Diversity in cognitive ability and mispricing in experimental asset markets*

Nobuyuki Hanaki† Eizo Akiyama‡ Yukihiko Funaki§ Ryuichiro Ishikawa¶

January 17, 2017

Abstract

Does diversity in cognitive ability among market participants enlarges mispricing? Does the common knowledge of heterogeneity among the cognitive ability of market participants further enlarges the mispricing? We investigated these questions by first measuring subjects' cognitive ability and categorizing them as 'H' type for those above median ability and 'L' type for those below median ability. We then constructed three kinds of markets with six traders each: 6H, 6L, and 3H3L. Subjects were informed of their own cognitive type and, depending on the treatment, also that of the others in their market. We found heterogeneous markets (3H3L) generated significantly larger mispricing than homogeneous markets (6H or 6L) regardless of whether subjects were informed about the cognitive type of others in the market. Thus, diversity in cognitive ability among market participants enlarged mispricing. However, common knowledge of heterogeneity or homogeneity did not have significant additional effects.

Keywords: Cognitive ability, Heterogeneity, Mispricing, Experimental asset markets

JEL Code: C90, D84

*Misaki Miyaji, Ayano Nakagawa, and Makoto Soga provided invaluable help in organizing the experiments. Comments and suggestions from Te Bao, Jürgen Huber, Daniel Kleinlercher, Fabrice Le Lec, Charles Noussair, Michael Razen, Eko Riyanto, Thomas Stöckl, and seminar participants at CREM, Edinburgh, Glasgow, GRIPS, GREDEG, Hitotsubashi, Hokkaido, Innsbruck, Keio, Osaka, Waseda, Ritsumeikan, Nanyang Technological University, as well as participants at SEF 2015 (Nijmegen), CEF 2016 (Bordeaux), FUR 2016 (Warwick), SEF North America 2016 (Tuscan) are gratefully acknowledged. Jeremy Mercer has provided an excellent proof-reading. This project is partly financed by a JSPS-ANR bilateral research grant "BECOA" (ANR-11-FRJA-0002), ORA-Plus research project "BEAM" (ANR-15-ORAR-0004), financial aid from the Japan Center for Economic Research, and financial aid from the Japan Securities Scholarship Foundation. A part of this work has been conducted when Hanaki was affiliated with Aix-Marseille Université (Aix-Marseille School of Economics). Hanaki thanks Aix-Marseille School of Economics for their support. The experiments reported in this paper have been approved by Institutional Review Board of Faculty of Engineering, Information and Systems, University of Tsukuba (No. 2012R25).

†Université Côte d'Azur, CNRS, GREDEG. Corresponding author. GREDEG, 250 rue Albert Einstein, 06560 Valbonne, France. E-mail: Nobuyuki.HANAKI@unice.fr

‡Faculty of Engineering, Information and Systems, University of Tsukuba. E-mail: eizo@sk.tsukuba.ac.jp

§School of Political Science and Economics, Waseda University. E-mail: funaki@waseda.jp

¶School of International Liberal Studies, Waseda University. E-mail: r.ishikawa@waseda.jp

1 Introduction

Historical episodes of economic bubbles are often characterized by euphoria and an inflow of new and possibly naïve investors (Kindleberger and Aliber, 2005). Indeed, Lopez (2015) reports that a non-negligible fraction of new investors in the China's stock market can't read, and for a majority of these new investors, the junior high school is the highest level of education completed. Such an inflow of new investors can amplify the heterogeneity among market participants regarding their belief about future prices of the asset being traded as well as their naivity in terms of financial knowledge and trading behavior, and thus may inflate the mispricing as noted by Xiong and Yu (2011) in their study of "bubbles" in a subset of China's warrant market.

Furthermore, several theoretical works (Allen and Gale, 1992; Aggarwal and Wu, 2006; Allen et al., 2006) build upon such an idea and show how heterogeneities in terms of (strategic) sophistication among traders can lead to large mispricing. Allen and Gale (1992), for example, show that sophisticated strategic traders try to generate an initial upward price trend to influence the belief of naïve trend follows in order to later profit from their naïveté.

Recent experimental studies have demonstrated the relationship between the cognitive abilities of subjects and the mispricing in asset market experiments *à la* Smith et al. (1988).[1] Breaban and Noussair (2015) and Cueva and Rustichini (2015) show that the average cognitive skills for subjects in market is negatively correlated with the magnitude of observed mispricing in the market. Cognitive skills are measured by the Cognitive Reflection Test (CRT, Frederick, 2005) in the former and by Raven's progressive matrix test (see Raven, 2008, for an overview) and Race to X game (Gneezy et al., 2010; Dufwenberg et al., 2010) in the latter. Corgnet et al. (2015) and Cueva and Rustichini (2015) demonstrate that subjects with higher cognitive skill earned more than their lower cognitive skill counterparts.

These experimental findings are in line with findings from empirical studies based on large-scale surveys that tend to report that people with high cognitive skills make better financial decisions (see, for example, Korniotis and Kumar, 2010, for a survey of the empirical literature).

[1] See Palan (2013) and Powell and Shestakova (2016) for recent surveys of the literature. However, the body of literature is expanding very quickly with many new papers being presented each year at the annual meeting of the Society of Experimental Finance. See `http://www.experimentalfinance.org/` for a list of papers presented at recent meetings.

However, the effect of interactions among traders with varying degrees of strategic sophistication has not been explicitly investigated very much either empirically or experimentally. The above-mentioned experimental studies relate cognitive skills and market outcomes only ex-post, and thus do not use cognitive skills as an experimental variable.[2] An exception is Bosch-Rosa et al. (2015). They investigate how the average cognitive ability among market participants influence mispricing in an experimental market by creating markets based explicitly on the subjects' cognitive abilities that were measured ex-ante. Bosch-Rosa et al. (2015) first conduct an experimental session consisting of the CRT, guessing games, and multiple rounds of the Race to 60 game to measure and to create a composite index of cognitive abilities of their subjects. Then, they later re-recruit subjects from the top 30% ("high sophistication") or from the bottom 30% ("low sophistication") of their subject pool according to the index, and conduct an asset market experiment with markets consisting only of high-sophistication subjects or those consisting only of low-sophistication subjects. They report a large mispricing in the markets consisting only of low-sophistication subjects but almost no mispricing in those consisting only of high-sophistication subjects.

These experimental and empirical studies led us to speculate that the mispricing observed in experimental and real financial markets is primarily due to bad decisions made by naïve market participants, and not so much by interactions (both strategic and non-strategic) among traders of varying degrees of cognitive sophistication. Yet, two recent experimental studies, Cheung et al. (2014) and Akiyama et al. (2015), convincingly demonstrate that this may not be the whole story.

Cheung et al. (2014) investigate the effect of the lack of the common knowledge about everyone's understanding of the fundamental value (FV) of the asset being traded. They

[2]There are increasing number of experimental studies on games and individual decisions that use cognitive skills as an experimental variables. For example, in game theoretic settings, Gill and Prowse (2016) studies 3-player p-beauty contest game by creating three types of groups based on the subjects' relative cognitive ability: all high cognitive ability group, all low cognitive ability group, and mixed groups. They find that subjects with higher cognitive ability are faster in learning to choose numbers close to the Nash equilibrium and earn more. They also find that those with higher cognitive ability respond to the cognitive ability of their counter parts, while those with low cognitive ability do not. Another recent game theoretic work is Proto et al. (2016). They study the evolution of cooperation in repeated games varying the cognitive ability of groups. They find that, although initial level of cooperation are similar, those groups of high cognitive ability subjects learn to achieve high or full cooperation, while cooperation decline in those groups of low cognitive ability. For the studies on individual decisions, **?**, in their study with German representative sample, find that subjects with higher cognitive ability take more calculated risk and more patient. **?** study the relationships between the score of CRT and various behavioral biases such as conjunction fallacy, conservatism in updating probabilities as well as time and risk preferences. They found a similar relationship between the CRT score and the risk and time preferences as **?**, as well as a negative correlation between the CRT score and the incidence of the two biases.

train (some of) their subjects extensively about the FV of the asset before the experiment, and compare the magnitude of mispricing in the three types of markets: (1) everyone is trained and everyone knows about it, (2) everyone is trained but they don't know about it, and (3) none is trained. Their results show that the magnitude of mispricing is small only when everyone is trained and that is common knowledge. When it is not common knowledge, the mispricing is as large as the markets where none is trained.

Akiyama et al. (2015) investigate how the presence of uncertainty about behavior of others in the market influences the (long-run) price forecasts by comparing the price forecasts in the two market environments: one in which one subject interacts with computer traders with known behavior, and another in which subjects interact among themselves. They find that subjects' initial long-run forecasts deviate more from the FV in the former case than in the latter case. Furthermore, subjects with the perfect CRT score reacted more strongly to the presence of uncertainty about behavior of others in the market (where they interact with other subjects) than those with lower CRT scores by forecasting future prices to deviate more from the FV compared to the market without such uncertainty (where they interact with computer with known behavior).

