E C O N O M I C S B U L L E T I N

Comparing Guessing Games with homogeneous and heterogeneous players: Experimental results and a CH explanation

Eugen Kovac
University of Bonn and CERGE-EI

Andreas Ortmann
CERGE-EI

Martin Vojtek
CERGE-EI and Czech National Bank

Abstract

We investigate the decisions of individuals in simple and complex environments. We use a version of the Guessing Game (Beauty-contest Game) as a vehicle for our investigation, employing mathematically talented students. We find that our subjects think in complex environments more carefully before making decisions. We rationalize our findings using the Cognitive Hierarchy (CH) model proposed by Camerer, Ho, and Chong (2002). We relate our results to the emerging literature on the decision making of collective actors.

The authors would like to thank Colin Camerer, Dirk Engelmann, Werner Gueth, Martin Kocher, Ondrej Rydval, Matthias Sutter, and anonymous referees for helpful comments, Peter Kosinar for his contributions in the early stages of this project, Alexander Erdelyi and Peter Novotny for allowing us to run the experiments in their math summer schools, and Matthias Sutter for providing us the data from their experiment. We also appreciate the help of Frantisek Brazdik with this project. We thank Bank Austria for providing the funds. All errors are ours. CERGE-EI is a joint workplace of the Center for Economic Research and Graduate Education at Charles University and the Economics Institute of the Czech Academy of Sciences, both in Prague.
Citation: Kovac, Eugen, Andreas Ortmann, and Martin Vojtek, (2008) "Comparing Guessing Games with homogeneous and heterogeneous players: Experimental results and a CH explanation." *Economics Bulletin,* Vol. 3, No. 9 pp. 1-9
Submitted: October 30, 2007. **Accepted:** February 12, 2008.
URL: http://economicsbulletin.vanderbilt.edu/2008/volume3/EB-07C90016A.pdf

> "Human rational behavior is shaped by a scissors whose two blades are the structure of task environments and the computational capabilities of the actor." (Simon 1990, p. 7)

1. Introduction

We are interested in the interaction of the two blades of Simon's scissors. Specifically, we would like to understand better whether a (relatively) complex environment triggers deeper thinking. Our working hypothesis — also proposed by Güth, Kocher, and Sutter (2002) and implicitly by Camerer, Ho, and Chong (2004) — suggests it does. Experimentally, the importance of the complexity of the environment for individuals' behavior has been demonstrated, for example, by Wilcox (1993, 2006), Palfrey and Prisbrey (1997), Camerer, Ho, and Chong (2004), Ballinger, Hudson, Karkoviata, and Wilcox (2005), and Slonim (2005).

The Guessing Game is our vehicle of choice for an experimental test of our working hypothesis. In this game (also called Beauty-contest Game) participants are asked to choose a number from a closed interval. The winner is the person who picks the number closest to a given proportion of the average of all chosen numbers.[1] The simplicity and flexibility of this game, and the fact that it captures nicely the strategic interaction of actors in financial markets, have made it a frequent topic of experimental studies of depth of reasoning (Camerer 2003).

Güth, Kocher, and Sutter (2002), henceforth GKS, investigated experimentally an interesting modification of the standard Guessing Game (see also Bosch-Domènech, Montalvo, Nagel, and Satorra 2002). They use a continuous payment scheme and allow individuals — students from the Humboldt-University Berlin — to make decisions in "homogeneous" (simple) and "heterogeneous" (complex) environments.

In contrast to GKS, we ran the experiment with mathematically talented students who were skilled in solving abstract problems. Our results conform to GKS's and our intuition (but contradict their results) that more complicated environments, here constructed by way of more heterogeneous players, trigger deeper thinking. In particular, the guesses of subjects in the heterogeneous environment are closer to the equilibrium than the guesses of subjects in the homogeneous environment. We rationalize our findings using an extension of the Cognitive Hierarchy (CH) model by Camerer, Ho, and Chong (2004), which we believe to be a contribution in its own right. Our estimates of the model suggest indeed that more complexity leads to a higher number of steps of thinking.

The source of the difference in results is unclear; we argue below that they are unlikely to stem from differences in implementation. In our view, the prime suspects are the different subject pools and the differences in mathematical abilities.

The structure of the paper is as follows. In the next section we describe our Guessing Game treatments and the working hypothesis. In Section 3 we discuss the implementation and design of our experiment. In Section 4 we describe the

[1] See, for example, Nagel (1995).

experimental results that we use in Section 5 to rationalize our findings. Section 6 concludes and relates our work to an important emerging literature that Camerer (2003) has identified as one of his top ten research questions.

