Journal of Economic Dynamics & Control 000 (2018) 1–32

Contents lists available at ScienceDirect

Journal of Economic Dynamics & Control

journal homepage: www.elsevier.com/locate/jedc

Cognitive ability and earnings performance: Evidence from double auction market experiments

Chung-Ching Tai[a,*], Shu-Heng Chen[b], Lee-Xieng Yang[c,d]

[a] Department of Economics, Tunghai University, No.1727, Sec. 4, Taiwan Boulevard, Xitun District, Taichung 40704, Taiwan, ROC
[b] AI-ECON Research Center, Department of Economics, National Chengchi University, NO.64, Sec. 2, ZhiNan Rd., Wenshan District, Taipei City 11605, Taiwan, ROC
[c] Department of Psychology, National Chengchi University, NO.64, Sec. 2, ZhiNan Rd., Wenshan District, Taipei City 11605, Taiwan, ROC
[d] Research Center for Mind, Brain, and Learning, National Chengchi University, NO.64, Sec. 2, ZhiNan Rd., Wenshan District, Taipei City 11605, Taiwan, ROC

ARTICLE INFO

Article history:
Received 14 August 2017
Revised 30 October 2017
Accepted 20 December 2017
Available online xxx

Keywords:
Cognitive ability
Working memory capacity
Double auction markets
Bidding behavior
Earnings performance
Artificial traders

ABSTRACT

Our goal in this paper is to understand how heterogeneity in people's cognitive ability leads to different market behavior, and thus different market performance. To do this, subjects with heterogeneous working memory capacity (WMC) were placed in a double-auction environment to compete against artificial traders. We considered two treatments which differ in the artificial traders. The artificial traders are truth-telling in the first treatment, but demonstrate adaptive trading behavior in the second one. Our results show that working memory capacity has a significantly positive effect on subjects' market performance, and the performance gap caused by cognitive ability, while narrowing over time, remains significant by the end of experiment. We find that differences in subjects' performance resulted from their behavior: high-WMC subjects were better at exploiting extra profit opportunities and avoiding unprofitable transactions, and they tended to underbid more than those with lower WMC. Among the five constituent abilities of WMC, we find that it is distinctive abilities which contribute to the overall significance in these two treatments. For the treatment involving truth-telling traders, the relevant factor is the ability of simultaneous processing and storing information; whereas, for the treatment involving adaptive traders, the only one that matters is subjects' ability to coordinate elements into structures.

1. Introduction

When characterizing economic agents in the markets, the classical approach in economic theory is to assume that investors are homogeneous and rational. Models based on rational and homogeneous (and therefore representative) agent is the backbone of macroeconomic and financial theories, such as the Efficient Market Hypothesis (EMH) and the Capital Asset Pricing Model (CAPM). However, such assumptions (rational and homogeneous) about agents have been greatly challenged by empirical studies in recent decades.[1] Acknowledging these findings, economists have commenced an important paradigm

* Corresponding author.
E-mail address: chungching.tai@thu.edu.tw (C.-C. Tai).
URL: http://web.thu.edu.tw/chungching.tai/www/ (C.-C. Tai)

[1] For example, Allen and Taylor (1990); Frankel and Froot (1987a, 1987b, 1990a, 1990b); Ito (1990), and Taylor and Allen (1992) all found that practitioners in foreign exchange markets employ different trading and forecasting strategies. Vissing-Jorgensen (2003) found that investors have heterogeneous beliefs,

https://doi.org/10.1016/j.jedc.2017.12.008
0165-1889/

shift from homogeneous and rational agent approach towards an approach where market participants are boundedly rational and heterogeneous (Hommes, 2006). For example, the most flourishing strand of research is to assume that people in the markets have different expectations (or beliefs).[2]

Heterogeneous agent models (HAMs) can explain stylized facts well. However, most of the models have agents differing only in their behavior, and their behavior is predetermined or depends on a mechanism which is uncorrelated with personal factors. For example, in heterogeneous expectation models, researchers have to either determine the compositions of agents or design a switching mechanism for agents to choose among different expectation heuristics. The underlying assumption of the latter is that agents are homogeneous in their tendencies toward different beliefs, or they have the same intensity of choice when choosing among different expectation heuristics. The result is that the fractions of agents is the outcome of a stochastic process and independent of agents' personal factors. Fortunately, there are increasing efforts trying to go deeper and ask whether heterogeneity in traders' behavior stems from more fundamental factors. For example, Bosch-Rosa et al. (2015) found that subjects with higher levels of cognitive sophistication predicted asset prices closer to the fundamental values.

In fact, the idea of linking behavioral heterogeneity and its economic outcomes to personal factors is not new in the history of economic thoughts. Vilfredo Pareto, who is well-known for his study of income and wealth distribution, when speculating about the causes of economic inequalities, submitted the notion of *social heterogeneity*:

> "Human society is not homogeneous; it is made up of elements which differ more or less, not only according to the very obvious characteristics such as sex, age, physical strength, health, etc., but also according to less observable, but no less important, characteristics such as intellectual qualities, morals, diligence, courage, etc." (Pareto, 1971, Chap II, 102)

In this paper, we focus on one of the factors Pareto has pointed out: *intellectual qualities*.

How do people's intellectual qualities, or cognitive abilities, affect their market behavior has been an attractive issue for many empirical and experimental economists in recent years. Empirically, by investigating people's cognitive ability and their financial portfolios, both Christelis et al. (2010) and Grinblatt et al. (2011) found that people with high cognitive ability invest more in stock markets. Specifically, Grinblatt et al. (2011) found that the stock market participation rate of individuals with IQs at the lowest end (lowest stanine) is 15.4% lower than that of individuals at the other end of the spectrum. This IQ effect is far larger than income's effect on participation. Grinblatt et al. (2012) found that high-IQ investors had better trading skills and were less susceptible to the disposition effect, and their portfolios displayed better performance results.

Experimental evidence also emerges rapidly. By comparing subjects' actual bids to the theoretical break-even bids, Casari et al. (2007) examined whether subjects suffer the winner's curse in common value auctions. They found that those whose SAT/ACT (American College Test) scores are below the median are more susceptible to the winner's curse. Bosch-Rosa et al. (2015) observed the bubble and crash patterns in experimental asset markets composed of subjects with low levels of cognitive sophistication, while no bubbles or crashes occurred in market sessions with sophisticated subjects. Cueva and Rustichini (2015) found that the higher the subject group's average cognitive skills, the lower the market volatility in an asset market. Corgnet et al. (2015) motivated subjects in asset markets with two different schemes–earned money v. house money. They found that no matter where the endowments came from, the higher subjects' CRT (Cognitive Reflection Test) scores, the more money they earned. To be more specific, subjects with lower CRT scores tended to buy shares when prices were above fundamental values, and sell shares when below the fundamental values. Breaban and Noussair (2015) had two treatments: a bear market and a bull market. They observed a significant correlation between CRT scores and being a fundamental value trader in Market 1; additionally, the greater the average CRT score, the smaller the differences between market prices and fundamental values. Noussair et al. (2016) had both a spot market and a futures market within which participants could trade. They found that the average CRT score of a trader cohort is negatively correlated with the degree of mispricing when no futures market was present, and this relationship disappeared if a futures market existed. They also found that traders with higher CRT scores had greater earnings no matter whether a futures market was presented or not.

Although the aforementioned findings are valuable, we are interested in knowing how cognitive ability influences people's trading behavior in a more fundamental way. We want to investigate whether the link exists and provide the basis for future heterogeneous agent modeling not only for asset markets but also for other market activities. Note that markets could have different pricing mechanisms and institutional factors, and trading against other market participants involves complex strategic interactions. If researchers want to understand whether inherent cognitive ability limits the benefits people can elicit from mutual exchanges, they must account for these factors. This is not an easily achievable task, however. We therefore think it is a good strategy to begin within a simple but standard environment. Based on our findings in this simple setting, economists, through continued studies, may fortify our understandings of how cognitive ability manifests its influences in people's market activities.

and Elliott and Ito (1999); MacDonald and Marsh (1996), and Bénassy-Quéré et al. (2003) fount that trader's expectations are also heterogeneous in foreign exchange markets.

[2] Examples of heterogeneous belief models include Brock and Hommes (1997a, 1997b, 1998); Bullard and Duffy (1999); Chiarella and He (2002); Chiarella et al. (2006, [2007,2013,2015]); Chiarella and He (2001, 2003a, 2003b); Chiarella et al. (2012, 2009); Day and Huang (1990); Franke and Nesemann (1999); Franke and Sethi (1998); Gaunersdorfer (2000); Hommes (2001); Hommes and Lux (2013); Lux (1995, 1997,1998); Lux and Marchesi (1999, 2000); Sethi (1996).

The present study responds to this quest using controlled experiments. By conducting individual-based market experiments where each human subject trades against computer agents, we were able to better control the factors influencing subjects' performance and therefore identify the effects of cognitive ability with higher confidence. Firstly, we directly assigned reservation prices to our subjects, so there is no room for judgment bias as in Casari et al. (2007), and also no need to calculate fundamental prices as in experiment stock markets. Secondly, the double auction mechanism is easier to understand in a relatively short time, so subjects should hardly have any confusion about how market works as observed in Noussair et al. (2016). Thirdly, our subjects competed with computer agents, so other-regarding preferences should play no roles here, and it is less likely that theory of mind would manifest its influence as in asset bubbles observed in De Martino et al. (2013). As a result, what we are interested in is purely cognitive ability's influences on subjects' ability to sense market situations and their trading skills. We also conducted two experiments with different trading agents so that we could discuss whether the effects dissipate or are amplified under strategic interactions. Finally, by employing a formal psychometric measurement of cognitive ability, that is, *working memory capacity* (WMC), we hope that the analytical results reported in this study will inspire advanced and in-depth exploration of the behavioral foundations of economic heterogeneity.

In this study, we want to thoroughly examine the link between cognitive heterogeneity and heterogeneity in people's market activities from the superficial level to the psychological level. We first inspect how market performance can be explained by cognitive ability, followed by discussions about the interaction between cognitive ability and learning. Secondly, we explore how cognitive ability manifests its influences in the behavioral level. Lastly, we go to the psychological level and ask which facets of cognitive ability play significant roles in different market circumstances.

The results of our experiments show that the larger a subject's WMC, the better their performances in the double auctions. Although all subjects did improve a lot due to learning, it could not entirely compensate for having a lower WMC. Most importantly, subjects with different levels of WMC indeed had different behavioral tendencies in double auction markets. To be exact, subjects with higher WMC had a stronger tendency to underbid, and this tendency did not seem to change over time. Finally, when markets are static, it is the storage and transformation function of WMC that can explain the differences in market performance. On the contrary, it is the coordination ability of WMC that plays a dominant role in explaining heterogeneous performance in a more sophisticated market environment.

The rest of the paper is organized as follows. Section 2 introduces our experimental design and procedures. Section 3 and Section 4 present the experimental results of the two treatments involving different types of artificial agents. Section 3 gives the results of the simple scenario, where human agents trade with 'honest' or 'simple' artificial agents. Section 4 gives the results of the complex scenario, where human agents trade with adaptive and strategic artificial agents. Section 5 gives a discussion of the constituents of WMC and their work within different market scenarios. Section 6 presents our conclusion.

2. Experimental designs and procedures

Our research has two distinguishing features: the individual-based market experiments and a formal psychometric test. The design of individual-based market experiments aims to insulate subjects' earning capability in markets from complex strategic interactions which are difficult to track in the standard human experiments.[3] It helps us to focus on the effect of individuals' cognitive ability on trading behavior.[4] Using a formal psychometric test can further help pinpoint the cognitive tasks required to perform certain bidding or asking decisions in a competitive environment. It enables us to ask not just whether cognitive ability matters, but also why, if it does matter. In this section, we will introduce how we achieved the above goals with a series of designs.

2.1. The double auction markets

In our double auction experiments, each market consists of four buyers and four sellers. Following the design of Rust et al. (1993, 1994), each trader is granted four tokens.[5] As a result, there will be sixteen buyer tokens and sixteen seller tokens in the market. These tokens determine the supply and demand schedules of the market.[6] Fig. 1 illustrates the supply

[3] The method of using computer agents as human subjects' companions is not new to experimental economics. For example, Davis and Williams (1991) and Davis et al. (1993)'s market experiments as well as Devetag and Warglien (2003)'s experimental games.

[4] One may argue that in real situations people interact with each other, and we should take this complicated interaction into account if the potential effects of cognitive ability worth any discussion. However, if human interactions are allowed in a market with only a few human subjects, it will be difficult to clearly identify cognitive ability's effects on earnings due to the complex interactions. Let alone the possibility that human subjects may choose strategies contingent on others' strategies, which is hard to know for experimenters. By putting a single human subject in a market with computer agents, we will have better knowledge of other market participants' goals and strategies and therefore have a better explanation for human subjects' behavioral differences.

[5] For buyers, the token value can be viewed as the *maximum willingness to pay* (WTP) for the token to acquire. For sellers, the token value can be viewed as the *minimum willingness to accept* (WTA) for the token to sell.

[6] Each trader's four tokens are evenly distributed along the demand or supply curve. In Market 1, for example, only eight out of sixteen pairs of tokens can make transactions. As a result, every trader in Market 1 will have two units bought (sold) if everyone bids/asks according to their token values. See Table B.14 in Appendix B for detailed numbers.

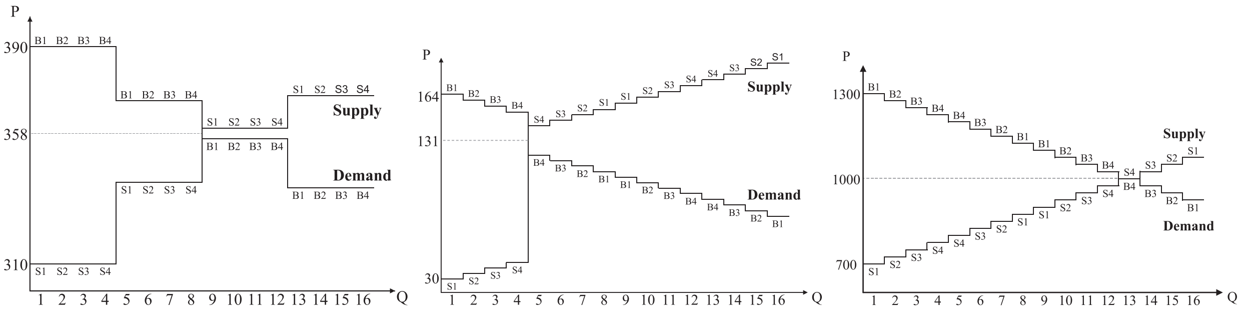

Fig. 1. Demand and supply schedules for the auction experiments. From the left to the right are Market 1 (M1), Market 2 (M2), and Market 3 (M3), respectively.

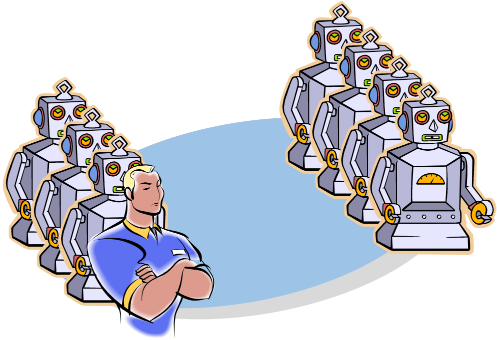

Fig. 2. Market participants: Human agents and robot traders.

and demand curves of the three markets used in our experiments. The three markets, named as Market 1 (M1), Market 2 (M2), and Market 3 (M3), differ in their trading opportunities as well as the potential surplus for buyers and sellers.[7]

As shown in Fig. 1, the trading opportunities as characterized by the number of intra-marginal tokens differ in these three markets; the number of intra-marginal tokens each trader has in M1, M2, and M3 are two, one, and three, respectively. The differences in trading opportunities make profitability in each market different from one another. Among the three, M3 is the easiest one because each buyer (seller) has three potentially profitable tokens to buy (sell). With this ease, the subject can make more attempts to buy or sell their tokens with profitably. This ease of process, however, does not exist in M2, where each subject has only one intra-marginal token. If the subject does not bid/ask properly and allows their competitor holding an extra-marginal token to steal a trade, then they will miss their only chance and earn nothing. Following this line of reasoning, one can see that M1 lies between M2 and M3, i.e., it is easier than M2 but harder than M3.[8]

Each subject was assigned as one of eight traders in the market, and their given position would remain the same throughout the experiments.[9] Seven robot traders (software agents) (Fig. 2) accompanied each human subject, who were informed of this arrangement before the experiments started. We used software agents to design the markets with naive opponents and the markets with sophisticated opponents. The ease or difficulty of markets and opponents were then combined to shape the market environment with which to test the significance of subjects' cognitive ability.

Subjects played the double-auction experiments following a sequence of three market sessions: M1 → M2 → M3. We defined this sequence from the medium to hard, and then to the easy (the 'MHE' sequence), mainly due to the following considerations. First, it helped us to distinguish the novice effect from the hardness effect. If we commence the task with the hard market, we may not have been able to answer why subjects did not perform well in this market. Was it because they did not have experience or they were simply given a particularly hard problem? Secondly, in a similar vein, it also helped us distinguish the experience effect from the cognitive-ability effect. If we arrange the markets at the sequence of easy-medium-hard, then the successes in these markets could either be because subjects gain experiences through time

[7] We can see that Market 1 and Market 2 are not completely symmetric. As a result, Market 1 and Market 2 might be more beneficial for sellers, while buyers and sellers have exactly the same opportunity in Market 3.

[8] We need to emphasize that the difficulty of markets is defined based on the difficulty of making profitable trades (the number of potential trades), and it is not necessarily related to the amount of profits subjects could earn in the markets. For example, a buyer subject in M2 can underbid the first token, thereby lowering the transaction price and earning a huge amount of profits. Although it is more difficult to do so in M2, once a subject succeeds, the potential payoff could be large. On the other hand, making profitable trades in M3 is easier, but the profits earned might not be as large as that in M2.

[9] For example, once a subject was assigned as the second seller in the market, they will keep playing the role of the second seller throughout the three markets.

or because their cognitive capacities are large enough to tackle all three market problems. Alternatively, if we applied the medium-hard-easy order, then the 'lack of experience' excuse presents no defensible position for the hard and easy markets, and subjects only needed to deal with a medium problem when they had no experience at the beginning.

Each market experiment lasts for six recurring periods, and each period consists of twenty five rounds. At the beginning of each round, subjects submit bids and asks, and as in the call market these bids and asks are matched at the end of the round (see also Section 2.2). Before each period starts, the market supply and demand have to be reset back to their given schedules (Fig. 1), and another twenty-five rounds (calls) follow. We have twenty-five rounds in a period because the number is large enough for all possible transactions to take place according to our preliminary tests. We only inform subjects of their own token values, and subjects can judge from the market information how many traders are in the market (see Appendix C, Fig. C.12).

