Osaka University Knowledge Archive

Title	Cognitive ability and observed behavior in laboratory experiments: implications for macroeconomic theory
Author(s)	Hanaki, Nobuyuki
Citation	The Japanese Economic Review. 2019, 71, p. 355-378
Version Type	AM
URL	https://hdl.handle.net/11094/89939
rights	
Note	

Osaka University Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

Osaka University

Cognitive ability and observed behavior in laboratory experiments: implications for macroeconomic theory*

Nobuyuki Hanaki†

September 1, 2019

Abstract

This paper discusses the relationships between the "measured" cognitive ability of participants and their behavior as observed during laboratory experiments. Based on such relationships, macroeconomic implications of micro-level "boundedly rational" individual behavior will be discussed. The paper also addresses potential problems that arise when insufficient attention is paid to large differences in the measured cognitive ability of participants across several experimental laboratories, influencing the replicability of existing experimental results but also the interpretation of results from cross-country experimental analyses, and proposes to complement participants' database with individual characteristics.

Keywords: Cognitive Ability, Laboratory Experiments, Strategic Environment Effect

JEL Code: C90, D84

*This article is based on an invited lecture that the author gave at the 2019 Japanese Economic Association Spring Meeting in Musashi University. I thank the members of the program and organizing committee of the conference for giving me this opportunity. I also thank co-authors of those articles discussed during this lecture and various researchers, in particular, Eizo Akiyama, Te Bao, Alessandra Casella, John Duffy, Ido Erev, Yukihiko Funaki, Eric Guerci, Cars Hommes, Keigo Inukai, Ryuichiro Ishikawa, Nicolas Jacquemet, Mamoru Kaneko, Alan Kirman, Stephane Luchini, Takehito Masuda, John Miller, Charles Noussair, Hideo Owan, Scott Page, Eko Riyanto, Rajiv Sethi, Angela Sutan, Jan Tuinstra, Naoki Watanabe, Duncan Watts, Martin Weber, Marc Willinger, and Adam Zylbersztejn, for stimulating discussions and inspirations. Abby Cuttriss provided an excellent proof-reading. Part of research discussed during this lecture was financed by the French National Research Agency (ANR) under an ORA-Plus project "BEAM" (ANR-15-ORAR-0004) and under Investissements d'Avenir UCA^{JEDI} (ANR-15-IDEX-01), in particular, UCAinACTION project, and the Behavioral Economics Research Center at the Institute of Social and Economic Research of Osaka University. The lecture as well as the original version of this manuscript have been completed while I was affiliated with GREDEG, CNRS, Université Côte d'Azur (UCA), France. I thank these institutions for their fantastic research environment where I have prepared the lecture and the original version of the manuscript.

†Institute of Social and Economic Research, Osaka University. E-mail: nobuyuki.hanaki@iser.osaka-u.ac.jp

1 Introduction

We have ample evidence showing that people differ in their intelligence (see, for example, Brouwers et al., 2009). Yet, many theoretical analyses in economics have assumed that decision makers are equally rational or that a representative rational decision maker. It is true that there have been many attempts in the past to introduce (heterogeneous) boundedly rational decision makers into analyses (see, Conlisk, 1996, for a review). However, it is only recently, thanks to developments in behavioral and experimental economics,[1] that "behavioral" models, which include boundedly rational decision makers, have been proposed and investigated extensively in such fields as game theory (Camerer, 2003), industrial organization (Spiegler, 2011), and finance (Shleifer, 2000).

As noted by Thaler (2015), however, macroeconomics has been the field least impacted by the recent behavioral and experimental revolution. Perhaps the main reason for this is a belief held by many economists that can be summarized by an old statement from Gary Becker: "households may be irrational and yet markets quite rational" (Becker, 1962, p.8). According to this view, because deviations from rational behavior by many boundedly rational individuals cancel out in aggregate, bounded rationality at the individual or household level does not matter at the macro level. Indeed, the famous zero-intelligence agents model studied by Gode and Sunder (1993, 1997) shows that even the markets consisting of zero-intelligence computer traders can exhibit high allocative efficiency when these traders respect their budget constraints.

There are, however, theoretical studies that show the existence of a few boundedly rational decision makers in a large population can have a larger-than-proportional impact on aggregate outcomes. In particular, Haltiwanger and Waldman (1985, 1989, 1991) demonstrate that the behavior of boundedly rational decision makers can have a large influence at the macro-level when the environment is characterized by strategic complementarity rather than strategic substitutability. This is what Hanaki et al. (2019) call the strategic environment effect (SEE). The SEE has been shown experimentally in price setting games (Fehr and Tyran, 2008), forecasting-games (Heemeijer et al., 2009; Bao et al., 2012), duopoly games (Potters and Suetens, 2009), and the beauty contest games (Sutan and Willinger, 2009; Cooper et al., 2017).

In this article, we first show the relationship between participants' measured cognitive ability and their observed behavior in laboratory experiments including a coordination game (Hanaki et al.,

[1]There are now many references, including popular books such as Ariely (2008) and Kahneman (2011).

2016) and an asset trading game (Akiyama et al., 2017). These results demonstrate that the main reasons why participants' behavior deviate from the Nash equilibrium or the rational expectations equilibrium differ depending on their cognitive ability. Such results are consistent with the view behind the level-K (Nagel, 1995) or the cognitive hierarchy model (Camerer et al., 2004) that explicitly assume heterogeneity in the depth of strategic thinking among decision makers.

We then discuss the strategic environment effect by referring to Hanaki et al. (2019), which shows both theoretically and experimentally that the SEE exists when group size is not too small (and is thus relevant at the macro-level) in beauty contest games. The main theoretical insight of Hanaki et al. (2019) is that the SEE is driven by the existence of heterogeneity in terms of the depth of strategic thinking among decision makers, and therefore, in the presence of strategic complementarity, one must investigate seriously the macroeconomic implications of the interactions among heterogeneous boundedly rational decision makers.

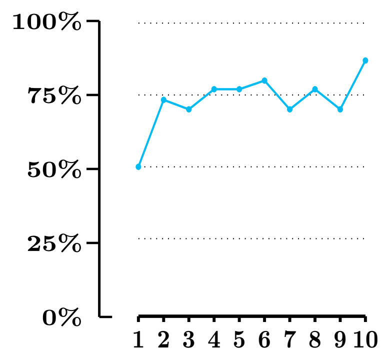
We complement the main discussion with some methodological considerations. Cognitive ability of participants varies not only within the pool of participants in each experimental laboratory, but more importantly, across these pools. Paying insufficient attention to such differences across pools of participants can not only impact the replicability of experimental results but also lead to misinterpreting the differences in the experimental results obtained in different countries. We conclude this paper by proposing to complement the database of participants used in each experimental laboratory with information on individual characteristics such as cognitive ability.

2 Cognitive ability and coordination failure

Let us first consider the following experiment of a very simple 2×2 coordination game. Participants are divided into two groups: A and B. We call those participants who belong to group A, A, and those who belong to group B, B. One A and one B will form a pair, and play a simultaneous move game. In this game, A can choose either L or R, and B can choose either U or D. The payoffs for A and B are as follow:

- If A chooses L, regardless of B's choice, A obtains ¥850 and B obtains ¥300.
- If A chooses R and B chooses U then A obtains ¥650 and B obtains ¥475.
- If A chooses R and B chooses D then A obtains ¥1000 and B obtains ¥500.

Figure 1: Fraction of As choosing R



Source: Hanaki et al. (2016, Fig. 2). The color of the line is modified from the original figure to make it consistent with Figure 2 below.

Imagine you are A. Which option do you choose, L or R?

In Hanaki et al. (2016), we report the results of an experiment with this game. In the experiment, each participant repeated the same game 10 times with different opponents. Let's call one play of the game a round. This perfect stranger matching was made possible by recruiting 20 participants, thus 10 As and 10 Bs, in each experimental session. After each play of a game, each participant was informed of the payoff s/he had obtained, but not the choice made by the opponent. Of course, in the above example, if A has chosen R, one could have guessed what the other has chosen based on the payoff, but this is not the case in case A has chosen L. At the end of the experiment, one of the 10 rounds was selected randomly and participants were paid according to the payoff they obtained in the selected round. Because the experiments were conducted in Paris, all the payoffs were shown in Euros with ¥100 = €1 conversion rate. Thus, ¥850 was converted into €8.50, for example.