2. The game and our working hypothesis

Let n (where $n > 2$) be the number of players participating in the game. A pure strategy (action) of player $i \in \{1, \ldots, n\}$ is a real number $s_i \in S_i = [0, 100]$, also called player i's choice or guess. Given a strategy profile $s = (s_1, \ldots, s_n)$ let $\overline{s} = \frac{1}{n}\sum_{i=1}^{n} s_i$ denote the average of numbers chosen by all players. The payoff function of player i is given by

$$u_i(s) = C - c|s_i - q_i\overline{s}|,$$

where $q_i \in (0, 1)$ is the proportion of $\overline{s}$ determining player i's target number, and C and c are positive constants. If player i guesses exactly $q_i\overline{s}$ (the target number), he receives the payoff C. If he does not, his payoff is reduced by c for each unit of the distance between s_i and $q_i\overline{s}$.

The above payoff function represents a continuous payment scheme, which differs from the standard version of the Guessing Game, used, for instance, by Nagel (1995) or in the experiments reported by Bosch-Domènech et al. (2002), where only the player who is closest to $q_i\overline{s}$ wins a fixed payoff. The continuous payment scheme seems most appropriate for implementing heterogenous environments.

Following the advice of Davis and Holt (1993) to facilitate comparison with previous experiments, we use both the payoff function, and in particular, the boundary equilibrium parametrization from GKS.[2] Specifically, we use the following parametrization:

$$n = 4, \qquad C = 50\,\text{SKK}, \qquad c = 1\,\text{SKK}.$$

Clearly, a player could potentially make a loss. In order to avoid the possibility of losses, we truncated payoffs at zero and informed the subjects accordingly. Thus, the actual payoff function was $\tilde{u}_i(s) = \max\{0,\ 50 - |s_i - q_i\overline{s}|\}$.

We study two treatments that we call HOM and HET. In the HOM treatment, all four players are given $q_i = 1/2$ (for $i = 1, 2, 3, 4$). We will call the players in that treatment *homogeneous*. In the HET treatment, two of the players are given $q_i = 1/3$ (for $i = 1, 2$), while two of the players are given $q_i = 2/3$ (for $i = 3, 4$). We will call the players in that treatment *heterogeneous*. It is easy to prove that iterated elimination of strictly dominated strategies yields a unique equilibrium $s_i^* = 0$ in both treatments (see GKS for details).

GKS conjectured that heterogeneous players think harder about other players' behavior and hence their guesses are closer to the equilibrium. We find this an intuitive and persuasive conjecture that was, however, not confirmed in their earlier experiment.

[2]More precisely, we set $d = 0$ in the notation of GKS.

3. Design and implementation

Our experimental sessions were run in two summer schools for young mathematically talented students. In one summer school we ran the HOM treatment with seven groups of four homogeneous players; in the other summer school we ran the HET treatment with seven groups of four heterogeneous players. While, ideally, we would have liked to run both treatments in each summer school, it was not possible. Nearly all participants were students of secondary schools (aged 14–18). None of them had taken a course on game theory before and it was their first experience with experimental economics. The participants of these summer schools were chosen in two independent national correspondence competitions in mathematics.[3] Both competitions were of comparable difficulty, quality, and reputation. Indeed almost identical numbers (84 and 86) of participants took part in the competitions. An overlap group of 28 participants took part in both competitions and 7 ended up participating in the first summer school and 10 in the second summer school. Since the two summer schools took place simultaneously, no student participated in both of them.

There is no obvious reason why there should be a selection bias in favor of one of the summer schools. For example, our subject pool included 12 (out of 37) students who reached the national round of the Mathematical Olympiad, equally distributed across the two summer schools. In addition, the choice behavior in a preceding Guessing Game redux was very similar.[4] The Guessing Game redux with which we kicked off our treatments in the two summer schools is arguably the simplest form of the Guessing Game: players in groups of two or three choose only between numbers 0 and 1, or among 0, 1, and 2.[5] After this experiment we paid all participants according to their choices without revealing any other information. Then we proceeded with the actual Guessing Game treatments. In every summer school, the participants were read the instructions (see Appendix B) at the beginning of the experiment. The instructions specified that in each of five rounds the subjects would be matched randomly so as to yield seven four-player groups. The groups were rematched in each round. In the HET treatment, the subjects were told that in each round there would be two subjects of each type in each group.

There was no time limit for making the choice. After each round we first collected the record sheets and calculated averages (using calculators) in each group. Then we provided each participant the average in his own group and announced publicly the averages in all groups. After the experiment, for each round, two out of seven groups were randomly selected and earnings from that round were paid out to every selected participant. This payment mode had been announced as a part of the instructions.

[3] The competitions were organized by the students of the Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava, Slovakia. E. Kováč and M. Vojtek were involved in the organization of these summer schools in previous years. The problems solved in the competitions were at a level similar to regional rounds of the Mathematical Olympiad (see, for example, `en.wikipedia.org/wiki/Mathematical_olympiad`).

[4] In this game, 15 out of 28 subjects in one summer school and 18 out of 28 subjects in the other summer school chose the equilibrium strategy. This difference is not significant.

[5] The results of this experiment will be reported in a different paper.

The maximum amount that participants could earn in each round was 50 SKK.[6] For every unit of difference from the target number they lost 1 SKK. However, taking into account the age of the participants, we excluded the possibility of negative payoffs. Participants were informed in the instructions that the minimal payoff in each round was 0 SKK.

