Behavior*, 44, 134–144. Cited on page 11.

Weizsäcker, G. (2003): "Ignoring the rationality of others: evidence from experimental normal-form games," *Games and Economic Behavior*, 44, 145–171. Cited on page 2.

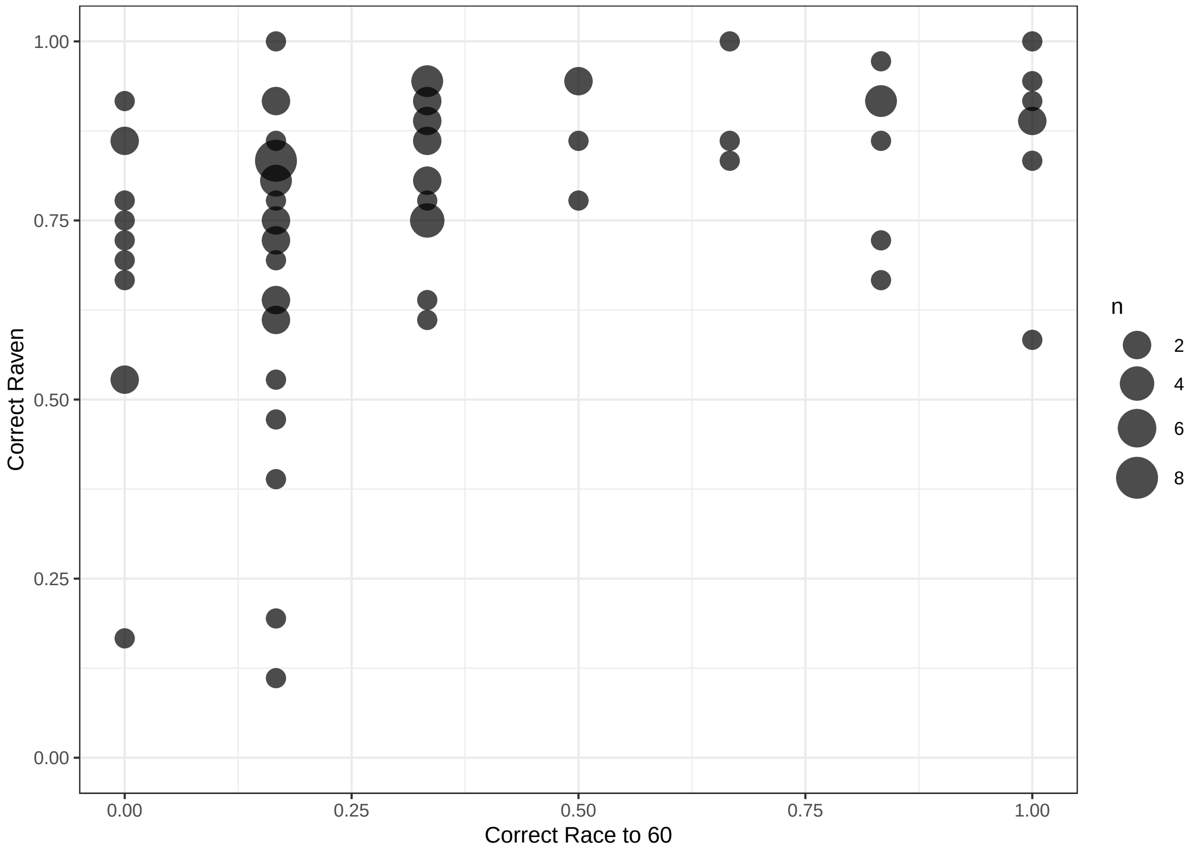

Figure 10: Normalized results for Race to 60 (horizontal axis) and Raven Test (vertical axis). The dashed lines mark the median result for each test.

A Cognitive Ability

As described in section A all subjects were asked to solve thirty-six matrices from the Raven Test, and took part in six rounds of Race-to-60. The normalized results of these two measures are plotted in Figure 10. Most subjects performed well at the Raven test (median of twenty nine solved matrices), but not in the Race-to-60 game (median of one won round). Interestingly, the lower right quadrant of the scatter plot is completely empty. This implies that while a good score in the Race-to-60 game requires a good score in the Raven Test, the opposite is not true. The correlation between both measures of cognitive ability is positive and significant, as well as with the payments for the 1PG and 2PG (see Table 2).