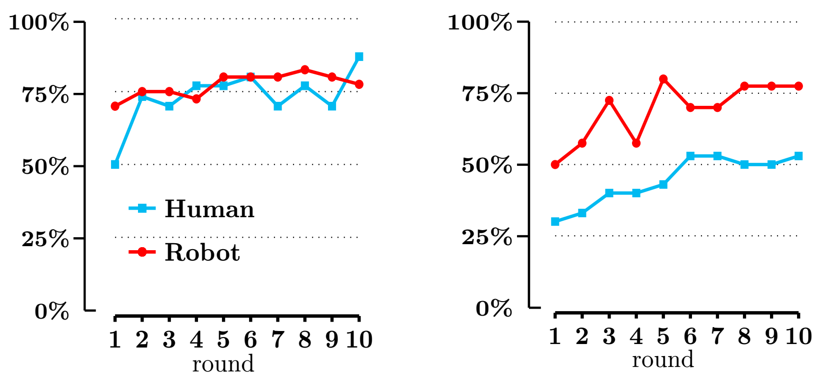
We focus on As' choices. The paper reports the behavior of Bs as well, so interested readers can refer to the paper. Figure 1 shows the dynamics of the fraction of 30 As who had chosen R over 10 rounds. As you can see, half of 30 participants who acted as As in our experiment chose R in the first round, and then about 75% of them did so in the remaining rounds. The first question we asked was why some As did not choose R.

If we look at the game, we notice that B has a dominant strategy, which is to choose D. If A believes that B will choose D, then it is better for A to choose R as well. Thus, the reason for A not choosing R is the uncertainty A faces about B's behavior. In particular, if A believes that

Figure 2: Fraction of As choosing R

B \ A	L	R
U	3.00, 8.50	4.75, 6.50
D	3.00, 8.50	5.00, 10.00

B \ A	L	R
U	8.50, 9.75	8.50, 3.00
D	8.50, 9.75	10.00, 10.00



Source: Hanaki et al. (2016, Fig. 1)

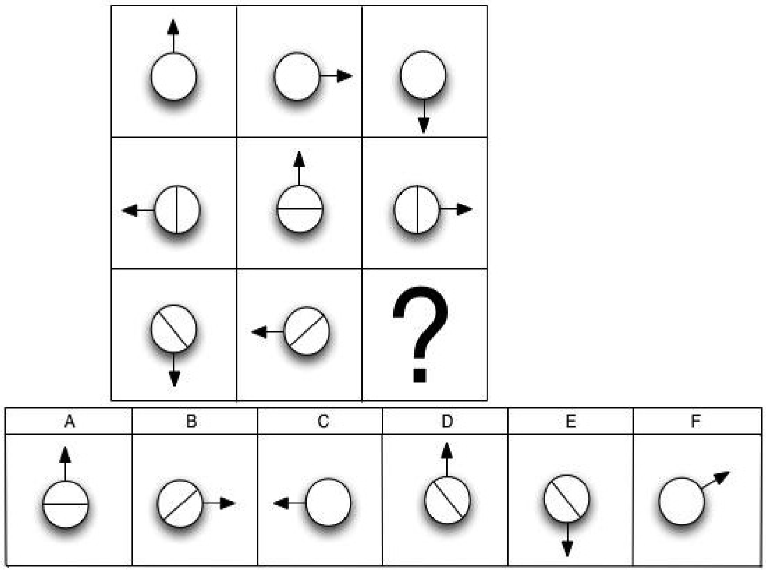
the likelihood of B choosing U is high enough (in this particular case, higher than 3/7, for a risk neutral A), it is better for him/her to choose L. If this is the main reason, then, if A knows that B will choose D for sure, s/he will choose R. To test this, we have conducted another experiment in which As face a computer program that is acting as B for 10 rounds. Participants were informed, in addition to the rules of the game, that **"the computer chooses D at each round, without exception."**

The left panel of Figure 2 shows the result. It shows, in red, the dynamics of the fraction of As (N=40) who have chosen R against the computer program known to choose D (robot B), as well as the fraction of As (N=30) who have chosen R against human Bs. The surprising result is that, not all the As have chosen R in any of the 10 rounds even when they play against robot B, and furthermore, except for round 1, there is no difference between the results from two settings.

This absence of difference between the fraction of As choosing R against the robot Bs and against human Bs is specific to this particular payoff setting as one can see from the result of another payoff matrix shown in the right panel of Figure 2. In this setting, the payoffs were such that

- If A chooses L, regardless of B's choice, A obtains €9.75 and B obtains €8.50.
- If A chooses R and B chooses U then A obtains €3.00 and B obtains ¥8.50.
- If A chooses R and B chooses D then A obtains €10.00 and B obtains ¥10.00.

Figure 3: An example of matrix reasoning quiz



Source: The International Cognitive Ability Resource Team (2014; https://icar-project.com/)

so that the payoff for A from choosing L is higher, while it is lower when A chooses R and B chooses U, than the previous one. As a result, against human Bs, the fraction of As (N=30) choosing R is lower compared to the case shown in the left panel. This results in the fraction of As (N=40) choosing R against the robot B being higher than the case against human B, although the former is similar to the level shown in the left panel, .

What this result suggests is that not all As choose L instead of R in response to the uncertainty of the behavior of Bs. Facing such a result, we wondered whether participants' cognitive ability had something to do with their decision to not choose R against the robot B.

We have used the advanced version of Raven's progressive matrix (RPM) test (Raven, 1998) to measure participants' cognitive ability. This test consists of series of tasks, with increasing difficulty, similar to the one shown in Figure 3. In each task, a participant is asked to pick one item from those shown in the bottom that best fits the space marked with "?" (just a blank space in the original test) in the 3×3 pictures above. This test is a non-verbal measure of fluid intelligence, that is "the capacity to think logically, analyze and solve novel problems, independent of background knowledge" (Mullainathan and Shafir, 2013, p.48), and is widely used by psychologists, educators and the military (Raven, 2008).[2] The full test consists of 48 tasks to be solved in 30 minutes or so, but we have chosen 1 out of every 3 questions, and asked our participants to solve them within 10 minutes. The test was conducted after participants had completed playing the game for 10 rounds.

In total, 140 participants who acted as A took the test. We have divided the participants, based on their score of RPM test, into three groups: Low, Medium, and High. Low group consists of those participants whose score of RPM test was less than 7. Those in High group scored 10 or above in RPM test. The remaining participants are in the Medium group.[3]

Figure 4 shows the dynamics of the fraction of As in three cognitive ability groups choosing R when playing against human Bs (in blue) and against the robot B (in red). Here the data from the experiments with two payoff matrices shown above are pooled.[4] What one can immediately notice is the absence of differences between the two conditions for Low group (shown in the left panel). For the other two groups, the fraction of As choosing R is higher when playing against the robot B compared to the case against human Bs.

Figure 5 shows the same information from a different perspective. It shows the empirical cumulative distributions (ECDs) of the frequencies (out of 10 rounds) of R choices in three groups under two conditions. Three observations can be made. (1) Against human Bs, the ECDs of the frequency of R choices are not significantly different across three cognitive groups.[5] (2) Against robot B, however, the ECD of the frequency of R choices for Low group is first order stochastically dominated by those of Medium and High groups, while there is no significant difference between the latter two groups.[6] The reason for these observations is that (3) subjects in Low group do not

[2]Carpenter et al. (2013) report a positive correlation between RPM score and the degree of strategic sophistication measured in terms of the number of wins in Race to 5, 10, and 15 games (also called Hit-5, -10, and -15 games). Gill and Prowse (2016) shows a positive relationship between the speed of learning (to choose a number of closer to Nash equilibrium) in a three-player p-beauty contest game and RPM score. Proto et al. (2019), as discussed in Section 5.1 below, show that those participants with high RPM scores tend to be better at resisting short-term temptation and better at best-responding to the other's strategy in an indefinitely repeated game than low RPM score counter parts.

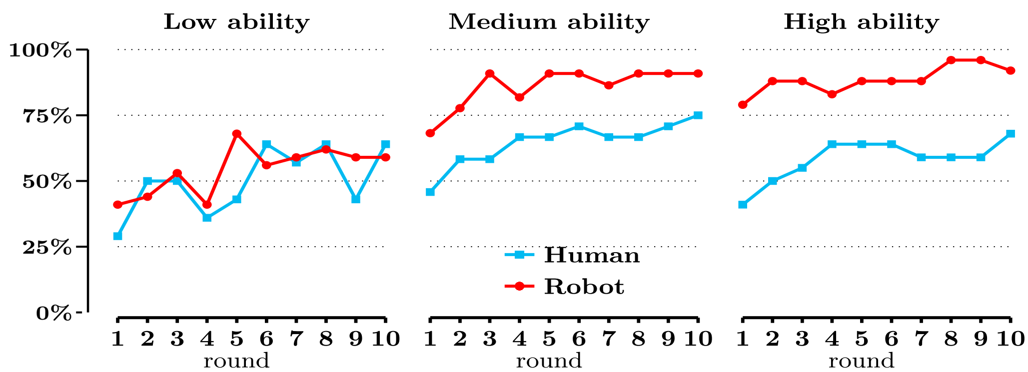
[3]This grouping corresponds to the top 1/3, middle 1/3, and bottom 1/3 of all the participants, including Bs, who took the RPM.