Our design and implementation differ from GKS in five aspects. First, our subject pool did not consist of undergraduate students of economics, but mathematically talented students of secondary schools skilled in solving abstract problems. Second, the averages in all groups became public information after each round, whereas GKS announced only the average of his own group to each participant after each round.[7] Third, our experiment was not computerized. Fourth, we conducted a simple Guessing Game redux before the actual experiment reported here. Fifth, our payoff function slightly differs from GKS. In particular, we did not allow for losses,[8] and we used the parametrization with proportion C/c equal to $50\,\text{SKK}/1\,\text{SKK} = 50$, as opposed to $2\,\text{DM}/0.05\,\text{DM} = 40$. We discuss now in detail the possible effects of these differences in design and implementation.

4. Results of the experiment

Table 1 and Figures 1 and 2 in Appendix C summarize the results of the experiment. In the HET treatment we split the asymmetric treatment data into two clusters, with $q = 1/3$ and $q = 2/3$. Figure 1 shows the distribution of first-round guesses and Figure 2 shows the pattern of average guesses across rounds. Table 1 shows the means, standard deviations, medians, and the minimum and the maximum of guesses, separately for each round and each treatment. In addition, Table 1 contains the average payoffs,[9] and the absolute and relative decreases in the average guess.

Result 1. In both treatments, the guesses of players "converge" to the equilibrium.

The average guess decreases substantially from round to round (by more than 25%). In addition, the standard deviation (with one exception),[10] as well as the median (again, with one exception) decrease and the average payoffs increases continually with the number of rounds. We conclude that the guesses converge to the Nash equilibrium; see also Figure 2 for illustration.

Result 2. Heterogeneity induces more equilibrium-like behavior.

[6] According to the official exchange rate 50 SKK was about 1.13 EUR at the time when we conducted the experiment. For that amount, it was possible to buy 2–3 beers or 3 loaves of bread. Given the age of the participants, the payoffs were not insignificant.

[7] Personal communication M. Sutter.

[8] Among all guesses, only 2 in the first round of the HOM treatment and 2 in the first round of the HET treatment (both with $q = 1/3$) would have lead to negative payoffs.

[9] These are potential payoffs. As specified in Section 3 in each round we selected randomly two groups and paid them their earnings. We paid off 1802 SKK in the HET treatment and 1769 SKK in the HOM treatment.

[10] This exception is the last round in the HOM treatment and is caused by an outlier, who guessed 8, 2, 0.9, and 0.9 in rounds 1–4, but increased his guess to 24 in the last round.

Heterogeneous agents guessed on average closer to the equilibrium in each round[11] (Figure 2). The only exception was the second round, where the guesses of players with $q = 2/3$ were slightly higher, on average, than those of homogeneous players. As expected, session averages for players with $q = 1/3$ were lower than averages for players with $q = 2/3$. Our results, thus, conform to our working hypothesis and reverse the findings of GKS who found that session averages with homogenous groups were significantly lower than average guesses in heterogeneous groups.

In addition, we may use the relative decrease as a proxy for the speed of convergence. Note that for both treatments, the successive elimination of strictly dominated strategies yields a relative decrease of 1/2 (see GKS for details). Table 1 shows that the relative decrease of heterogeneous players with $q = 1/3$ is in each round higher than the relative decrease of the homogeneous players. The relative decrease of heterogeneous players with $q = 2/3$ is almost the same (although slightly higher; with the exception of the change from second to third round, where it is substantially lower) as the relative decrease of the homogeneous players.

In addition to analyzing our own results, we compare them to those of GKS. Table 2 in Appendix C shows summary statistics (namely, average guesses, standard deviations, and medians) of their data.[12]

Result 3. The behavior of our subjects in the HOM treatment is similar to the behavior of GKS's subjects, whereas our subjects' guesses in the HET treatment are closer to the equilibrium than the guesses of GKS's subjects.

We see that both means and standard deviations of the guesses in the HOM treatments follow similar patterns, with our subjects' average guesses being lower in the first, fourth, and fifth round, but higher in the second and third round.[13] In the HET treatment, however, the average guesses of our subjects are well below the average guesses of the GKS subjects.[14] This reverses the key qualitative result of GKS. There is also an interesting quantitative result, namely that the average guesses in our experiment are similar to those in GKS's experiment in the HOM treatment, but are lower in the HET treatment. The first-round guesses deserve special attention, since we use them in the next section to estimate the CH-model. These guesses reflect subjects' initial understanding of the game and strategic thinking. In addition, they represent independent observations. Comparing our data to GKS, we obtain that our

[11] We used the Mann-Whitney test (1-sided test with the null hypothesis that the distributions of guesses in HOM and HET treatments are the same against the alternative that they differ with lower guesses in the HET treatment) for each round. Starting from the second round, the distributions were significantly different and heterogeneous players' guesses were closer to the equilibrium (with $p < 0.1$ in round 2, $p < 0.05$ in round 3, and $p < 0.01$ in rounds 4 and 5). The detailed statistics are available at `www.uni-bonn.de/~kovac/papers/xp`.