Because diversity (or heterogeneity) in cognitive ability among market participants can be an important source of heterogeneous belief about future prices, as well as behavioral uncertainty in case the heterogeneity is commonly known, these two experimental studies hint at the possibility that such diversity can indeed amplify the mispricing of the asset being traded. Therefore, in this paper, we investigate *whether the heterogeneity in cognitive ability among market participants enlarges the mispricing.* In addressing this question, we also investigate the relationship between the average cognitive ability of market participants and degree of mispricing. Furthermore, we ask *whether common knowledge of heterogeneity (or homogeneity) among the cognitive ability of market participants further enlarges the mispricing.* The last question is motivated by the theoretical literature on strategic manipulation we have cited above. We conjecture that knowing the presence of naive traders in the market create a room for more sophisticated traders to manipulate the prices, and thus enlarges the mispricing.

We approach these research questions by first measuring subjects' cognitive ability by employing a part of the advanced version of Raven's progressive matrix (RPM) test,[3] and grouping subjects based on their relative RPM test scores within an experimental session. Namely, those subjects with above the median RPM score in the session are called H type and others are called L types. We consider three types of markets: those consist of only H types, those consist of only L types, and those consist of an equal number of H and L types. By comparing the outcomes of these three types of market, we investigate the influences not only of the average cognitive ability among market participants, but also of their diversity on market outcomes. Furthermore, we investigate the impact of subjects being informed about the composition of market participants to the market outcomes by conducting experiments with and without informing our subjects about the type composition of the market they participate. Our main focus is to investigate the effect of heterogeneity of cognitive ability among market participants by creating market that mix both high and low sophistication subjects, which Bosch-Rosa et al. (2015) do not address.

We found that heterogenous markets, i.e., those markets consisting of the equal number of H and L type subjects resulted in larger mispricing than two homogeneous markets, i.e., those consisting of only H or L type subjects, regardless of whether the composition of types within markets is ex-ante known or not. Our results show, therefore, that not only average cognitive ability of market participants but also their heterogeneity, regardless of whether the heterogeneity is ex-ante known or not, enlarges mispricing. We do not, however, find significant difference in terms of mispricing between treatments with and without subjects being ex-ante informed about the type compositions within markets. Thus, at least in our experimental setup, we did not observe significant additional effect of common knowledge of cognitive heterogeneity on mispricing beyond the effect of the existence of cognitive heterogeneity.

[3]The RPM test measures what is called "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 its score has shown to be correlated with a degree of strategic sophistication which is measured in terms of the number of wins in the Race to 5, 10, and 15 game (Carpenter et al., 2013), or the deviation from the equilibrium in a three-player Beauty Contest game (Gill and Prowse, 2016). "Fluid intelligence" should be distinguished from what is called "executive control." The latter is an ability to control one's impulsive behavior or response. The CRT, from this point of view, can be interpreted as a measure of one's executive control and not their fluid intelligence.

2 Experiment

In each session of 24 subjects, we first ask subjects to answer a part of Raven's advanced progressive matrix test (24 questions to be answered in 15 minutes).[4] We do not inform our subjects why we ask them to answer the RPM test (which we called a quiz during the experiment), nor what kind of experiments will follow after completing the test. Thus, when our subjects are answering the RPM test, they are not aware that their relative scores will be later used to place them into different groups in an asset market experiment. Following the standard practice in administrating the RPM test, we do not offer monetary incentives to our subjects for answering as many questions correctly as possible.

Once the RPM ends, we divide our subjects into two types based on their relative scores on the RPM test. We call those above the median score 'H type' and those below the median score 'L type'. We then create three versions of a 20-period call asset market with six traders: in version one, all six traders are H type (6H markets); in version two, all the six traders are L type (6L markets); and in version three, there is an equal number of H and L types (3H3L markets). In one experimental session of 24 subjects, we create two 6H markets and two 6L markets. In another session, we create four 3H3L markets.[5] In all our treatments, we inform our subjects of their own type (H or L), but not how many questions of the quiz (out of 24) they answered correctly.

In order to separately investigate the effect of (1) the compositions of the market participants (in terms of their relative cognitive ability), and (2) the fact that such compositions are common information to the market participants, we consider two information treatments. In a half of our treatments, we do not inform our subjects of the composition of the six traders in the same market (unknown composition), while in another half, we inform them of the composition (known composition). In unknown composition treatment, therefore, subjects are informed only of their own relative type, H or L, but not the composition of

[4] The full advanced RPM test consists of 48 questions to be answered in 30-40 minutes. We used all the odd-numbered questions from the full test in the original ordering to keep the questions becoming progressively more difficult.

[5] Groups are created according to the rankings on the RPM test of the participants in that session. In the 24-subject 6H and 6L sessions, the first 6H market consist of subjects with rankings of $\{1, 3, 5, 7, 9, 11\}$ and the second consists of subjects with rankings of $\{2, 4, 6, 8, 10, 12\}$; for the two 6L markets, the first consists of subjects with rankings of $\{13, 15, 17, 19, 21, 23\}$ and the second of subjects with rankings of $\{14, 16, 18, 20, 22, 24\}$. For the 3H3L markets, we have four markets that consist of those subjects with rankings of $\{1, 5, 9, 13, 17, 21\}$, $\{2, 6, 10, 14, 18, 22\}$, $\{3, 7, 11, 15, 19, 23\}$, and $\{4, 8, 12, 16, 20, 24\}$, respectively. In cases where subjects have identical scores, rankings are assigned randomly.

five other traders' types in their market.[6] In known composition treatment, on the other hand, if an H-type subject is in a 6H market, s/he is informed that s/he is H type and all the other five traders in the market are also H type. If an H-type subject is in a 3H3L market, s/he is informed s/he is H type and that the other five traders consist of two H-type traders and three L-type traders. Similarly, if an L-type subject is in a 6L market, s/he is informed that s/he is L type and all the other five traders in the market are also L type.[7]

In all the markets, traders are initially given 4 units of the asset and 1040 experimental currency units (ECUs) which they can use to trade over 20 periods. Each unit of asset pays a dividend of 12 ECUs at the end of each period, which will be added to traders' cash holdings and can be used for trading in the future periods. After the final dividend payment at the end of period 20, all the assets lose their value. Under these conditions, the fundamental value of a unit of the asset during period t ($t = 1, 2, ..., T$), FV_t, is the sum of the remaining dividend payments, that is, $FV_t = 12(21 - t)$. For example, a unit of asset initially has a value of 240 ECUs. Thus the value of initial endowment is 2000 ECUs for all the market participants (1040 ECUs in initial currency plus 960 ECUs for the four units of the asset). We have eliminated the uncertainty in dividend payments in order to minimize the presence of uncertainty beyond the uncertainty caused by the behavior of market participants. Even with fixed and known dividend payments, mispricing has been observed in these markets (Porter and Smith, 1995; Akiyama et al., 2014, 2015).

We employ a call market mechanism as in van Boening et al. (1993); Haruvy et al. (2007); Akiyama et al. (2014, 2015) instead of continuous double auction as in many other

[6] We inform our subjects in unknown composition treatment as follows. In the beginning of the instruction of the asset market experiment, we state, "You are divided into the top 12 scorers and the bottom 12 scorers of 24 people from the previous quiz. Before starting the game, you will know which your rank is, the top or the bottom. The 24 people in the room are divided into four groups, each of which consists of six people." And we display, for each subject, their own type (H or L) in the first screen of the asset market experiment.

[7] More specifically, we inform our subjects in known composition treatment as follows. (1) In the beginning of the asset market experiment, we state, in 6H and 6L treatment, "You are divided into the top 12 scorers and the bottom 12 scorers of 24 people from the previous quiz. Before starting the game, you will know which your rank is, the top or the bottom. The 24 people in the room are divided into four groups, each of which consists of six people: Two of the four groups consist of the top scorers, and the other two groups consist of the bottom scorers." In case of 3H3L, the last sentence reads "Each of four groups consists of three top scorers and three bottom scorers." And (2) at the end of the instruction, we repeat the same information. In 6H or 6L treatment, we state "There are six people in a market. All the people of the group are in this room. Each group consists of only the top scorers or of only the bottom scorers from the previous quiz. Your rank, the top or the bottom, from the previous quiz is noted on the first screen." In case of 3H3L treatment, the third sentence is replaced by "Each group consists of three top scorers and three bottom scorers from the previous quiz." See Appendix for an English translation of the instruction as well as examples of the first screen of asset market experiment in which subjects' type and the group composition are displayed.

studies. In our call market, in each period, each trader can submit at most one buy order and one sell order.[8] An order consists of a pair of values: a price and a quantity. When submitting a buy order in period t, trader i must specify the *maximum price*, b_t^i, at which s/he is willing to buy a unit of asset, and the *maximum quantity*, d_t^i, s/he is willing to buy at that price. In the same manner, when submitting a sell order in period t, trader i must specify the *minimum price*, a_t^i, at which s/he is willing to sell a unit of asset, and the *maximum quantity*, s_t^i, s/he is willing to sell at that price. We attached three constraints: the admissible price range, a budget constraint, and the relationship between b_t^i and a_t^i in the case that a subject submits both buy and sell orders. The admissible price range is set so that, when $d_t^i \geq 1$ ($s_t^i \geq 1$), b_t^i (a_t^i) must be an integer between 1 and 2000, i.e., $b_t^i \in \{1, 2, ..., 2000\}$ ($a_t^i \in \{1, 2, ..., 2000\}$). The budget constraint simply means that neither borrowing of cash nor short-selling of an asset is allowed.[9] The final constraint is such that when a trader is submitting both buy and sell orders, i.e., $d_t^i \geq 1$ and $s_t^i \geq 1$, the maximum buying price must not be greater than the minimum selling price, i.e., $a_t^i \geq b_t^i$. Once all the traders in the market have submitted their orders, the price that clears the market is calculated,[10] and all transactions are processed at that price among traders who submitted a maximum buying price no less than, or a minimum selling price no greater than, the market clearing price.[11]

The entire experiment is computerized with z-Tree (Fischbacher, 2007). Each session lasted about 1.5 hours including a post-experimental questionnaire. We have also administered CRT as a part of post-experimental questionnaire without monetary incentive for correct answers. On average, subjects earned 3000 yen ($\approx$ 22 euros at the average exchange rate during the time of experiment) including a 1000-yen participation fee. See Appendix for English translation of the instructions.