In the first column of Table 3 we report a linear probability model where the binary dependent variable is having solved the 1PG. In the second column, we report a linear

	Raven	Race	Π^{1PG}	$\bar{\Pi}_i^{2PG}$	Δz_i^*	ΔB_i	Solved 1PG
Raven	1						
Race	0.454***	1					
Π^{1PG}	0.433**	0.462***	1				
$\bar{\Pi}_i^{2PG}$	0.299***	0.203*	0.454***	1			
Δz_i^*	-0.322***	-0.307***	-0.531***	-0.476***	1		
ΔB_i	0.254	-0.042	0.262	0.208	-0.132	1	
Solved 1PG	0.451***	0.415***	0.818***	0.335***	-0.366***	0.065	1

Table 2: Spearman correlation between cognitive ability measures, payoffs, and beliefs.

regression with the payment in the 1PG as the dependent variable.[23] In both cases, the independent variables are the (normalized) cognitive ability test scores. The results show an interesting asymmetry; subjects that fully solve the 1PG scored high both at the the Race-to-60 and the Raven test, while only backward induction seems to have an effect on the payoff for the 1PG.[24]

Result: *Both a high score at the Raven test and Race-to-60 game are important to fully solve the one-player guessing game, while only Race-to-60 seems to matter for the payoff in the one-player guessing game.*

Additionally, in column 3 $\bar{\Pi}_i^{2PG}$ is regressed on both measures of cognitive ability, and only Race-to-60 is marginally important to score high in the 2PG. In column 4 Δz_i^* is the dependent variable, and this time the coefficient for Race-to-60 is highly significant and negative. This means that backward induction abilities are central for a subject to be able to best respond to her own beliefs.

[23] A Tobit model for the 1PG payoff gives the same results.

[24] The Variance Inflation Factor discards any severe case of multicollinearity between both regressors.

	(1)	(2)	(3)	(4)
	Correct 1PG	Π_i^{1PG}	$\bar{\Pi}_i^{2PG}$	Δz_i^*
Raven	0.706**	0.390	1.217	-9.371
	(0.279)	(0.326)	(0.756)	(8.084)
Race	0.505***	0.548***	0.814*	-9.594**
	(0.163)	(0.190)	(0.440)	(4.708)
Constant	-0.409*	1.176***	-0.172	20.90***
	(0.208)	(0.243)	(0.563)	(6.014)
N	80	80	80	80
adj. R^2	0.226	0.133	0.086	0.074

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Linear regression of cognitive ability measures on Correct 1PG, Π_i^{1PG}, $\bar{\Pi}_i^{2PG}$, Δz_i^* and ΔB_i.

B Normalization

The way in which we normalize the distribution for each subject (i) is by adding for each bin (b) all the 2PG choices made across all sessions by all subjects (z_j), except her own choice (z_i) . This results in a distribution of 79 choices spread across all possible 20 bins 0-4, 5-9,..., 95-100. We then normalize this distribution by multiplying the resulting number in each bin by 19/79. Therefore, each bin (b) of the distribution for subject i is: $Bin_{b,i} = \sum_{j \neq i}^{80} \mathbb{1}_{z_j \in b} \frac{19}{79}$, where $\mathbb{1}$ is an indicator function that takes a value of one whenever the subscript is true, zero otherwise.