[12] We thank the authors (M. Sutter) for providing the data.

[13] According to the Mann-Whitney test (both 1-sided and 2-sided), the distributions were not significantly different.

[14] According to the Mann-Whitney test (1-sided), the distributions differ significantly ($p = 0.11$ in round 1, $p < 0.1$ in round 2, $p < 0.005$ in round 3, and $p < 0.001$ in rounds 4–5 for subjects with $q = 1/3$; $p < 0.01$ in round 1 and $p < 0.001$ in rounds 2–5 for subjects with $q = 2/3$).

subjects' average guesses are closer to the equilibrium in both treatments and the difference is higher in the HET treatment.

In order to explain the reversal in Result 2 and Result 3, we review the differences in the experimental design and discuss their relevance for our results.

1. Subject pool. As we have argued, our subjects represented a sample from the best mathematical talents in the Slovak Republic. Since the subject pool is the same for both treatments, there is no reason why the effects in these treatments should be different, unless (due to the complexity of the HET treatment) the cognitive constraints of our subjects were not binding, whereas they were binding for the subjects in the GKS experiment. Indeed, we conjecture that this factor is likely to account for the reversal in the results.

2. Public information. Providing public information may have caused a positive effect on players' guesses (in the sense that they are closer to the Nash equilibrium) by speeding up the process of learning. Since public information was provided after the subjects made their first-round guesses, it cannot account for the difference in the first round averages. In later rounds, by obtaining more information players could make better inferences about the average guesses and recognize that lower guesses lead to higher payoffs. Since public information was provided in both treatments, it cannot be responsible for the reversal in results.

3. Experimental procedure. Our experiment was not computerized and the calculation of averages took about 3–5 minutes. The subjects could use this delay for deeper thinking about the game, which might have generated a positive effect on guesses. There was no such delay before the first round, and this factor can therefore be responsible only for differences in later rounds. In line with the previous argument, the change in the experimental procedure was the same in both treatments. Therefore, experimental procedure cannot be responsible for the reversal in results.

4. Guessing Game redux. By playing the Guessing Game redux, participants could see that the "optimal" strategy in a related game is 0. This may have induced a positive effect on players' guesses, by carrying over the idea of iterated elimination of dominated strategies to the Guessing Game. There is no obvious reason why the experience with the Guessing Game redux should affect the treatments differently.

5. Changes in the payoff function. Analogously to the above arguments, the changes in the payoff function concerned both HOM and HET treatments and therefore cannot be responsible for the reversal in results. In addition, the changes were only marginal and strike us as insignificant.

Summing up, there is no obvious reason why the last factor should affect the results. The second, third, and fourth factors are expected to induce positive effects in both treatments (compared to GKS), and there is no factor with a negative effect. The similar patterns in the HOM treatment indicate that all these factors have only marginal effects (note also that the second and third factor may have affected only guesses in later rounds). In addition, since the changes in both treatments are the same, there is no reason why they should have affected the HOM and HET treatments differently. Therefore, we conclude that the difference in subjects' cognitive constraints, which we conjecture became binding in GKS's HET treatment but not in ours, accounts for more of the equilibrium-like behavior of our subjects in the HET treatment.

5. Application of the CH model

In this section we rationalize our findings using the Cognitive Hierarchy (CH) model by Camerer, Ho, and Chong (2004), henceforth CHC. In this model, each player is able to perform only a limited number of steps of reasoning and assumes that his strategy is the most sophisticated. Following CHC, we consider *Poisson distribution* with parameter τ as the distribution of players' steps of reasoning (see Appendix A for details). A higher value of τ reflects a population of more players with better thinking capabilities or players thinking deeper about their strategies. As a consequence, with increasing τ the prediction of the CH model for the Guessing Game converges to the Nash equilibrium.

CHC argue that "...nothing in the Poisson-CH model, per se, requires τ to be fixed across games or subject pools, or across details how games are presented or choices are elicited" (p. 876). However, if two similar games with players from the same population yield different estimates of τ, it seems reasonable to conclude that the one with higher value reveals more complex reasoning and deeper thinking. It follows that, if our working hypothesis is valid, the estimated parameter for the HET treatment should be higher than the one for the HOM treatment. Since observations in later rounds are no longer independent, we use the first-round guesses in order to estimate the value of τ for each treatment separately.[15] Consistent with our hypothesis, the estimated value of τ in the HET treatment is higher than the one in the HOM treatment. Because the subjects in both summer schools were comparable, a higher value of τ means that those in the HET treatment think more carefully about their strategy.