[8] Of course, a trader can choose not to submit any orders by specifying zero as the quantities to buy and sell. We imposed a 60-second, non-binding, time limit for submitting orders. When the time limit was reached, the subjects were simply informed, though a flashing message in the upper right corner of their screen, to submit their orders as soon as possible.

[9] Thus, the budget constraint implies (i) $d_t^i \times b_t^i \leq$ cash holding at the beginning of the period t, and (ii) $s_t^i \leq$ units of asset on hand at the beginning of the period t.

[10] Following the previous experiments (Haruvy et al., 2007; Akiyama et al., 2014, 2015), when there are several such prices, the lowest one is chosen as the market clearing price. This is important to ensure the price does not jump up in the absence of transactions at the market clearing price.

[11] Any ties among the last accepted buy or sell orders are resolved randomly. It is possible that no transaction will take place given the computed market clearing price.

Table 1: Summary of treatments

Treatment	No. of subjects	No. of markets
Unknown composition, 6H	48	8
Unknown composition, 6L	48	8
Unknown composition, 3H3L	48	8
Known composition, 6H	72	12
Known composition, 6L	72	12
Known composition, 3H3L	72	12

3 Result

The experiment was conducted at Waseda University (in Tokyo, Japan) between November 2014 and July 2016. A total of 360 subjects (144 subjects in unknown composition treatments of which 96 subjects in 6H/6L sessions and 48 subjects in 3H3L sessions, and remaining 216 subjects in known composition treatments of which 144 subjects in 6H/6L sessions and 72 subjects in 3H3L sessions) participated. Subjects were recruited within the main campus of the university by e-mails and flyers. These subjects had never participated in a similar experiment before and each subject only participated in one session. We therefore have 8 markets of 6 subjects each for unknown composition treatment, and 12 markets of 6 subjects each for known composition treatment. Table 1 summarizes the treatments considered and number of subjects participated in each.

Figure 1 shows the empirical cumulative distributions of the scores from the RPM test (RPM scores) among the participants in unknown composition (left) and known composition (right) treatments. In each panel, three types of markets are shown separately: 6H (red, thick-solid), 6L (blue, thick-dashed), and 3H3L (black, thin). By construction, the empirical cumulative distribution of RPM scores are ordered as 6L, 3H3L, and 6H from the left (lowest) to the right (highest). The distribution of RPM scores between unknown and known composition treatments are not statistically significantly different for each of the market type (p-values are 0.123 for 6H, 0.592 for 6L, and 0.22 for 3H3L, according to two-sample permutation test (PT), two-tailed).

The results from the previous studies mentioned in the introduction suggest a negative relationship between the average level of cognitive ability among traders and mispricing in markets. If the diversity (or the heterogeneity) of cognitive ability among traders does not

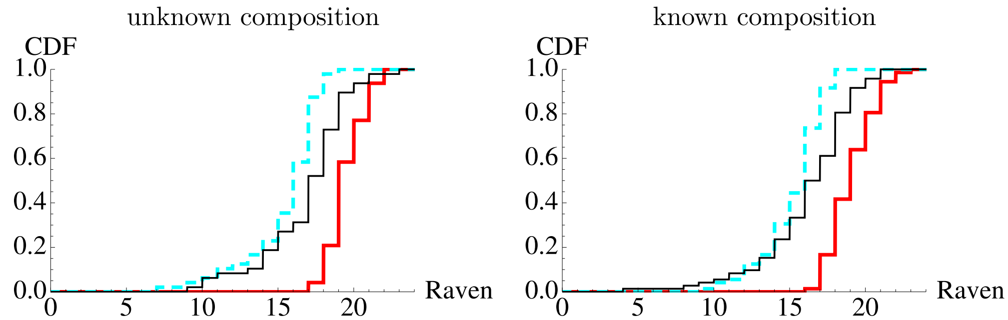


Figure 1: Distribution of scores from Raven's (RPM) test in 6H (black, thick-solid), 6L (gray, thick-dashed), and 3H3L (black, thin), in unknown composition treatment (left) and known composition treatment (right). For unknown composition treatment, there are 48 subjects in each group: 6H, 6L, and 3H3L. For known composition treatment, there are 72 subjects in each group: 6H, 6L, and 3H3L. The highest score obtainable is 24.

have a strong effect on the magnitude of mispricing, then we would expect there to be a larger mispricing in 6L than in 3H3L, and, in turn, in 3H3L than in 6H. On the other hand, if the diversity has a significant effect, then we may observe a larger mispricing in the 3H3L markets than in the two homogeneous markets.

3.1 Prices

Figure 2 shows the observed price dynamics from the unknown (top) and known (bottom) composition treatment. Three types of markets, 6H (left), 3H3L (center), and 6L (right) are shown separately. The results for three types of markets look very similar regardless of whether compositions of cognitive types within a market are ex-ante known or not. In both unknown and known composition treatments, while prices follow the fundamental value very closely in most of the 6H and 6L markets, they deviate substantially from the fundamental value in the 3H3L markets.

To systematically analyze the magnitude of mispricing in various markets, we employ the relative absolute deviation (RAD) proposed by Stöckl et al. (2010). For each market m, RAD is defined as

$$RAD^m = \frac{1}{20} \sum_{p=1}^{20} \frac{|P_p^m - FV_p|}{|\overline{FV}|} \tag{1}$$

where P_p^m is realized price in period p in market m. FV_p is fundamental value of asset in

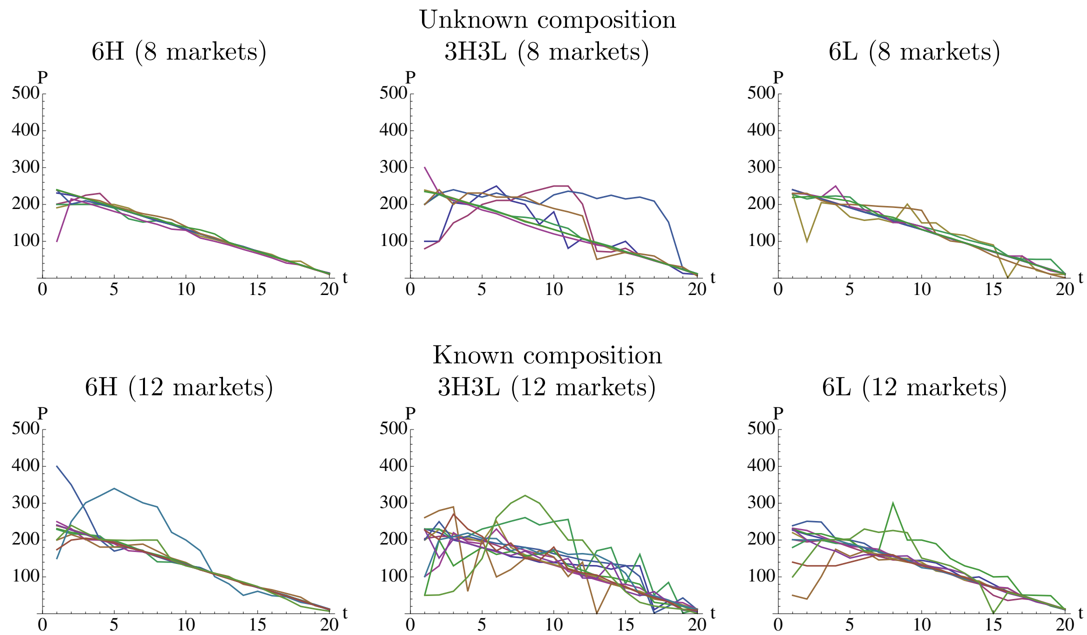


Figure 2: Realized price dynamics in unknown (top) and known (down) treatments for three market types: 6H (left), 3H3L (center), and 6L (right).

period p. $|\overline{FV}| = |\frac{1}{20}\sum_{p=1}^{20} FV_p|$. We supplement the analyses by computing the positive and negative deviations separately. Namely, we define, relative positive deviation (RPD) and relative negative deviation (RND) as follow:[12]

$$RPD^m = \frac{1}{20}\sum_{p=1}^{20} \frac{\max(P_p^m - FV_p, 0)}{|\overline{FV}|} \tag{2}$$

$$RND^m = \frac{1}{20}\sum_{p=1}^{20} \frac{\max(FV_p - P_p^m, 0)}{|\overline{FV}|} \tag{3}$$

Figure 3 shows the empirical cumulative distribution (CDF) of RAD (top), RPD (middle), and RND (bottom) observed in unknown (left) and known (right) composition treatments.[13] In each panel, the outcomes from the three types of markets are shown: 6H (red, thick-solid), 6L (blue, thick-dashed), and 3H3L (black, thin).