C Extra Graphs and Figures

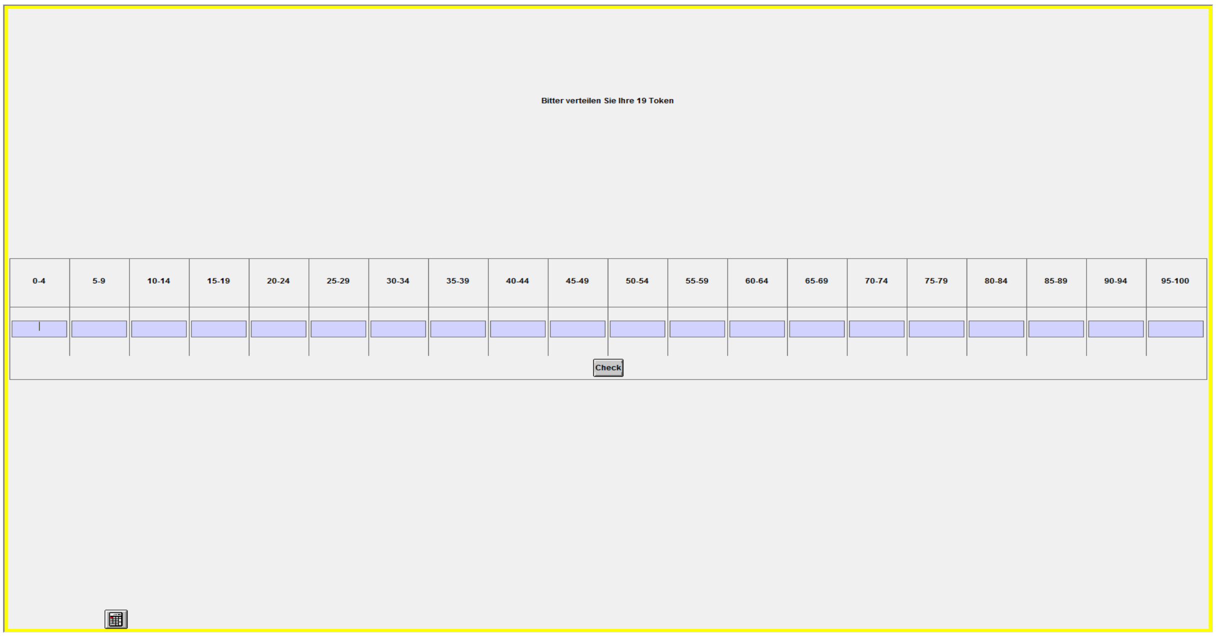

Figure 11: Distribution Screen. Subjects were provided with N-1 tokens to distribute across the 20 bins provided in the screen. The check button counted for them the amount of tokens they had deposited at any moment. Subjects could use less than 19 tokens, but never more.