We estimate the value of τ from our data separately for each treatment (see Appendix A for details). The estimation in the HOM treatment is straightforward: following CHC, we find the value of τ that minimizes the (absolute) difference between the predicted mean and the sample mean. For our sample mean 20.28 (standard deviation 22.56), we obtain $\tau = 1.64$ (yielding mean 20.29 and standard deviation 20.26). The estimation of τ in the HET treatment is more difficult. Due to the assumption that the distribution of players of both types is the same, we need to fit two means (one for the players with $q = 1/3$, the other for the players with $q = 2/3$) with a single value of τ. In order to estimate the value of τ we use two methods: MM and LS. Method MM minimizes the *absolute value of the sum of differences* between the actual means and predicted means, whereas method LS (least squares) minimizes the *sum of squares of differences* between the actual means and predicted means.[16]

[15]In the HET we assume that the steps follow the same distribution, i.e., we apply only one value of τ for both types of players.

[16]Formally, method MM minimizes $|(m_{1/3} + m_{2/3}) - (d_{1/3} + d_{2/3})|$ and method LS minimizes

Table 3 in Appendix C shows the results of estimation as well as predicted means and standard deviations (columns denoted "s.d."). The table also shows these statistics from the data and predictions for the value $\tau = 1.64$ estimated in the HOM treatment. In addition, the last two columns contain the boostrapped standard deviations of the estimate of τ and its 90% confidence interval. Using methods MM and LS, we obtain similar estimates of τ, namely 1.86 and 1.89. Both are higher than the value 1.64 estimated in the HOM treatment, which suggests that the players in the HET treatment indeed do think more carefully about the game. Further supporting this finding, we obtain a medium effect size.[17] These results conform to our hypothesis and rationalize our results.

Result 4. Comparison of the estimated values of τ suggests that players in heterogenous groups use more steps of reasoning than players in homogeneous groups.

In order to compare our results to GKS, we use their data (see Table 2) and estimate the value of τ for both treatments. Specifically, we estimate $\tau = 1.42$ in GKS's HOM treatment, which is 0.22 lower than the estimate in our experiment and yields a medium effect size.[18] For the HET treatment, we estimate $\tau = 0.80$ using method MM and $\tau = 0.82$ using method LS.[19] Both of these values are substantially (by more than 1) lower than the estimates for our subjects and yield a large effect size.[20] The comparison of HET treatments seems to confirm our intuition that the heterogenous environment pushed subjects to their cognitive limits in the GKS experiment, while it induced deeper thinking in our experiment.

Result 5. Comparison of the estimated values of τ suggests that in the HOM treatment our subjects used 0.22 more steps of reasoning than GKS's subjects, whereas in the HET treatment they used at least one step more.

6. Discussion and conclusion

Recently an important piece of literature on the decision making of collective actors has emerged (Camerer 2003, p. 475). One key finding is that collective actors in Guessing Games converge, in a statistically significant manner, faster to the Nash equilibrium than individual actors (Kocher and Sutter 2005). As we do, these authors do not find a statistically significant difference in first-round choices although the mean and median of chosen numbers is consistently lower for collective actors (groups of these individuals) than it is for individuals. Sutter (2005) finds furthermore that teams with four members outperform teams with two members and single persons significantly, whereas the latter two types of decision makers do not differ. This

$(m_{1/3} - d_{1/3})^2 + (m_{2/3} - d_{2/3})^2$, where d_q and m_q denote the sample mean and the predicted mean for $q = 1/3,\ 1/2,\ 2/3$.

[17] Using the bootstrapped standard deviations, we find Cohen's d to be 0.57 and 0.63 for MM and LS methods, respectively.

[18] With value of Cohen's d equal to 0.64.

[19] Bootstrapped standard deviations are 0.26 in the HOM treatment and 0.23 and 0.22 in the HET treatment using methods MM and LS, respectively.

[20] With values of Cohen's d of 2.53 and 2.49 for methods MM and LS, respectively.

suggests that groups of a certain size are, like heterogeneous groups, more efficient in their convergence behavior. Interestingly, subjects in our groups were not allowed to communicate, therefore the increased efficiency that we found seems due to harder thinking, triggered by the more complex environment. In other words, the complexity of an environment to some extent seems able to substitute for group deliberation, at least in Guessing Games.

Relatedly, Slonim (2005) has explored experimentally, and by way of a standard Guessing Game, how subjects' behavior depends on competitors' levels of experience. By also reviewing related evidence from other games, he finds that subjects indeed (learned to) take into account their competitors' experience, or "sophistication" (p. 68). Morone, Sandri, and Uske (2006) have also demonstrated, by way of the Guessing Game proposed by GKS, that subjects converge faster when prompted accordingly. In our case, that prompting — in Morone et al. (2006) done by the experimenter — is done by the complexity of the environment. Essentially, the depth of reasoning, or τ is endogenized.

CHC point out that such an endogenization would be desirable and make a case for it by reporting that subjects at CalTech think about the same degree deeper relative to students at a nearby community college, as our mathematically talented students do relative to the students employed by GKS. Our result is also in line with results reported in Palacios-Huerta and Volij (2006), who find that professional chess players engage in more equilibrium-like behavior than typically employed student subjects.