[4]The reason for pooling the data from two payoff matrices is to secure large enough sample size for three cognitive groups in both robot and human treatments. If we have separated the two, then there would have been only around 10 subjects in each condition in each cognitive group, which have not been enough for conducting meaningful statistical analyses. However, even if we analyze the two games separately, we obtain the same result regarding the difference between Robot and Human conditions for Low (p= 0.255 and 0.304, respectively, for the first (shown on the left panel of Figure 2) and the second (shown on the right panel of Figure 2) payoff matrices) and High groups (p=0.041 and 0.053 for the first and the second payoff matrices, respectively.) For the Medium group, the difference between Robot and Human conditions is statistically significant for the first payoff matrix (p=0.003) but not so for the second payoff matrix (p=0.173).

[5]Low vs Medium (p=0.288), Low vs High (p=0.599), and Medium vs High (p=0.695) all based on two-sided bootstrap Kolmogorov-Smirnov test.

[6]Low vs Medium (p=0.003), Low vs High (p=0.001), and Medium vs High (p=0.318) all based on two-sided bootstrap Kolmogorov-Smirnov test.

Figure 4: Fraction of As in three cognitive ability groups choosing R



Source: Hanaki et al. (2016, Fig. 5)

respond to the absence of behavioral uncertainty in Robot condition, compared to Human condition, by increasing frequency of R choices (p=0.415), while those in remaining two groups do so (p=0.001 for both groups).

Thus, for those in the Low group, the choice of L against human Bs is not at all caused by the uncertainty related to the behavior of human Bs (behavioral uncertainty). Instead, it is mainly due to their own bounded rationality, regardless of its cause. For those in Medium and High group, the main cause of the choice of L against human Bs is the behavioral uncertainty, although there are other contributing factors.

These observations suggest that without taking an additional step, as we have done in this set of experiments, in order to better understand the reasons behind the choice of L, we would not have discovered the differences across the cognitive ability groups.

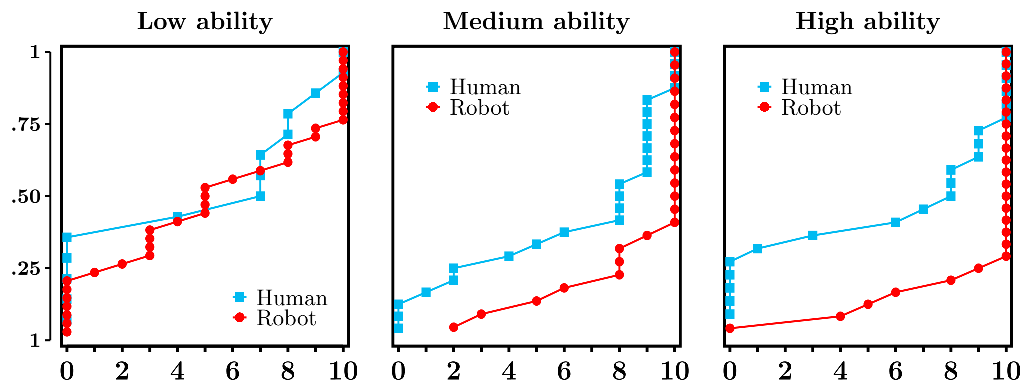
In the next section, we discuss a similar finding related to the cognitive ability in an experimental asset market based on Akiyama et al. (2017).

3 Cognitive ability and mis-pricing in experimental asset markets

Akiyama et al. (2017) investigated "irrational" bubbles observed in experimental asset markets pioneered by Smith et al. (1988).[7] In the experimental paradigm of Smith et al. (1988), N participants

[7]The literature is extensive. See Palan (2013) for a review.

Figure 5: Distribution of the frequencies of R choices in three cognitive ability groups.

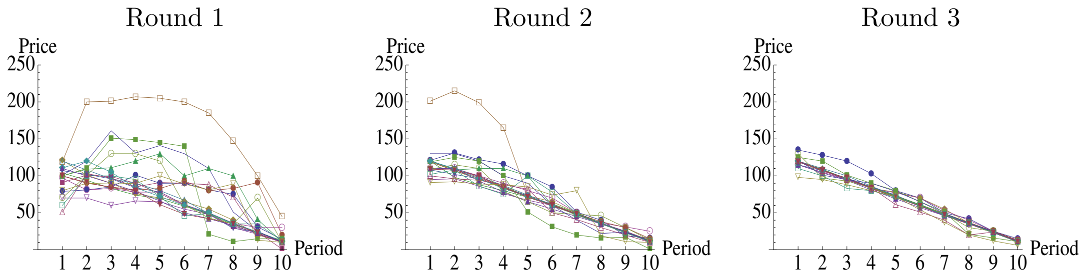


Source: Hanaki et al. (2016, Fig. 6). The color of lines are modified from the original figure to make them consistent with Figure 2 and 4 above.

trade a hypothetical asset over T periods. Before the first period, each participant, say i, receives some units of the hypothetical asset (A^i units of asset) as well as some amount of hypothetical cash (M^i ECUs of cash) that can be used to trade. At the end of each trading period (t), a unit of asset pays a dividend, d_t, and the cash held (before receiving the dividend payment) generates r_t % of interest revenue. The asset and cash held are carried over to the next trading periods. At the end of period T, after the dividend and interest payment, each unit of the asset held will be converted to B ECUs of cash. Participants receive the monetary reward according to the final amount of cash they hold according to the pre-specified rule. In the simplest possible case where $r_t = 0$ and $d_t = d > 0$ for all t and $B = 0$, the fundamental value (FV) of a unit of asset at the beginning of period t, FV_t, is simply a sum of remaining dividend payment, $FV_t = (T - t + 1)d$. Given that the experiment normally lasts no more than a couple of hours, we are ignoring the discounting factor. Under this setting, the common knowledge of rationality implies no trading and the price in each period, if any, should follow the FV. The common experimental finding, however, is that transactions occur at the prices different from FV.

Figure 6 shows the result of the experiment reported in Akiyama et al. (2014). In this experiment, the parameters were set so that $N = 6$, $T = 10$, $d_t = 12$ and $r_t = 0$ for all t, $B = 0$. Each participant received 4 units of asset and 520 ECUs of cash as their initial endowment. Participants repeated the same experiment, with the same group of 6 participants under the same parameter values three times (each repetition is called a round below). Each connected dot represents the outcome of one

Figure 6: Dynamics of realized price in the asset market experiment of Akiyama et al. (2014)



Source: Akiyama et al. (2014, Fig. 1)

group, and there are 20 groups. As one can observe in the figure, there are many groups that trade at prices different from (in many case, higher than) the FV in the first round (Round 1). We call this price deviation from the FV "mis-pricing."[8] Another common finding is that mis-pricing becomes smaller if the same group of participants repeat the experiment under the same conditions, and the mis-pricing disappears by the third round.[9]

Previous studies suggest two main reasons for this mis-pricing. One was individual confusion or bounded rationality (Lei et al., 2001; Kirchler et al., 2012), and the other was uncertainty regarding others' behavior or understanding of the experimental set-up that we call "behavioral uncertainty" (Smith et al., 1988; Cheung et al., 2014). We believed, however, it was not just one or the other, but both individual bounded rationality and behavioral uncertainty that played a role, and their importance depended on participants' cognitive ability.

To investigate this, we conducted the following experiment. Just as in Akiyama et al. (2014), the parameters were set so that $N = 6$, $T = 10$, $d_t = 12$ and $r_t = 0$ for all t, $B = 0$. Each participant received 4 units of asset and 520 ECUs of cash as their initial endowment. There were two conditions: one in which a market consisted of six human participants (6H), and another in which a market consisted of one human participant and five computer traders (1H5C). Participants were informed of the condition in which they participated (i.e., 6H or 1H5C), and those who participated to 1H5C were also informed about the behavioral rule of the five computer traders. As discussed in detail below, we were interested in comparing participants' initial price forecasts in these two conditions.

[8]Because prices often deviate upward from FV in the middle and drop back to FV toward the end, this pattern is often called a "bubble."

[9]It is important to note that this convergence of prices to FV after a couple of repetitions happens only when *the same group of subjects repeat the experiment under the same conditions.* Prices deviate from FV even among the same group of experienced subjects if market conditions are altered (see, Hussam et al., 2008).

Table 1: Table regarding the dividend and FV given to participants

	Remaining Periods	Dividend	Next Value (in ECUs)
At the end of 1st period	9	12	108
At the end of 2nd period	8	12	96
At the end of 3rd period	7	12	84
At the end of 4th period	6	12	72
At the end of 5th period	5	12	60
At the end of 6th period	4	12	48
At the end of 7th period	3	12	36
At the end of 8th period	2	12	24
At the end of 9th period	1	12	12
At the end of 10th period	0	12	0

* The next value at the beginning of Period 1 is remaining periods (10) $\times$ 12 ECUs = 120 ECUs.