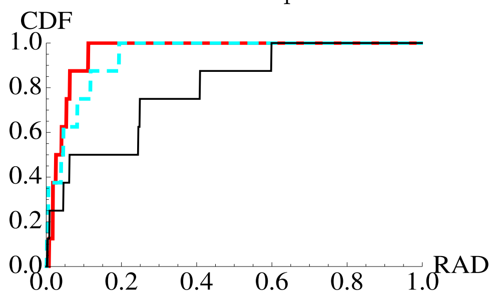
As one could expect from the price dynamics shown in Figure 2, the distribution of RAD from 3H3L lies on the right of 6H and 6L markets in both unknown and known composition

[12] RPD and RND are defined based on the positive and negative deviations often used in the literature. We call them relative deviation because we normalize them by $20|\overline{FV}|$ to ease the comparison with RAD.

[13] See Appendix A for values of these measures for each market.

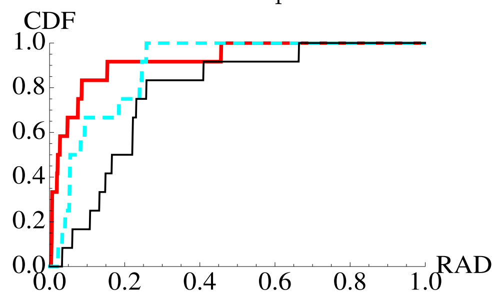
Relative Absolute Deviation (RAD)

unknown composition



	6L	3H3L
6H	0.504	0.055
6L		0.098

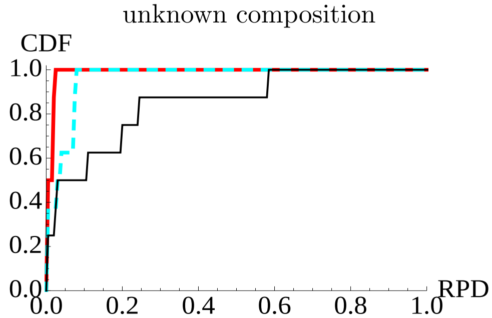
Known composition



	6L	3H3L
6H	0.440	0.021
6L		0.057

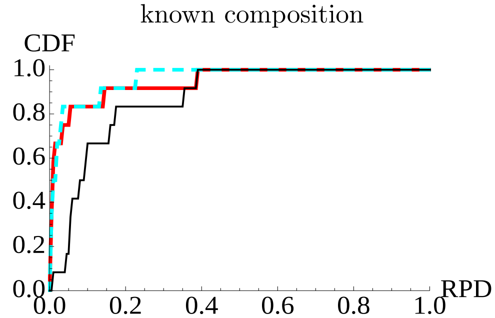
Relative Positive Deviation (RPD)

unknown composition



	6L	3H3L
6H	0.064	0.025
6L		0.117

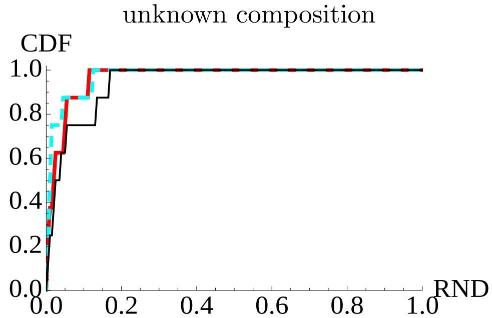
known composition



	6L	3H3L
6H	0.716	0.140
6L		0.033

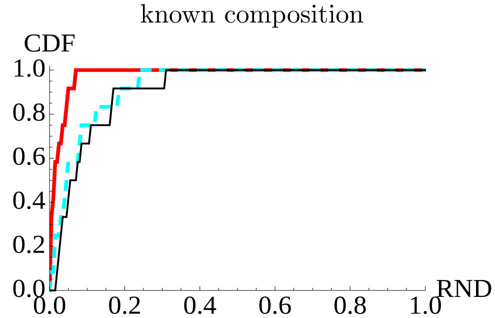
Relative Negative Deviation (RND)

unknown composition



	6L	3H3L
6H	0.699	0.373
6L		0.248

known composition



	6L	3H3L
6H	0.015	0.002
6L		0.595

Figure 3: Empirical cumulative distribution of *RAD* (top), *RPD* (middle), and *RND* (bottom) in unknown (left) and known (right) composition treatments. In each panel, three types of markets are shown with 6H (red, thick-solid), 6L (blue, thick-dashed), and 3H3L (black, thin). The table below each panel reports the p-values from pair-wise comparisons based on two-sample permutation tests, two-sided.

treatments. The pair-wise comparisons show that the RAD in 3H3L is significantly greater than those in 6H and 6L (p-values are 0.056 for 6H vs 3H3L and 0.098 for 6L vs 3H3L in unknown composition treatment, and 0.022 for 6H vs 3H3L and 0.055 for 6L vs 3H3L in known composition treatment, according to two-sample permutation test (PT), two-tailed). Between the two homogeneous market, 6H and 6L, RAD are not significantly different (p-values are 0.502 and 0.441 for unknown and known composition treatments, respectively, according to two-sample PT, two-tailed.).

The similar results are obtained for RPD but not in RND. The distribution of RPD in 3H3L lies on the right of those of 6H and 6L in both known and unknown composition treatments, although RPDs are no longer statistically significantly different between 3H3L and 6H in unknown composition treatment or between 3H3L and 6L in known composition treatment at 10% level. The distributions of RNDs observed in three types of markets in unknown composition treatment are almost on top of each other. For the known composition treatment, RND in 6H is significantly smaller than both 6L and 3H3L. The distributions of RNDs in latter two are on top of each other. This suggests that the significantly larger mispricing in 3H3L markets compared to 6H and 6L markets are mainly due to the positive deviations.

Is there a significant effect of ex-ante common information about the composition of cognitive types within markets on the magnitude of mispricing? Contrary to our expectation, we did not find such an effect. For each market type, RADs and RPDs are not significantly different between the two information treatments (p-values are 0.629 for 6H, 0.175 for 6L, and 0.846 for 3H3L markets, two-sample permutation test (PT), two-tailed for RAD. For RPD, they are 0.258, 0.947, and 0.788, for 6H, 6L, and 3H3L, respectively). RND for 6L is significantly different at 10% level (p=0.093, PT) between the two information treatments, but not significantly so for 6H and 3H3L (p=0.457 and p=0.318, respectively). Therefore, in our experiment, subjects being informed about the composition of cognitive types (H or L) of other participants in the same market does not have a significant effect on the magnitude of mispricing.

The small mispricing in our 6L markets may seem quite surprising to many readers in light of the existing literature that has demonstrated systematically larger mispricing for markets that consist of subjects with low cognitive ability. It should be noted, however, our experiment is much simpler compared to other studies in that there is no uncertainty in the amount of dividend payment. Furthermore, cognitive ability of subjects in our 6L markets are still high from the point of view of a wide pool of experimental subjects. One of the authors has been conducting a shorter version of advanced RPM test (16 questions to be answered in 10 minutes) in various experimental laboratories, and found that, not surprisingly, the distributions of the scores vary greatly across laboratories. The distribution of the scores for our subjects recruited at Waseda university is highest among the subjects pools the author has the data. Thus, the low mispricing of 6L is, in addition to our simple structure of dividend payments, likely due to our particular subjects pool.

The larger mispricing observed in 3H3L markets compared to two homogeneous markets (both 6H and 6L), however, is very surprising in light of above remark about the pool of subjects we are dealing with as well as our simple dividend payment process. Below, we provide further analyses aiming to better understand this result.