During the experiments, we reveal the past history of every trader's bids/asks and transaction prices on the screen. The details of this information disclosure policy is presented in Appendix C (Fig. C.12). Certainly, what we have tested here is not whether subjects can recite the historical numbers, but whether they can discover and exploit useful patterns from the data. To this end, subjects were given neither paper and pencil nor calculators to aid their decisions.

2.2. The auction process

We adopted discrete-time double auctions in our experiments. Each market experiment consisted of six trading periods, and each trading period consisted of twenty-five rounds (steps). Each trader could choose to bid/ask or pass in each round. During each round, both the buyer who had the highest bid (the current buyer) and the seller who had the lowest ask (the current seller) had an opportunity to reach a transaction.[10] If the current buyer's bid (the current bid) was higher than current seller's ask (the current ask), a transaction would take place and the price is the average of the bid and the ask. If the current bid was lower than the current ask, nothing happened and the auction would proceed to the next trading round.

Note that there are two important features of our auction rule. First, speed is not important, because in each round we only determine who the current buyer and seller are after we receive all the decisions from traders. This feature prevents our subjects to be outperformed by computer agents simply because of the differences in their decision and action time. Secondly, the price is the mean of the current bid and the current ask.

Traders' bids or asks are not constrained by their token values, although their profits are calculated by the differences between the transaction prices and their token values. A buyer's profit is defined as their token value minus the transaction price; a seller's profit is defined as the transaction price minus their token value.

2.3. The trading agents

In order to delineate different degrees of sophistication among opponents for our human subjects, several types of software agents were utilized to compete within the markets. We commenced with a market environment in which the constituent opponents were simple (the simple opponents). Hence, in Experiment 1, a kind of software agent named *the truth teller* was used. A truth teller will always bid/ask according to maximum (minimum) WTP (WTA), i.e., their token values. A truth-telling robot trader is not adaptive and will bid/ask in the same way from period to period.

In Experiment 2, the opponents were adaptive (sophisticated opponents). These sophisticated opponents can adapt or learn over time. To make the auction environment more realistic, we employed seven different software agents, and all human subjects would face the same software agents in the experiments. These software agents are borrowed from the existing literature on double auctions: the Easley–Ledyard agent (Easley and Ledyard, 1993), the Skeleton agent, the Kaplan agent, the Ringuette agent (Rust et al., 1993), the zero-intelligence plus agent (Cliff and Bruten, 1997), the belief-based learning agents (Gjerstad and Dickhaut, 1998), and the Markup agent (Zhan and Friedman, 2007). These agents were chosen from the top seven performing software agents in the agent-based double auction tournament conducted in Chen et al. (2010).

Before the experiment, we made it clear to the subjects that all of their opponents would be software agents rather than human subjects. However, we did not reveal any further information about the software agents' strategies or their exact origins.

2.4. The measurement of performance

As mentioned earlier, we wanted to examine whether cognitive ability has a general impact on human subjects' market performance. One natural measure for market performance is the profit earned from trading. However, the raw profit data is not invariably comparable among different subjects because their profit potentials vary with the assigned trading positions. Therefore, the raw profit has to be normalized before suitable comparisons can be made. Our proposed performance index is then based on the following normalization.

$$\text{Performance index} = \frac{\text{The actual profit earned}}{\text{The potential profit}} \times 100 \tag{1}$$

[10] If there is a tie between traders, the system will randomly choose one as the current buyer/seller.

The potential profit is derived based on what we call the *fair-share price*. The fair-share price is simply the middle point of the token values of the corresponding buyers and sellers along the supply and demand schedules (Fig. 1), unit by unit, from the left to the right until match is no longer possible. The potential profit of each token is then derived using the fair-share price of that token, and the potential profit is the sum of the potential profit of all tokens that a trader has. In other words, the potential profit is based on an anchor to the 50% division of the potential surplus. Everyone no matter where they are placed, their performance shall be compared to how much better they can do when compared to this 'fair share'. Therefore, this index provides a level playing field for all subjects. An index of 100 implies that the performance of the subject was just good enough to win their fair share; an index greater than 100 indicates that subject was active enough to win more than a fair share, and vice versa.

The fair-share prices are the prices one observes when all market participants are truth-tellers. The truth-telling strategy is a kind of minimalist trading strategy, very much in the same vein of the buy-and-hold strategy in the financial market. This feature makes truth-telling naturally stand out as a benchmark.[11] In what follows, when referring to subject's "profit" we actually mean "performance index" in each period, and "wealth" means cumulated performance indexes over six periods.

2.5. Working memory capacity

There are many options for researchers to assess subjects' cognitive ability. In the present study, we chose *working memory capacity* (WMC) instead of using a general IQ or other test scores (such as SAT scores). Working memory is a psychological construct which is thought of as a complex system used not only for storing information but also processing information (Baddeley and Hitch, 1974). WMC is not simply a measurement of short-term memory capacity, but a "conceptual ragbag for everything that is needed for successful reasoning, decision making, and action planning" (Oberauer et al., 2003). To be more precise, Oberauer et al. (2000) summarized the functions of WMC as follows:

1. *Storage and transformation*–how people simultaneously process and store information.
2. *Supervision*–how humans monitor and control ongoing mental operations and actions, including selectively activating relevant representations and processes and inhibiting irrelevant ones.
3. *Coordination*–how humans coordinate information elements into structures.

These functions presumably are generally required when making decisions.

WMC is positively correlated with the performance in many cognitive tasks, including reading comprehension (Baddeley and Hitch, 1974; Daneman and Carpenter, 1983; King and Just, 1991), decision making (Hinson et al., 2003), categorization (Craig and Lewandowsky, 2012; Lewandowsky, 2011; Lewandowsky et al., 2012), Berlin Intelligence Structure Model (BIS) (Süß et al., 2002), and fluid intelligence (Engle et al., 1999; Kane and Engle, 2002). The reasons why we chose WMC instead of IQ or personality are as follows. First, WMC is a relatively micro/precise index than IQ to address the individual differences in cognitive tasks. Second, there is a large deviation in the statements about the construct of intelligence among psychologists. Some think only one factor is enough to represent intelligence, whereas others might support an account of three factors or even more. However, psychologists have a larger consensus about WMC such that it can be viewed as a general cognitive resource for doing different tasks. Third, we do not argue that the performance in our double auction experiments has nothing to do with personality. However, in this study, we would like to focus on the cognitive facet more than personality or motivation.

Working memory is the extension of short-term memory (STM). Traditionally the efficiency of working memory is defined in terms of the number of items which a person can correctly recall under the interruption of another cognitive processing. Although there are many different WMC tasks in the literature, the way of measuring WMC makes no difference in predicting individuals performance in cognitive tasks (Turner and Engle, 1989). Following the general capacity hypothesis, we can expect that different WMC tasks can be viewed as parallel tests measuring the same psychological construct. Lewandowsky et al. (2010) published their working memory battery with four different tasks (OS, SS, MU, and SSTM) and showed via structural equation modeling that these tasks can be linked to a single latent variable which is referred to WMC. Generally speaking, a person with a larger WMC also has a larger attention span to undertake cognitive processing (Conway and Engle, 1996) and is better able to ignore the irrelevant but potentially distracting information (Neath and Surprenant, 2003).

There are five WMC tasks employed to measure subjects WMC in this study. The details about each of them are described as follows.

- The operation span task (OS): In this task, subjects are given a series of consonants and asked to recall them in the correct order. In the learning phase, the presentation of consonants is intersected by a mathematic equation verification task, in which the subject has to verify a mathematic equation (e.g., $1 + 4 = 7$, correct or wrong?). After they verify the equation, a corrective feedback will be presented to them. The subjects are asked to try their best to be as accurately as possible in this equation verification task. In total, there are 15 trials, 3 trials per list length (from 3 consonants to 7 consonants). The task accuracy is the measurement of WMC.

[11] Another popularly used anchor is the theoretical competitive equilibrium price. When the market is symmetric, the theoretical competitive equilibrium price is identical to the fair-share price. However, when the market is not symmetric, then two shall deviate to each other. We have actually tried both benchmarks, but found no substantial difference. Therefore, in this paper, we only show the results using the fair-share price as the benchmark.

- The sentence span task (SS): This task is almost the same as the OS task, except that the equation verification task now is changed to the sentence verification task. Again, subjects will be given a sentence in between two consonants and asked to judge whether this sentence is syntactically correct. After they make their response, a corrective feedback will be presented to them. The task accuracy is the measurement of WMC.
- The memory-updating task (MU): In this task, on each learning trial, subjects will see a number of boxes on the computer monitor. The number of boxes varies from 2 to 6, each for 3 trials. Thus, in total, there are 15 trials. At the onset of a trial, all boxes are empty. Take the number of boxes = 2 as an example. At the onset of a trial, two boxes on the computer screen are empty. Subsequently, one digit appears in one of the two boxes, say 3, and then disappears. Another digit appears in another box, say 2, and then disappears. In the first box, there appears an operator, say +3. Now the subjects need to update the digit in the first box from 3 to 6. Again, +3 disappears and −1 appears in the second box. Now the subjects have to update the digit in the second box from 2 to 1. When the symbol ? appears in the first box, the subjects have to report the latest digit in this box. For instance, if ? appears in the first box, the latest digit in this box is 6, so the subject should report 6. Again, the accuracy in this task is the measurement of WMC.
- The spatial short-term memory task (SSTM): Different from other tasks used in this study, there is no digit nor consonant in this task. The subjects will be presented a 10×10 grids at the onset of every trial, followed by the presentation of a series of dots one after one each lasting for 0.9 s. When all dots are presented, the screen will be cleared and an empty 10×10 grid appears. The subjects are asked to use the mouse cursor to point in the locations of the previously seen dots. There is no need to point the dots in the same location nor the same presenting order. The similarity between the pattern of restored dots and that of previously seen dots is the measurement of WMC. In order to be parallel to the previous tasks, the similarity is normalized between 0 and 1.
- The backward digit span task (BDS): In this task, a series of digits are presented to the subjects who are asked to recall them in a backward order. The list length is from 4 to 8, and there are 6 trials for each length. Thus, in total, there are 30 trials. Again the subjects' accuracy in this task is the measurement of their WMC.

There were several trials for each test battery, and a score was computed for each test. The score for each subtest was then normalized with the mean and standard deviation of the scores derived from our subject pool, i.e., the 346 subjects who completed the tasks. Then a single measure of WMC was derived by averaging these five normalized scores.[12]

2.6. The experimental procedures

Our experiments were conducted from 2009 to 2010 in the Experimental Economics Laboratory (EEL) of National Chengchi University. Both the experiments and WMC tests were computer-based. The double auction environment and the computerized trading agents were programmed using Java, while the WMC test ran in Matlab with psychtoolbox.

At the beginning of the experiments, subjects were asked to be seated and the computers they used determined their roles in the markets. We set-up the computers according to the Latin Square Design in order to distribute the subjects evenly to every role in the market.[13] When reading the instructions, we emphasized the fact that they would be playing against software agents instead of other human subjects in the room.

A three-period trial run was conducted before formal experiments, and subjects' opponents in the trial run were zero-intelligence traders with constraint (ZI-C). The ZI-C trader was proposed by Gode and Sunder (1993). These computerized traders sent bids (asks) randomly in a range constrained by their token values. Subjects were told that the computer agents in the trial run would send random bids and asks, and we reminded them not to pay attention to agents' behavior in the trial run due to agents' randomness. We used ZI-C traders in the trial run simply to familiarize our subjects with the pricing mechanism and the required operation procedures.

The experiments always started in the morning as it took about two hours to finish the double auction experiments. Once the auctions were done, we provided lunch boxes and subjects had one-hour of free time at their disposal. At 1:00 p.m., they returned to the lab for the WMC tests. After the WMC tests, every subject was asked to fill out a questionnaire before collecting the monetary reward.

The participation fee for the double auction experiment is NT$200.[14] A subject will be rewarded an extra NT$250 if they are the champion in their designated market (beating the other seven robot traders), NT$150 if they are the first runner-up,

[12] Our research starts when Lewandowsky et al. (2010) is still developing their test batteries, so we use five test batteries instead of the four finally appear in Lewandowsky et al. (2010).

[13] The Latin Square Design has been used in experimental economics (Davis and Holt, 1993).

> "When there are a large number of nuisance variables, it is possible to economize on the number of observations per cell by using partial blocking designs, such as Latin square and/or lattice procedures." (Ibid, p. 524)

Using the Latin Square Design, the numbers of subjects assigned to the trading positions at B1, B2, B3, B4, S1, S2, S3, and S4 in Experiment 1 are 26, 23, 15, 23, 17, 26, 21, and 19, respectively, and they are 26, 23, 15, 24, 20, 22, 20, and 19 in Experiment 2. The numbered trading positions are trader indexes. For example, B2 denotes the second buyer, and S4 denotes the fourth seller. Each trader had four units to trade and was assigned four token values, as can be seen in Fig. 1.

[14] In terms of hourly wage rate, the participation fee was roughly the same as the minimum wage rate in Taiwan, which was NT$95 per hour in 2009 and NT$98 per hour in 2010. We also paid subjects a fixed fee of NT$300 for the WMC test in the afternoon.

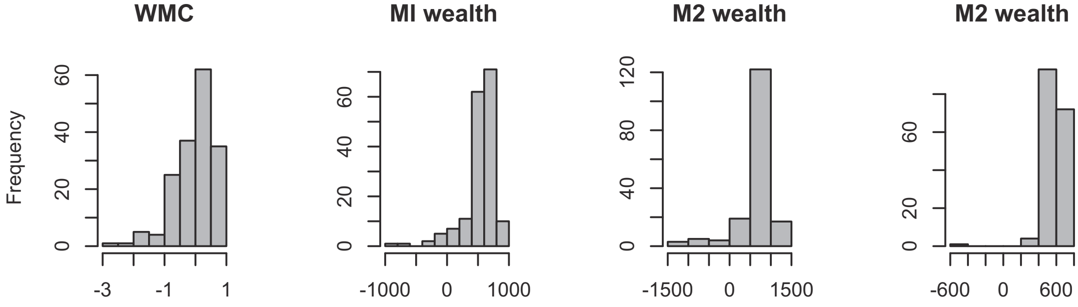

Fig. 3. The distribution of subjects' WMC scores and their cumulated performance (wealth) in Experiment 1.

and NT$75 if they are the second runner-up. We used subjects' rankings to pay bonuses instead of calculating bonuses with an exchange rate because our subjects could earn a good profit even if they lazily did nothing but sent simple orders equal to their token values. In other words, we wanted to encourage our subjects to earn extra profit as much as possible in order to beat other market participants.[15]

Most of our subjects were undergraduates, and some of them were graduate students. We conducted the recruiting job through the Internet, and our subjects were from different departments and different universities. For each treatment (simple-opponent markets and sophisticated-opponent market) we successfully recruited 173 subjects (a total of 346 subjects). We did not accept experienced subjects, so all of these 346 subjects attended only one of the two treatments and only once for that treatment. For Experiment 1, we had 82 males and 91 females; for Experiment 2 we had 97 males and 76 females. Detailed information about our subjects can be found in Table D.15 in Appendix D.

3. Experiment 1: Trading with simple agents

We started our analysis from Experiment 1, where human subjects' opponents in the market were non-adaptive truth-telling agents (easy opponents). There were a total of 173 subjects, each attending one of the 12 easy-opponent sessions. Not all of the data from our subjects was valid.[16] After removing the invalid cases, we ended up with a size of 170 subjects; among them, 80 are males and 90 are females.

Before looking into the relationship between WMC and subjects' performance, we first check whether the aggregate outcomes of our experimental auctions are consistent with the convergence phenomenon repeatedly observed over the past few decades (the detailed results are presented in Appendix E).[17] The result is that our markets indeed exhibit rapid convergence toward equilibria since period 1. Additionally, the prices in asymmetric markets (M1 and M2) converge to equilibria from below (Fig. E.13), and this is exactly what Smith (1962) observed in his pioneering double auction experiments. Market efficiencies are also close to 100% throughout the experiments for all three markets, suggesting that the results of our experimental auctions are consistent with past studies and no abnormal patterns exist in the aggregate level.

To infer the possible relationship between WMC and subject's market performance, we first compared the distributions of these variables. Fig. 3 demonstrates these distributions, and Table 1 gives the descriptive statistics. The distributions show that both WMC and subjects' profits in each market are left-skewed, and the statistics confirm that the distributions are far from normal. One may wonder why the wealth distributions are left-skewed instead of a commonly seen right-skewed Pareto distribution. The reason being that there is an upper limit on a trader's profit in our auction market, yet it is also easy to lose and the potential losses one may incur are theoretically unlimited. The similarity of these distributions further propels our observations of what may connect these variables.

3.1. Can WMC predict subjects' performance?

A multiple regression model is proposed to examine the effect of WMC on subjects' cumulated profits (wealth) in each market. The regression, which we shall call the earning equation, is given in Eq. (2),

$$E_i = \alpha + \beta_1 WMC_i + \beta_2 Male_i + \beta_3 Buyer_i + \beta_4 EX1_i + \beta_5 EX2_i + \beta_6 EX3_i + \beta_7 Tool_i + \epsilon_i \tag{2}$$

[15] In our two treatments, the percentage of students won these addition rewards is high. For Experiment 1, 52% of the subjects won the 'award'–11% won the champion, 21% won the first runner-up, and 20% won the second runner-up. For Experiment 2, 69% of the subjects won it–20% won the champion, 23% won the first runner-up, and 27% won the second runner-up.

[16] This normally happened when they mistakenly keyed in a number by adding or missing a digit. Such mistakes would make their profit abnormally low. Subjects with these mistakes easily become outliers of the data; their performance index can be below −1000 in the period that these mistakes occurred. A total of three subjects of this kind were then dropped from our final dataset.

[17] Ever since Smith (1962)'s pioneering experiments, price behavior in the double auction mechanism have been tested repeatedly over the past decades. Although experimental economists observed a few factors that could influence the speed or direction of price convergence, convergence toward the competitive equilibrium in double auction institution has been recognized as a robust phenomenon over "at least a thousand sessions in a variety of designs" (Holt, 1995).

Table 1
Descriptive statistics of Experiment 1.

	WMC	Wealth (M1)	Wealth (M2)	Wealth (M3)
Mean	−0.03	538.36	591.99	576.19
Median	0.105	597.5	640	591
Maximum	1	862	1157	723
Minimum	−2.66	−990	−1458	−500
Std. Dev.	0.64	260.61	427.54	105.13
Skewness	−1.10	−2.50	−2.51	−6.73
Kurtosis	1.63	8.83	7.57	63.65
p-value of Shapiro–Wilk Test	3.77e-07	5.50e-15	$<$ 2.20e-16	$<$ 2.20e-16

Table 2
Explaining variables of the multiple regression model of Experiment 1.