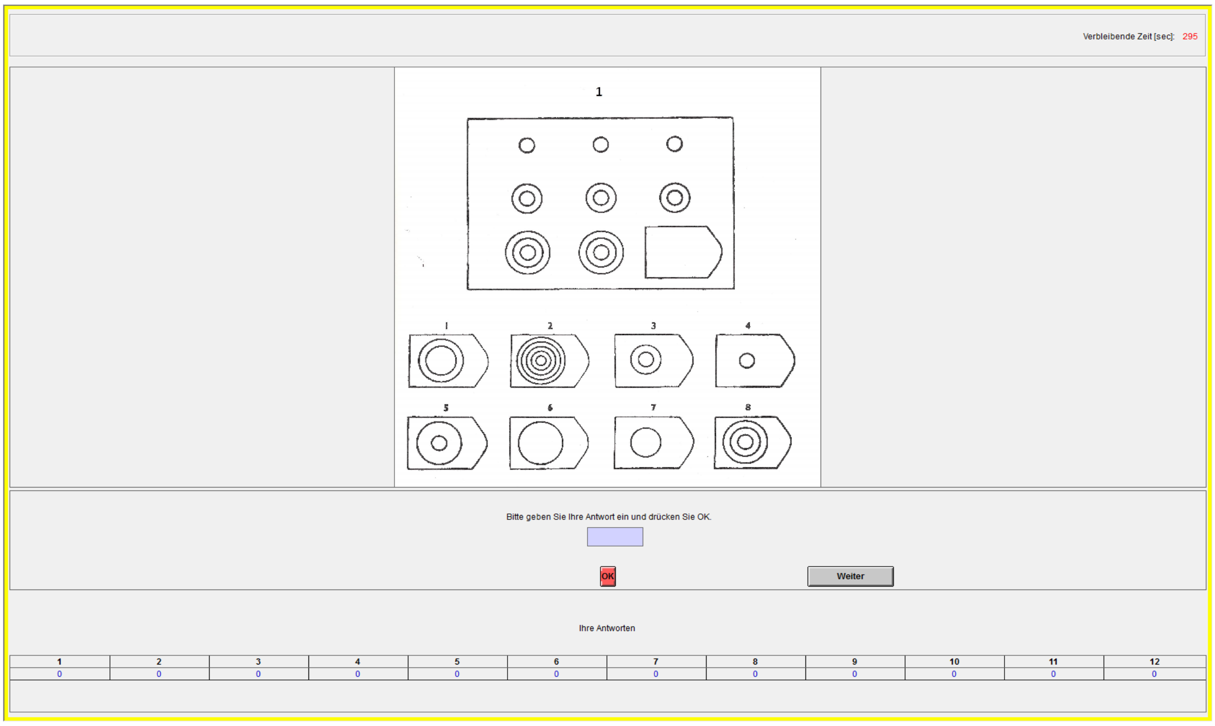

Figure 12: Raven Test Screen

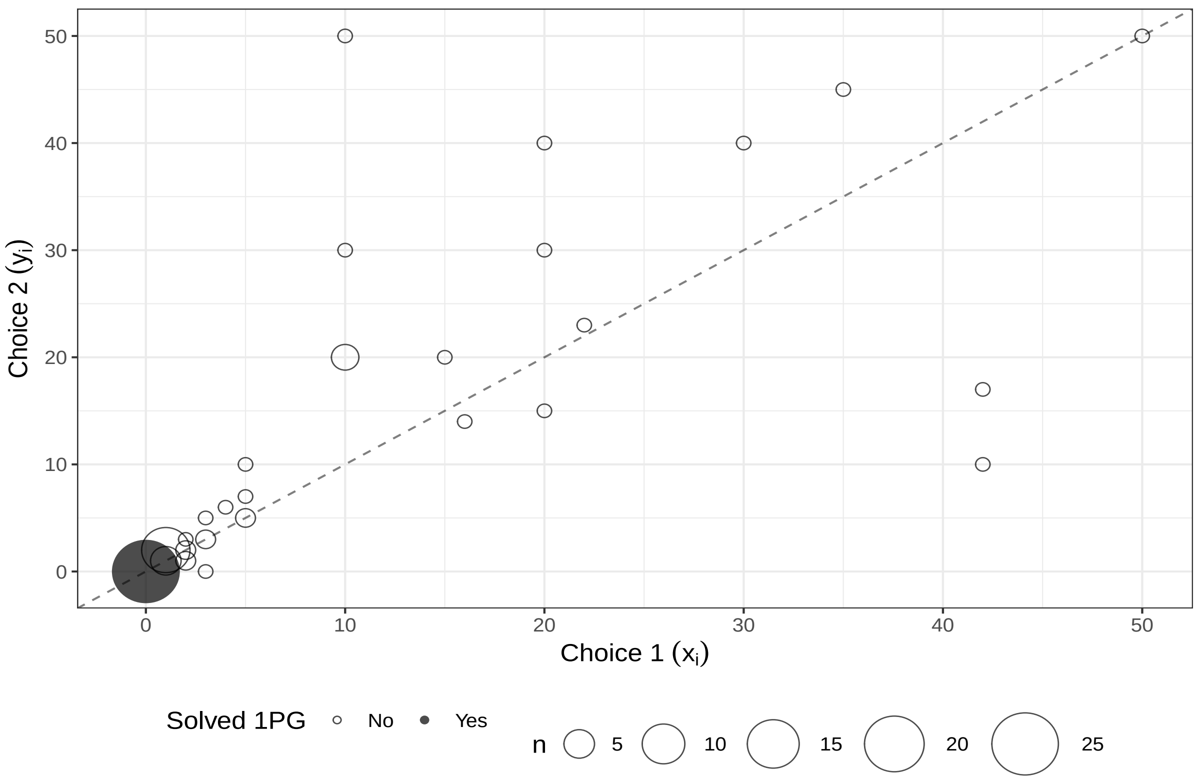

Figure 13: Zoom-in depicting only choices [0, 50] of Figure 1. In black the choices made by those that fully solved the 1PG.