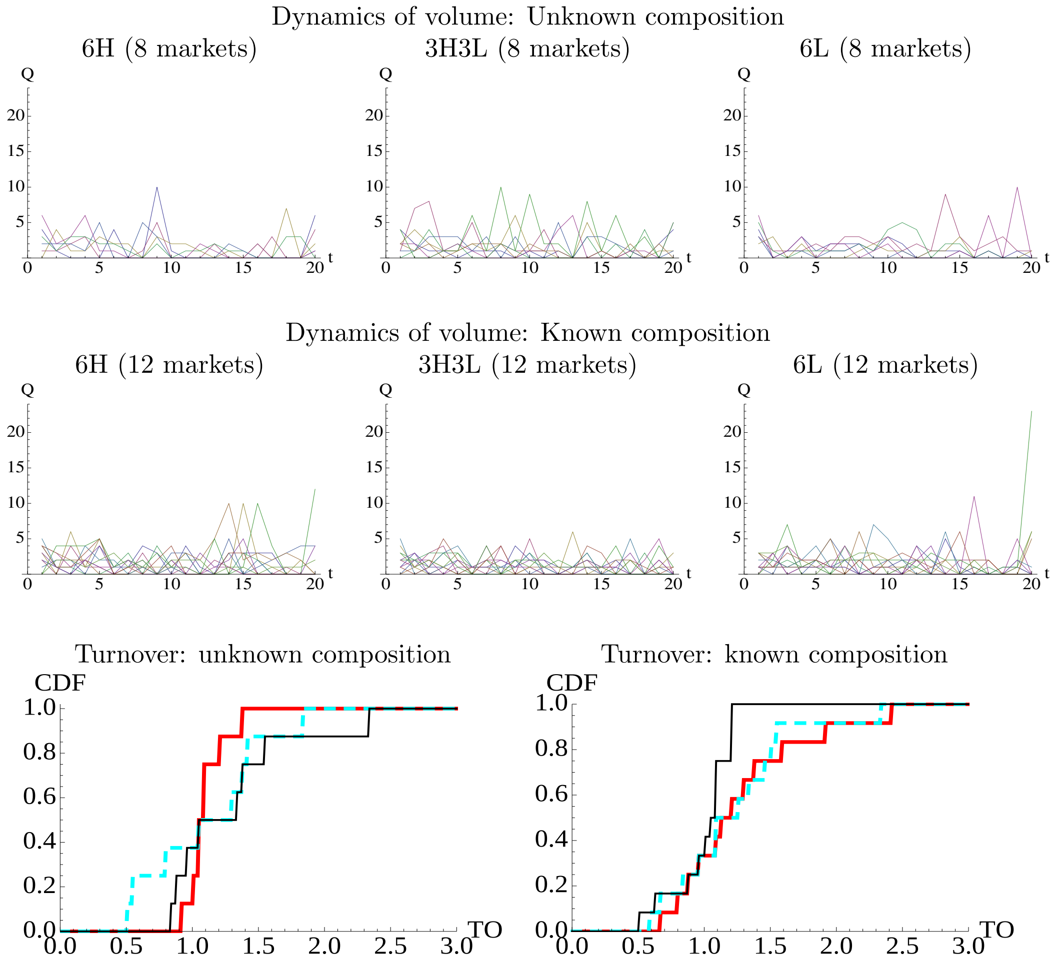
3.2 Trading volumes and volume adjusted mispricing

The top two panels of Figure 4 shows the observed dynamics of trading volume from the unknown (top) and known (middle) composition treatment. Three types of markets, 6H (left), 3H3L (center), and 6L (right), are shown separately for each treatment. While for the unknown composition treatment shown in the top row, the trading volumes seem to be higher in 3H3L markets compared to 6H and 6L market, the opposite seems to be the case in the known composition market shown in the bottom row. We also note that in many markets, there are periods with zero transaction.[14]

The bottom panel of Figure 4 shows the empirical cumulative distribution of turnover, $\sum_p Q_p^m/24$, where Q_p^m is the realized trade volume in period p of market m. As one can see that there is statistically significant differences across the three types of markets in two information treatments, except between 6H and 3H3L of known composition treatments. Furthermore, except for 3H3L markets, there is no statistically significant differences in turnover between two information treatments (p-values are 0.367, 0.592, and 0.095 for 6H, 6L, and 3H3L markets, respectively).

It is possible that the significantly larger mispricing in the heterogeneous markets (3H3L)

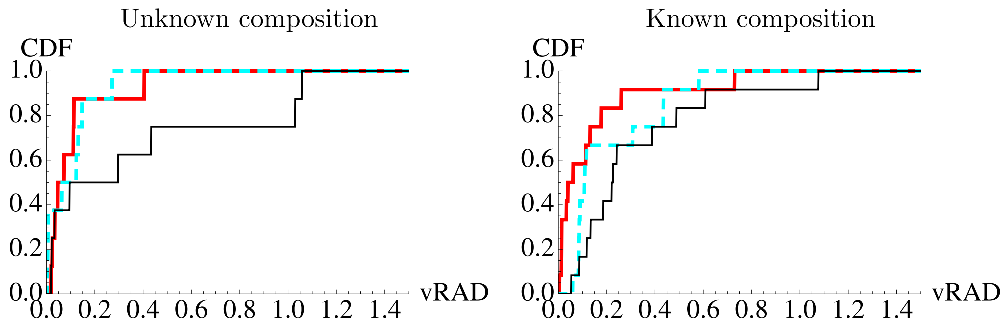
[14]Our price determination procedure returns a price even in the absence of transaction.



	6L	3H3L
6H	0.994	0.369
6L		0.464

	6L	3H3L
6H	0.779	0.088
6L		0.152

Figure 4: Top: Realized trade volume dynamics in unknown (top) and known (bottom) composition treatments. Three types of markets are shown separately: 6H (left), 3H3L (center), and 6L (right). Bottom: Empirical cumulative distribution of turnover in unknown (left) and known (right) composition treatments. In each panel, three markets types, 6H (rad, thick-solid), 6L (blue, thick-dashed), and 3H3L (black, thin) are shown. The table below each panel reports the p-values from pair-wise comparisons based on two-sample permutation tests, two-sided.



	6L	3H3L
6H	0.871	0.147
6L		0.109

	6L	3H3L
6H	0.357	0.080
6L		0.290

Figure 5: Distribution of $vRAD$ in unknown (left) and known (right) composition treatments. In each panel, three markets types, 6H (rad, thick-solid), 6L (blue, thick-dashed), and 3H3L (black, thin) are shown. The table below each panel reports the p-values from pair-wise comparisons based on two-sample permutation tests, two-sided

compared to the homogeneous markets (6H and 6L) we have observed above is due to mispricing that took place only when trading volume was zero or very low. If this is the case, the straight measure of mispricing, RAD, we have considered above over-represents the effective magnitude of mispricing. To address this potential problem, we define volume-adjusted RAD for market m, $vRAD^m$, as follows:

$$vRAD^m = \frac{1}{20}\sum_{p=1}^{20} Q_p^m \left(\frac{|P_p^m - FV_p|}{|\overline{FV}|}\right) \tag{4}$$

Figure 5 shows the empirical cumulative distributions of $vRAD$ in unknown (left) and known (right) composition treatment. In each panel, three market types, 6H (red, thick-solid), 6L (blue, thick-dashed), and 3H3L (black, thin) are shown. These distributions, however, are ordered in a similar manner to the one of RAD reported in Figure 3 above. For the unknown composition treatment, the distribution of $vRAD$ for two homogeneous markets are almost on top of each other, and that of 3H3L lies on the right of them. For the known composition treatment, the distribution of $vRAD$ for 6H markets lies on the left of that of 6L markets, which in tern lies on the left of that of 3H3L. As in the case with RAD, there is no statistically significant difference in $vRAD$ between two information treatments (p-values are 0.833 for 6H, 0.116 for 6L, and 0.728 for 3H3L, two-sample PT, two-tailed) in each market type.

Unlike the case of RADs, however, $vRAD$s are no longer statistically significantly different between heterogeneous markets and homogeneous markets except between 3H3L and 6H markets with known composition (p-values are 0.080 for 6H vs 3H3L, 0.290 for 6L vs 3H3L, and 0.357 for 6H vs 6L for known composition treatment, and 0.147 for 6H and 3H3L, 0.109 for 6L and 3H3L, and 0.871 for 6H and 6L for unknown composition treatment based on two-sample PT, two-tailed). This suggests that the mispricing in periods with no or low transaction does indeed explain a part of the larger mispricing in the heterogeneous market compared to the homogeneous market, but that is not the whole story.

Note, however, that if we pool the known and unknown composition treatments (because there is no statistically significant difference between the two treatments in any of the three types of the markets), then, the $vRAD$s for 3H3L markets are significantly greater than those in 6H and 6L markets (p-values are 0.014 for 6H vs 3H3L, 0.043 for 6L vs 3H3L, and 0.445 for 6H vs 6L)

3.3 Gender composition

In our analyses above, we have not controlled for possible effects of gender compositions on mispricing. Eckel and Füllbrunn (2015) found that the experimental asset markets with a larger fraction of female subjects resulted in a smaller mispricing. Cueva and Rustichini (2015) compared all female, all male, and mixed gender markets, and found that the mixed gender markets resulted in smaller mispricing than two other types of markets. In this sub-section, we report the results of linear regression analyses to investigate the relationship between the cognitive ability of market participants and the magnitude of mispricing while controlling for the gender composition. Because we did not find significant difference between known and unknown composition treatments, we pool the data from these two treatments in the following analyses.