* After the dividend payment after the end of Period 10, the value of the asset will be 0.

To explain how participants were informed about the behavioral rule of the computer traders, we need to describe a few other aspects of the experiment first. All the participants received an explanation about the dividend payment process, and the also received the table regarding the FV shown in Table 1. Note that we called FV "Next Value" in the instructions for the experiment. Participants could refer to the table anytime during the experiment.

Participants also received a detailed explanation about the market mechanism employed in each period. We employed a call market setup. In each period, each trader can submit a buy and a sell order by specifying a price-quantity pair. To submit a buy order, a trader has to specify the maximum price at which they are willing to buy (bid) and the quantity demanded. To submit a sell order, a trader needs to set the minimum sell price at which they are willing to sell (ask) and the quantity supplied. The orders need to respect the budget constraint so that the trader must hold enough cash to fully execute the buy order at their bid, as well as enough asset to fully execute the sell order. Once all the traders submit their orders, market clearing price is computed and transactions take place among those traders whose bid were no less than the market clearing price and those traders whose ask was no greater than the market clearing price. Ties were broken randomly.

After receiving these explanations regarding the dividend, FV, and the market mechanism employed, those participants in 1H5C condition were told the exact behavioral rule of computer traders as follows: "In each period, each computer trader places buy and sell orders by setting both the maximum price it is willing to pay and the minimum price it is willing to accept to the next value at the beginning of the period."

As noted above, we were interested in comparing the price forecasts submitted by participants under two conditions to quantify the effect of individual bounded rationality and behavioral uncertainty behind the mis-pricing observed in experimental asset market. How does the current experiment allow us to do so? Imagine a participant in 1H5C condition who perfectly understands the experimental setting. Such a participant should forecast the prices to follow FV in each period. Thus, any deviation of price forecasts from FV observed in 1H5C should be due to some kind of individual bounded rationality. Now, imagine a participant in 6H condition who perfectly understands the experimental setting. Such a participant may not forecast the prices to follow FV in each period because of the uncertainty regarding the behavior of other participants in the market. Of course, the real participants in 6H may suffer from some kind of bounded rationality just as those in 1H5C do. Thus, deviation of price forecasts from FV observed in 6H should be due to both behavioral uncertainty and individual bounded rationality. Taking the difference between the two conditions, therefore, we can quantity the effect of each.

The forecasts were elicited as per Haruvy et al. (2007) as follows: At the beginning of each period, before submitting their orders, participants were asked to submit their price forecasts for all the remaining periods. Thus, at the beginning of period 1, each participant submitted his/her forecasts for prices in period 1, 2, .., 10. At the beginning of period 2, they submitted forecasts for prices in period 2, 3, .., 10. If a forecast was within plus-minus 10% of the realized price, it generated a bonus payment equal to the 0.5% of their final cash holding at the end of Period 10. Thus, the total bonus payment was equal to 0.5% $\times$ Number of forecasts that fell within the 10% range of the realized price $\times$ the final cash holding at the end of Period 10.[10]

Finally, we employed the Cognitive Reflection Test (CRT, Frederick, 2005) to measure participants' cognitive ability. CRT consists of a few questions such that the first "intuitive" answer that comes to our mind is wrong. To arrive at the right answer, one has to reflect upon the intuitive answer. And thus, CRT is often discussed in relationships with System-1 and System-2 thinking (Kahneman, 2011). CRT is short and easy to implement and its score, the number of questions correctly answered, is positively correlated with the score of RPM test (Corgnet et al., 2018).[11] We

[10]In Hanaki et al. (2018), we report the result of comparing different ways of incentivizing forecasting performance in asset market experiments. We found that the bonus scheme employed in Akiyama et al. (2017) resulted in a significantly larger mis-pricing compared to the experiment without forecast elicitation. However, since the same incentive scheme is used in both 6H and 1H5C conditions, it does not influence the conclusion of Akiyama et al. (2017).

[11]Oechssler et al. (2009) report the negative correlation between CRT score and the incidences of the conjunction

Table 2: Distribution of CRT scores

	CRT score ≤ 1	CRT score $= 2$	CRT score $= 3$
1H5C (N=101)	27	26	48
6H (N=72)	25	26	21

have conducted CRT after the asset market experiment.

Experiments were conducted at the University of Tsukuba between May and July, 2013. A total of 173 students, both undergraduate and graduate, recruited from all over the campus participated to the experiment. Table 2 shows the distribution of CRT scores.[12]

We focus on the initial forecasts submitted by participants, and measure the deviation, from FV, of initial 10 forecasts submitted by participant i with the relative absolute forecast deviations, $RAFD_1^i$,

$$RAFD_1^i = \frac{1}{10} \frac{|f_{1,p}^i - FV_p|}{\overline{FV}} \tag{1}$$

where $f_{1,p}^i$ is the forecast for period p price submitted by participant i at the beginning of period 1, FV_p is the FV of the asset in period p, and $\overline{FV} = \frac{1}{10}\sum_t FV_t$.

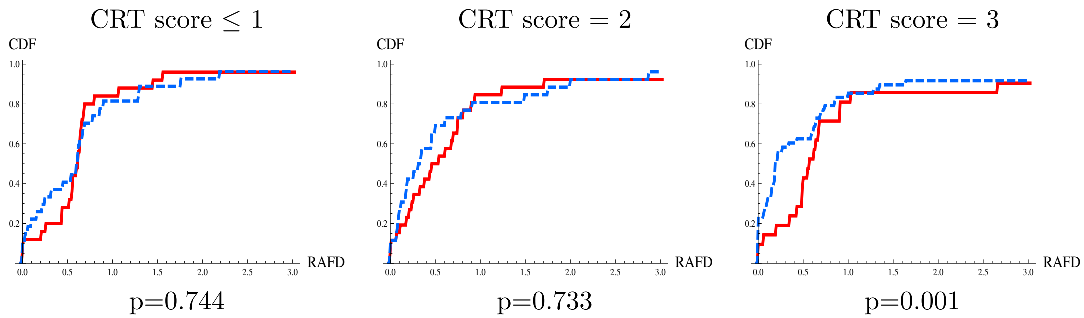
Figure 7 shows the distribution of $RAFD_1^i$ observed in 6H (red) and 1H5C (blue) depending on CRT score. For those subjects whose CRT scores are 0 or 1 (low CRT score), the two distributions are on top of each other. This means that their forecast deviations are a result of their individual bounded rationality. For those with a higher CRT score, the distribution from 1H5C lies to the left of 6H, although for those with CRT score = 2 (medium CRT score) the two distributions cross and, thus, are not statistically significantly different (p=0.733, Kolmogorov-Smirnov, KS, test, two-sided). For those with a perfect CRT score (high CRT score), the two distributions are significantly different (p=0.001, KS test). For these participants, the forecast deviations are results of both individual bounded rationality and behavioral uncertainty. This matches what we have observed for the very simple 2×2 coordination game in the previous section.

Furthermore, Figure 8, which compares the distribution of RAFD_1^i in 6H (left) and 1H5C (right)

fallacy and conservatism in updating probabilities. Brañas-Garza et al. (2012) show the negative correlation between the CRT score and the deviation of chosen number from the Nash equilibrium in beauty contest games. In the context of the asset market experiment similar to the one studied here, Corgnet et al. (2015) report that subjects with low CRT scores tend to buy (sell) an asset at prices above (below) FV while the opposite is true for those with high CRT scores.

[12]The average CRT score of these 173 participants is 2.1. This is very high, close to that of MIT, in light of the average scores from 8 universities reported in Frederick (2005). Thoma et al. (2015) report the average CRT score of traders, bankers, and non-financial professionals. The average score of traders are the highest among them and around 2.4.