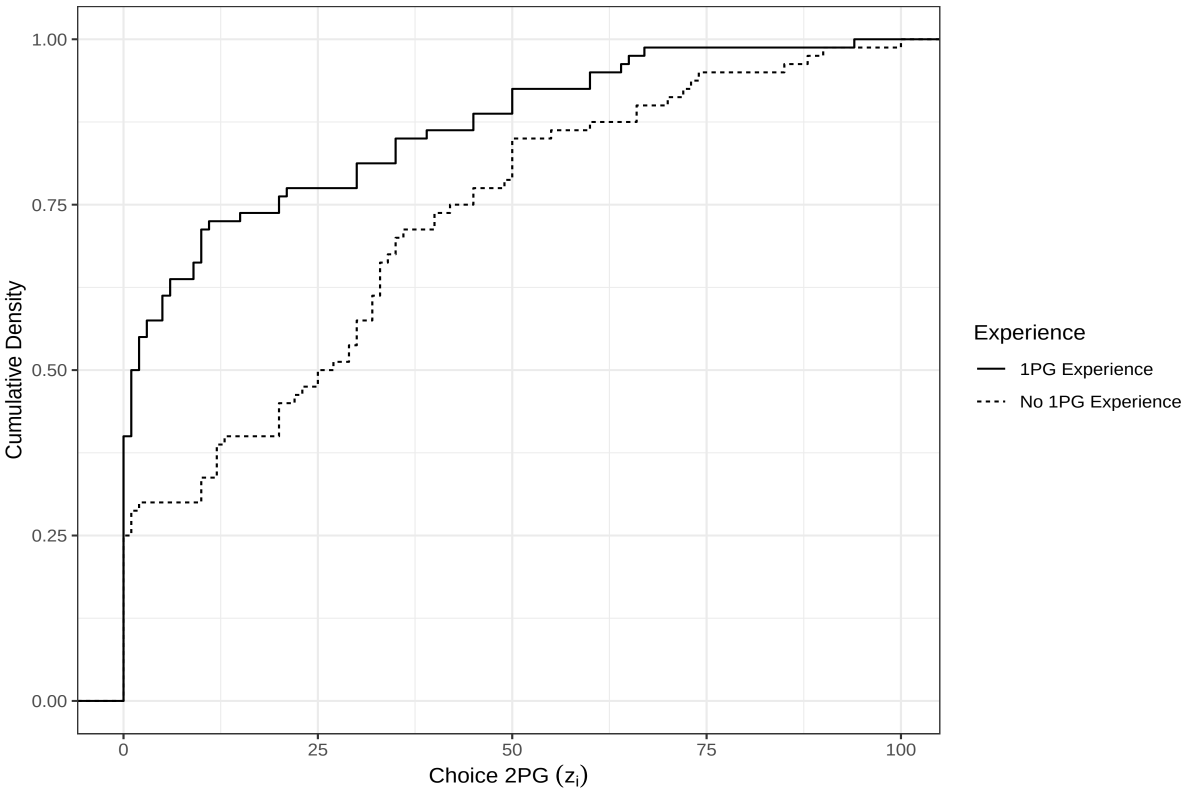

Figure 14: Cumulative density plot of 2PG play, with and without prior experience in the 1PG.