In sum, we have experimentally explored a version of the Guessing Game. Our subjects were mathematically talented students that were skilled in solving abstract problems. Our data confirm the hypothesis that heterogeneous players guess closer to the equilibrium. This conclusion is rationalized by way of the Cognitive Hierarchy Model proposed by CHC. We also find that heterogeneous players' guesses converge faster to the equilibrium. Our results contribute both to the ongoing debate about the depths of reasoning and the debate about the limitations of traditional subject pools.

Appendix A: CH model

The CH model builds on a probability distribution $f(k)$ of players where k denotes the number of steps of thinking the player takes. A step 0 player randomizes his guesses (uniformly among all strategies) and does not assume anything about his opponents. A step k player thinks that he is the smartest (i.e., there are no other players capable of higher steps of thinking than $k-1$). He assumes that other players' steps of reasoning are distributed according to some probability distribution on $\{0, 1, \ldots, k-1\}$, and plays the best response to this distribution. The player is aware of the proportions of players with less steps of reasoning and his subjective probabilities (frequencies in the population) follow the truncated distribution.[21]

CHC assume that the distribution $f(k)$ of steps of thinking follows the *Poisson distribution* with parameter τ, i.e., $f(k) = e^{-\tau}\tau^k/k!$. This distribution is described by a single parameter τ representing both its mean and variance. The authors claim that the Poisson distribution is able to fit data almost as well as more parametrized models, but it is easier to compute and work with. Thus, it represents a parsimonious model of the distribution of depth of reasoning in a subject pool. The particular value of the parameter τ can be estimated from data.

In order to estimate the parameter τ from the data, we consider a grid of values of τ and for each τ from the grid we compute numerically the strategies (guesses) of players resulting from various steps of thinking. In particular, we use a grid with difference 0.01 on the interval $[0, 4]$ (for comparison, the highest value of τ estimated by CHC is 3.7). Using the actual payoff function $\tilde{u}$ (see Section 2), we compute the strategies of players up to the 10-th step and for certain values of τ (of our interest) up to the 100-th step.[22] In this way we obtain for each τ a (predicted) distribution of guesses which allows us to predict the mean and the standard deviation of guesses depending on τ. Following CHC (estimation procedure for their Table II), we then choose the value of τ that fits our data best. In the HOM treatment, we minimize the absolute difference between the predicted mean and the actual mean. In the HET treatment, we propose two methods: minimize sum of absolute differences (of the two groups) or minimize sum of squares of differences (see Section 5 for details).

Let us denote $b_q(k)$ the strategy of a k-step player i with $q_i = q$ (where $k = 1, 2, \ldots$).[23] The predicted strategies show several interesting patterns. The strategies of 1-step players are obviously independent on the distribution. In the HOM treatment we obtain $b_{1/2}(1) = 21$; in the HET treatment $b_{1/3}(1) = 14$ and $b_{2/3}(1) = 30$. For any fixed k, the strategies are non-increasing in τ. This reflects the fact that for the same step of thinking, higher values of τ are associated with guesses closer to the equilibrium. In addition, for any fixed τ, the strategies are non-increasing in k and become stationary when k is sufficiently high. This means that players capable

[21] Then, for any $i = 0, 1, \ldots, k-1$, the distribution of players capable of i steps is $f(i)/\sum_{j=0}^{k-1} f(j)$.

[22] The source code (in *Mathematica*) and the detailed results of the computations are available at `www.uni-bonn.de/~kovac/papers/xp`.

[23] Although we allowed the players to guess non-integers (up to two decimals), we restrict the theoretical analysis of the CH model to integer strategies. Only 7 players (2 in HET, 5 in HOM) out of 56 submitted non-integers in the first round. Hence the restriction seems justified.

of more steps guess closer to the equilibrium. For the values of τ in our grid we conjecture that they do not converge to zero as $k \to \infty$. This is intuitively clear, because it is obviously not optimal to play zero (Nash equilibrium), when there is a high probability of some of the opponents being a 1-step player. For example, for $\tau = 1.64$ we obtain $b_{1/2}(2) = 9$, $b_{1/2}(3) = b_{1/2}(4) = 7$, $b_{1/2}(5) = b_{1/2}(6) = \cdots = 6$. Using these strategies we can compute statistics of the distribution of guesses.

Appendix B: Instructions

The instructions were written in Slovak. Based on the instructions by GKS, we first created an English version which was later translated into Slovak. The instructions below are for heterogeneous players with $q = 1/3$. The instructions for heterogeneous players with $q = 2/3$ differ only in the last sentence of paragraph 4 and in the formal expression for the payoff. In the instructions for homogeneous players (i.e., $q = 1/2$) the whole of paragraph 3 was replaced by: "The target number for you (and everyone else in your group) is one-half of the average of all 4 chosen numbers in your group." Additionally, the formal expression for the earnings contained 1/2 instead of 1/3.

Sample instructions

Welcome to our experiment and thank you for participating.