Figure 7: Distribution of RAFD_1^i in 6H (red) and 1H5C (blue)



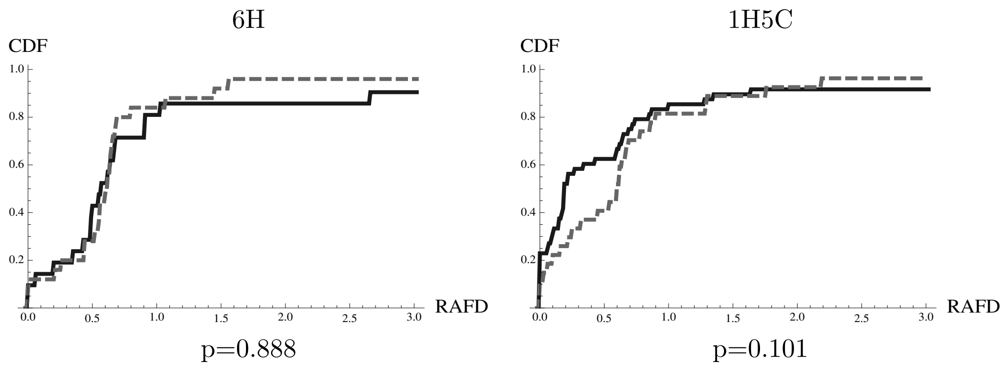
Source: Akiyama et al. (2017, Fig. 2)

between low CRT (dashed gray) and high CRT (solid black) participants, reveals another finding that is similar to that observed in the previous section. In 6H, the two distributions are on top of each other. This means that, when facing other human participants, forecasts of both low and high CRT participants deviate from FV at a similar magnitude. That is, the total effect of individual bounded rationality and behavioral uncertainty in causing the forecast deviations are the same between low and high CRT participants. If we had restricted our analyses to human only experiments, we would not have observed the differences in the outcome depending on the participants' cognitive ability. But for 1H5C, the distribution for low CRT participants lies to the right of the high CRT participants, although the two distributions are not statistically significantly different at the 5% significance level. This shows that high CRT participants, while they are less boundedly rational than low CRT participants, are impacted by the behavioral uncertainty to such a large extent that their forecast deviations become as large as those of low CRT participants in 6H.

4 Cognitive ability, cognitive hierarchy model, and the strategic environment effect

In the two experimental results presented above that address cognitive ability, individual bounded rationality, and behavioral uncertainty can be interpreted in terms of the cognitive hierarchy (Camerer et al., 2004) or the level-K (Nagel, 1995) model. These models assume that decision makers (agents) can be categorized into several levels. It is typically assumed that Level-0 agents either make a random choice or pick an intuitive choice following their impulsive drive. Level-1 agents, assuming

Figure 8: Distribution of RAFD_1^i for low and high CRT participants



Source: Akiyama et al. (2017, Fig. 3)

that others are Level-0, best (or, at least, better)[13] respond to the expected behavior of Level-0. Level-2 agents, in turn, assume that others are at lower levels (i.e., Level-0 or 1) and best (or better) respond to their expected behavior, etc.. In terms of the results presented above, we can say that those with low cognitive ability can be considered to be Level-0, and those with high cognitive ability to be Level-1 and above.[14]

This section presents a recently published study (Hanaki et al., 2019) that employs the framework of the cognitive hierarchy model to investigate macroeconomic implications of micro-level "boundedly rational" individual behavior. In particular, it shows a strategic environment in which micro-level individuals' boundedly rational behavior are not cancelled out through interactions among them and thus can have an important macro-level effect in the beauty contest games first experimentally studied by Sutan and Willinger (2009).

Let us first define the strategic environment effect (SEE). We say that "*the strategic environment effect arises if the expected absolute deviation of choices from the Nash equilibrium is larger when players' actions are strategic complements than when they are strategic substitutes.*"(Hanaki et al., 2019, Definition 1). The paper shows, both theoretically and experimentally, that the SEE arises in beauty contest games as long as the number of players, n, involved is not too small. While the theoretical analyses suggest that the SEE arises for $n > 2$, the SEE was observed experimentally for

[13] Here "better reply" means that an agent makes noisy choices in such a way that choice probabilities are positively correlated with the expected payoffs from each option. See, among others, Rogers et al. (2009) and Goeree and Holt (2004) for better-reply models.

[14] Whether these levels should be interpreted as types of decision makers so that the estimated level of a participant is consistent across various setting, or not is debated. For example, Heap et al. (2014) claim the estimated level should not be considered as a type because the estimates depend crucially on the assumption about Level-0 behavior. Here, we are not trying to estimate the levels from observed behavior.

$n > 4$.

Let us describe the beauty contest games (BCGs): n players simultaneously choose a number between 0 and 100. The player who has chosen the number closest to the target number (to be defined below) is the winner and obtain a prize. In case of a tie, one of them will be chosen randomly to receive the prize. We consider two BCGs, BCG+ and BCG−, that differ in the way the target number is defined. Let T^i_+ and T^i_- be the target number for player i in BCG+ and BCG−, respectively, and let x_j be the number chosen by player j.

$$T^i_+ = 20 + \frac{2}{3}\frac{\sum_{j\neq i} x_j}{n-1} \tag{2}$$
$$T^i_- = 100 - \frac{2}{3}\frac{\sum_{j\neq i} x_j}{n-1} \tag{3}$$

As one can observe, BCG+ is a game of strategic complementarity, and BCG− is a game of strategic substitutability. Note that if i believes others are going to choose higher numbers, then in BCG+, i also tries to choose a higher number to win, while in BCG−, i will try to choose a lower number. Both BCG+ and BCG− have the same Nash equilibrium (i.e., to choose 60), and the slope of the best reply functions are the same in absolute value between the two games.

Because theoretical analyses show that the SEE arises for $n > 2$, Hanaki et al. (2019) experimentally test $n \in \{2, 3, 4, 5, 6, 8, 16, \text{uncertain}\}$ to identify the minimum n above which the SEE is consistently observed.

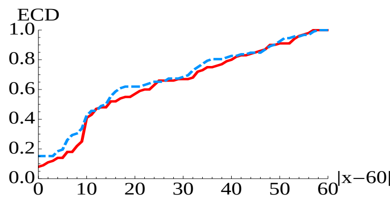
Figure 9 shows part of the experimental results reported in Hanaki et al. (2019). The four panels of Figure 9 show the distributions of the absolute deviation of the numbers chosen by participants from the Nash Equilibrium, $|x_i - 60|$, for four values of $n \in \{2, 5, 8, 16\}$. The distributions observed in BCG− are shown in red (solid) and those from BCG+ are shown in blue (dashed). For $n = 2$ (top left), the two distributions are on top of each other. Thus, the SEE is not observed as predicted by the theory. For the other three values of n, however, the distribution from BCG− lies significantly to the left of the distribution from BCG+, thus demonstrating SEE.

An intuition behind the theorem proved in Hanaki et al. (2019) is as follows. Let's assume the following version of the cognitive hierarchy model: Level-0s all choose 100.[15] Level-1s, assuming that all the others are Level-0s, best reply to the choice of Level-0s (i.e., 100). Level-Ks (K >1)

[15] As long the the absolute deviation of the number chosen by Level-0 from the Nash equilibrium is the same between BCG+ and BCG−, it does not matter for the theorem whether Level-0 tend to choose a number above, below or on the opposite side of the Nash equilibrium in the two BCGs.

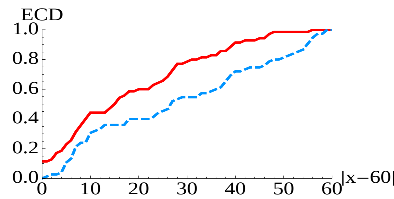
Figure 9: The empirical cumulative distribution $|x-60|$ for BCG_n+ (dashed) and BCG_n- (solid).

$n=2$



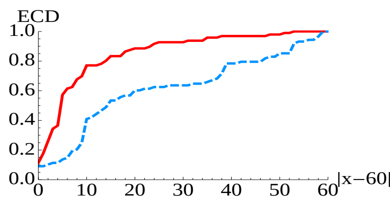
BCG_2-: 21.87 (17.42)
BCG_2+: 21.17 (17.64)
$p=0.506$

$n=5$



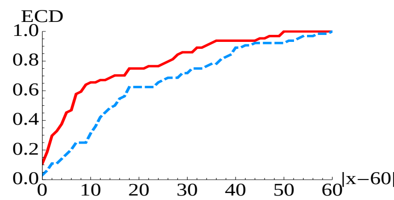
BCG_5-: 18.40 (14.90)
BCG_5+: 28.33 (18.79)
$p<0.001$

$n=8$



BCG_8-: 9.16 (11.15)
BCG_8+: 23.41 (19.10)
$p<0.001$

$n=16$



$BCG_{16}-$: 12.55 (14.00)
$BCG_{16}+$: 25.29 (17.10)
$p<0.001$

Note: Mean $|x-60|$ and its standard deviation (in the parentheses) is also reported. p-values are based on a two-sample permutation test (two-tailed) with the null hypothesis that $|x-60|$s are the same between BCG_n+ and BCG_n- for each n.

Source: Hanaki et al. (2019, Fig. 3).

and above assume that others are between Level-0 and Level-(K-1), with equal probability, and best reply to their expected choices.[16] The choices, as well as their absolute deviation from the Nash equilibrium for Level-1, 2, and 3 in two BCGs under this model, are summarized in Table 3.