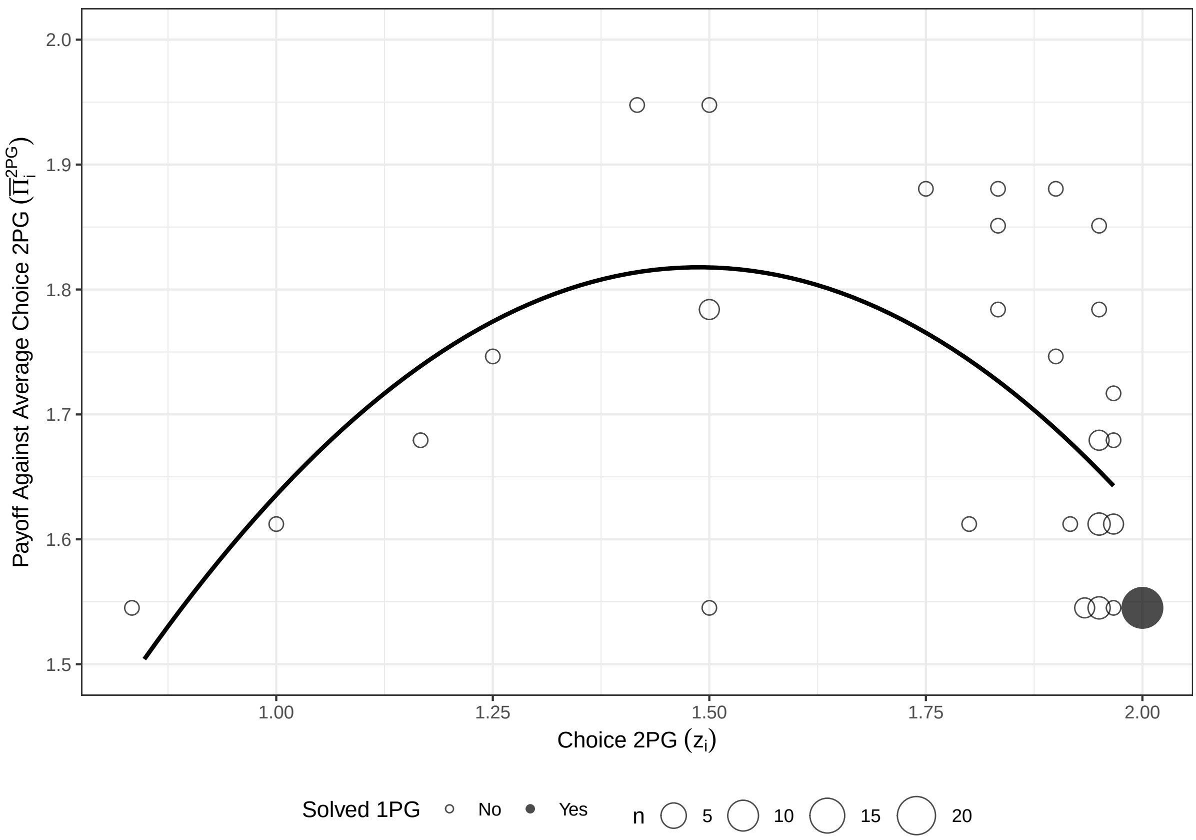

Figure 15: Zoom-in depicting payoffs [1.5, 2] of Figure 4. In black the choices made by those that fully solved the 1PG.