Variable	Definition
Male	An indicator variable equal to 1 if subject i is male.
Buyer	An indicator variable equal to 1 if subject i is a buyer.
EX1	An indicator variable equal to 1 if subject i has experiences in online auction markets (such as eBay or Yahoo auctions, etc.).
EX2	An indicator variable equal to 1 if subject i has experiences in financial markets (stock, futures, or exchange markets).
EX3	An indicator variable equal to 1 if subject i has any other experiences in auctions (such auctions of antiques or agricultural products).
Tool	An indicator variable equal to 1 if subject i expressed the need of paper and pencil during the auctions.

where E_i is the six-period cumulated profit (wealth) earned by subject i. In addition to WMC, we include six additional explanatory variables. The definition of these explanatory variables is listed in Table 2. They are included based on the following considerations. The literature of experimental economics suggests some possible relevancy of gender (Croson and Gneezy, 2009; Dato and Nieken, 2014; Deaves et al., 2008; Dechenaux et al., 2015; Niederle and Vesterlund, 2007; 2011) and experience. Therefore, we added these two variables into the earning regression. For the latter, based on the questionnaire at the end of the experiments we differentiate participants' experiences in on-line auctions (Yahoo, ebay, etc.) (*EX1*), financial investments (stock, futures, currencies) (*EX2*), and other kinds of auctions (antiques, commodities) (*EX3*).

In addition to these four, we also considered the role of the subject in the market, i.e., whether a buyer or a seller. This consideration is due to the fact that not all of the three markets are symmetric; Market 2 being particularly evident (See Fig. 1, the middle panel). This asymmetry may be more favorable for one side of the market when using specific strategies.[18] Therefore, to balance out this asymmetry, a role variable (*Buyer*) is also included. Finally, very much motivated by Simon (1996), we have a variable on tool (*Tool*).[19] Hence, in sum, we have six additional variables which provide information on gender, experiences, trading roles, and tool dependence to feed into the earning equation (Eq. 2).

Table 3 reports the multiple regression results. From Table 3, we can see that WMC had a positive effect on market performance. In all the three markets, the coefficient of WMC is significant at the 0.1% level. Its magnitude varies monotonically with the hardness of the market; M2, the hardest market, has the largest estimate, and M3, the easiest market, has the smallest estimate. This evidence further suggests that the harder the environment (the market) the more cognitive ability is a prime determinant in decision making, a result which is quite intuitive.

Apart from WMC, the other variable which deserves our attention is gender. As mentioned earlier, gender has become a factor of top interest in experimental economics; however, to the best of our knowledge, there is no formal study comparing the performance of male and female subjects in the context of double auction experiment. Our finding sheds light on this previously ignored area. From Table 3, the male dummy variable is significant at a level of 5% in the first two markets, the two relatively harder markets. Does this show implications for males' market competitiveness as compared to females'? We shall come back to this issue with the results from Experiment 2.

For variables of market side, market experience, and tool dependence, significance of their coefficients are only sporadically found, and a pervasive influence is not observed. Recall that we have asymmetric supply and demand curves in M2,

[18] For example, when opponents are all truth-tellers there is a strategy called the *optimal procrastination* (Chen et al., 2012; Chen and Yu, 2011). The optimal procrastination strategy dictates the subject to delay their participation in the market transaction so as to avoid early competition and become a monopsonist or monopolist in the later stage. Once after they stand in a monopsony (monopoly) position, they can fully exercise the monopsony (monopoly) power by bidding (asking) with full price discrimination (third-degree price discrimination). This strategy is feasible once the subject learns the maximum WTP and the minimum WTA of all their opponents' tokens from their constantly truth-telling behavior. However, when applying this strategy in Market 2, sellers can earn more than buyers due to its asymmetry.

[19] In the questionnaire after the experiments, we asked whether they considered the auctions difficult, and whether they had the desire to use paper and pencil to help decisions during the experiments. Notice that subjects were not allowed to use any decision-support facilities during the auctions. This question simply sought to identify the subjects who were desperate for this kind of facilities.

Table 3
Regression results of Experiment 1.

Variable	Wealth (M1)	Wealth (M2)	Wealth (M3)
Constant	489.28****	574.04****	593.62****
	(39.95)	(66.72)	(16.63)
WMC	142.07****	187.51****	43.72****
	(29.73)	(49.65)	(12.38)
Male	76.48**	143.74**	7.64
	(38.32)	(63.99)	(15.95)
Buyer	20.25	−124.56*	−18.65
	(38.53)	(64.34)	(16.04)
EX1	−46.86	−34.72	−21.54
	(39.57)	(66.07)	(16.47)
EX2	64.86	50.19	−43.92*
	(54.49)	(90.99)	(22.68)
EX3	22.03	−14.58	34.01
	(64.96)	(108.47)	(27.04)
Tool	85.72**	122.28*	21.43
	(40.64)	(67.87)	(16.92)
$\bar{R}^2$	0.1476	0.1167	0.09242

Note: Standard errors are in parentheses.
Significant at the 0.1% level: **** Significant at the 1% level: *** Significant at the 5% level: ** Significant at the 10% level: *

and this creates a potentially advantageous position for sellers. In M2, a seller could gain huge profits by optimal withholding. To see this, one must remember that within Experiment 1, all market participants were truth-telling, except the human subjects. If the seller subject withheld their first unit in the first three market transactions, the low-cost units owned by other sellers were sold out and the remaining units were non-competitive high-cost ones. Then this human seller could make an optimal offer just below the first high-cost unit, then matching the fourth highest buyer units. These two actions: initial withholding, then making an aggressive but reasonable ask are the keys of success in M2. As a result, being a seller is more advantageous and we therefore observed a negative coefficient of the variable *Buyer* in M2.

For the variable of financial market experience *EX2*, we have an unintuitive negative coefficient in M3, which is the easiest one in making trades. The fact that experiences in financial markets had a weak effect only in M3 was also hard to foresee in advance. Our explanation is that subjects with financial market experiences tend to make abundant but unprofitable transactions in M3. Note that extra-marginal units in M3 are unprofitable. It turns out that subjects with financial market experiences made less extra-marginal transactions than inexperienced subjects in M1 and M2, but not in M3.[20]

The *Tool* variable has a significant effect in M1 and a weakly significant effect in M2. Our conjecture is that those who wanted decision support tools are people who want to base their decisions on calculations, and this is the reason why they performed better in the beginning. After that, all the subjects gradually learned suitable techniques to make successful transactions, so this effect decays over time. One way to verify this hypothesis is to measure subjects' CRT scores. However, this test was not included in our design so we leave this as an open question for future research.

3.2. Explaining performance distribution

In the previous section, we examined the *mean effect* of WMC on earnings. While knowing how on average WMC affects earnings provides us a primary point of reference, it could be equally important to know how WMC affects earnings differently at different points of the conditional earnings distribution; especially when the distribution of subjects' performance does not have a "standard" shape. Being fully aware that increases in WMC may impact earnings performance, it would be useful to know whether its effect on earnings at the bottom of the conditional distribution is different from its effect on earning at the middle or at the top of the conditional distribution. In other words, we not only ask whether WMC matters, but for whom it matters. Hence, in this section, we apply quantile regression to estimate the effect of WMC on the earnings at different points of the earnings' conditional distribution.

Table 4 reports the regression results from the 10th to the 90th percentiles. We have two observations. Firstly, the effect of WMC on earnings performance is ubiquitous. Its influence is not necessarily restricted to a specific subject subset. If we take a significance level of 10%, then in all percentiles, while not completely uniform, it is at least significant in two out of the three markets. For the low quantiles, particularly, the 10th percentile, the WMC effect is statistically significant in all three markets. Secondly, in terms of its magnitude, in all three markets the WMC quantile effect is large initially for low quantiles and then declines all the way up to the median. Beyond the median, the quantile effect becomes more or less stable. In Fig. 4, we plot the coefficient estimates with respect to the nine quantiles. This figure clearly reiterates the

[20] The average number of trades made by subjects with and without financial market experiences are 2.43 versus 2.51 in M1, 1.45 versus 1.61 in M2, and 3.24 versus 3.23 in M3. All these differences are not statistically significant.

Table 4
Estimated coefficients of WMC scores on market performance from quantile regressions of Experiment 1.

Percentile	Wealth (M1)	Wealth (M2)	Wealth (M3)
10th	212.42775**	518.00000****	73.24561****
	(84.72969)	(150.56601)	(14.61504)
20th	152.40000*	197.14286*	46.139***
	(83.64901)	(115.85105)	(15.67518)
30th	108.25688**	80.31496	36.42857****
	(41.72964)	(73.16394)	(10.76415)
40th	98.58156***	42.85714	34.11765****
	(30.02118)	(28.05415)	(9.42353)
50th	77.77778***	26.92308	16.81818*
	(27.48703)	(29.10307)	(9.09175)
60th	55.93220**	42.02128**	20.0409*
	(27.45539)	(16.79462)	(10.48584)
70th	56.83453***	38.23529****	16.84533**
	(22.20533)	(9.99514)	(8.50958)
80th	50.21645***	32.40741**	10.000
	(16.26988)	(13.07728)	(6.17487)
90th	10.09174	47.20280***	15.71429***
	(20.03938)	(19.14003)	(5.9413)

Note: Standard errors are in parentheses.
Other factors used in multiple regressions are also used here as the explanatory variables.
Significant at the 0.1% level: ****
Significant at the 1% level: ***
Significant at the 5% level: **
Significant at the 10% level: *

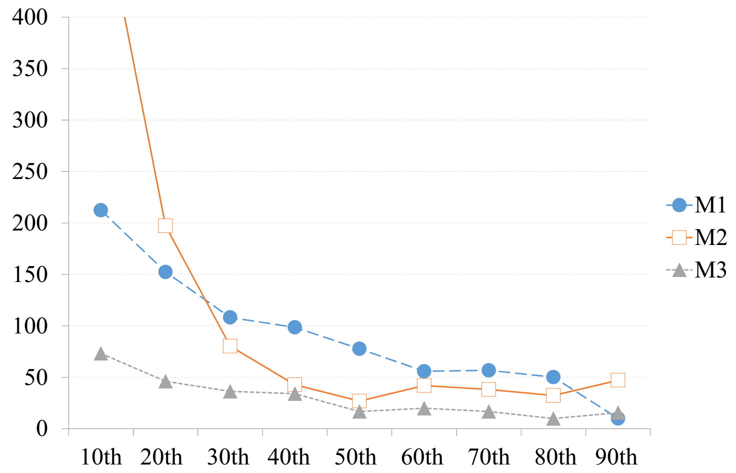

Fig. 4. Coefficients of WMC scores from quantile regressions for different percentiles–Experiment 1.

aforementioned pattern, which suggests that, for Experiment 1, the WMC effect appears to be stronger in magnitude for the subject whose conditional earning performance is in the bottom half (the bottom-half subjects).

If we look at the magnitude of the coefficients more carefully, it seems that the higher the percentile, the smaller the coefficient. The first observation suggests that the influence of WMC is not constrained to a specific subject group. But the second observation is to a degree, very intriguing. It is obviously that, regardless of the market structure, the coefficient drops drastically from low percentiles to higher ones. The fact that working memory has greater influence on performance for subjects in lower quantiles means that most poor performers are those who have low working memory capacity, while high performers have a more diverse range of working memory capacity. See Fig. G.19 in Appendix G for the relationship between WMC and wealth in each market.

3.3. The dynamics of subjects' performance

At this point, we have addressed the central issue: whether WMC is of consequence, for whom, and to what degree. It seems that it matters and, while the effect is contextual, by and large, it matters for all subjects from the bottom to the

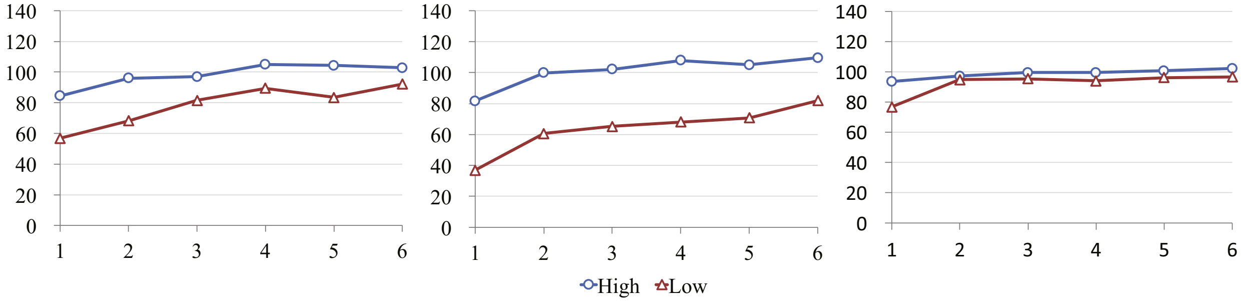

Fig. 5. The evolution of average performance for High and low groups–Experiment 1. From the left to the right are average performance in M1, M2, and M3, respectively.

top. WMC or cognitive ability may have already developed in childhood, and hence in a sense it is innately determined. The question is whether this innate personal trait should inevitably have a deterministic implication on earnings and any other related performance. Where is the 99% perspiration alluded to by Albert Einstein?[21] This question not only concerns economists but also educators in general. Our repeated market experiment puts subjects in a 'groundhog day'.[22] Will the innate effect gradually disappear when 'the same day' occurs over and over again? In this section, we address the earning performance from a dynamic perspective and examine the effect of learning.

To answer the above questions, we have to examine WMC's contribution from a dynamic perspective: we want to know how WMC's influence persists over time. We divided the subjects into two groups–the High WMC group and the Low WMC Group–according to their WMC scores. Subjects with WMC scores above zero are marked as 'high', others as 'low'. Among the 170 subjects, 97 are in the high group and 73 are in the low group. Fig. 5 illustrates the evolution of each group's average performance throughout the experiment.

From Fig. 5, we clearly observe several features: (1) It is obvious that the high group outperformed the low group in every period of every market. (2) There are obvious learning periods for both the high and low groups. (3) The gap between the high and low groups shrinks overtime, suggesting that the advantage of having larger WMC weakens within the learning process. (4) Subjects' performances drop when the supply-demand schedule changes. We will explore these visual features and determine their significance in what follows.

For the first observation, a quick look seems to indicate that subjects with high WMC tend to perform better than subjects with low WMC. To make sure that these differences are significant, we performed nonparametric tests for each period. The null hypothesis assumes that the high group and the low group have the same performances, while the alternative hypothesis states that the high group has higher performances on average. Table F.16 in Appendix F reports the results of the Wilcoxon rank sum tests. The null hypothesis is rejected in most periods in every market; therefore, it validates our first observation.[23]

Further inspection of Table F.16 reveals that while the performance index of the high group eventually surpassed 100 in all three markets, the low group failed to reach this level. Notice that, based on the interpretation of the performance index (Eq. 1), with all opponents being truth-tellers, a subject can earn an index of 100 simply by bidding/asking according to the assigned token values. Therefore, an index higher than 100 indicates that the high group managed to develop effective strategies to take advantage of the opponents' unchanging behavior and successfully made deals with 'unfair' prices. On the other hand, the low group might intend to deviate away from truth-telling as well; however, their efforts for added benefits were to no avail. They either suffered losses with aggressive offers or failed to make deals with greedy offers.[24]

As to the second finding, we first notice that the upward trend is graphically evident for both groups in all three markets. For closer examination, we define the *effect of learning* as the improvement made by the subject from period 1 to the last period, and then conduct Wilcoxon Signed Rank Test to see whether this improvement is greater than zero. The test results confirm that subjects did make significant improvement throughout the experiments (see Table F.17 in Appendix F). Now, if both groups can learn and make improvements over time, would they then be able to learn at a different rate wherein the Low group can eventually catch up to the high group? In other words, will the significance of cognitive ability decrease over time and learning becomes a dominant factor as times goes on? This question pertains to our third observation.

Fig. 5 shows that the disparity in the earnings performance between the two groups shrinks throughout the experiments, but statistical tests in fact indicate no significant change (Table F.17, last column). Hence, we cannot conclude that the low group is catching up to the high group. The rejection of the convergence pattern comes from the high variation of the low

[21] This corresponds to the famous quotation from Albert Einsten, "Genius is 1% talent and 99% percent hard work."
[22] See Thaler (2000) for the use of this movie '*Groundhog Day*' as a description of the environment which will repeat itself constantly without any change.
[23] Those are not significant are periods 4, 5, and 6 in M2 as well as periods 1 and 2 in M3 at the 5% significance level.
[24] From the performance of these two groups, we might also be surprised by the fact that the seemingly naive truth-teller strategy is too formidable to beat. One of the possible reasons is that when all market participants (artificial agents) become truth-tellers, the competition could be keener than we might think. With the presence of these truth-telling opponents, any greedy bid/ask may end up with the loss of a profitable trading opportunity.

Please cite this article as: C.-C. Tai et al., Cognitive ability and earnings performance: Evidence from double auction market experiments, Journal of Economic Dynamics & Control (2018), https://doi.org/10.1016/j.jedc.2017.12.008

Table 5
Number of transactions by High and Low groups in Experiment 1.

	M1		M2		M3	
	High	Low	High	Low	High	Low
Min	0	0	0	0	2	1
Max	4	4	4	4	4	4
Median	2	2	1	1	3	3
Mean	2.4124	2.6461	1.5052	1.7055	3.2131	3.2557
Std.dev	0.7078	0.9002	0.8677	1.0382	0.4686	0.5226
Wilcoxon rank-sum test	$W = 107160$, p-value = 1.006e-06		$W = 115630$, p-value = 0.003317		$W = 121590$, p-value = 0.1123	

group. In fact, we find that the learning capability of the low-group subjects are much more heterogeneous than the learning capability of the high-group subjects. We also conducted regression analysis to examine whether subjects' WMC scores can predict the effects of learning. The results are far from significant for all three markets, indicating that WMC cannot explain the amount of improvement that subjects had made.

Even though there is no significant evidence to show that the gaps get narrower, the gaps seem to vary among different markets. From Fig. 5, one observes that it is widest for M2, and narrowest for M3. This ranking shows consistency with hardness levels of the three markets as we discussed in Section 2.1, where we argue that by our design the strategies to earn may be less sophisticated in M3 and more sophisticated in M2. Therefore, the third finding of this experiment may be rephrased as follows: the earning performance of the low group constantly falls behind that of the high group, and the size of the performance gap depends upon the hardness of the decision problem implied by the market topology. One may argue that learning in sequence of the three market experiments in fact may make the last market experiment the easiest one and the first market the most difficult. It is exactly because of this concern we put M1 before M2, as we have explained in Section 2.1. Therefore, one can say that even though the cross-market learning might be expected, its real effect could still crucially depend on the hardness of the problem.