Dependent variables are the four mispricing measures we have considered above: RAD, RPD, RND, and $vRAD$. Independent variables are the mean and the standard deviation of the CRT score of subjects in the market (mCRT and sdCRT), the mean and the standard deviation of the RPM score of subjects in the market (mRPM and sdRPM), and the number of male subjects (out of 6) in the market and its square (gender and gender^2). We have included the squared term of the gender composition to capture the non-linear effect of

Table 2: Descriptive statistics

Variable	No. Obs	Mean	Std. Dev.	Min	Max
RAD	60	0.124	0.145	0.002	0.664
RPD	60	0.071	0.115	0	0.583
RND	60	0.052	0.063	0	0.309
vRAD	60	0.208	0.257	0.003	1.077
mCRT	60	2.019	0.483	0.667	3
sdCRT	60	0.975	0.293	0	1.506
mRPM	60	16.967	1.857	14	20.333
sdRPM	60	2.157	0.982	0.632	5.404
gender	60	4.167	1.107	2	6
gender^2	60	18.567	9.101	4	36

the gender composition reported by Cueva and Rustichini (2015). Table 2 summarizes the descriptive statistics for these variables.

We consider both CRT and RPM scores because these two tests capture different aspects of subjects' abilities. It should be noted, however, that CRT and RPM scores are positively correlated in our data (the correlation coefficients are 0.40 for the mean and 0.15 for the standard deviations). Therefore, we also report the results of regressions that only have either CRT or RPM but not both. Because the results of estimation with only either CRT or RPM, however, are not very different from the one that we include both CRT and RPM scores, we only comments on the results that include both.

Table 3 reports the results. Let us look at the result of RAD, the main mispricing measure we consider. The mean CRT score, but not its standard deviation, is negatively and significantly correlated with the RAD. This is consistent with previous findings (such as Breaban and Noussair, 2015). On the other hand, the standard deviation of RPM score, but not its mean, is positively and significantly correlated with the RAD as we have reported above. We also find that the larger the number of male subjects in the market, the higher the RAD becomes as in Eckel and Füllbrunn (2015). Furthermore, as we can see from a negative and significant coefficient of gender^2, the marginal effect of additional male subjects on RAD decreases with the number of male subjects in the market. The magnitudes of estimated coefficients of gender and its squared term show that there is an inverse U-shaped relationship between the number of male subjects in the market and the extent of mispricing as in Cueva and Rustichini (2015).

We find very similar results for RPD (relative positive deviation). The difference from

Table 3: Results of linear regression

	Dependent variables											
	RAD			RPD			RND			vRAD		
mCRT	-0.111***		-0.103**	-0.053		-0.037	-0.058***		-0.065***	-0.191**		-0.173**
	(0.042)		(0.041)	(0.033)		(0.033)	(0.019)		(0.018)	(0.077)		(0.074)
sdCRT	0.042		0.032	0.095*		0.078	-0.053*		-0.046	0.045		0.027
	(0.066)		(0.067)	(0.053)		(0.055)	(0.029)		(0.029)	(0.122)		(0.121)
mRPM	0.020	0.004		0.024**	0.012		-0.004	-0.008		0.030	0.005	
	(0.013)	(0.012)		(0.010)	(0.010)		(0.006)	(0.006)		(0.023)	(0.023)	
sdRPM	0.049**	0.044*		0.043**	0.039**		0.006	0.005		0.062	0.538	
	(0.022)	(0.023)		(0.017)	(0.018)		(0.010)	(0.010)		(0.040)	(0.042)	
gender	0.210*	0.19	0.170	0.196**	0.187**	0.154*	0.012	0.003	0.016	0.409**	0.379*	0.355*
	(0.108)	(0.116)	(0.109)	(0.087)	(0.092)	(0.090)	(0.048)	(0.052)	(0.047)	(0.200)	(0.211)	(0.198)
$gender^2$	-0.027**	-0.026*	-0.022*	-0.025**	-0.024**	-0.020*	-0.002	-0.001	-0.003	-0.050**	-0.048*	-0.044*
	(0.013)	(0.014)	(-0.013)	(0.011)	(0.011)	(0.011)	(0.006)	(0.006)	(0.006)	(0.024)	(0.026)	(0.024)
Const	-0.499	-0.361	0.003	-0.773**	-0.554*	-0.207	0.271*	0.190	0.209*	-0.867	-0.682	-0.130
	(0.361)	(0.360)	(0.243)	(0.290)	(0.287)	(0.199)	(0.160)	(0.161)	(0.106)	(0.668)	(0.657)	(0.441)
R^2	0.273	0.136	0.203	0.256	0.135	0.157	0.252	0.107	0.217	0.211	0.093	0.172
N. Obs.	60	60	60	60	60	60	60	60	60	60	60	60

[i] Standard errors are in the parentheses.
[ii] gender : number of male subjects (out of 6) in the market
[iii] ***, **, and *: statistically significant at 1%, 5%, and 10% level, respectively.

RAD is that now, mean CRT score loses its statistical significance, while the mean RPM scores and the standard deviation of CRT scores become statistically significant. The standard deviation of RPM scores remain to be positive and significantly correlated with the RPD.[15] Thus, even after controlling for the potential effects of gender composition, the heterogeneity in cognitive ability among market participants does increase the mispricing, especially, the positive deviation of prices from the FVs.

The results are quite different for RND (relative negative deviation). For RND, either the mean or the standard deviation of RPM scores are statistically significant. The gender compositions (both the level and its square) are not statistically significant. Mean CRT scores, on the other hand, is negative and significant in this regression.

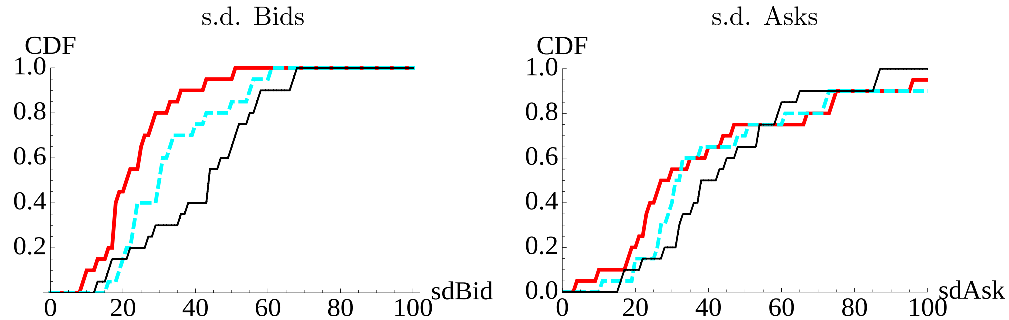
Finally, for the volume adjusted RAD ($vRAD$), while the mean CRT scores remains to be negative and statistically significant, RPM scores, both the mean and the standard deviation, become not statistically significant once the gender compositions are controlled for. The gender composition effects remain to be statistically significant with the same sign as in the case of RAD.

3.4 Heterogeneity in trading behavior and mispricing

Why does the heterogeneity in cognitive ability enlarge the mispricing? We hypothesize that the heterogeneity in cognitive ability results in the heterogeneity in trading behavior, which results in larger price variations and mispricing.

To capture the heterogeneity in trading behavior among market participants, we compute the standard deviation of bids and asks submitted by market participants for each period in each market, and take the average across 20 periods. Figure 6 shows the empirical cumulative distributions of the within market standard deviations of bids (left) and of asks (right) that are computed as described. From the left panel of the figure, it is clear that the within market heterogeneity in bids are the largest in 3H3L markets, while they are the smallest in 6H markets. As the table below the left panel shows that these are statistically significantly different (p-values are 0.02 for 6H vs 6L, less than 0.001 for 6H vs 3H3L, and 0.078 for 6L vs 3H3L, two-sample permutation tests, two tailed). The right panel of the

[15] It is also interesting to note the positive and significant coefficient of mRPM. But we do not have a very clear interpretation of this result.



	6L	3H3L
6H	0.020	< 0.001
6L		0.078

	6L	3H3L
6H	0.863	0.875
6L		0.976

Figure 6: Distribution of within market standard deviations of bids (left) and of asks (right). In each panel, three markets types, 6H (rad, thick-solid), 6L (blue, thick-dashed), and 3H3L (black, thin) are shown. The table below each panel reports the p-values from pair-wise comparisons based on two-sample permutation tests, two-sided

figure shows, however, that within market heterogeneity in asks are quite similar across three market types. Indeed, the table below the panel shows that they are not statistically significantly different (p-values are 0.863 for 6H vs 6L, 0.875 for 6H vs 3H3L, and 0.976 for 6L vs 3H3L, two sample permutation test, two-tailed).

Table 4 reports the result of linear regressions that take mispricing measures as dependent variables and the average within market heterogeneity of bids (sdBids) and of asks (sdAsks) as independent variables. Except for the RND, the within market heterogeneity of bids are positively and statistically significantly correlated with the mispricing measures. The heterogeneity of asks, on the other hand, is not statistically significantly correlated with any of the mispricing measures we consider.