You will be randomly divided into groups of 4 persons. Each person in your group chooses a number (denote it x) from the closed interval $[0, 100]$. It is not necessary to choose an integer, however, numbers with more than two decimals are excluded.

Your potential earnings in this experiment depend on how close your chosen number is to a target number. The closer your chosen number is to the target number, the higher are your earnings.

Your group consists of two participants of type A and two participants of type B. Target numbers of type A and type B participants are different. If you are type A, your target number is one-third of the average of all 4 numbers chosen in the group. If you are type B, your target number is two-thirds of the average of all 4 numbers chosen in the group. **You are type A**, so the target number in your case is **one-third of the average of all 4 chosen numbers** in your group.

The potential earnings in each round depend on the difference between your chosen number and the target number. If your chosen number in that round is identical with the target number, your earnings will be 50 crowns. If the two numbers differ, their distance will be deducted from the 50 crowns. Formally, your potential earnings per round are calculated as follows:

$$\text{earnings} = 50 - \left| x - \tfrac{1}{3}\,\text{average} \right| .$$

If your earnings are negative, we will treat them as zero.

The experiment will last 5 rounds. Groups will be rematched in each round.

After each round, you will receive information on the average in your group and your target number. In addition, the averages in all groups will be announced by writing on the blackboard. Because of time constraints the earnings will not be computed. We recommend that you calculate your earnings after each round (using the above formula).

After the experiment, for each round, we will draw randomly two groups that will be paid their earnings from that round. All earnings will be paid in cash and privately at the end of the experiment.

Round	**1**	**2**	**3**	**4**	**5**
Chosen number					
Average					
Earnings					

Appendix C: Figures and tables

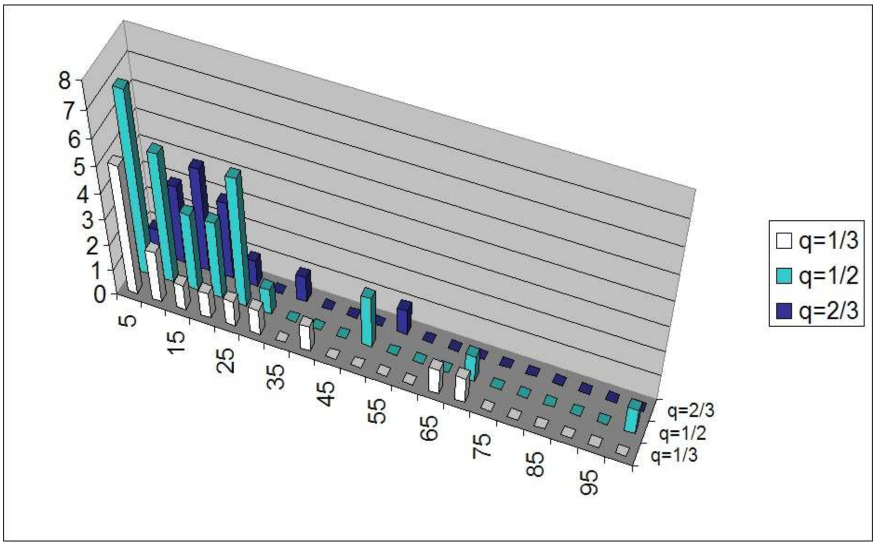

Figure 1: Distribution of first-round guesses (number of individuals who have chosen a number in particular interval)