Level-1s' average choice is above the Nash equilibrium (NE) in BCG+, while it is below the NE in BCG−. The absolute deviation of Level-1's choice from the NE, however, is the same between the two games. The difference between the absolute differences of choices and the NE appears between two BCGs for Level-2 and above. In particular, the absolute difference between the choices of Level-2 and above and the NE is much smaller under BCG− than under BCG+, which results in SEE.

The reason for this difference is simple. In BCG+, the choices made by Level-0 and 1 are both on the same side of the NE (in this particular case, above) due to the strategic complementarity. On the other hand, their choices are on the opposite sides of the NE in BCG− because of the strategic

[16]The assumption about the distribution among agents of lower levels are not very important here as long as it is not degenerate.

Table 3: Choices and their absolute deviation from the Nash equilibrium for Level-1, 2, and 3

		Level-1	Level-2	Level-3
BCG+	x_i	86.666	82.222	79.753
	$\lvert x_i - 60\rvert$	26.666	22.222	19.753
BCG-	x_i	33.333	55.555	58.025
	$\lvert x_i - 60\rvert$	26.666	4.445	3.025

Note: We assume that (a) Level-0s choose 100 in both games, (b) Level-K ($K > 0$) assumes that others are between Level-0 and Level-(K-1), with each level having the equal likelihood.

substistutability. Thus, the average number chosen by Level-0 and -1 is necessarily closer to the NE under BCG$-$ than under BCG+. And since Level-2s best respond to the average number chosen by Level-0 and 1, their choices will also be closer to the NE under BCG$-$ than BCG+. The same reasoning applies to all the higher levels.

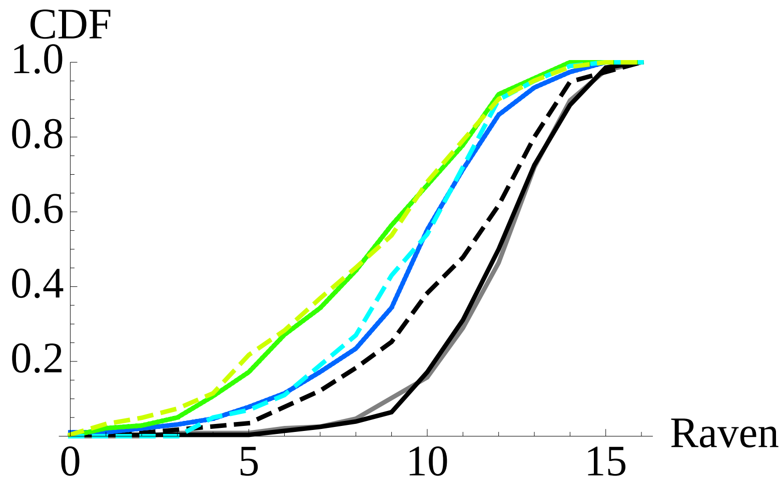
What are the environments characterized by the strategic complementarity? One can easily think of such examples as asset markets, various coordination problems, and maybe the economy as a whole as it is driven by expectations of various decision makers. Because people are heterogeneous in terms of their cognitive ability, and thus, their reasoning processes, we believe we should start seriously investigating the various macro-implications of interactions among heterogeneous boundedly rational decision makers, especially, for those environments characterized by strategic complementary.

5 Differences in cognitive ability measured across several laboratories.

Let me now turn to a discussion regarding large differences in the measured cognitive ability of participants across several experimental laboratories, and the potential problems associated with not paying sufficient attention to such differences for conducting and interpreting results of experimental analyses.

Figure 10 shows the cumulative distributions of RPM scores from seven laboratories where I have gathered data so far. These laboratories are all located at a university or an institution of the equivalent level. In all the laboratories, participants were asked to answer the same set of 16 questions within 10 minutes. The timings in which the RPM test were conducted differ across locations, however. In some cases, it was at the end of the experiment, while in other cases, it was

Figure 10: Distributions of RPM scores from seven experimental laboratories



The number of observations for each distribution: 280 (black solid), 235 (gray solid), 115 (black dashed), 192 (blue solid), 100 (light blue dashed), 140 (green solid), and 244 (green dashed).

at the beginning of the experiment. Because it is possible that participants who took the RPM test at the end of the experiment did worse than those who did it at the beginning, please interpret the score with a grain of salt. However, even considering such a possibility, the differences in the distributions of RPM scores across these seven locations are striking. On one hand, in the laboratory with the left-most distribution of RPM score, less than 30% of participants scored *more* than 11 in the RPM test. In the laboratory with the right-most distribution of RPM score, on the other hand, less than 30% of participants scored *less* than 11. Furthermore, as previously noted, the CRT scores are correlated with RPM score. Let's take the two extreme distributions of RPM score. The average CRT score of the participants in the laboratory with the left-most RPM score distribution is less than 1, while it is more than 2 in the laboratory with the right-most RPM score distribution.

Such a large difference in the cognitive ability of participants across laboratories raises two major concerns: one is the replicability of experimental findings, and another is the interpretation of cross country differences in the experimental findings. Let us first discuss the initial concern by taking the recent experimental findings regarding the cooperation in an indefinite repeated game and mispricing in the asset market experiment of Smith et al. (1988).

Table 4: PD game considered by Proto et al. (2019)

	C	D
C	48,48	12,50
D	50,12	25, 25

5.1 Replicability of experimental findings

Proto et al. (2019) investigate the relationships between the cognitive ability of pairs of participants and their behavior in indefinitely repeated games. They do so by first measuring participants' cognitive ability using an RPM test,[17] and split them in half based on participants' relative RPM scores within their group: High (top half) and Low (bottom half).[18] Then, three types of pairs are created: High-High, Low-Low, and random. Once a play of the repeated game ends, pairs are randomly re-created, respecting their types, and a new play of the same repeated game starts. This process is repeated until the 45-minute time limit is reached.

They consider several payoff matrices and two continuation probabilities, but here we focus on the Prisoners' dilemma game with the payoff matrix shown in Table 4 with the continuation probability $\delta = 0.75$ that was shown to result in high cooperation by Dal Bó and Fréchette (2011).

Figure 11 shows the average frequency of cooperation in a block of 5 plays of the game, averaged across super-games and groups, for High-High and Low-Low pairs (left panel) and High, Low, and average in random pairs (right). The left panel clearly shows that those in High-High pairs managed to achieve and sustain high cooperation, while those in Low-Low pairs fail to do so, although the fractions of cooperative play are initially the same between the two types of pairs because participants do not know the way they are paired. The right panel shows also that if matched with a High type, Low type participants can also learn to cooperate more, although Low types cooperate less than High types.

The difference in the rate of cooperation between High-High and Low-Low pairs suggest that the results of experiment of the same repeated game will be quite different depending on the laboratory. Namely, cooperation will be much less likely to be observed in those laboratories where most of participants have low RPM scores.

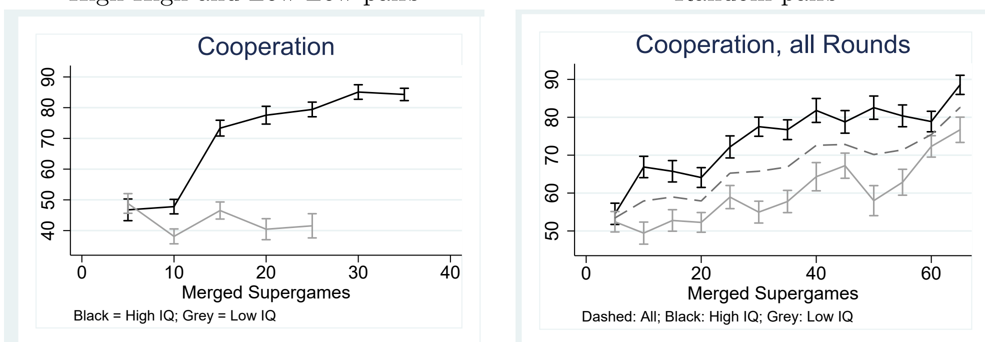
Another example where experimental results differ drastically depending on the participants'

[17] 30 questions, with a strict time limit of 30 seconds per question.

[18] Note that such within-session grouping is not perfect in that participants with the same score can be categorized as High or Low depending on the session.