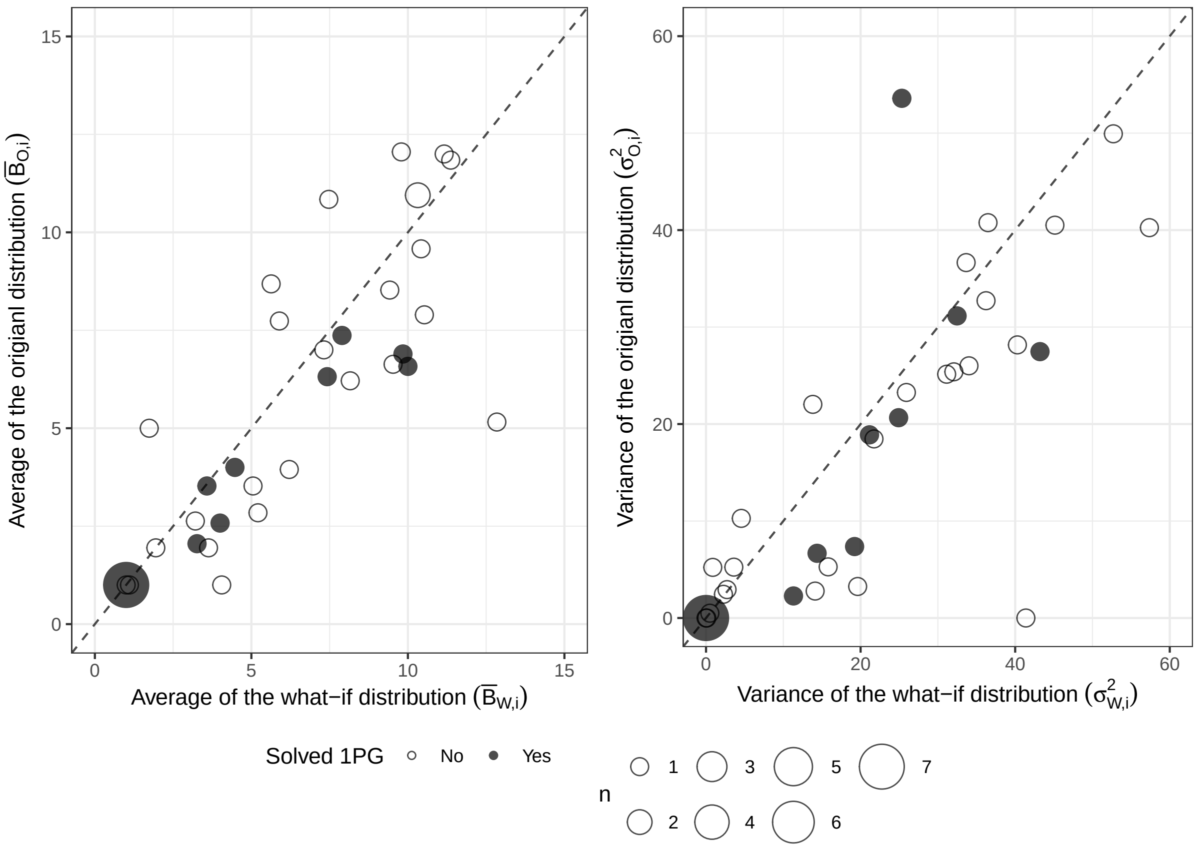

Figure 16: Figure depicting the change in distribution across the First and Second belief elicitation tasks. In the left(right) panel on the vertical axis the average value(variance) of the First distribution for subject i, in the horizontal axis the average value(variance) for the Second distribution for subject i. In both cases the dotted line is a 45 degree line; any value to the right of it represents an increase in the parameter in the Second distribution with respect to the First distribution.

D Instructions

The instructions below are translated from the original German instructions. The instructions were distributed sequentially, each of the following subsections at a time. After reading instructions, subjects completed the tasks, and then received instructions for the following task. Subjects were given time to carefully read the instructions and ask questions.

D.1 Overview

Today's experiment consists of a sequence of questionnaires and games. After each game or questionnaire, you receive instructions for the next game or questionnaire. In total, the experiment will last about 1.5 hours. For your participation your will receive a show-up fee of 5€. Depending on how you answer/play the questionnaires/games you can earn money on top of that. After you have finished all questionnaires and games, your final payoff will be shown on your screen. You will then receive a receipt, which you please fill out with the shown payoff and your name and address. After this you will be called to the adjoining room to receive your payment.

You will now receive the specific instructions for the first game.

D.2 Task 1

In this task, you have to pick two numbers (named X and Y) between 0 and 100 (both inclusive). Your payment for this task depends on how close these two numbers (X and Y) are to the so called "target number". The closer your numbers are to the target number, the higher your payment.

Caclulation of the target number The target number depends on your picked numbers. It is calculated as the mean of both numbers (X and Y), multiplied by two thirds:

$$\text{target number} = \frac{2}{3}\left(\frac{X+Y}{2}\right)$$

Your payment Your Payment depends on absolute distance of your numbers to the target number. For each number, if the number is exactly equal to the target number, you receive 1€. Should you not hit the target number exactly, some money will be deducted.

For each absolute unit distance of your numbers to the target number, 0.05€ will be deducted:

$$\text{payment for X} = 1€ - 0.05€\,|X - \text{target number}|$$

$$\text{payment for Y} = 1€ - 0.05€\,|Y - \text{target number}|$$

Because only the absolute difference between your number and the target number is used to calculate your payment (indicated with by the vertical bars "|"), it does not matter whether you deviate upwards or downwards - only the absolute difference counts. Please note that your payment can not become negative. Should your payment according to the above formula become negative, you will receive 0€ instead.

You have four minutes time to pick both numbers. After you have chosen your numbers, please confirm your choice by clicking the red "continue" button. Otherwise, your answer will not be recorded.

D.3 Task 2

This task is similar to the previous task. This time, however, you will play against another person in this room, who will be matched with you randomly. This time, you choose only one number between 0 and 100 (both inclusive). The other person also chooses a number between 0 and 100. Your payment now depends on how close your number is to the so called "target number". The closer your number is to the target number, the higher your payment.

Caclulation of the target number The target number depends your number, and the number stated by the other person. It is calculated as the mean of both numbers, multiplied by two thirds:

$$\text{target number} = \frac{2}{3}\left(\frac{\text{your number} + \text{other player's number}}{2}\right)$$

Your payment Your Payment depends on absolute distance of your numbers to the target number. If your number is exactly equal to the target number, you receive 2€.

Should you not hit the target number exactly, some money will be deducted. For each absolute unit distance of your number to the target number, 0.1€ will be deducted:

$$\text{payment} = 2€ - 1€\,|\text{your number} - \text{target number}|$$

Because only the absolute difference between your number and the target number is used to calculate your payment (indicated with by the vertical bars "|"), it does not matter whether you deviate upwards or downwards - only the absolute difference counts. Please note that your payment can not become negative. Should your payment according to the above formula become negative, you will receive 0€ instead.