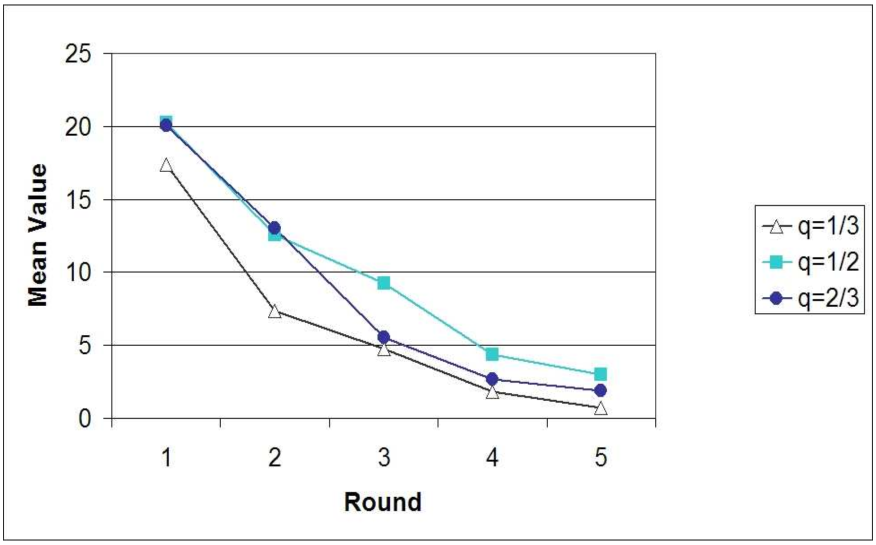

Figure 2: Treatment averages

Group	Statistics (variable)	Round 1	2	3	4	5
$q = 1/3$	mean	17.39	7.35	4.76	1.85	0.72
$(n = 14)$	standard deviation	13.67	6.22	3.65	1.98	1.61
	median	14.50	5.50	4	1.06	0.10
	min	0	1	0	0	0
	max	54.50	21	12	7	6
	average payoff (SKK)	38.41	42.04	45.57	48.21	49.10
	change in mean: absolute		10.04	2.59	2.90	1.13
	relative		0.58	0.35	0.61	0.61
$q = 1/2$	mean	20.24	12.55	9.24	4.36	3.00
$(n = 28)$	standard deviation	22.56	9.70	8.04	4.15	4.77
	median	13.50	10	7.51	4	2
	min	0	0	0	0	0
	max	100	38	35	19.73	24
	average payoff (SKK)	37.63	42.54	44.10	47.17	47.54
	change in mean: absolute		7.69	3.31	4.88	1.36
	relative		0.38	0.26	0.53	0.31
$q = 2/3$	mean	20.05	13.04	5.51	2.70	1.89
$(n = 14)$	standard deviation	22.46	14.08	7.32	4.54	4.14
	median	12.50	8.77	2.54	1.14	0
	min	0	0	0	0	0
	max	66	47	26	15	15
	average payoff (SKK)	34.82	44.14	45.38	47.98	48.11
	change in mean: absolute		7.01	7.53	2.81	0.81
	relative		0.35	0.58	0.51	0.30

Table 1: Results of the experiment

Group	Statistics (variable)	Round 1	2	3	4	5
$q = 1/3$	mean	27.27	17.23	14.30	9.74	6.76
$(n = 20)$	std. deviation	23.61	12.16	11.00	7.80	6.28
	median	15.50	13.57	11.26	8.10	5
$q = 1/2$	mean	22.79	12.00	7.54	4.73	3.03
$(n = 40)$	std. deviation	20.54	9.67	8.71	7.04	5.53
	median	19.75	10	5	1.28	0.63
$q = 2/3$	mean	38.42	30.79	20.07	16.30	12.00
$(n = 20)$	std. deviation	28.05	20.80	11.87	13.49	10.80
	median	36.50	25.28	20	16	10.25

Table 2: Results of the experiment by GKS

		Predicted statistics		Boostrap results for τ	
		$q = 1/3$	$q = 2/3$	standard	90% conf.
Method	τ	mean (s.d.)	mean (s.d.)	deviation	interval
Data		17.39 (13.67)	20.05 (22.46)		
HOM	1.64	17.11 (20.92)	24.53 (19.81)	0.41	(1.11, 2.43)
MM	1.86	15.11 (19.23)	22.32 (18.63)	0.37	(1.33, 2.55)
LS	1.89	14.86 (19.01)	22.03 (18.47)	0.39	(1.40, 1.60)

Table 3: Results of estimation methods for the HET treatment

References

Ballinger, P., E. Hudson, L. Karkoviata, and N. T. Wilcox (2005). Saving performance and cognitive abilities. Mimeo.

Bosch-Domènech, A., J. G. Montalvo, R. Nagel, and A. Satorra (2002). One, two,(three), infinity,...: Newspaper and lab beauty-contest experiments. *The American Economic Review 92*(5), 1687–1701.

Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press.

Camerer, C. F., T.-H. Ho, and J.-K. Chong (2004). A cognitive hierarchy theory of games. *The Quarterly Journal of Economics 119*(3), 861–898.

Davis, D. D. and C. A. Holt (1993). *Experimental Economics*. Princeton, NJ: Princeton University Press.

Güth, W., M. Kocher, and M. Sutter (2002). Experimental 'beauty contest' with homogeneous and heterogeneous players and with interior and boundary equilibria. *Economic Letters 74*, 219–228.

Kocher, M. G. and M. Sutter (2005). The decision makers matters: Individual versus group behavior in experimental beauty-contest games. *Economic Journal 115*, 200–223.

Morone, A., S. Sandri, and T. Uske (2006). On the absorbability of the guessing game theory: A theoretical and experimental analysis. Mimeo.

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *American Economic Review 85*(5), 1313–1326.

Palacios-Huerta, I. and O. Volij (2006). Field centipedes. Mimeo.

Palfrey, T. R. and J. Prisbrey (1997). Anomalous behavior in linear public goods experiments: How much and why? *American Economic Review 87*(5), 829–846.

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology 41*(1), 1–19.

Slonim, R. L. (2005). Competing against experienced and inexperienced players. *Experimental Economics 8*(1), 55–75.

Sutter, M. (2005). Are four heads better than two? An experimental beauty-contest game with teams of different size. *Economics Letters 88*, 41–46.

Wilcox, N. T. (1993). Lottery choice: Incentives, complexity and decision time. *Economic Journal 103*, 1397–1417.

Wilcox, N. T. (2006). Theories of learning in games and heterogeneity bias. *Econometrica 74*(5), 1271–1292.