Figure 11: Result of repeated PD games depending on cognitive ability of pairs by Proto et al. (2019)

High-High and Low-Low pairs | Random pairs



Source: Proto et al. (2019, Fig.2) | Source: Proto et al. (2019, Fig.3)

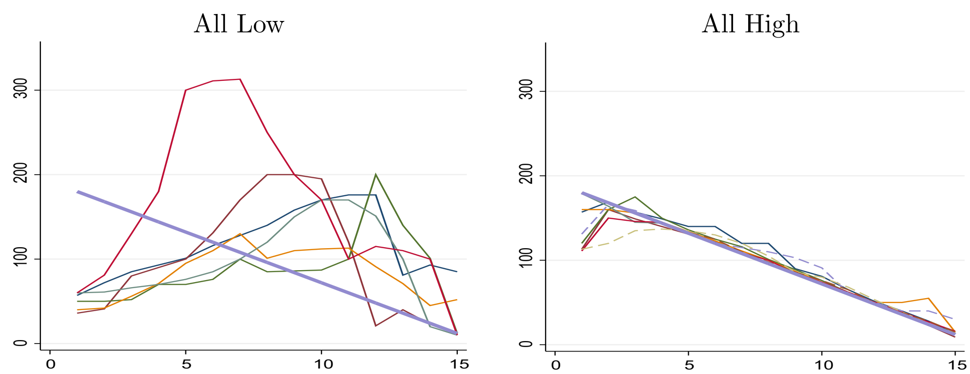
cognitive ability is the asset market experiment we have seen above. Bosch-Rosa et al. (2018) measured the cognitive ability of participants by CRT, the beauty contest game, and hit-60 games. According to the composite score computed based on these measures, they divided the participants into three groups, Low (bottom 1/3), High (top 1/3), and Medium (remaining), and conducted the asset market experiment of Smith et al. (1988) by recruiting participants either only from the Low group (All Low) or the High group (All High). Figure 12 shows the price dynamics observed in All Low markets (left) and All High markets (right). It is clear that a large mis-pricing is observed only in All Low markets. In All High markets, prices follow the FV of the asset very closely. This again clearly suggests the dependence of experimental results and its replicability on the pool of participants.

5.2 Cross-cultural differences in experimental findings

Now, let's turn to the interpretation of the results of cross cultural experimental comparisons. One of the most famous cross-cultural experimental works is that of Herrmann et al. (2008) and Gächter et al. (2010), which considers public good games (PGGs) with and without costly punishment. These authors have conducted the same experiment in 16 cities from 15 countries around the world covering 6 cultural areas (English speaking, Protestant Europe, Orthodox/ex-Communist, Confucian, Southern Europe, and Arabic speaking).[19] In order to minimize the differences in the participants

[19] Please refer to Herrmann et al. (2008) and Gächter et al. (2010) and citations therein for this grouping of locations into different cultural areas.

Figure 12: Results of asset market experiment based on cognitive ability groups by Bosch-Rosa et al. (2018)



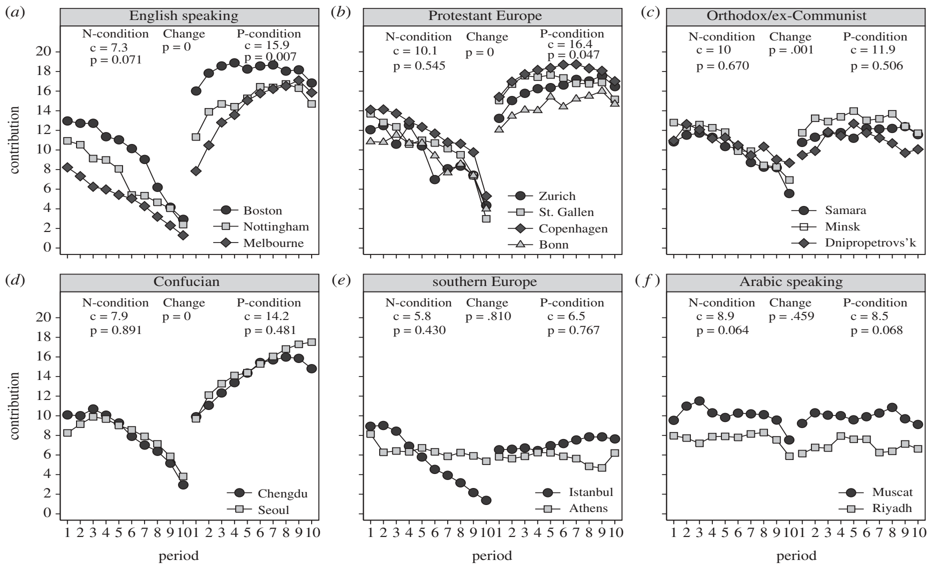
Source: Bosch-Rosa et al. (2018, Fig.1)

across these locations, the authors have recruited university students in these cities. But as we have already demonstrated, large differences in participants' cognitive ability exist across universities, and it is not clear if the participants from universities in these 16 locations had a similar cognitive ability.

The PGG without costly punishment is as follows: a group of 4 players play the same game for 10 times. In each play of the game, each participant receives 20 points which they can keep or invest to a public project. Let's call x_t^i the amount participant i invests in the public project in t-th play of the game (period t). The payoff for i in period t is $\pi_t^i = 20 - x_t^i + 0.5 \sum_j x_t^j$. After each play of the game, participants are informed of the payoff they have obtained in that period.

The PGG with costly punishment adds a punishment stage to the PGG without costly punishment (described above). Namely, after participants make their investment decisions, they are informed of the amount invested by each member of their group. After observing this information, each participant can decide to reduce the payoff obtained by other members of the group by paying a price. Namely, by paying 1 point, a participant can reduce the payoff of another member of the group by 3 points, and the maximum number of points that one can reduce from another is 30. Let $p_{j,t}^i$ be the points i pays to reduce j's payoff in period t. This "punishment" decision is done simultaneously without knowing others' decisions. The final payoff for subject i will be $\max[0, \pi_t^i - 3\sum_j p_{i,t}^j] - \sum_k p_{k,t}^i$. Note that, the minimum point subject i can obtain after the punishment, if any, is zero minus the total cost of the punishment i has decided to pay to "punish"

Figure 13: Average amount of contribution to public goods in 16 locations reported by Gächter et al. (2010).



Source: Gächter et al. (2010, Fig.1)

others.

The experiments were conducted so that the same group of 4 participants experienced 10 periods of PGG without punishment first, and then 10 period of PGG with punishment. Figure 13 shows the average amount invested in public projects in 16 locations. The results are shown separately for each cultural area. In four out of six cultural areas, (a) English speaking, (b) Protestant Europe, (c) Orthodox/ex-Communist, and (d) Confucian, the possibility of punishing others by paying a cost does help to increase the amount contributed to the public project. While the contributions to the public project tend to decline as participants repeatedly play the PGG without punishment, the possibility of punishment prevents it from declining. There are three locations, Athens (in (e) Southern Europe), Muscat and Riyadh (in (f) Arabic speaking), that demonstrate different outcomes. In these three locations, the average contributions do not decline with repetition in PGG without punishment, and the possibility of punishment does not change the average amount of contribution at all. Herrmann et al. (2008) shows that, in these three locations, participants have punished those who have contributed more than themselves, and considers that such "anti-social punishments" prevented the punishment possibility from increasing the amount of contribution compared to the case without such a possibility. As the title of Gächter et al. (2010) "Cultural and Cooperation" suggests, the authors seem to interpret this difference as a result of cultural difference, but there are other potential differences across subjects pools. Thus, without conducting the experiment with participants of different cognitive ability (to see the effect of cognitive ability in these games) as well as cross culturally while controlling for participants cognitive ability (to see the effect of cultural differences), we should be cautious in making such inferences about the effect of cultural differences on observed behavior.[20]

6 Summary and conclusion

Participants in laboratory experiments differ in terms of their cognitive ability and other characteristics. This is also the case for the wider population. In this paper, we have demonstrated how a similar behavior can emerge for different reasons depending on the decision makers' cognitive ability. We have also shown how explicitly considering such differences in our model would be important in understanding macroeconomic phenomena in environments characterized by strategic complementarity.

We have also argued for the importance of understanding the characteristics of the participant pool in various experimental laboratories both for replicating experimental results and for better interpretations of the differences in observed behavior across countries and cultural zones. For this purpose, databases of participants should be complemented with information about participants' characteristics including cognitive ability. As a first step, we, together with Keigo Inukai and Takehito Masuda, have started to complement a participant database at ISER, Osaka University, with various individual characteristics, such as cognitive ability, personality traits, risk preference, and theory of mind. The effort is not yet complete, but once it is done, we will be able to recruit participants based on their characteristics and conduct experiments to better understand the behavior among groups of participants with various characteristics. We very much hope that other experimental groups around the world will start doing the same.

[20] Another approach to investigate the cultural differences is to conduct a larger scale "on-line" experiment with a representative sample of population as participants. See, for example, OECD trust-lab project `https://www.oecd.org/sdd/trustlab.htm`. I thank Masao Ogaki for pointing me out about this OECD initiative.

References

Akiyama, E., N. Hanaki, and R. Ishikawa (2014): "How do experienced traders respond to inflows of inexperienced traders? An experimental analysis," *Journal of Economic Dynamics and Control*, 45, 1–18.

——— (2017): "It is not just confusion! Strategic uncertainty in an experimental asset market," *Economic Journal*, 127, F563–F580.