Let us continue examining the cross-market learning or adaptation. For both groups of subjects, when presented with a new market, there was always a drop in earning performance (Table F.16). For the high group, it dropped from 103 to 91 (when M1 was replaced by M2) and from 122 to 94 (when M2 is replaced by M3). A similar pattern of earning performance drop also held for the low group: 91 to 44 at the first transition, and 97 to 77 at the second transition. Among the two transitions, the first one was more challenging because it was a transition from a relatively simple decision problem to a more difficult one. Interestingly, the low group declined more severely than the high group during the transitions: about a 50% decline for the low group as opposed to about a 10% decline for the high group. This result shows contrast in subjects' ability to adapt to and solve problems in less familiar situations, which is known as *fluid intelligence* (Cattell, 1963). WMC has been found to be correlated with fluid intelligence (Unsworth et al., 2014), and that could explain the observed difference in adaptability between the two groups during market transitions. In sum, our evidence supports that cognitive ability is of consequence for cross-market learning and adaptations to new environment.

3.4. WMC and bidding behavior

In order to understand why subjects with higher WMC outperformed those with lower capacity, we examined their behavior in terms of trading quantities and bid values. Table 5 reports the descriptive statistics of both groups' number of transactions. As this table shows, both groups' median transaction numbers in each market are identical to the theoretical numbers of transactions (numbers of intra-marginal tokens) in equilibria. However, high-WMC subjects had significantly fewer transactions on average then low-WMC subjects in M1 and M2. Since the means are larger than the equilibrium quantities, this means high-WMC subjects had fewer extra-marginal units traded in M1 and M2.

Note that computer agents are truth-telling in Experiment 1, having extra-marginal units exchanged with other agents will incur losses most of the time unless the subject can steal a trade from intra-marginal units with a transaction price happens to be lower than the token value. To be exact, the ratios of the first extra-marginal transactions (the 3rd unit in M1, the 2nd unit in M2, and the 4th unit in M3) which results in positive profit in Period 6 are 40%, 78%, and 0%. Compared to the 95%–100% positive-profit ratios for intra-marginal units, taking advantage of truth-telling opponents by stealing deals from them is not easy in Experiment 1, especially in M1 and M3. As a result, we found that making fewer extra-marginal trades when competing with die-hard truth-telling opponents is the first distinction between high-WMC subjects and low-WMC ones.

In additional to the difference in extra-marginal trading, we investigate how subjects made their bids (and asks) for intra-marginal units. In general, we can categorize subjects' behavior into three different types: overbidding, truth-telling, and underbidding.[25] Take buyers for example, overbidding means their bids are higher than their token values (reservation

[25] In what follows, we will mention only buyers for illustration purpose.

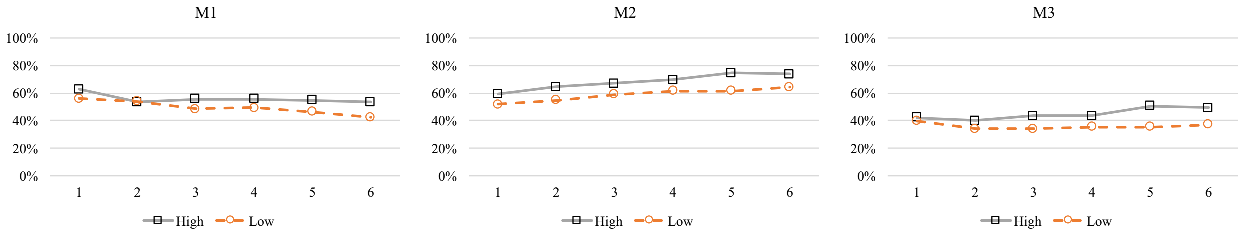

Fig. 6. The evolution of underbidding ratios for High and low groups–Experiment 1.

Table 6
Bid deviations for High and Low groups in Experiment 1.

	M1		M2		M3	
	High	Low	High	Low	High	Low
Min	−16.13	−10.7	−373.33	−373.33	−25.99	−26.79
Max	23.62	88.39	47.37	235.37	37.77	410.47
Median	−1.76	−0.36	−7.89	−2.44	0.91	2.96
Mean	−1.64	−0.09	−67.69	−49.78	1.68	3.79
Std.dev	4.68	8.95	115.45	103.83	8.29	21.36
Wilcoxon rank-sum test	W = 112670, p-value = 0.0136		W = 111830, p-value = 0.008035		W = 113690, p-value = 0.003117	

prices); truth-telling means the bids are equal to the token values; underbidding mean the bids are lower than the token values. An overbidding buyer has the advantage that they have better chance to surpass other buyers and win the opportunity to trade earlier. Note that with our pricing mechanism, the transaction price will be in the middle of the winning bid and winning ask (providing that the winning bid is higher than the winning ask), sending a bid higher than the token value will not necessarily incur a loss. On the other hand, a underbidding buyer postpones the trade and may get a lower transaction price.

Overbidding or underbidding, which one is better? It turns out the key relies on the shapes of the supply and demand curves and the positions of the token values. Referring to Fig. 1, and let's take Buyer 2 (B2) for illustration purpose. Providing that other market participants are truth tellers, for B2's first token (at the value of 390), overbidding will not bring any extra benefit and therefore a underbidding strategy should be used to lower the transaction price. However, the bid cannot be too low (lower than 370, to be exact) for it will be outranked by other buyers' bids and has to face high-cost sellers. As a result, the best action for B2's first token is to underbid to the boundary delineated by other participants' bidding behavior. For B2's second token, it appears that the best strategy is to overbid just a little bit and surpass B1's second unit so as to trade with a much lower price. However, if B2 already underbids for the first token (and therefore made the fourth transaction in the market), overbidding the second token becomes meaningless because it is not allowed to trade the second token unless the first one is bought. Consequently, underbidding is ideal for all intra-marginal tokens and will bring more profits.

To evaluate subjects' bidding behavior, we used a measure called *bid deviation* defined as follows:

$$\text{Bid deviation} = \begin{cases} \dfrac{\text{Actual bid} - \text{Token value}}{\text{Token value}} & \text{For buyers,} \\[2ex] \dfrac{\text{Token value} - \text{Actual ask}}{\text{Token value}} & \text{For sellers.} \end{cases} \tag{3}$$

A positive bid deviation means that the buyer (seller) subject sends a bid (ask) which is larger (lower) than the token value and therefore stands for the overbidding strategy; a negative bid deviation stands for the underbidding strategy.

Fig. 6 shows the evolution of underbidding behavior for intra-marginal units. It is clear that subjects with higher WMC are more inclining towards the underbidding strategy than subjects with lower WMC. The difference between both groups' tendencies to underbid are either consistent or increased over time. The underbidding ratios for the High and Low groups in the last periods are 54%:42%, 74%:64%, and 49%:37% in M1, M2, and M3, respectively. Not only so, Table 6 reports the descriptive statistics of the bid deviations for both groups, and it is clear that the high-WMC subjects has significantly more negative bid deviations. In sum, the second distinction between high-WMC subjects and low-WMC ones is their tendencies to underbid, and this distinction prevails in all the three market structures tested in our experiments.

4. Experiment 2: Trading with adaptive agents

In the previous section, an analysis was done to the experimental results in a market environment where human traders' opponents are all truth-telling agents. Although WMC does exhibits its salient influence on subjects' market performance, an inevitable question follows of what role WMC would play in a more realistic environment. Experiment 2 commenced

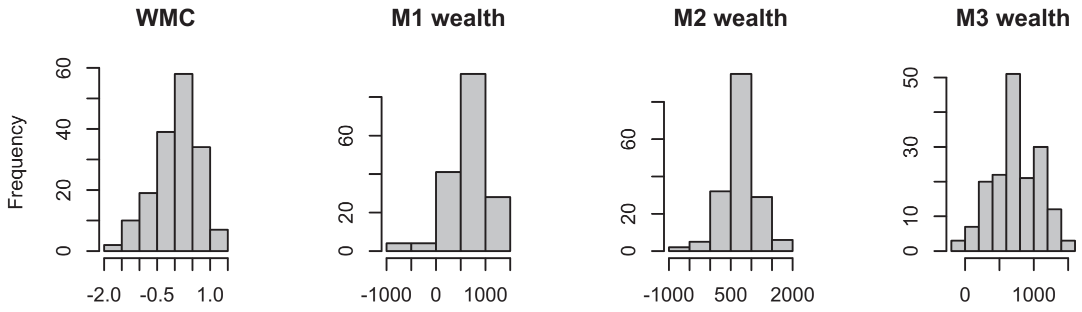

Fig. 7. The distribution of subjects' WMC scores and their cumulated performance (wealth) in Experiment 2.

Table 7
Descriptive statistics of Experiment 2.

	WMC	Wealth (M1)	Wealth (M2)	Wealth (M3)
Mean	0.05	616.82	716.28	745.78
Median	0.13	656	736	755
Maximum	1.29	1261	1755	1495
Minimum	−1.84	−938	−892	−122
Std. Dev.	0.61	379.78	392.29	345.01
Skewness	−0.57	−1.28	−0.63	−0.15
Kurtosis	−0.06	3.03	2.03	−0.53
p-value of Shapiro–Wilk Test	2.62e-03	3.09e-08	1.54e-04	5.50e-02

to address this question by hiring a series of computer agents as market traders aside from human subjects. These agents were chosen because they provided more variant behaviors than truth-telling agents.[26] In this section, we follow the same analytical process as in Experiment 1. Comparisons of the results from both experiments will also be delineated.

Before examining the relationship between WMC and subjects performance, we checked whether the aggregate outcomes of our experimental auctions are consistent with the convergence phenomenon repeatedly observed in the literature (Appendix E). We find that there are convergences toward equilibria in all three markets, although they seem to be slower than those in Experiment 1. Additionally, the direction of convergence is still consistent with previous studies. Both facts suggest that we can be confident in our experimental results because they do not exhibit any abnormal pattern different from the robust results observed in the literature.

4.1. Can WMC predict subjects' performance?

We recruited 173 subjects from fifteen experimental sessions. Four subjects were dropped due to the same reasons as described in Section 3. Therefore, we only analyze the data from 169 subjects (96 males, 73 females). Fig. 7 illustrates the frequency distributions of subjects' WMC and profit in each market. It is not hard to see that the distribution of profit in each market is less left-skewed than in Experiment 1. Table 7 reports the descriptive statistics of the results.

The next step is to examine whether WMC can predict subjects' performance as in Experiment 1. We adopted the same regression model (Eq. 2) as in Section 3.1. The results are given in Table 8.

We have the following observations from Table 8. Firstly, WMC is still important in M1 and M2, but it is insignificant in M3. Secondly, different from Experiment 1, a buyer's position is advantageous in M3.[27] Thirdly, while financial experience is of little help for earnings in Experiment 1, it becomes much more relevant in Experiment 2, which can be seen by the magnitudes of coefficients as well as their statistical significance in each market.

Although WMC fails to show its impact in M3, one may suspect that this is because traditional regression methods concerns only the mean. Table 9 presents more convincing evidence from quantile regressions. From these results, we clearly see that WMC scores can predict performances for most of the percentiles in M1 and M2, but not in M3, except the 80th one. Fig. 8 demonstrates the coefficients of WMC from quantile regressions. Recall that in Experiment 1, the coefficients decreased in size as we moved to higher percentiles. We observe the same pattern here in M1 and M2, but not in M3. In

[26] Interested readers can refer to Chen et al. (2010) for a detailed comparison among these agents. From that study, we choose the most promising agents (except the GP traders) as the company for our subjects.

[27] This is a seemingly unreasonable result because M3 is a symmetry market where consumer surplus and producer surplus are exactly the same under competitive equilibrium price. However, this is actually a common observation that buyers tend to have better performance in experimental double auctions using human subjects. This is known as the *weak seller hypothesis* (Smith and Williams, 1982). Weak seller hypothesis says that, for some reason, subjects perform better when assigned as buyers than as sellers. Some thought it could be because normal people are used to playing the roles as buyers in daily life. This hypothesis is in fact consistent with Chamberlin (1948)'s guess when explaining the lower prices observed in his search and haggle experiments. Stronger evidence for the weak seller hypothesis can also be found in Smith and Williams (1990)'s box design (parallel demand and supply) experiments.

Table 8
Estimated coefficients on market performance from Regressions of Experiment 2.

Variable	Wealth (M1)	Wealth (M2)	Wealth (M3)
Constant	533.69****	771.476****	565.99****
	(64.09)	(65.930)	(58.11)
WMC	152.49***	119.999**	57.99
	(47.07)	(48.424)	(42.68)
Male	67.45	45.559	74.96
	(57.52)	(59.168)	(52.15)
Buyer	84.96	−179.917***	162.92***
	(56.78)	(58.416)	(51.49)
EX1	−28.73	−6.077	62.76
	(56.49)	(58.117)	(51.23)
EX2	204.82**	192.005**	133.72*
	(83.75)	(86.154)	(75.94)
EX3	−146.46	11.334	31.02
	(100.90)	(103.803)	(91.49)
Tool	−43.26	−113.745	−49.85
	(75.95)	(78.134)	(68.87)
$\bar{R}^2$	0.09061	0.098	0.09402

Note: Standard errors are in parentheses.
Significant at the 0.1% level: **** Significant at the 1% level: *** Significant at the 5% level: ** Significant at the 10% level: *

Table 9
Estimated coefficients of WMC scores on market performance from quantile regressions of Experiment 2.

Percentile	Wealth (M1)	Wealth (M2)	Wealth (M3)
10th	322.68041****	238.12950****	26.13636
	(38.40133)	(44.50261)	(46.23830)
20th	289.09091****	210.16949****	56.05701
	(62.92596)	(34.49620)	(46.31035)
30th	194.37500***	222.41379****	50.00000
	(68.94788)	(51.22668)	(48.41892)
40th	124.04580**	179.46429****	22.79412
	(61.15371)	(47.67211)	(55.50203)
50th	116.77419**	140.25974****	11.35135
	(51.89780)	(40.05737)	(49.82646)
60th	108.04150****	85.52632*	45.43568
	(30.07585)	(43.94345)	(47.31506)
70th	104.34783*	81.95266*	70.10870
	(57.08791)	(46.29515)	(51.91277)
80th	100.70175	33.55705	138.12950**
	(77.32583)	(45.14631)	(58.38153)
90th	48.18841	30.33708***	35.34483
	(46.78011)	(11.01390)	(68.32928)

Note: Standard errors are in parentheses. Other factors used in multiple regressions are also used here as the explanatory variables. Significant at the 0.1% level: **** Significant at the 1% level: *** Significant at the 5% level: ** Significant at the 10% level: *

M3, the coefficients fluctuate and are marginal for low percentiles (see Fig. G.20 in Appendix G for the relationship between WMC and wealth in each market). In short, we conclude that WMC still plays a role in Experiment 2, but not in every market. Within the easiest market (M3), WMC displays no primacy over other factors. More interestingly, *experiences in financial market transactions were more important in this more realistic experiment than they were observed in Experiment 1.*

4.2. The dynamics of subjects' performance

One may also inquire as to how WMC contributes dynamically to subjects' performances. We followed the same logic in Section 3.3 and divided our subjects into two groups–the high group and the low group. In Experiment 2, the high group consists of 99 subjects, while the low group has 70 subjects. Fig. 9 portrays the specific dynamics of the average performance for both groups.

From Fig. 9, it is apparent that the high group performed better. Even in M3, where WMC appears not to be a decisive factor in earning performance, the high group still has an edge over the low group. Table F.18 in Appendix F endorses this

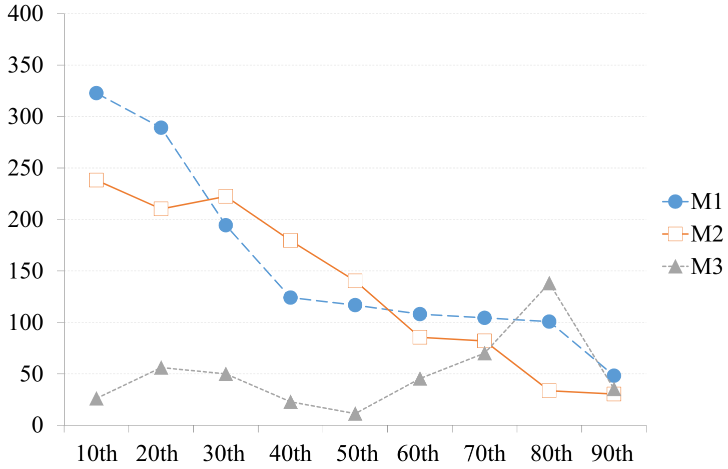

Fig. 8. Coefficients of WMC scores from quantile regressions for different percentiles–Experiment 2.

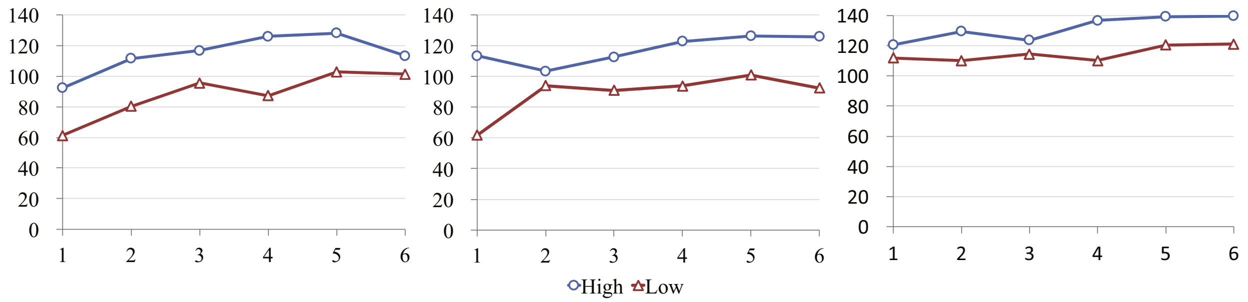

Fig. 9. The evolution of average performance for High and low groups–Experiment 2. From the left to the right are average performance in M1, M2, and M3, respectively.

observation by a series of Wilcoxon rank sum tests. The results show that the high-low differences are significant most of the time, even in M3.

Although in our regression analysis WMC appears to be insignificant in M3, our grouping analysis here indicates a qualifier to this result. We can reconcile these two results from a temporal perspective: the phenomena of the high group outperforming the low group in M3 is not as pervasive as it is in Experiment 1, especially for the last few periods, in which learning could play a role. Note that if we take the significance level of 5% as a threshold, according to Table F.16, the high-low differences are significant in 4 out of 6 periods in Experiment 1. However, the High-Low differences are also significant in 4 (almost 5) out of 6 periods in Experiment 2, based on the figures in Table F.18. To provide more evidence supporting this view, we must consider subjects' learning behavior over time.