These results support our hypothesis. That is, the heterogeneity in the cognitive ability among market participants results in the the heterogeneity in trading behavior, in particular, the bids they submit, which in tern results in a larger mispricing.

3.5 Profits and cognitive ability

Finally, we report the relationship between the cognitive ability and profit (at the end of period 20) controlling for the gender. Table 5 report the results of linear regressions. The dependent variable is the individual profit at the end of period 20 in all the regressions. We

Table 4: Heterogeneity in orders and mispricing

	Dependent variables			
	RAD	RPD	RND	vRAD
sdBids	0.0056***	0.0047***	0.0009	0.008***
	(0.0011)	(0.0008)	(0.0005)	(0.002)
sdAsks	-0.0003	-0.0002	-0.0001	-0.0003
	(0.0004)	(0.0003)	(0.0002)	(0.0008)
Const	-0.046	-0.074**	0.028	-0.052
	(0.038)	(0.029)	(0.020)	(0.071)
R^2	0.330	0.374	0.04	0.239
N. Obs.	60	60	60	60

[i] Standard errors are in the parentheses.
[ii] ***, **, and *: statistically significant at 1%, 5%, and 10% level, respectively.

consider three types of market, 3H3L, 6H, and 6L, separately. For each market type, we pool the data from two information treatments (known and unknown compositions), and report the results of three regressions: the one includes both RPM and CRT scores, and two others that consider RPM and CRT separately. We consider these three specifications because, although RPM and CRT captures different aspects of cognitive skills, RPM and CRT scores tend to be positively correlated in our sample (correlation coefficient is 0.345 taking at individual level).

The results of 3H3L and 6L markets show that considered separately, both RPM and CRT scores are positively and significantly related to the profit. This is in line with previous findings, such as Corgnet et al. (2015) and Cueva and Rustichini (2015). The male dummy is not statistically significantly correlated with the profit in 3H3L markets, but it is positively and significantly correlated with profit in 6L markets. In the 6H markets, however, neither cognitive ability nor gender is significantly correlated with the profit.

4 Summary and conclusion

How does the average cognitive ability among market participants, as well as their diversity, influence mispricing in an experimental market? We investigated this question by first measuring an aspect of cognitive ability of our subjects with the advanced version of Raven's progress matrix test, and then constructing markets by grouping subjects based on their

Table 5: Total profit and cognitive ability

	3H3L			6H			6L		
RPM	19.34	23.38*		2.66	7.66		18.50**	25.01**	
	(12.53)	(12.12)		(12.61)	(15.82)		(8.90)	(8.83)	
CRT	34.71		53.77**	32.57		33.59	46.68		58.09*
	(20.64)		(22.84)	(31.99)		(34.11)	(31.31)		(30.12)
Male	8.12	31.57	4.94	81.30	89.73	80.95	131.22	130.80**	130.78**
	(61.67)	(59.81)	(59.55)	(48.11)	(50.83)	(48.13)	(61.84)	(62.48)	(60.82)
Const	1614.50	1595.10	1899.43	1912.51	1886.15	1961.52	1629.89	1621.00	1891.05
	(195.65)	(195.87)	(44.65)	(297.94)	(326.89)	(99.14)	(125.23)	(126.40)	(69.35)
R^2	0.06	0.05	0.03	0.02	0.01	0.02	0.06	0.05	0.05
N. Obs.	120	120	120	120	120	120	120	120	120

[i] Robust standard errors corrected for group clustering effects are in the parentheses.
[ii] Male : = 1 if male, and = 0 if female.
[iii] ***, **, and *: statistically significant at 1%, 5%, and 10% level, respectively.

relative test scores. We define those subjects whose scores are above and below the median score in the session to be H type and L type, respectively. We have considered three kinds of markets: all 6 traders were H type (6H), all 6 traders were L type (6L), and H and L types were equally mixed (3H3L).

In order to investigate whether the heterogeneity of cognitive ability being known to market participants can have additional effects on mispricing compared to the case where it is not known, we considered two information treatments: the known composition treatment and the unknown composition treatment. In both treatments, we informed our subjects of their own type (H or L). In known composition treatment, we also informed subjects the types of the other five traders in the market. Thus, for example, those in 6H markets were informed that they were H type and all the other five traders in the same market were also H type. In unknown composition treatment, this type of information was not given to our subjects.

Contrary to what one may infer from the results of earlier experimental studies which report the negative relationship between the average cognitive ability of subjects in a market and the magnitude of mispricing, the magnitude of mispricing observed in 3H3L was significantly larger than that observed in 6H and 6L markets both in known and unknown composition treatments. Thus, it is not only the average cognitive ability of traders in the market but also their diversity that matters when it comes to the magnitude of mispricing. Contrary to our expectation, however, we did not find any significant additional effect of heterogeneity being ex-ante known on the magnitude of mispricing.

We hypothesized the reason for the larger mispricing in 3H3L markets compared to 6H and 6L markets was due to a positive correlation between the heterogeneity in cognitive ability among market participants and the heterogeneity in their trading behavior, and such a heterogeneity in trading behavior among market participants generated a larger mispricing. Our analysis supported this hypothesis. The within market heterogeneity of submitted bids was significantly larger in 3H3L markets than in 6H or 6L markets, and such heterogeneity was positively and significantly correlated with mispricing measures. We believe a more in-depth analyses on the heterogeneity in dynamics of trading behavior can be a fruitful future research. For this purpose, however, it may be useful to conduct experiments under continuous double auctions to gather more observations for the dynamics of trading behavior.

Recently, several researchers have investigated the effect of other types of heterogeneity on mispricing in similar experimental set up. Levine et al. (2014) report that the known ethnic diversity among market participants reduces the magnitude of mispricing. Their interpretation of the data is that participants do not think critically of others' decisions in ethnically homogeneous markets, and thus tend to ride on "bubbles," compared to those participants in ethnically diverse markets.

Eckel and Füllbrunn (2015) and Cueva and Rustichini (2015) study the effect of gender composition on mispricing. These two studies somewhat disagree in term of the effect of the gender composition. While Eckel and Füllbrunn (2015) find that all male markets generate higher mispricing than all female markets, and the mispricing observed in mixed gender markets fall between the two, Cueva and Rustichini (2015) find that the mispricing in the mixed-gender markets are larger than all -male or all-female markets. Our regression result is in line with that of Cueva and Rustichini (2015) in that while increasing male fraction of participants in a market increases the magnitude of mispricing, its marginal effect is negative, which results in an inverted U-shape relationship between the fraction of male participants in the market and the magnitude of mispricing.

Hefti et al. (2016) consider the effect of heterogeneity in two distinct capabilities: analytical capability (cognitive skill) and mentalizing capability (theory of mind). They found that, consistent with their hypotheses, to be successful in the asset market experiments, one has to have both high analytical and mentalizing capabilities because one's success depends not only in understanding market fundamental (which requires analytical capability) but also the dynamics of prices that results from the behavior of other participants (which requires mentalizing capability). It will be of a very interesting future research to construct various types markets by explicitly grouping subjects based on heterogeneities in these various dimensions, such as gender, ethnic identity, analytical and mentalizing capabilities, in order to investigate how heterogeneities in these various dimensions interact among themselves and determine the aggregate market outcomes.

References

AGGARWAL, R. K. AND G. WU (2006): "Stock Market Manipulations," *Journal of Business*, 79, 1915–1953.

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.

——— (2015): "It is not just confusion! Strategic uncertainty in an experimental asset market," *Economic Journal*, forthcoming.

ALLEN, F. AND D. GALE (1992): "Stock-Price Manipulation," *Review of Financial Studies*, 5, 503–529.

ALLEN, F., S. MORRIS, AND H. S. SHIN (2006): "Beauty Contests and Iterated Expectations in Asset Markets," *The Review of Financial Studies*, 19, 719–752.

BOSCH-ROSA, C., T. MEISSNER, AND A. BOSCH-DOMÈNECH (2015): "Cognitive Bubbles," Mimeo, Berlin University of Technology.

BREABAN, A. AND C. N. NOUSSAIR (2015): "Trader characteristics and fundamental value trajectories in an asset market experiment," *Journal of Behavioral and Experimental Finance*, 8, 1–17.

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.

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.

CUEVA, C. AND A. RUSTICHINI (2015): "Is financial instability male-driven? Gender and cognitive skills in experimental asset markets," *Journal of Economic Behavior and Organization*, 119, 330–344.

Dufwenberg, M., R. Sundaram, and D. J. Butler (2010): "Epiphany in the Game of 21," *Journal of Economic Behavior and Organization*, 75, 132–143.

Eckel, C. C. and S. C. Füllbrunn (2015): "Thar SHE Blows? Gender, competition, and bubbles in experimental asset markets," *American Economic Review*, 105, 906–920.

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

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

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.

Gneezy, U., A. Rustichini, and A. Vostroknutov (2010): "Experience and insight in the Race game," *Journal of Economic Behavior and Organization*, 144–155.

Haruvy, E., Y. Lahav, and C. N. Noussair (2007): "Traders' Expectations in Asset Markets: Experimental Evidence," *American Economics Review*, 97, 1901–1920.