Ariely, D. (2008): *Predictably Irrational: The Hidden Forces That Shape Our Decisions*, New York, NY: HarperCollins.

Bao, T., C. Hommes, J. Sonnemans, and J. Tuinstra (2012): "Individual expectations, limited rationality and aggregate outcomes," *Journal of Economic Dynamics and Control*, 36, 1101–1120.

Becker, G. S. (1962): "Irrational Behavior and Economic Theory," *Journal of Political Economy*, 70, 1–13.

Bosch-Rosa, C., T. Meissner, and A. Bosch-Domènech (2018): "Cognitive Bubbles," *Experimental Economics*, 21, 132–153, doi:10.1007/s10683-017-9529-0.

Brañas-Garza, P., T. García-Muñoz, and R. Hernán (2012): "Cognitive effort in the Beauty Contest Game," *Journal of Economic Behavior and Organization*, 83, 254–260.

Brouwers, S. A., F. J. V. de Vijver', and D. A. V. Hemert' (2009): "Variation in Raven's Progressive Matrices scores across time and place," *Learning and Individual Differences*, 19, 330–338.

Camerer, C. F. (2003): *Behavioral Game Theory: Experiments in Strategic Interaction*, New York: Russell Sage Foundation.

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

Carpenter, J., M. Graham, and J. Wolf (2013): "Cognitive ability and strategic sophistication," *Games and Economic Behavior*, 80, 115–130.

Cheung, S. L., M. Hedegaard, and S. Palan (2014): "To See is To Believe: Common Expectations in Experimental Asset Markets," *European Economic Review*, 66, 84–96.

Conlisk, J. (1996): "Why bounded rationality?" *Journal of Economic Literature*, 34, 669–700.

Cooper, K. B., H. S. Schneider, and M. Waldman (2017): "Limited rationality and the strategic environment: Further theory and experimental evidence," *Games and Economic Behavior*, 106, 188–208.

Corgnet, B., M. DeSantis, and D. Porter (2018): "What makes a good trader? On the role of intuition and reflection on trader performance," *Journal of Finance*, 73, 1113–1137, economic Science Institute, Chapman University.

Corgnet, B., R. H. Gonzalez, P. Kujal, and D. Porter (2015): "The Effect of Earned vs. House Money on Price Bubble Formation in Experimental Asset Markets," *Review of Finance*, 19, 1455–1488.

Dal Bó, P. and G. R. Fréchette (2011): "The Evolution of Cooperation in Infinitely Repeated Games: Experimental Evidence." *American Economic Review*, 101, 411–429.

Fehr, E. and J.-R. Tyran (2008): "Limited Rationality and Strategic Interaction: The Impact of the strategic environment on nominal inertia," *Econometrica*, 76, 353–394.

Frederick, S. (2005): "Cognitive reflection and decision making," *Journal of Economic Perspectives*, 19, 25–42.

Gächter, S., B. Herrmann, and C. Thöni (2010): "Culture and Cooperation," *Philosophical Transactions of The Royal Society B*, 365, 2651–2661.

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.

Gode, D. K. and S. Sunder (1993): "Allocative Efficiency of Markets with Zero-Intelligence Traders: Market as a Partial Substitute for Individual Rationality," *Journal of Political Economy*, 101, 119–137.

——— (1997): "What makes markets allocationally efficient?" *Quarterly Journal of Economics*, 112, 603–630.

Goeree, J. K. and C. A. Holt (2004): "A model of noisy introspection," *Games and Economic Behavior*, 46, 365–382.

HALTIWANGER, J. AND M. WALDMAN (1985): "Rational Expectations and the Limits of Rationality: An Analysis of Heterogeneity," *American Economics Review*, 75, 326–340.

——— (1989): "Limited Rationality and Strategic Complements: The implications for macroeconomics," *Quarterly Journal of Economics*, 104, 463–484.

——— (1991): "Responders versus non-responders: A new perspective on heterogeneity," *Economic Journal*, 101, 1085–1102.

HANAKI, N., E. AKIYAMA, AND R. ISHIKAWA (2018): "Effects of different ways of incentivizing price forecasts on market dynamics and individual decisions in asset market experiments," *Journal of Economic Dynamics and Control*, 88, 51–69.

HANAKI, N., N. JACQUEMET, S. LUCHINI, AND A. ZYLBERSZTEJN (2016): "Cognitive ability and the effect of strategic uncertainty," *Theory and Decision*, 81, 101–121.

HANAKI, N., Y. KORIYAMA, A. SUTAN, AND M. WILLINGER (2019): "The strategic environment effect in beauty contest games," *Games and Economic Behavior*, 113, 587–610.

HARUVY, E., Y. LAHAV, AND C. N. NOUSSAIR (2007): "Traders' Expectations in Asset Markets: Experimental Evidence," *American Economics Review*, 97, 1901–1920.

HEAP, S. H., D. R. ARJONA, AND R. SUGDEN (2014): "How portable is level-0 behavior? A test of level-k theory in games with non-neutral frames," *Econometrica*, 82, 1133–1151.

HEEMEIJER, P., C. HOMMES, J. SONNEMANS, AND J. TUINSTRA (2009): "Price stability and volatility in markets with positive and negative expectations feedback: An experimental investigation," *Journal of Economic Dynamics and Control*, 33, 1052–1072.

HERRMANN, B., C. THÖNI, AND S. GÄCHTER (2008): "Antisocial Punishment Across Societies," *Science*, 319, 1362–1367.

HUSSAM, R. N., D. PORTER, AND V. L. SMITH (2008): "Thar She Blows: Can Bubbles Be Rekindled with Experienced Subjects?" *American Economic Review*, 98, 927–934.

KAHNEMAN, D. (2011): *Thinking, Fast and Slow*, New York, NY: Farrar, Straus and Giroux.

KIRCHLER, M., J. HUBER, AND T. STÖCKL (2012): "Thar She Bursts: Reducing Confusion Reduces Bubbles," *American Economic Review*, 102, 865–883.

Lei, V., C. N. Noussair, and C. R. Plott (2001): "Nonspeculative Bubbles in Experimental Asset Markets: Lack of Common Knowledge of Rationality vs. Actual Irrationality," *Econometrica*, 69, 831–859.

Mullainathan, S. and E. Shafir (2013): *Scarcity: Why Having Too Little Means So Much*, New York, NY: Times Books, Henry Holt and Company, LLC.

Nagel, R. (1995): "Unraveling in Guessing Games: An Experimental Study," *American Economics Review*, 85, 1313–1326.

Oechssler, J., A. Roider, and P. W. Schmitz (2009): "Cognitive abilities and behavioral biases," *Journal of Economic Behavior and Organization*, 72, 147–152.

Palan, S. (2013): "A Review of bubbles and crashes in experimental asset markets," *Journal of Economic Surveys*, 27, 570–588.

Potters, J. and S. Suetens (2009): "Cooperation in experimental games of strategic complements and subsitutes," *Review of Economic Studies*, 76, 1125–1147.

Proto, E., A. Rustichini, and A. Sofianos (2019): "Intelligence, Personality and Gains from Cooperation in Repeated Interactions," *Journal of Political Economy*, Forthcoming.

Raven, J. (2008): "General introduction and overview: the raven progressive matrices tests: their theoretical basis and measurement model," in *Uses and abuses of intelligence*, ed. by John and J. Raven, Edinburgh, Scotland: Competency Motivation Project, chap. 1, 17–68.

Raven, J. C. (1998): *Raven's Advanced Progressive Matrices (APM)*, San Antonio, TX: Pearson, 2003 ed.

Rogers, B. W., T. R. Palfrey, and C. F. Camerer (2009): "Heterogeneous quantal response equilibrium and cognitive hierarchies," *Journal of Economic Theory*, 144, 1440–1467.

Shleifer, A. (2000): *Inefficient Markets: An introduction to behavioral finance*, New York, NY: Oxford University Press.

Smith, V. L., G. L. Suchanek, and A. W. Williams (1988): "Bubbles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets," *Econometrica*, 56, 1119–1151.

Spiegler, R. (2011): *Bounded Rationality and Industrial Organization*, New York, NY: Oxford University Press.

Sutan, A. and M. Willinger (2009): "Guessing with negative feedback: An experiment," *Journal of Economic Dynamics and Control*, 33, 1123–1133.

Thaler, R. H. (2015): *Misbehaving: The Making of Behavioral Economics*, New York, NY: W. W. Norton & Company.

The International Cognitive Ability Resource Team (2014) `https://icar-project.com/`.

Thoma, V., E. White, A. Panigrahi, V. Strowger, and I. Anderson (2015): "Good Thinking or Gut Feeling? CognitiveReflection and Intuition in Traders, Bankersand Financial Non-Experts," *Plos One*, 10, e0123202. doi:10.1371/journal.pone.0123202.