As in Section 3.3, we can further examine whether the learning capability of subjects is related to their WMC. Table F.19 in Appendix F presents the average improvement in each market for both groups. Not surprisingly, our subjects learned how to gain more profit over time in most cases. However, a notable exception is the low group in M3. According to the test results, the low group showed no significant signs of learning in M3. This result is consistent with our quantile regression results, which indicate that the only significant effect takes place in a high percentile. This also explains why a sizable high-low difference in the middle of the M3 auctions was observed, yet regression results suggest no importance for WMC at the same time.

Finally, we compared the average improvement made by both groups. From Table F.19, we observed that the low group had larger average improvement in M1 and M2, but not in M3. However, the low group's improvement displays large variance as well, making the high-low comparisons insignificant as seen in the last column of Table F.19.

4.3. WMC and bidding behavior

In order to understand why high-WMC subjects outperformed low-WMC ones in Experiment 2, we examined subjects' behavior in terms of trading quantities and bid values. Table 10 reports the descriptive statistics of the High and Low groups' number of transactions. In contrast to Experiment 1, high-WMC subjects appeared to have more units traded than low-WMC

Table 10
Number of transactions by High and Low groups in Experiment 2.

	M1		M2		M3	
	High	Low	High	Low	High	Low
Min	0	0	0	0	1	0
Max	4	4	4	4	4	4
Median	3	2	1	1	4	3
Mean	2.5707	2.4929	1.3973	1.3429	3.5320	3.3143
Std.dev	0.6915	0.9126	0.6285	0.7322	0.6226	0.8719
Wilcoxon rank-sum test	W = 129910, p-value = 0.2207		W = 131750, p-value = 0.06917		W = 139950, p-value = 0.0001787	

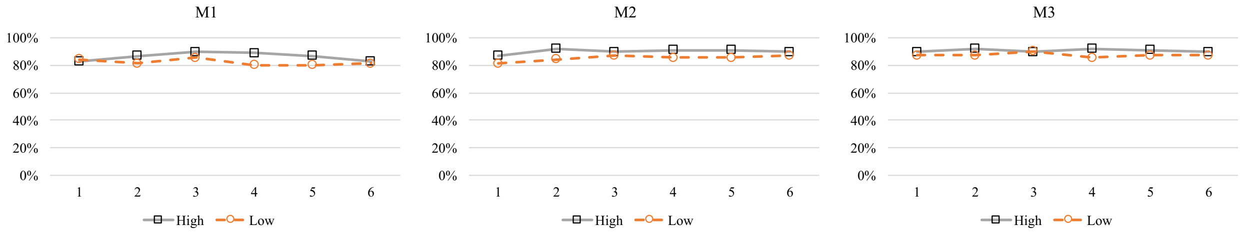

Fig. 10. The evolution of underbidding ratios for High and low groups–Experiment 2.

Table 11
Bid deviations for High and Low groups in Experiment 2.

	M1		M2		M3	
	High	Low	High	Low	High	Low
Min	−18.39	−19.35	−396.67	−383.33	−44.01	−57.14
Max	41.94	84.58	31.58	201.28	31.38	85.52
Median	−6.48	−5.47	−34.94	−26.22	−14.33	−12.89
Mean	−6.32	−4.70	−112.04	−99.80	−14.80	−12.57
Std.dev	5.75	9.55	120.46	119.92	11.32	13.35
Wilcoxon rank-sum test	W = 105470, p-value = 0.002036		W = 110650, p-value = 0.09582		W = 111340, p-value = 0.02889	

ones in Experiment 2. The differences are significant in M3, and almost significant in M2. Even in M1, the median of the High group is still larger than the theoretical number of transactions (number of intra-marginal tokens) in equilibrium.

The reason behind this phenomenon is that computer agents are adaptive in Experiment 2, and their bidding behavior might create some space for human subjects to exploit. On the contrary, truth-telling agents in Experiment 1 did not leave much space for human subjects to exploit extra opportunities. To see the difference, recall that the ratios of profitable transactions for the first extra-marginal unit in Period 6 are 40%, 78%, and 0% in Experiment 1. In Experiment 2, these numbers are 64%, 71%, and 82% from M1 to M3–they are much higher than those in Experiment 1 for M1 and M3 and very close for M2. This observation suggests that exploiting extra profit opportunity by stealing deals from non-truth-telling computer agents is more feasible in Experiment 2.[28] As a result, subjects with higher WMC took advantage of this feature and made more transactions than those with lower WMC.

In additional to the difference in extra-marginal trading, we also investigate subjects' intra-marginal bidding behavior in terms of bid deviations. Fig. 10 demonstrates the ratios of underbidding behavior over time for both High and Low group subjects. One sharp distinction from Experiment 1 is that both groups' tendencies to underbid are much stronger in Experiment 2 and are closer to each other. It seems that the High group only has a slightly higher ratios of underbidding. However, if we look at the degree of underbidding as presented in Table 11, it is clear that the High group still has significantly more negative bid deviations than the Low group. In brief, high-WMC subjects still underbid more than low-WMC subjects in Experiment 2.

4.4. Comparisons and contrasts

Experiment 1 was a simple environment in the sense that all artificial agents were 'honest' and did not change their behavior over time. It can be shown that this static environment allows us to deliver an analytical solution for the optimal

[28] We are not saying that all computer agents in Experiment 2 are pushovers. In fact, agents such as the belief-based learning agents were very difficult to defeat for our subjects.

Table 12
The effects of experiences in both experiments.

	Experiment 1			Experiment 2		
Market	Experienced	Inexperinced	p-value	Experienced	Inexperienced	p-value
M1	597	529	0.2981	776	589	0.02592
M2	640	584	0.5049	887	687	0.03019
M3	534	583	0.8562	900	719	0.01543

Numbers are cumulative wealth earned by different groups of subjects after six periods.

bids/asks as well as the submission time through integer programming. As said before, the availability of this optimum facilitates experimenters' observations of subjects' learning processes (Chen et al., 2012). Nonetheless, such trading environments do not exist in reality, specifically when in the case of robot traders (Furse et al., 2011). In this sense, Experiment 2 was more realistic because all of the artificial agents were adaptive and would react to the changing dynamics of the market. In addition to adjusting their bids/asks over time, the robot traders could choose not to trade by passing in a trading round, which further restricts the information disclosure. These factors together made the flow of Experiment 2 less predictable, simultaneously serving as a better 'virtual reality' for our subjects. This fundamental difference between the two treatments show three interesting contrasts, respectively, on earnings performance, the experience effect, and the cognitive effect.

Let us begin with earnings performance. A glance over Tables F.16 and F.18 shows that subject performances in Experiment 2 were superior to those in Experiment 1, when comparing each congruent period within each market session. At first inspection, this overwhelming dominance seemed to be at odds with the relative complexity of the two experiments. Nonetheless, the seemingly complex nature of Experiment 2 was not necessarily more difficult than Experiment 1. This is explained by the fact that most of the artificial adaptive agents were not omniscient; they are programmed to behave in every manner but optimally. Their 'bounded rationality' caused them to perform well on some occasions, but perform marginally in others (Chen and Tai, 2010; Chen et al., 2010). Their imperfections made the early profitable but unattainable opportunities become more feasible, for example, 'stealing a trade' with extra-marginal tokens, and helped to enhance human subjects' earnings performances.[29]

Secondly, we compare the experience effects for both experiments. Earlier in Experiment 1, we found the self-reported experiences on similar auction environments have no effect on earnings performance. The only exception is the financial market experience (*EX2*) in M3; however, its coefficient is even negative. In Experiment 2, we still fail to see the effect of experience from on-line auctions (*EX1*) and other auctions (*EX3*), but the financial market experience had now shown a significantly positive effect on earnings performance (Table 8). One possible explanation for this disparity is that the flow of Experiment 2 shows more characteristics of an actual trading room with human-like agents rather than a mathematics lab with truth tellers; which makes experienced subjects familiar, effectively pulling their experiences from real-life markets.

Upon closer examination, Table 12 gives the average performances of experienced and inexperienced subjects in each case. In Experiment 1, financial market experiences contribute to earnings in M1 and M2, but not to a significant degree. In Experiment 2, it paid to have experiences in financial markets, and experienced subjects gained up to 25% more profit than inexperienced ones.

Thirdly, it is the WMC effect at work. The comparison made here is to see whether the WMC effect can be altered when strategic interaction was introduced in Experiment 2. As already seen in Sections 3 to 4.2, the WMC effect can be examined from a comparative static or a dynamic (learning) perspective. The former can be further distinguished by the mean effect and the quantile effect. Juxtaposing Table 3 with Table 8, one sees that the mean effect of WMC on earning performance in Experiment 2 seems to be weaker than that in Experiment 1; in particular, its mean effect is absent in M3 in Experiment 1. However, from the quantile regressions, one observes a stronger effect for the *bottom-half* quantiles in both Experiment 2's M1 and M2.

In terms of learning, the high-low difference already appeared at the beginning of Experiment 1, specifically, the first period of M1, but this beginning-of-the-beginning effect did not appear in Experiment 2. With regards to the low-high difference (the catching-up effect), most gaps remained at the end of each market session; the only one which disappears is M2 of Experiment 1. Finally, of improvement rate, substantial improvement was observed in all groups, excepting M3 of Experiment 2, which statistically, had zero improvement. Clearly, this compiled evidence shows that the original WMC effect was altered by the introduction of an interactive environment, further complicated by the ease or hardness of the markets. It is, therefore, not yet entirely clear to have a systematic view of these differences. Conversely, at this point, a more salient question may be whether the difference in the two experiments indeed causes different cognitive effects. Can we actually identify their differences in terms of cognitive tasks required? This is the question to which we now focus our attention.

[29] Even though the result is based on a sample size of 170 subjects for each experiment, one may still wonder whether the difference in earnings can be caused by a higher WMC of the second population (Table 1 and Table 7). We, therefore, ran both a Wilcoxon rank sum and a t test to examine whether the WMC of the two populations is the same, and the p-values of the null hypothesis of the equality of the two are 0.3363 and 0.243, respectively. Hence, the null cannot be rejected. Therefore, the observed difference in earnings is not because of a difference in WMC, which increases the plausibility of the interaction effect caused by different types of robot trader.

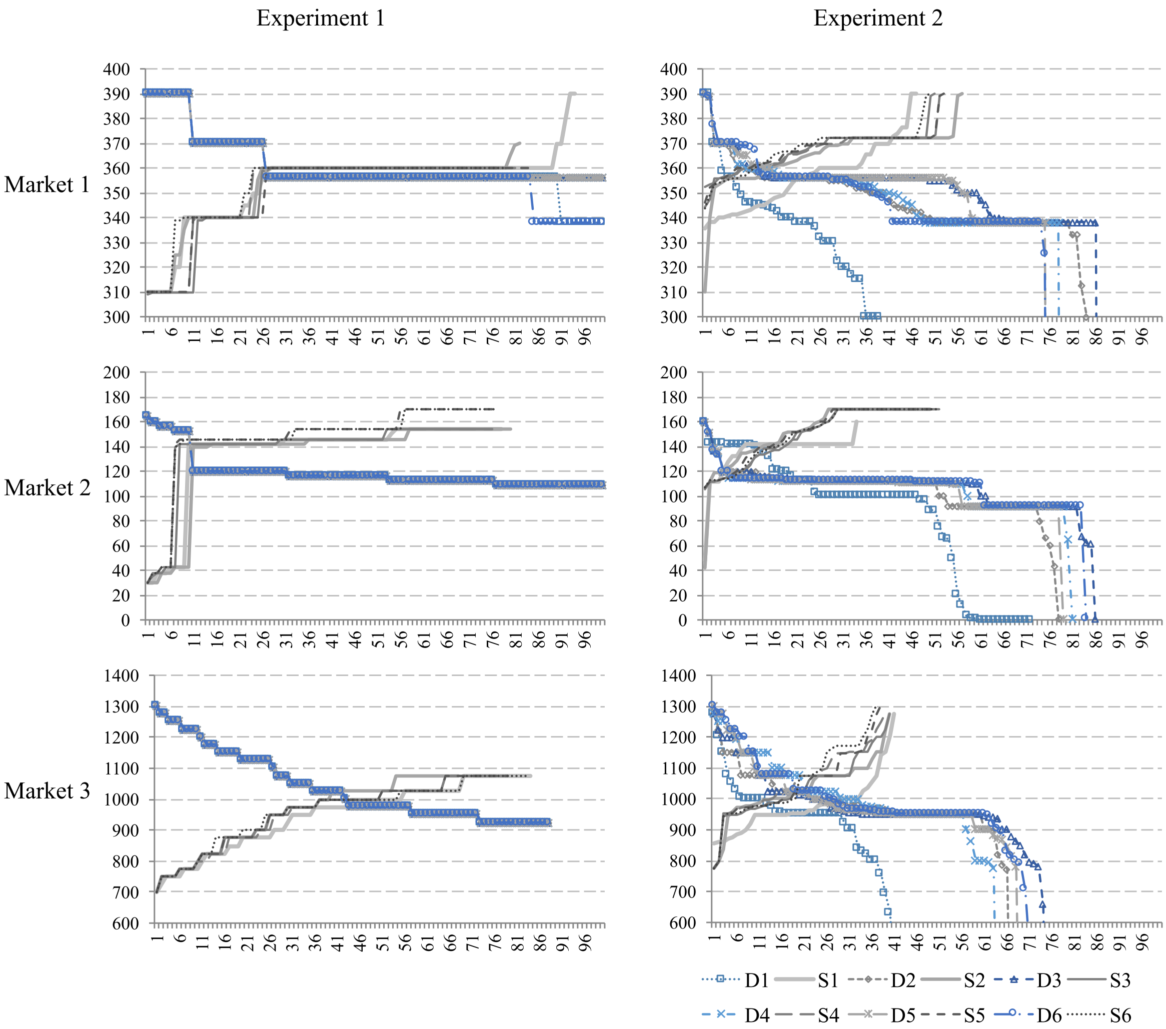

Fig. 11. A typical evolution of actual bids and asks from period 1 to period 6. 'D' denotes the demand curve based on actual bids; 'S' denotes the supply curve based on actual asks; the number attached to 'D' or 'S' indicates the period.

5. Critical components of working memory

Our experimental evidence suggests that WMC is a crucial factor in determining subjects' earning performance in double auction markets (the cognitive effect). We also observe that the influences of WMC seem to be dependent on the behavior of other artificial traders (the interaction effect). These observations motivate us to consider two further questions:

1. What is the difference between these two experiments?
2. How did WMC influence subjects' performances? In which aspects of subjects' mental processes did WMC manifest its effect?

As mentioned in Section 2, Experiment 1 was a simple environment where all market participants truthfully revealed their WTP or WTA. By contrast, Experiment 2 was a more sophisticated environment where participants adapted themselves to the constantly changing market dynamics. We took a typical auction tournament in our experiments and depicted the supply and demand curves based on market participants' actual bids and asks in Fig. 11. This figure illustrates typical market dynamics in both Experiment 1 and Experiment 2. By examining the evolution of actual bids and asks sent by all traders, we could experience the market exactly how the human trader would navigate the market during the specified series of transactions.

Actual bids and asks in Experiment 1, as expected, did not change a lot from Period 1 to Period 6. On the opposite, a human trader in Experiment 2 could find it difficult to sort out the underlying demand-supply structure, since artificial agents' bid/ask prices varied a lot as time passed. From this point of view, Experiments 1 and 2 would require subjects to solve different types of problems, which in turn involved different mental processes.

Table 13
Variables selected by stepwise regressions.

Treatment	Environment	Variables
Experiment 1	Market 1	**SS****, **OS****, SSTM, male**, tool*
	Market 2	**OS*****, **MU****, male**, buyer**, tool**
	Market 3	**OS*****, **MU****, financial**, tool
Experiment 2	Market 1	**SSTM******, male, buyer*, financial***, other
	Market 2	**SSTM******, buyer***, financial***
	Market 3	**SSTM****, male, buyer****, online, financial**

Significant at the 0.1% level: **** Significant at the 1% level: *** Significant at the 5% level: ** Significant at the 10% level: *

To examine this possibility, we first treated the five task scores of WMC as different variables, and then we perform stepwise regressions using Akaike information criterion to select the predictive variables. Table 13 reports the models selected following this procedure. Based on the selected variables and their significance in this table, we find that Experiment 1 and Experiment 2 are distinguishable according to the employed cognitive tasks. The abilities to solve SS, OS, and MU tasks were important for participating in the markets with simple agents, whereas SSTM was the key for successfully participating in the markets with adaptive agents.

Why do these two experiments correspond to different components of WMC? To the best of our knowledge, this is likely the first time that this issue, which integrates psychology and economics, has been raised in this specific context. We can give a sketch based on our understanding, but will certainly require more research to perfect the full picture.

According to the classic theory of working memory proposed by Baddeley (2000), working memory system consists of four components, including three different buffers–*visual-spatial sketchpad, phonological loop*, and *episodic buffer*–and the *central executive* component, which mediates the operation and storage of items. The BDG, OS and SS tasks all ask participants to recall the learned consonants. Presumably, these tasks should be relevant to the phonological loop. The SSTM task only requires participants to redraw the locations of dots. There is no digit nor consonant but the visual-spatial information. Thus, the efficiency of the visual-spatial sketchpad is thought to be relevant to be this task. The MU task requires participants continuously update the digits. That is, when seeing an operation (e.g., +3), participants need to immediately change the digit in the box to the result generated by that computation. Presumably, comparing with other tasks, the MU task is more directly linked to the function of central executive. As a consequence, performances in the first experiment mainly related to the phonological loop and central executive components of working memory, and those in the second experiment related to the visual-spatial sketchpad.

In addition to Baddeley (2000)'s model, another theory of working memory emphasizes its functional facet instead of the domain facet. According to Oberauer et al. (2000), working memory is associated with three main functions: (1) *simultaneous storage and manipulation*, (2) *supervision*, and (3) *coordination* (see Section 2.5). The five WMC tasks (BD, MU, OS, SS, and SSTM) in our experiments have already been identified with these three functions (Oberauer et al., 2000): the former three tasks are associated with one's ability of simultaneous storage and manipulation, whereas the last task is associated with the ability of coordination (the ability of coordinating the elements into a structure). What is more important is that SSTM is also found to be strongly related to the reasoning ability (Oberauer, 1993).

First of all, before the statistical analysis, we already knew that the experiments were different in the designs, and from our regression and non-parametric analyses, we also observed behavioral differences between these two markets. Hence, the only question remaining is the missing connection between the design and behavior (performance). Our sketch is as follows.