Hefti, A., S. Heinke, and F. Schneider (2016): "Mental Capabilities, Trading Styles, and Asset Market Bubbles: Theory and Experiment," Working Paper Series 234, Department of Economics, University of Zurich.

Kindleberger, C. P. and R. Z. Aliber (2005): *Manias, Panics, and Crashes. A History of Financial Crises. 5th Edition*, Hoboken, New Jersey: John Wiley & Sons, Inc.

Korniotis, G. M. and A. Kumar (2010): "Cognitive abilities and financial decisions," in *Behavioral Finance: Investors, Corporations, and Markets*, ed. by H. K. Baker and J. R. Nofsinger, Hoboken, New Jersey: John Wiley & Sons, Inc., chap. 30, 559–576.

Levine, S. S., E. P. Apfelbaum, M. Bernard, V. L. Bartelt, E. J. Zajac, and D. Stark (2014): "Ethinic diversity deflates price bubbles," *Proceedings of the National Academies of Sciences*, 111, 18524–18529.

Lopez, L. (2015): "6can't read," *Business Insider*.

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

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

Porter, D. P. and V. L. Smith (1995): "Futures contracting and dividend uncertainty in experimental asset markets," *The Journal of Business*, 68, 509–541.

Powell, O. and N. Shestakova (2016): "Experimental asset markets: A survey of recent developments," *Journal of Behavioral and Experimental Finance*, 12, 14–22.

Proto, E., A. Rustichini, and A. Sofianos (2016): "Intelligence, personality and gains from cooperation in repeated interactions," Workinng paper, University of Warwick.

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.

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.

Stöckl, T., J. Huber, and M. Kirchler (2010): "Bubble measures in Experimental Asset Markets," *Experimental Economics*, 13, 284–298.

van Boening, M. V., A. W. Williams, and S. LaMaster (1993): "Price bubbles and crashes in experimental call markets," *Economics Letters*, 41, 179–185.

Xiong, W. and J. Yu (2011): "The Chinese warrants bubble," *American Economic Review*, 101, 2723–2753.

Table 6: Definition of the measures of mispricing

Relative absolute deviation (RAD)	$\frac{1}{20}\sum_{p=1}^{20}\frac{\lvert P_p-FV_p\rvert}{\lvert\overline{FV}\rvert}$
Relative deviation (RD)	$\frac{1}{20}\sum_{p=1}^{20}\frac{P_p-FV_p}{\lvert\overline{FV}\rvert}$
Relative Positive Deviation (RPD)	$\frac{1}{20}\sum_{p=1}^{20}\frac{\max(P_p-FV_p,0)}{\lvert\overline{FV}\rvert}$
Relative Negative Deviation (RND)	$\frac{1}{20}\sum_{p=1}^{20}\frac{\max(FV_p-P_p,0)}{\lvert\overline{FV}\rvert}$
Boom Duration	the greatest number of consecutive periods that prices are above fundamental values
Bust Duration	the greatest number of consecutive periods that prices are below fundamental values
Turnover	$\sum_{p=1}^{20} Q_p/24$

A Various measures of mispricing

In this appendix we reports the values of various measures of mispricing for each market. In addition to RAD, RD, RPD, RND, and turnover, we also report Boom and Bust Durations. Table 6 summarizes the definitions of these measures. Table 7 reports the results for 36 markets in known composition treatment and Table 8 reports the result for 24 markets in unknown composition treatment

Table 7: Known composition

Composition	group	RAD	RD	RPD	RND	Boom Duration	Bust Duration	Turnover
6H	1	0.007	0.006	0.006	0.000	12	1	1.292
6H	2	0.004	0.002	0.003	0.001	3	2	0.667
6H	3	0.006	-0.001	0.002	0.004	3	4	1.375
6H	4	0.028	-0.014	0.007	0.021	2	3	0.958
6H	5	0.153	0.130	0.142	0.012	3	3	1.583
6H	6	0.019	-0.008	0.006	0.014	2	4	0.792
6H	7	0.076	-0.010	0.033	0.043	5	5	1.917
6H	8	0.022	0.004	0.013	0.009	7	4	2.417
6H	9	0.456	0.317	0.387	0.070	10	7	0.875
6H	10	0.006	-0.001	0.003	0.004	5	2	1.208
6H	11	0.048	-0.043	0.002	0.045	3	5	1.125
6H	12	0.086	0.016	0.051	0.035	5	2	1.083
average		0.076	0.033	0.055	0.021	5	4	1.274
s.d.		0.128	0.098	0.112	0.022	3.191	1.679	0.500
6L	1	0.053	-0.002	0.026	0.027	3	7	0.958
6L	2	0.083	-0.080	0.002	0.081	1	8	0.833
6L	3	0.055	-0.025	0.015	0.040	3	5	1.083
6L	4	0.053	-0.045	0.004	0.049	4	5	0.583
6L	5	0.034	0.027	0.030	0.004	3	6	1.083
6L	6	0.022	-0.006	0.008	0.014	3	5	1.542
6L	7	0.245	-0.233	0.006	0.239	3	9	1.333
6L	8	0.239	0.215	0.227	0.012	13	2	2.333
6L	9	0.041	-0.039	0.001	0.040	2	5	1.500
6L	10	0.094	-0.059	0.017	0.076	3	7	0.667
6L	11	0.185	-0.176	0.004	0.180	4	7	1.250
6L	12	0.257	0.011	0.134	0.123	10	4	1.458
average		0.113	-0.034	0.040	0.074	4	6	1.219
s.d.		0.091	0.109	0.069	0.073	3.499	1.899	0.473
3H3L	1	0.149	0.045	0.097	0.052	7	9	1.000
3H3L	2	0.034	-0.022	0.006	0.028	3	7	1.042
3H3L	3	0.060	0.024	0.042	0.018	10	2	0.875
3H3L	4	0.410	0.366	0.388	0.022	8	2	1.208
3H3L	5	0.220	0.128	0.174	0.046	13	3	0.500
3H3L	6	0.165	-0.047	0.059	0.106	5	6	0.958
3H3L	7	0.257	-0.068	0.094	0.163	4	7	0.625
3H3L	8	0.221	-0.112	0.055	0.167	10	6	1.208
3H3L	9	0.231	0.084	0.157	0.073	12	3	1.208
3H3L	10	0.133	-0.032	0.050	0.083	4	7	1.083
3H3L	11	0.108	0.043	0.075	0.032	8	2	1.083
3H3L	12	0.664	0.046	0.355	0.309	9	5	1.083
average		0.221	0.038	0.129	0.092	7.750	4.917	0.990
s.d.		0.171	0.124	0.123	0.085	3.251	2.429	0.226

Table 8: Unknown composition

Composition	group	RAD	RD	Positive Deviation	Negative Deviation	Boom Duration	Bust Duration	Turnover
6H	1	0.007	-0.004	0.002	0.006	2	2	1.083
6H	2	0.040	-0.008	0.016	0.024	3	2	0.917
6H	3	0.062	-0.028	0.017	0.045	5	5	1.375
6H	4	0.054	-0.046	0.004	0.050	4	7	1.208
6H	5	0.018	-0.014	0.002	0.016	2	4	1.000
6H	6	0.112	-0.112	0.000	0.112	0	17	1.042
6H	7	0.026	0.022	0.024	0.002	10	2	1.083
6H	8	0.017	0.017	0.017	0.000	10	0	1.042
average		0.042	-0.022	0.010	0.032	4.500	4.875	1.094
s.d.		0.034	0.043	0.009	0.037	3.703	5.357	0.140
6L	1	0.006	-0.006	0.000	0.006	0	11	1.042
6L	2	0.002	0.002	0.002	0.000	2	0	1.833
6L	3	0.194	-0.048	0.073	0.121	7	8	0.792
6L	4	0.083	0.061	0.072	0.011	12	2	1.292
6L	5	0.004	-0.002	0.001	0.003	2	5	0.500
6L	6	0.046	0.030	0.038	0.008	5	2	1.417
6L	7	0.117	0.033	0.075	0.042	6	2	1.375
6L	8	0.039	0.015	0.027	0.012	7	2	0.542
average		0.061	0.011	0.036	0.025	5.125	4.000	1.099
s.d.		0.067	0.032	0.034	0.041	3.796	3.742	0.466
3H3L	1	0.244	-0.025	0.110	0.135	6	4	0.833
3H3L	2	0.408	0.072	0.240	0.168	8	4	1.542
3H3L	3	0.009	-0.003	0.003	0.006	6	3	1.333
3H3L	4	0.046	0.004	0.025	0.021	4	4	1.042
3H3L	5	0.598	0.567	0.583	0.016	18	1	1.375
3H3L	6	0.062	-0.012	0.025	0.037	2	11	0.958
3H3L	7	0.248	0.144	0.196	0.052	9	4	0.875
3H3L	8	0.003	-0.003	0.000	0.003	0	1	2.333
average		0.202	0.093	0.148	0.055	6.625	4.000	1.286
s.d.		0.215	0.200	0.198	0.063	5.476	3.117	0.495