Since the payoff of the other player is calculated in the same way, s/he also has the incentive to state the closest number possible to the target number.

You have four minutes time to pick your number. After you have chosen your number, please confirm your choice by clicking the red "continue" button. Otherwise, your answer will not be recorded.

D.4 Task 3

Reminder: in the previous task every participant was asked to pick a number between 0 and 100 (both inclusive). The payment for this game depended on how close this number was from the target number (mean of the numbers of both players, multiplied by two thirds).

In this task we would like to ask you to estimate how the other 19 participants who are participating in this experiment, behaved in the previous game. For this purpose, you receive 19 "token", that you can distribute in a number of bins (see screenshot below [Figure 11]).

Each token represents your estimate for one participant. The bins represent an interval for your estimation. To distribute tokens into bins, you type the number of participants, who you estimate to play a number in the interval of the respective bin, into the respective field on your screen. If you believe that no participant played a number in any interval, you can leave the field empty (or type in 0).

Example: You believe that 5 participants chose numbers between 75 and 79, and 14 participants chose numbers between 20 and 24. In this case you type "14" in the field 20-24 and "5" in the field 75-79.

For each token that coincides with a decision of a participant, you receive 0.20€.

Example: You have put 5 tokens in a bin, and 2 participants have actually played numbers in this bin. In this case, you receive 2*0.20€. If 5 people have actually played numbers in this bin, you receive 5*0.20€. If you estimate all 19 participants correctly, you receive 19*0.2€=3.80€.

Note: because it is difficult to keep track of how many tokens you have already distributed, we have built in a "check-button", that sums up all distributed tokens. Should not all 19 tokens be distributed, or more 19 tokens distributed, a pop up will tell you this. If you are happy with your distribution, and have distributed 19 token, you can finish this task by clicking "Ok".

D.5 Task 4

[Only a subset of 40 subjects participated in this task]

A few weeks ago, we conducted an experiment with 20 participants, who also solved Task 2. The participants received the exact same instructions as you, however, they have not solved Task 1 before they solved Task 2. That means, they played directly against another player, without picking two numbers before in Task 1.

We would like to ask you, similarly to the previous task, to estimate how these participants have played in Task 2. Again you receive 19 token, that you can distribute in a number of bins.

Since 20 participants took part in this experiment, but you have 19 tokens, we will randomly choose 19 participants from these 20 participants.

Again, for each correctly placed token, you receive 0.20€.

D.6 Task 5

In this task, you see a puzzle on every screen: a matrix containing 8 graphical entries and one empty field. There are 8 given option to complete the puzzle - only one is correct. For each puzzle, you have to select one of the 8 options. To this, you type in the corresponding number, and confirm by clicking OK.

[screenshot of Figure 12]

In total you will complete 36 such puzzles, divided into 3 blocks of 12 puzzles. Within each block you can move back and forth between puzzles, and correct previously stated answers. You have 5 minutes to solve each of the first two blocks, and 8 minutes for the remaining block. The remaining time will be displayed in the header of your screen. After you have completed all puzzles, or the time runs out, you have to confirm your answers by clicking "send answers". If you do not confirm your choices, they will not be recorded.

For each correctly answered puzzle, you earn 0.10€.

D.7 Task 6

This task consist of the game "Race to 60", that you play repeatedly against the computer. Your task is to win this game as often as possible against the computer.

In this game, you and the computer player state numbers between 1 and 10 sequentially. The numbers are added up, and whoever is first to push the sum of all stated numbers to or above 60 wins the game.

The game functions as follows: You start the game by stating a number between 1 and 10 (both inclusive). Then the game continues as follows:

1. The computer chooses a number between 1 and 10. This number wil be added to your number.
2. The screen shows the sum of all numbers chosen so far. If the sum is smaller than 60, you again choose a number between 1 and 10, which again will be added to all previously stated numbers.

This will be repeated until the sum of all numbers is greater or equal to 60. Whoever states the number that pushes the sum to or above 60, wins the game. You will play this game 6 times against the computer, and you have 90 seconds time for each repetition. For each game won, you receive 0.50€.

Your payment This is the last game for today's experiment. After completing this game, you will receive a questionnaire that you please complete. After that, your payments will be summarized on the screen. You then receive a receipt, that you please fill out. After you have done that, an experimenter will come to your place to check if everything is in order. After that you can go to the adjoining room to receive your payment.