In Experiment 1, subjects faced truth-telling agents whose behavior was consistent and easier to predict, which helped shape a static environment (Fig. 11, the left panel). This 'groundhog-day' environment, to some extent, may had caused the subjects to develop a mental representation of the auction as an optimization problem, such as a scheduling problem (Chen et al., 2012). Given that there was no strategic interaction between subjects and truth-telling agents, game-playing strategies were not required. Subjects simply needed to figure out some "numbers", such as the *best timing* to bid the second token, or the *best bid price* so as to 'steal' the third transaction, etc. During this mental process, what is required is the ability to focus on a narrow set of numbers and make counter-factual calculations. Solving this optimization-like problem, therefore, is related to the "simultaneous processing and storage of information" which corresponds to the ability to solve SS, OS, and MU tasks.

In Experiment 2, subjects were situated in a fundamentally different environment. With their adversaries being the adaptive agents, our subjects interacted, received feedback, and engaged in games with a feeling of strategic uncertainty. The 'trading room' accommodation was more dynamic, noisy, and less predictable, which has been well demonstrated in Fig. 11 (the right panel). The shape of the problem is not well-defined, and subjects first needed to explore the potential forces of supply and demand via the ever-changing bids and asks. Based on that knowledge, subjects further needed to develop their own strategies. As Oberauer et al. (2000) stated, "Many reasoning tasks demand the integration of elements into new structures. … the models constructed for each premise in a deductive reasoning task must be integrated into a single, more complex model." Hence, 'coordination of elements into structures' becomes the essential task in Experiment 2.

An initial sketch is offered here as the first attempt to identify the key cognitive facets of WMC operating in different market environments. Whether our sketch can be completed and meaningfully generalized or extended into other markets or economic environments is an issue warranting further research.

6. Concluding remarks

In this paper, we try to link behavioral heterogeneity and its economic outcomes to personal factors in terms of cognitive ability. Abundant past research suggests that differences in economic decisions and performances can be attributed to the heterogeneity in humans' innate cognitive ability, but few have attempted to test this in a general context closely resembling many of our daily market experiences. Our study takes a marked step to fulfill this gap by engaging our cognitively heterogeneous subjects in double-auction activities. Our aim is to thoroughly examine the link between cognitive heterogeneity and heterogeneity in people's market activities from the superficial level to the psychological level.

In the experiments, we measured subjects' cognitive ability in terms of their working memory capacity with a psychometric test battery. We adopted working memory capacity because it is a cognitive ability construct which captures critical facets of human decision making processes. The WMC test battery adopted in this study provides the opportunity to observe different channels through which working memory influences subjects' market performances.

We conducted two double auction experiments following the single-subject design. In Experiment 1, subjects traded with truth-telling bidding agents, while subjects in Experiment 2 had to compete with adaptive trading agents. Because the "behavior" of subjects' opponents was controlled, there was no room for additional factors such as subjects' other-regarding preferences or the complex interactions among subjects' unknown strategies to get involved.

Our unique experimental design rules out the possibility of many biases and mistakes that people usually had, such as judgment bias, inability to compute fundamental values, or failure to understand market mechanism such as a futures market. As a result, what we are interested in is purely the influences of cognitive ability on subjects' trading skills and their ability to sense market situations. Trading skills and the ability to sense market situations are essential for any market activity that one can imagine, so our results are not confined to asset markets but have wider market applications. On the other hand, we can also combine our discoveries with findings from other asset market experiments and yield a more complete picture of how heterogeneity in cognitive ability impacts financial markets.

Can cognitive ability explain the differences in people's market performance? Our results show that WMC is an important factor in predicting subjects' performance in almost every market experiment. However, the only exception reminds us that WMC's influence may not be universal. If making transactions is easy in the market (Market 3), WMC's contribution is either less in magnitude (in Experiment 1) or even insignificant (in Experiment 2).

Is cognitive ability, measured as WMC, really a decisive factor even when people can learn and gain experiences? Although we observe that learning does occur, our evidence indicates that it can only partially eliminate the differences in performance resulting from the heterogeneity in innate capability. However, experiences in financial markets appear to be another decisive element when we try to figure out why some subjects were more successful than the others in a more realistic market environment.

In the behavioral level, there are two distinct features regarding how subjects made their bids and conducted transactions. Firstly, we find that subjects with high WMC were better at identifying profitable extra-marginal trading opportunity. When exchanging extra-marginal units was less profitable when facing truth-telling agents (Experiment 1), high-WMC subjects had fewer extra-marginal transactions than low-WMC subjects. Instead, when profitable extra-marginal transactions were more feasible due to the fact that computer agents deviated from truth-telling behavior (Experiment 2), high-WMC subjects took the opportunity and made more extra-marginal transactions than low-WMC subjects. That is to say, high-WMC subjects were better at stealing deals and making positive profits from other market participants. Secondly, we find that high-WMC subjects tend to underbid more than low-WMC subjects in both Experiment 1 and Experiment 2. Underbidding will lead to postponed transactions with more favorable prices, and this is another reason why high-WMC subjects outperformed low-WMC ones.

In the psychological level, we notice that diverse trading environments demand different functions of working memory to make successful transactions. In a simple environment (Experiment 1), it is the simultaneous storage and manipulating capability which predicts subjects' performance. In a more realistic environment (Experiment 2), it is the coordination (reasoning) capability explaining subjects' performance. This is a very important step towards the ultimate goal of unfolding the black box and revealing the true mental processes behind people's market activities. Within the experimental methodology, the first step is usually to identify the critical mental function associated with a specific task (see, for example, Bao and Yu (2016)'s study regarding memory and discount rate). While the difference between our two experiments is salient, we still need more experimental evidence, probably from neuroeconomic studies, to pinpoint exactly how cognitive ability determines and directs the mental processes during market activities.

Compared with former studies which also examined the relationship between cognitive ability (or cognitive sophistication) and market behavior, our results confirmed that the higher subjects' ability, the better their performance. Additionally, our study offers a deeper insight into how cognitive ability influences people's trading behavior. Note that in our design, subjects could retrieve all the information from the experimental interface. The values of the goods were given, so there is no need to evaluate asset values as in the asset market experiments. The market mechanism is simple, so it is unlikely that subjects would be confused as those in experiments involving futures and spot markets. The differences in behavior observed in our study are unrelated to subjects' expectations of future prices or understandings of the market environment, and they purely reflect the differences in subjects' ability to sense market situations and develop decent trading tactics. Consequently, what we have discovered is a more fundamental phenomenon which may exist in a wide range of markets.

With the aforementioned discoveries, we hope that this study lays down the basis of heterogeneous agent models where agents' behavioral heterogeneity can be connected with their personal factors. If this connection can be well studied and fully understood, economists can add behavioral factors into HAMs in a natural and solid way. One advantage of this kind of HAMs is that economists can start with parameters consistent with target groups' actual distribution of personal factors and therefore make their HAMs more transparent.

There are limitations to our study. Firstly, our subjects only had few repetitions to learn how to gain profits in each market, and therefore our results cannot give a decisive conclusion about whether the differences caused by innate cognitive ability could be eliminated through persistent learning. We argue that our results could be interpreted as the initial effects of cognitive ability, when people make market decisions without experiences. How exactly cognitive ability and learning intertwine is very important, but it will take further experiments to fully unfold the details of this relationship. Secondly, it could be that our results apply only to the specific demand-supply arrangements used in our experiments. To generalize our results, we need further experiments utilizing various types of markets, such as dynamic (random or cyclical), imbalanced (Swastika design or marker power), and additionally, other pricing mechanisms. Thirdly, one may wonder whether WMC still plays a role when human subjects with different WMC interact directly with one another. Our findings cannot directly infer the results, however our findings here are an essential step toward more complex experimental studies.

The results of our analysis also bring about an important issue regarding bounded rationality. Herbert Simon stated clearly that cognitive limitations and the structure of the environment, like a pair of scissors, shape our bounded rationality (Simon, 1996), and we have to look at them simultaneously. On the other hand, one must use caution in interpreting our findings, in light of limited data about how subjects actually determine their bid values. We simply need more detailed evidence to clarify the role of cognitive ability in complex market decisions. Further exploration into subjects' bidding/asking behavior, perhaps with the help of neuroscientific methods, will enlighten and illuminate our field immeasurably.

Acknowledgments

We gratefully acknowledged the research support in the form of Taiwan National Science Council grant no. NSC 101-2410-H-029-010-MY2 and Ministry of Science and Technology grant MOST 103-2410-H-004-009-MY3.

Appendix A. Instructions for subjects

Welcome to the NCCU Experimental Economics Lab! You are going to participate in a double auction experiment. During the experiment, your profit is calculated based on your decisions, and your accumulated profits determines your monetary bonus. If you want to earn bonuses, please carefully read and follow the experiment's rules and procedures as follows.

Monetary payoffs

Your monetary bonuses are determined by your ranking position in your own market. You can earn an NT$250 cash bonus if you are first in your own market; you can earn an NT$150 cash bonus if you are second in your own market; you can earn an NT$75 cash bonus if you are third in your own market. No matter what your ranking positions are, you will receive NT$200 cash after participating in today's double auction experiment.

Members and grouping

Every participant who comes to the lab will play the role of a trader in the markets. During the experiment, every participant competes with software traders. That is to say, your competitors are all computer agents. At the beginning of the experiment, you will see the assigned identification numbers of all buyers and sellers, so you may know the exact number of traders in your market.

Your bonus is determined by your ranking position in your own market. After the experiments, we take your average ranking position within all of your markets to calculate your bonus.

Conversations with other participants are strictly prohibited during the experiment. Any attempt to influence other participants by articulating or making noise is also forbidden. Those who disobey the rules will be warned twice, and then asked to leave. In this event, they forgo any cash payoffs.

Trading rules

Double auction is a trading mechanism similar to online auctions, it will select those who submit the best bids and asks to conclude a transaction. What's different from online auction is that there are multiple buyers who want to buy with their bids and multiple sellers who want to sell with their asks in the double auction market.

Every trader in the market must send bids or asks, making trades so as to gain the maximum profit that they can. Next, we introduce how buyers and sellers earn profits in the markets.

If you are a buyer, we will assign you several tokens as the basis of your transactions. You can regard buyer tokens as the values a certain commodity means to you. When you successfully buy a commodity, your profit is the difference between the token value and the commodity's price. The assigned tokens may have different values. For a buyer, the tokens are ordered from high-to-low value and you must make transactions from the highest token first in descending order.

Exercise 1
If a buyer has three tokens, of which the values are 100, 50, and 30, respectively, he has to start trading with the token of 100. Suppose he buys the first commodity at the price of 60, the profit earned is 40: $100 - 60 = 40$. Next, he must trade based on the second token. Suppose that he buys the second commodity at 60, then the profit made is $50 - 60 = -10$. If he buys the third commodity at 29, what is his profit from the third transaction? How much is his total profit in this period?

If you are a seller, the situation is the exact opposite of a buyer's. We assign several tokens as the basis of transactions. Treat seller tokens as the production cost of the commodities. When you sell a commodity, your profit will be the difference between transaction price and the token value. The tokens we assigned may have different values. For a seller, the tokens will be ordered from low-to-high value and you must make transactions from the lowest token in ascending order.

Exercise 2
If a seller has three tokens, of which the values are 10, 60, and 80, respectively, he has to start trading with the token of 10. Suppose he sells the first commodity at the price of 40, the profit earned is 30: $40 - 10 = 30$. Next, he must trade based on the second token. Suppose he sells the second commodity at 40, then the profit made is $40 - 60 = -20$. If he sells the third commodity at 79, what is his profit from the third transaction? How much is his total profit in this period?

If you don't send bids/asks to the market, you lose the opportunity to trade and gain profits in that period. The tokens assigned to you in every period will not accumulate nor count towards profits. They are completely worthless if you keep them.

Auction procedures

In today's experiment, each one of you will participate in three market experiments. Each market experiment consists of 6 trading periods, and each trading period has 25 trading steps.

Your buyer/seller market identity as well as your opponents will remain the same throughout each market experiment. Once you advance to another market experiment, you might be appointed a different buyer/seller role, and your opponents' strategies might also be altered.

Every trader must send a bid/ask/pass decision to the market in every step of a trading period. The auction mechanism will start to match bids and asks after receiving decisions from all the traders. Consequently, it doesn't matter how fast you send your decisions. It is the number you send to the market that ultimately decides whether you win or lose the auction and how much profit you realize.

If you have quickly sent your bid/ask/pass decision to the market while other traders deliberate, you will need to wait for them. There is no time limit to make decisions, so you can think carefully before submitting your bids/asks. However, let us not unnecessarily prolong the whole experiment, so please make your own decisions in a reasonable timeframe!

Remember, when making your bids/asks, you must start from the first token value. We will enumerate all the token values on the screen for your reference, but you need to keep track of the current token on your own. You have only one chance to submit your bid/ask in each trading step, and your submission cannot be withdrawn or modified. Please enter your decisions carefully and enter all values as an integer.

If you have bought up/sold out all the tokens you have, you are finished for that period, and can only watch others playing until the end of that period. Your tokens are replenished as the next trading period starts, and you may submit bids/asks again at this point.

Transaction prices

The transaction prices are calculated based on buyers' bids and sellers' asks. First, the auction picks the buyer who submits the highest bid as well as the seller who submits the lowest ask; and these two traders will proceed to the next process. Whether these two traders can make a trade depends on the requirement that the buyer's bid is higher than the seller's ask. If this requirement is met, the mean of their bid and ask is taken as the transaction price of the commodity.

Exercise 3
In a market consisting of three buyers and three sellers: if during a certain trading step the buyers' bids are (90, 100, 70), respectively, and the sellers' asks are (20, 50, 35), respectively, which buyer has the highest bid? Which seller has the lowest ask? Can they reach a transaction? What will be the transaction price?

All traders except the highest bidder and the lowest seller will not have chances to make trades in that trading step. What they can do is simply wait until the next trading step, when they can submit bids/asks again and try to seize profit opportunity.

Logging procedures and introduction to the computer interface

(omitted here due to screenshots in Chinese)

Practices

Now, let us commence some trial practices. The results of the trial practices will not be recorded and used in payoff calculations. We hope you can familiarize yourselves with the computer interface operations and thoroughly understand the auction market mechanism.

Raise your hand should you have any question regarding the computer interface or the trading procedures during the trial periods. We will answer you immediately.

Appendix B. Subjects' redemption prices (token values) for each market

The following table (Table B.14) presents the reservation prices of buyers and costs of sellers (the redemption values, or the token values) for our three markets. The values are decreasing for buyers, obeying the rule of decreasing marginal benefits for buyers. The values are increasing for sellers to imply increasing marginal costs.

Table B.14
Subjects' Redemption Prices (Token Values) for Each Market.

		Buyer 1	Buyer 2	Buyer 3	Buyer 4	Seller 1	Seller 2	Seller 3	Seller 4
M1	Token 1	390	390	390	390	310	310	310	310
	Token 2	370	370	370	370	340	340	340	340
	Token 3	356	356	356	356	360	360	360	360
	Token 4	338	338	338	338	372	372	372	372
M2	Token 1	164	160	156	152	30	34	38	42
	Token 2	108	112	116	120	154	150	146	142
	Token 3	104	100	96	92	158	162	166	170
	Token 4	76	80	84	88	186	182	178	174
M3	Token 1	1300	1275	1250	1225	700	725	750	775
	Token 2	1125	1150	1175	1200	875	850	825	800
	Token 3	1100	1075	1050	1025	900	925	950	975
	Token 4	925	950	975	1000	1075	1050	1025	1000

Appendix C. The information revealed to subjects

Fig. C.12 is a snapshot of what our subjects would see on their screens. Subjects enter their decisions in the left part of the window by entering their bid/ask prices. The "pass" button means subjects can choose to skip this trading step without submitting any price. There is a small table on the top of the window, which reports the token values for our human subjects in each market. Another small table in the bottom reports the raw profit earned during each trading period.

The large table in the main body of the screen contains all market information during the experiments. The information specifics are as follows:

- Column 1: The index of trading period.
- Column 2: The index of trading step.
- Column 3–6: Past bidding prices from buyers.
- Column 7–10: Past asking prices from sellers.
- Column 11: The winning buyer and their bid price ("- -" to indicate no buyer won because of a failure of reaching a transaction at the last step).
- Column 12: The winning seller and their ask price ("- -" to indicate no seller won because of a failure of reaching a transaction at the last step).
- Column 13: The transaction price. ("-1" to indicate no transaction took place at the last step).

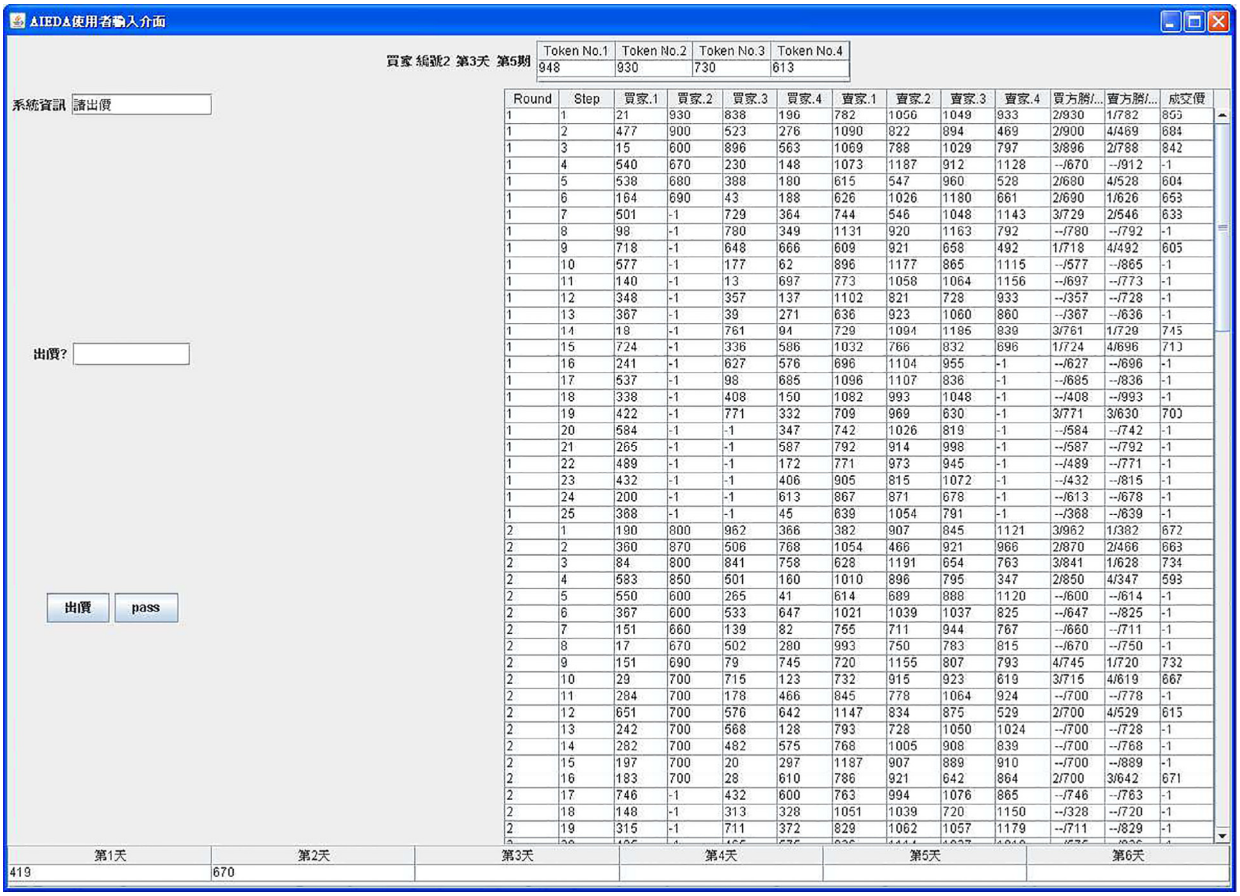

Fig. C.12. A sample snapshot of the auction information presented to our subjects during experiments.

Appendix D. Taxonomy of subjects and numbers

Table D.15
Taxonomy of Subjects and Numbers.

		Exp1	Exp2
Total	Original	173	173
	Valid	170	169
Gender	Male	80	96
	Female	90	73
Position	Buyer	87	83
	Seller	88	81
Experience	Online auction	104	95
	Financial markets	24	25
	Other	17	16
WMC	High group	97	99
	Low group	73	70
Bonus	Champion	11%	20%
	First runner-up	21%	23%
	Second runner-up	20%	27%

The numbers shown in this table are restricted to only the valid observations. Three subjects from Experiment 1 and four subjects from Experiment 2 are dropped out of the sample mainly due to data errors.

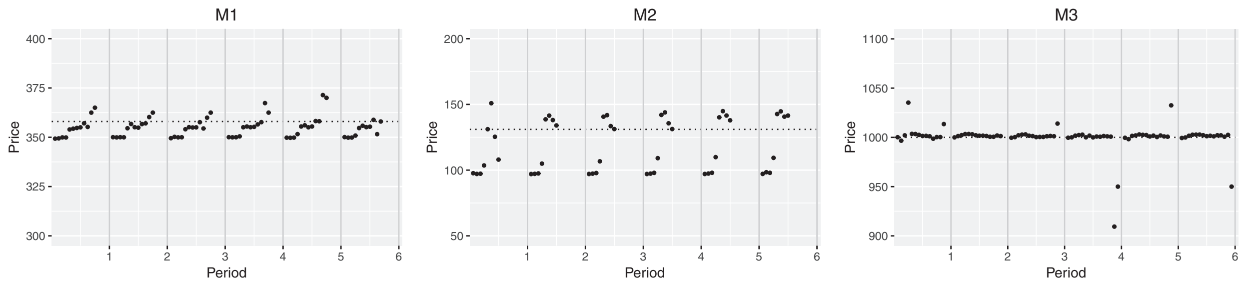

Fig. E.13. The evolution of average prices in Experiment 1.

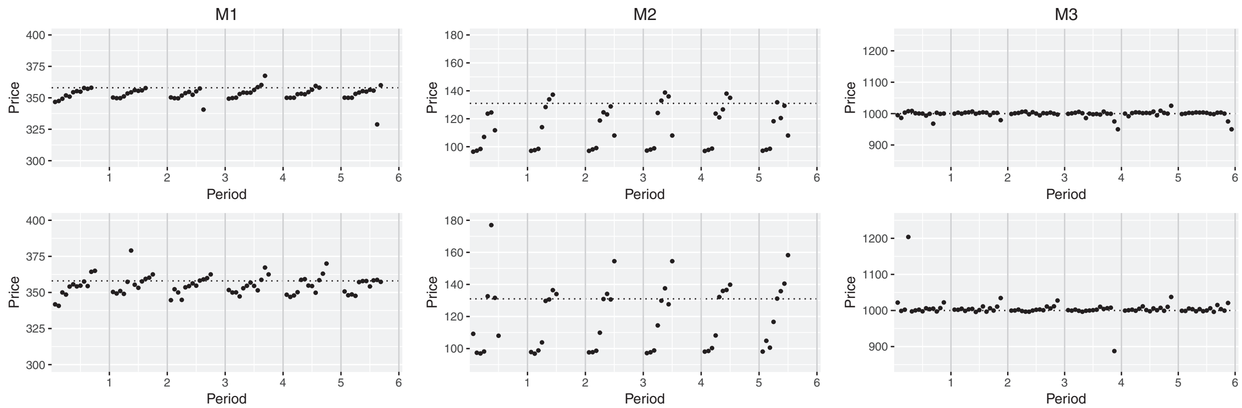

Fig. E.14. The evolution of average prices in Experiment 1: High versus Low group.

Appendix E. Convergence and efficiency

The average transaction prices in Experiment 1 are presented in Fig. E.13. The *i*th dot in a period represents the average price of the *i*th transactions. For example, the third dot in Period 1 plots the average price of the third transactions occurred in all auctions in Period 1, regardless they were conducted by human subjects or not. The order of the dots represents the order of transactions and has nothing to do with the actual time. Furthermore, the numbers of dots in each period usually exceeds the equilibrium quantities because whenever there is a transaction, we have to calculate the average price. Take M1 for example, the equilibrium quantity is eight, but there are twelve dots in Period 1. This doesn't mean that we had twelve transactions in all experimental auctions. The actual situation is that only a few auction markets had transactions up to twelve units. This explains why the final few dots in each period are sometimes far away from the equilibrium price.

Because the supply and demand curves are asymmetric in M1 and M2, the prices converge to equilibrium from below. This phenomenon was first observed by Smith (1962), who found that the price will converge from the side that has a larger surplus. In our case, producer surplus is larger than consumer surplus in M1 and M2, so the prices converge from the seller side.

One may wonder whether subjects with different levels of WMC will lead to different price patterns. We divided the subjects into two groups–the High WMC group and the Low WMC Group–according to their WMC scores. Subjects with WMC scores above zero are marked as 'High', others as 'Low'. Fig. E.14 presents the average prices for High and Low groups. The upper part is the price dynamics for High group, and lower part is that of the Low group. We only include transactions where human subjects were involved.

There seems to be no obvious differences between the High and Low group. In fact, not only the standard deviations of prices are very close for these two group of subjects, Wilcoxon rank-sum test shows that the mean absolute deviation from equilibrium prices of both groups are not significantly different (with p-values of .2156, .9487, and .2638 for M1, M2, and M3, respectively). Fig. E.15 demonstrates the average market efficiencies (total realized surplus) for High and Low group in each market. It is clear that no matter how large WMC a human subject has, the market efficiency in their market is close to one since the beginning of the experiment. Again, this is consistent with that observed in the literature and shows how double auction mechanism is capable in generating equilibrium results. Despite so, we cannot regard the finding here as an opposition to the assertion that market participants' cognitive ability will influence market dynamics. The reasons is that we have only one human subject in each market, so their influences on market dynamics is quite limited. What we want to discuss in this paper is to study how cognitive ability influences market participants' behavior and performance in a fundamental environment, and the current experimental design meets our need.

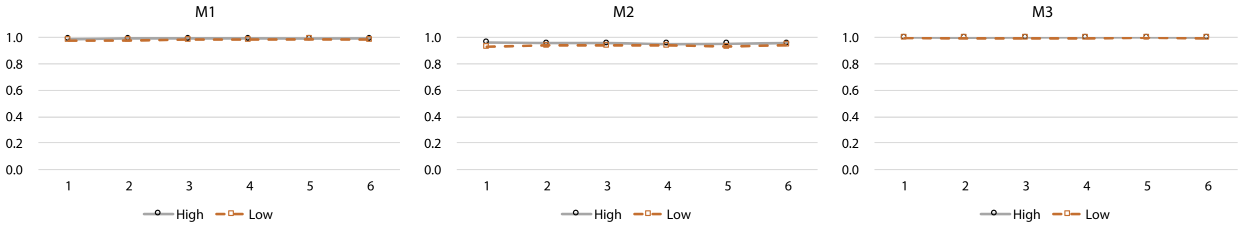

Fig. E.15. Market efficiency in Experiment 1: High versus Low group.

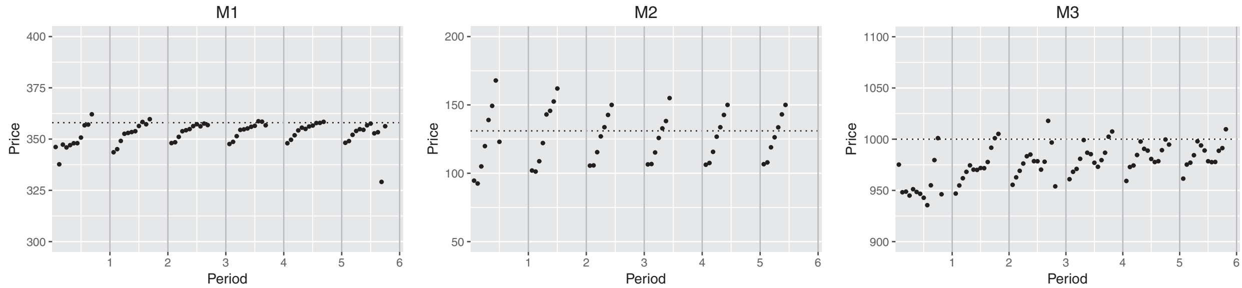

Fig. E.16. The evolution of average prices in Experiment 2.

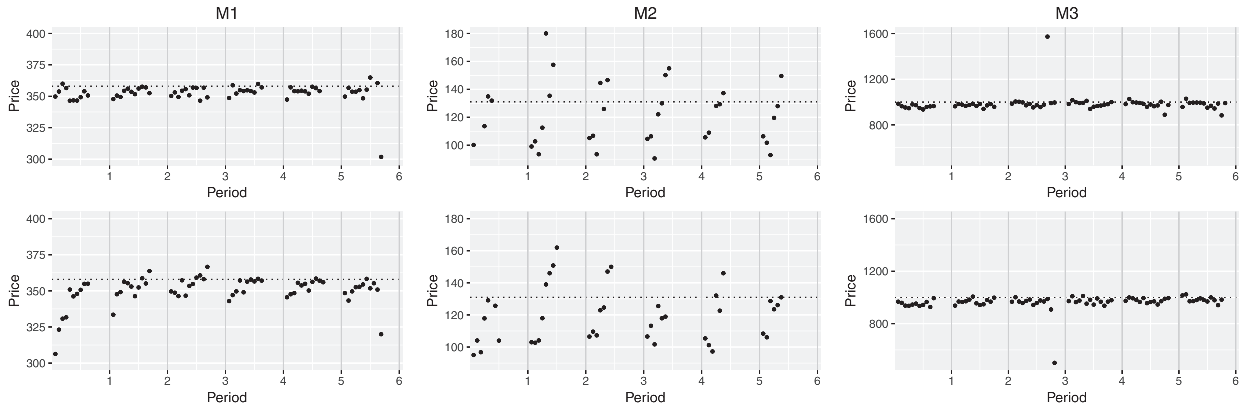

Fig. E.17. The evolution of average prices in Experiment 2: High versus Low group.

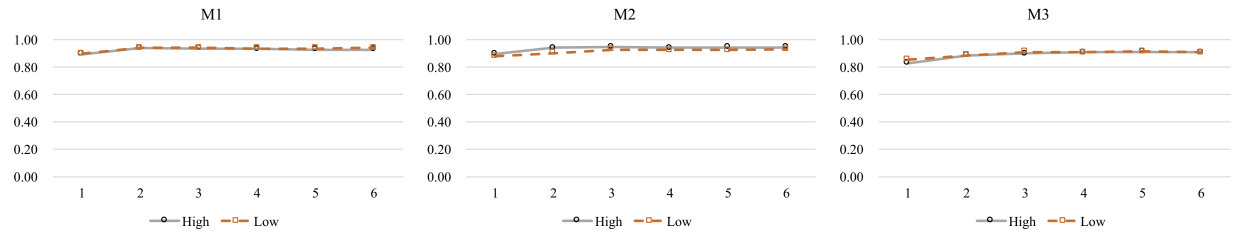

Fig. E.18. Market efficiency in Experiment 2: High versus Low group.

Similarly, Fig. E.16 presents the average prices for Experiment 2, and Fig. E.17 demonstrates the High-Low contrast of average prices. We find that the convergence toward equilibrium is slower than that in Experiment 1, but the direction of convergence is still consistent with previous studies. Again, there is no obvious differences in the price patterns of the High and Low groups. Wilcoxon rank-sum test shows that the mean absolute deviation from equilibrium prices of both groups are not significantly different (with p-values of .8414, .2256, and .4908 for M1, M2, and M3, respectively).

Fig. E.18 demonstrates the average market efficiencies for High and Low groups in each market. It is clear that High group and Low group have almost the same market efficiency, although market efficiencies in Experiment 2 are lower than those in Experiment 1.

Table F.16
Results of Wilcoxon rank sum tests for the High-Low comparisons–Experiment 1.

	M1			M2			M3		
Period	High	Low	p-value	High	Low	p-value	High	Low	p-value
1	85	55	.0119	91	44	.0082	94	77	.1332
	(41.71)	(100.01)		(62.44)	(135.70)		(19.85)	(122.14)	
2	97	68	.0002	111	78	.0286	97	95	.1279
	(32.88)	(85.29)		(56.10)	(87.72)		(11.40)	(11.66)	
3	98	80	.0055	114	84	.0555	100	95	.0113
	(42.71)	(68.63)		(69.05)	(93.79)		(8.47)	(12.18)	
4	106	88	.0083	119	89	.0938	100	94	.0023
	(30.71)	(49.74)		(54.86)	(96.93)		(11.89)	(16.00)	
5	105	82	.0180	117	91	.3074	101	96	.0017
	(36.38)	(75.82)		(63.60)	(94.45)		(11.24)	(12.86)	
6	103	91	.0500	122	97	.4364	102	97	.0001
	(47.06)	(53.79)		(56.79)	(103.78)		(9.05)	(11.94)	

Note: Standard deviations are in parentheses.

Table F.17
The effects of learning in Experiment 1.

	High			Low			High v.s. Low?
Market	Average	S.D.	p-value	Average	S.D.	p-value	p-value
M1	17.96	53.53	4.942E-07	35.33	95.61	1.371E-06	.4689
M2	30.62	56.90	4.273E-10	53.46	139.16	3.615E-07	.1318
M3	8.64	18.86	6.803E-10	19.77	121.55	.00023	.2677

Table F.18
Results of Wilcoxon rank sum tests for the High-Low comparisons–Experiment 2

	M1			M2			M3		
Period	High	Low	p-value	High	Low	p-value	High	Low	p-value
1	91	56	.1588	133	69	.0009	121	112	.2332
	(125.42)	(194.17)		(97.57)	(163.37)		(63.03)	(70.70)	
2	111	79	.0026	117	105	.0551	130	110	.0278
	(54.80)	(81.00)		(118.38)	(78.42)		(57.09)	(62.31)	
3	116	96	.0080	128	101	.0011	124	115	.0376
	(57.81)	(56.49)		(84.16)	(62.19)		(99.71)	(60.15)	
4	126	87	.0010	139	105	.0019	137	109	.0349
	(51.79)	(77.20)		(69.44)	(67.48)		(54.46)	(81.16)	
5	128	102	.0037	142	114	.0024	139	119	.0439
	(56.14)	(59.94)		(60.81)	(68.00)		(55.39)	(65.89)	
6	112	100	.0301	144	100	.0028	140	120	.0545
	(109.42)	(82.48)		(67.30)	(126.14)		(58.58)	(58.21)	

Note: Standard deviations are in parentheses.

Table F.19
The effects of learning in Experiment 2.

	High			Low			High v.s. Low?
Market	Average	S.D.	p-value	Average	S.D.	p-value	p-value
M1	21.49	136.09	.00316	44.58	196.73	.05418	.7436
M2	10.35	80.78	.02931	30.63	194.04	.01924	.3959
M3	19.02	51.21	.00084	8.22	56.26	.26370	.2092

Appendix F. The High-Low comparisons

The 2nd to 7th columns of Table F.17 report the average and standard deviation of the effect of learning as well as the p-values of the Wilcoxon signed rank tests for whether the mean is different from zero. These statistics confirm the improvement made by our subjects.

The last column of Table F.17 presents the p-values of Wilcoxon rank sum Tests on the effects of learning of both groups under 5% significance level. We therefore cannot conclude that the low group exhibits larger improvement although its average improvement is indeed larger.

Appendix G. The results of quantile regression

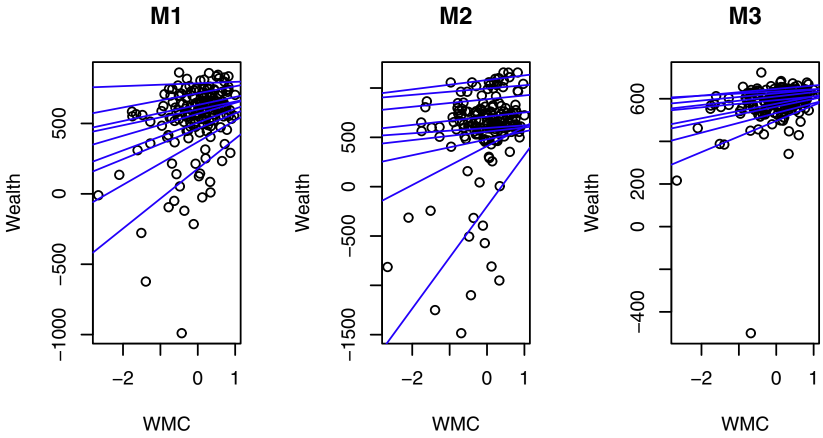

Fig. G.19. Results of quantile regressions–Experiment 1.

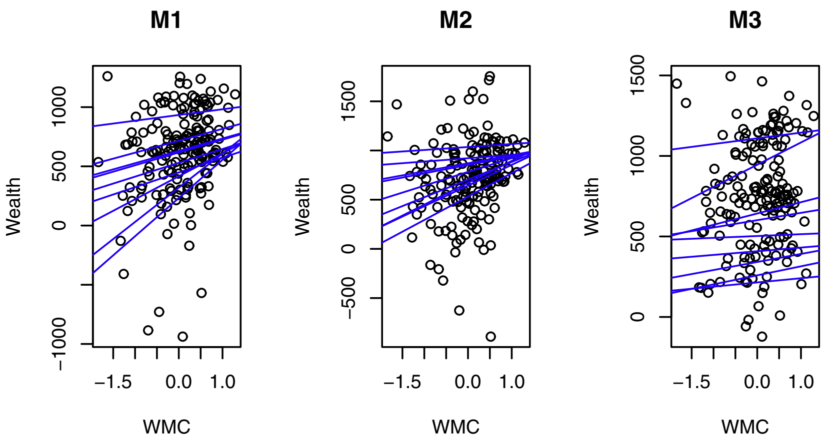

Fig. G.20. Results of quantile regressions–Experiment 2.

References

Allen, H., Taylor, M.P., 1990. Charts, noise and fundamentals in the london foreign exchange market. Econ. J. 100 (400), 49–59.
Baddeley, A.D., 2000. The episodic buffer: a new component of working memory? Trends Cogn. Sci. (Regul. Ed.) 4, 417–423.
Baddeley, A.D., Hitch, G., 1974. Working memory. Psychol. Learn. Motivation 8, 47–89.
Bao, T., Yu, X., 2016. Memory and discounting: theory and evidence. Working paper. https://doi.org/DOI:10.13140/RG.2.2.27675.46883.
Bénassy-Quéré, L.S., Larribeau, S., MacDonald, R., 2003. Models of exchange rate expectations: how much heterogeneity? Int. Fin. Markets Inst. Money 13, 113–136.
Bosch-Rosa, C., Meissner, T., Bosch-Domènech, A., 2015. Cognitive bubbles, 1464. Department of Economics and Business, Universitat Pompeu Fabra. Economics Working Papers. Available at https://ideas.repec.org/p/upf/upfgen/1464.html.
Breaban, A., Noussair, C.N., 2015. Trader characteristics and fundamental value trajectories in an asset market experiment. J. Behav. Exp. Fin. 8, 1–17.
Brock, W., Hommes, C., 1997a. Models of Complexity in Economics and Finance. In: Heij, C., Schumacher, J., Hanzon, B., Praagman, C. (Eds.), Systems Dynamics in Economic and Finance Models. Wiley, New York, pp. 3–44. Chapter 1
Brock, W., Hommes, C., 1997b. A rational route to randomness. Econometrica 65, 1059–1095.
Brock, W., Hommes, C., 1998. Heterogeneous beliefs and routes to chaos in a simple asset pricing model. J. Econ. Dyn. Control 22, 1235–1274.
Bullard, J., Duffy, J., 1999. Using genetic algorithms to model the evolution of heterogeneous beliefs. Comput. Econ. 13, 41–60.
Casari, M., Ham, J., Kagel, J., 2007. Selection bias, demographic effects, and ability effects in common value auction experiments. Am. Econ. Rev. 97 (4), 1278–1304.
Cattell, R.B., 1963. Theory of fluid and crystallized intelligence: a critical experiment. J. Educ. Psychol. 54, 1–22.
Chamberlin, E.H., 1948. An experimental imperfect market. J. Pol. Econ. 56 (2), 95–108.
Chen, S.-H., Shih, K.-C., Tai, C.-C., 2012. Can Artificial Traders Learn and Err like Human Traders? a New Direction for Computational Intelligence in Behavioral Finance. In: Doumpos, M., Zoupounidis, C., Paralos, P. (Eds.), Financial Decision Making using Computational Intelligence, Series in Optimisation and its Applications. Springer-Verlag, Berlin, pp. 35–69.
Chen, S.-H., Tai, C.-C., 2010. The Agent-based Double Auction Markets: 15 Years on. In: Takadama, K., Cioffi-Revilla, C., Deffuant, G. (Eds.), Simulating Interacting Agents and Social Phenomena. Springer, Tokyo, pp. 119–136.

Please cite this article as: C.-C. Tai et al., Cognitive ability and earnings performance: Evidence from double auction market experiments, Journal of Economic Dynamics & Control (2018), https://doi.org/10.1016/j.jedc.2017.12.008

Chen, S.-H., Tai, C.-C., Wang, S.G., 2010. Does Cognitive Capacity Matter When Learning Using Genetic Programming in Double Auction Markets? In: Di Tosto, G., Parunak, H. (Eds.) Multi-Agent-Based Simulation X: International Workshop, MABS 2009, Budapest, Hungary, 2009. Springer-Verlag, Berlin/Heidelberg, Revised Selected Papers, pp. 37–48. May 11-12.
Chen, S.-H., Yu, T., 2011. Agents learned, but do we? knowledge discovery using the agent-based double auction markets. Frontiers Electr. Electron. Eng. (FEE) China 6 (1), 159–170.
Chiarella, C., Dieci, R., Gardini, L., 2006. Asset price and wealth dynamics in a financial market with heterogeneous agents. J. Econ. Dyn. Control 30 (9–10), 1755–1786.
Chiarella, C., Dieci, R., He, X.-Z., 2007. Heterogeneous expectations and speculative behavior in a dynamic multi-asset framework. J. Econ. Behav. Org. 62 (3), 408–427.
Chiarella, C., Dieci, R., He, X.-Z., Li, K., 2013. An evolutionary CAPM under heterogeneous beliefs. Annals Finance 9 (2), 185–215.
Chiarella, C., ter Ellen, S., He, X.-Z., Wu, E., 2015. Fear or fundamentals? heterogeneous beliefs in the european sovereign CDS market. J. Empir. Finance 32 (C), 19–34.
Chiarella, C., He, X.-Z., 2001. Asset price and wealth dynamics under heterogeneous expectations. Quant. Finance 1 (5), 509–526.
Chiarella, C., He, X.-Z., 2002. Heterogeneous beliefs, risk and learning in a simple asset pricing model. Comput. Econ. 19 (1), 95–132.
Chiarella, C., He, X.-Z., 2003a. Dynamics of beliefs and learning under a_l-processes – the heterogeneous case. J. Econ. Dyn. Control 27 (3), 503–531.
Chiarella, C., He, X.-Z., 2003b. Heterogeneous beliefs, risk and learning in a simple asset pricing model with a market maker. Macroecon. Dyn. 7 (4), 503–536.
Chiarella, C., He, X.-Z., Huang, W., Zheng, H., 2012. Estimating behavioural heterogeneity under regime switching. J. Econ. Behav. Org. 83 (3), 446–460.
Chiarella, C., Iori, G., Perelló, J., 2009. The impact of heterogeneous trading rules on the limit order book and order flows. J. Econ. Dyn. Control 33 (3), 525–537.
Christelis, D., Jappelli, T., Padula, M., 2010. Cognitive abilities and portfolio choice. Eur. Econ. Rev. 54 (1), 18–38.
Cliff, D., Bruten, J., 1997. Zero is not enough: On the lower limit of agent intelligence for continuous double auction markets. Technical Report no. HPL-97-141, Hewlett-Packard Laboratories. Available via CiteSeer:. http://citeseer.ist.psu.edu/cliff97zero.html.
Conway, A.R.A., Engle, R.W., 1996. Individual differences in working memory capacity: more evidence for a general capacity theory. Memory 4, 577–590.
Corgnet, B., Hernán-González, R., Kujal, P., Porter, D., 2015. The effect of earned versus house money on price bubble formation in experimental asset markets. Rev. Finance 19, 1455–1488.
Craig, S., Lewandowsky, S., 2012. Whichever way you choose to categorize, working memory helps you learn. Q. J. Exp. Psychol. 65, 439–464.
Croson, R., Gneezy, U., 2009. Gender differences in preferences. J. Econ. Lit. 47 (2), 448–474.
Cueva, C., Rustichini, A., 2015. Is financial instability male-driven? gender and cognitive skills in experimental asset markets. J. Econ. Behav. Org. 119, 330–344.
Daneman, M., Carpenter, P.A., 1983. Individual differences in integrating information between and within sentence. J. Exp. Psychol. 9, 561–583.
Dato, S., Nieken, P., 2014. Gender differences in competition and sabotage. J. Econ. Behav. Org. 100, 64–80.
Davis, D.D., Harrison, G.W., Williams, A.W., 1993. Convergence to nonstationary competitive equilibria: an experimental analysis. J. Econ. Behav. Org. 22 (3), 305–326.
Davis, D.D., Holt, C.A., 1993. Experimental Economics. Princeton University Press, New Jersey.
Davis, D.D., Williams, A.W., 1991. The hayek hypothesis in experimental auctions: institutional effects and market power. Econ. Inq. 29 (2), 261–274.
Day, R., Huang, W., 1990. Bulls, bears and market sheep. J. Econ. Behav. Org. 14 (3), 299–329.
De Martino, B., O'Doherty, J.P., Ray, D., Bossaerts, P., Camerer, C., 2013. In the mind of the market: theory of mind biases value computation during financial bubbles. Neuron 79 (6), 1222–1231.
Deaves, R., Luders, E., Luo, G.Y., 2008. An experimental test of the impact of overconfidence and gender on trading activity. Rev. Finance 13 (3), 555–575.
Dechenaux, E., Kovenock, D., Sheremeta, R.M., 2015. A survey of experimental research on contests, all-pay auctions and tournaments. Experimental Economics 18 (4), 609–669.
Devetag, G., Warglien, M., 2003. Games and phone numbers: do short-term memory bounds affect strategic behavior? J Econ Psychol 24, 189–202.
Easley, D., Ledyard, J., 1993. Theories of Price Formation and Exchange in Double Oral Auction. In: Friedman, D., Rust, J. (Eds.), The Double Auction Market–Institutions, Theories, and Evidence. Reading. MA, Addison-Wesley.
Elliott, G., Ito, T., 1999. Heterogeneous expectations and tests of efficiency in the yen/dollar forward exchange rate market. J. Monet. Econ. 43, 435–456.
Engle, R.W., Kane, M.J., Tuholsky, S.W., 1999. Individual Differences in Working Memory Capacity and What They Tell Us about Controlled Attention, General Fluid Intelligence, and Functions of Prefrontal Cortex. In: Miyake, A., Shah, P. (Eds.), Models of Working Memory: Mechanisms of Active Maintenance and Executive Control. Cambridge University Press, New York, pp. 102–134.
Franke, R., Nesemann, T., 1999. Two destabilizing strategies may be jointly stabilizing. J. Econ. 69 (1), 1–18.
Franke, R., Sethi, R., 1998. Cautious trend-seeking and complex asset price dynamics. Res. Econ. 52 (1), 61–79.
Frankel, J.A., Froot, K.A., 1987a. Using survey data to test standard propositions regarding exchange rate expectations. Am. Econ. Rev. 77, 133–153.
Frankel, J.A., Froot, K.A., 1987b. Short-term and long-term expectations of the yen/dollar exchange rate: evidence from survey data. J. Jpn. Int. Econ. 1, 249–274.
Frankel, J.A., Froot, K.A., 1990a. Chartists, Fundamentalists and the Demand for Dollars. In: Courakis, A., Taylor, M. (Eds.), Private Behaviour and Government Policy in Interdependent Economies. Oxford University Press, New York, pp. 73–126.
Frankel, J.A., Froot, K.A., 1990b. The rationality of the foreign exchange rate. chartists, fundamentalists and trading in the foreign exchange market. Am. Econ. Rev. AEA Papers Proc. 80 (2), 181–185.
Furse, C., Haldane, A., Goodhart, C., Cliff, D., Zigrand, J. P., Houstoun, K., Linton, O., Bond, P., 2011. The future of computer trading in financial markets. Working paper, Foresight, Government Office for Science, London (UK). http://www.bis.gov.uk/foresight/our-work/projects/current-projects/computer-trading.
Gaunersdorfer, A., 2000. Endogenous fluctuations in a simple asset pricing model with heterogeneous agents. J. Econ. Dyn. Control 24, 799–831.
Gjerstad, S., Dickhaut, J., 1998. Price formation in double auctions. Games Econ. Behav. 22, 1–29.
Gode, D., Sunder, S., 1993. Allocative efficiency of markets with zero intelligence traders: market as a partial substitute for individual rationality. J. Pol. Econ. 101, 119–137.
Grinblatt, M., Keloharju, M., Linnainmaa, J.T., 2011. IQ and stock market participation. J. Finance 66 (6), 2121–2164.
Grinblatt, M., Keloharju, M., Linnainmaa, J.T., 2012. IQ, trading behavior, and performance. J. Financial Econ. 104, 339–362.
Hinson, J.M., Jameson, T.L., Whitney, P., 2003. Impulsive decision making and working memory. J. Exp. Psychol. 29, 298–306.
Holt, C.A., 1995. Industrial Organization: A Survey of Laboratory Research. In: kagel, J., Roth, A. (Eds.), The Handbook of Experimental Economics. Princeton University Press, p. 349443.
Hommes, C., 2001. Financial markets as nonlinear adaptive evolutionary systems. Quant. Finance 1, 149–167.
Hommes, C., Lux, T., 2013. Individual expectations and aggregate behavior in learning-to-forecast experiments. Macroecon. Dyn. 17 (02), 373–401.
Hommes, C.H., 2006. Heterogeneous Agent Models in Economics and Finance. In: Tesfatsion, L., Judd, K. (Eds.), Handbook of Computational Economics, Vol. 2. Elsevier, pp. 1109–1186. Chapter 23
Ito, K., 1990. Foreign exchange rate expectations. Am. Econ. Rev. 80, 434–449.
Kane, M.J., Engle, R.W., 2002. The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. Psychonomic Bull. Rev. 9, 637–671.
King, J., Just, M.A., 1991. Individual differences in syntactic processing: the role of working memory. J. Mem. Lang. 30, 580–602.
Lewandowsky, S., 2011. Working memory capacity and categorization: individual differences and modeling. J. Exp. Psychol. 37, 720–738.
Lewandowsky, S., Oberauer, K., Yang, L.-X., Ecker, U.K.H., 2010. A working memory test battery for MATLAB. Behav. Res. Methods 42 (2), 571–585.

Lewandowsky, S., Yang, L.-X., Newell, B.R., Kalish, M., 2012. Working memory does not dissociate between different perceptual categorization tasks. J. Exp. Psychol. 38, 881–904.
Lux, T., 1995. Herd behaviour, bubbles and crashes. Econ. J. 105 (431), 881–896.
Lux, T., 1997. Time variation of second moments from a noise trader/infection model. J. Econ. Dyn. Control 22 (1), 1–38.
Lux, T., 1998. The socio-economic dynamics of speculative markets: interacting agents, chaos, and the fat tails of return distributions. J. Econ. Behav. Org. 33 (2), 143–165.
Lux, T., Marchesi, M., 1999. Scaling and criticality in a stochastic multi-agent model of a financial markets. Nature 397, 498–500.
Lux, T., Marchesi, M., 2000. Volatility clustering in financial markets: a micro-simulation of interacting agents. Int. J. Theor. Appl. Finance 3 (4), 675–702.
MacDonald, R., Marsh, I., 1996. Currency forecasters are heterogeneous: confirmation and consequences. J. Int. Money Finance 15, 665–685.
Neath, I., Surprenant, A.M., 2003. Human Memory, (2nd ed.) Wadsworth / Thomson Learning, CA.
Niederle, M., Vesterlund, L., 2007. Do women shy away from competition? do men compete too much? Q. J. Econ. 122 (3), 1067–1101.
Niederle, M., Vesterlund, L., 2011. Gender and competition. Annu. Rev. Econ. 3 (1), 601–630.
Noussair, C.N., Tucker, S., Xu, Y., 2016. Futures markets, cognitive ability, and mispricing in experimental asset markets. J. Econ. Behav. Org. 130, 166–179.
Oberauer, K., 1993. Die koordination kognitiver operationen. eine studie zum zusammenhang von 'working memory' und intelligenz [the coordination of cognitive operations. a study on the relationship of working memory and intelligence]. Zeitschrift für Psychologie 201, 57–84.
Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., Wittmann, W.W., 2000. Working memory capacity–facets of a cognitive ability construct. Personality Individual Diff. 29, 1017–1045.
Oberauer, K., Süß, H.-M., Wilhelm, O., Wittmann, W., 2003. The multiple faces of working memory–storage, processing, supervision, and coordination. Intelligence 31, 167–193.
Pareto, V., 1971. Manual of Political Economy. A. M. Kelley, New York. A. S. Schwier, Ed./Trans. (Original work published 1906)
Rust, J., Miller, J., Palmer, R., 1993. Behavior of Trading Automata in a Computerized Double Auction Market. In: Friedman, D., Rust, J. (Eds.), Double Auction Markets: Theory, Institutions, and Laboratory Evidence. Addison-Wesley, Reading, MA, pp. 155–198.
Rust, J., Miller, J., Palmer, R., 1994. Characterizing effective trading strategies: insights from a computerized double auction tournament. J. Econ. Dyn. Control 18, 61–96.
Sethi, R., 1996. Endogenous regime switching in speculative markets. Struct. Change Econ. Dyn. 7 (1), 99–118.
Simon, H., 1996. The Sciences of the Artificial. MIT Press, Cambridge, MA.
Smith, V.L., 1962. An experimental study of competitive market behavior. J. Pol. Econ. 70 (2), 111–137.
Smith, V.L., Williams, A.W., 1982. The effects of rent asymmetries in experimental auction markets. J. Econ. Behav. Org. 3 (1), 99–116.
Smith, V.L., Williams, A.W., 1990. The Boundaries of Competitive Price Theory: Convergence, Expectations, and Transaction Costs. In: Green, L., Kagel, J. (Eds.), Advances in Behavioral Economics, Vol. 2. Ablex Publishing, Norwood, NJ, pp. 31–53.
Süß, H.-M., Oberauer, K., Wittmann, W.W., Wilhelm, O., Schulze, R., 2002. Working memory capacity explains reasoning ability and a little bit more. Intelligence 30, 261–288.
Taylor, M.P., Allen, H., 1992. The use of technical analysis in the foreign exchange market. J. Int. Money Finance 11, 304–314.
Thaler, R., 2000. From homo economicus to homo sapiens. J. Econ. Perspect. 14 (1), 133–141.
Turner, M.L., Engle, R.W., 1989. Is working memory task dependent? J. Mem. Lang. 28, 127–154.
Unsworth, N., Fukuda, K., Awh, E., Vogel, E.K., 2014. Working memory and fluid intelligence: capacity, attention control, and secondary memory retrieval. Cogn. Psychol. 71, 1–26.
Vissing-Jorgensen, A., 2003. Perspective on Behavioral Finance: Does 'Irrationality' Disappear with Wealth? Evidence from Expectations and Actions. In: Gertler, M., Rogoff, K. (Eds.), NBER Macroeconomics Annual. MIT Press.
Zhan, W., Friedman, D., 2007. Markups in double auction markets. J. Econ. Dyn. Control 31, 2984–3005.