Journal of Economic Dynamics & Control 135 (2022) 104293

Contents lists available at ScienceDirect

Journal of Economic Dynamics & Control

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

Number sense, trading decisions and mispricing: An experiment

Tristan Roger[a,*], Patrick Roger[b], Marc Willinger[c]

[a] CEREFIGE, ICN Business School, Université de Lorraine, 86 rue Sergent Blandan, 54000 Nancy, France
[b] LaRGE Research Center, EM Strasbourg Business School, University of Strasbourg, 61 avenue de la Forêt Noire, 67085 Strasbourg Cedex, France
[c] CEE-M, Univ Montpellier, CNRS, INRAE, SupAgro, Avenue Raymond Dugrand, CS 79606, 34 960 Montpellier Cedex 2, France

ARTICLE INFO

Article history:
Received 2 May 2021
Revised 10 November 2021
Accepted 14 December 2021
Available online 16 December 2021

Keywords:
Experimental markets
Number line estimation (NLE)
Number sense
Trading performance
Mispricing

ABSTRACT

We show that the acuity of the Approximate Number System (ANS), a cognitive system that allows humans and many animal species to evaluate quantities without using exact calculations, is a strong predictor of subjects' earnings in experimental markets. We measure ANS acuity with a bounded number line estimation (NLE) task and find that subjects who perform better on the NLE task, obtain higher earnings in a continuous double auction experimental market. We underline two channels through which high ANS acuity subjects achieve better performance: they are rewarded for offering liquidity and are faster at exploiting trading opportunities. We also show that, in a given market, the distribution of NLE scores influences mispricing. Our results are unchanged when we control for differences in trading intensity, risk aversion, background education or demographic characteristics.

1. Introduction

Since Markowitz (1952) and Sharpe (1964), conventional finance models assume that economic agents are rational. The rationality assumption implies that agents have unlimited cognitive skills. These rational agents are able to instantaneously perform complex calculations that lead to optimal investment decisions. In practice, however, agents have limited cognitive abilities. They rely on heuristics to compensate for their cognitive limitations (Kahneman, 2011). Since heuristics involve performing approximate calculations, agents skilled at quickly performing such approximate calculations are likely to make better economic and financial decisions. This ability to perform such approximate calculations is linked to the acuity of the Approximate Number System (ANS).

The ANS is an important feature of the brain, allowing humans and many animal species to approximately evaluate and discriminate quantities without counting or using a symbolic representation of numbers (Odic and Starr, 2018 and Feigenson et al., 2004 for a review). The ability to estimate numbers (albeit imprecisely) is deeply rooted in the human brain and precedes the acquisition of formal mathematical education. Evidence of the existence of the ANS has been found in babies (Hyde and Spelke, 2011; Izard et al., 2009; Libertus and Brannon, 2009; 2010; Lipton and Spelke, 2003; Xu and Spelke, 2000; Xu et al., 2005) and even in animals (Agrillo et al., 2008, 2009; Gallistel and Gelman, 1992; Garland et al., 2012; Mehlis et al., 2015; Nieder and Dehaene, 2009; Pepperberg, 2006; Perdue et al., 2012).

* Corresponding author.
E-mail addresses: tristan.roger@icn-artem.com (T. Roger), proger@unistra.fr (P. Roger), marc.willinger@umontpellier.fr (M. Willinger).

https://doi.org/10.1016/j.jedc.2021.104293
0165-1889/

A popular approach for measuring the acuity of the ANS is the Number Line Estimation (NLE) task (Booth and Siegler, 2006; Siegler and Booth, 2004; Siegler and Opfer, 2003), sometimes called Symbolic-number Mapping task (SMAP). It involves a mapping between numerical and spatial representations of numbers. The task consists of locating a given number on an empty bounded number line with 0 at the left end and 100 (or 1000) at the right end. Because of its simplicity, the NLE task is suited to everyone (the literature in neuropsychology shows that even young children are able to perform the NLE task) and requires only a few minutes to be completed.

Recent empirical research on the ANS acuity (and more specifically on the NLE task) provides growing evidence of the link between ANS acuity and mathematical competence (Chen and Li, 2014; Fazio et al., 2014; Halberda et al., 2008; Schneider et al., 2017; 2018).[1] ANS acuity also has implications for economic valuation and decision making (Peters and Bjalkebring, 2015; Peters et al., 2008; Schley and Peters, 2014; Sobkow et al., 2020). For instance, Schley and Peters (2014) show that, when it comes to estimating the economic value of a given good, incorrect valuations mainly come from an inexact mapping of symbolic numbers onto mental magnitudes, that is, a lack of acuity of the ANS. In four decision-making problems, Sobkow et al. (2020) show that the ANS acuity is the strongest predictor of optimal choices, beyond cognitive reflection and fluid intelligence.

Today, individuals face more complicated financial choices and have greater exposure to financial markets. It is therefore paramount to understand the drivers of financial decisions. Since differences in ANS acuity are linked to mathematical abilities and decision-making, it is likely that approximate numeracy also matters for financial decisions. In this paper, we test whether approximate numeracy is a predictor of financial performance in a continuous double auction experimental market. Experimental markets are an ideal setting to test the influence of approximate numeracy on subjects' earnings. Indeed, experimental markets are usually organized as a sequence of short trading periods lasting 2 or 3 minutes. Trading and managing a portfolio in such a setting requires to take quick buy/sell decisions (to adjust one's orders for changes in fundamental value (FV hereafter) or to take advantage of trading opportunities arising from others' mistakes). Taking decisions under time pressure involves the use of heuristics. We therefore expect individuals exhibiting acute approximate numeracy to perform better. We also analyze whether approximate numeracy has an influence on mispricing. We expect that the presence of subjects with high ANS acuity contributes to reducing the deviations between prices and FV.

The design of our experimental asset market is similar to the one of Smith et al. (1988). Subjects engage in two consecutive double auction markets. In contrast to the design of Smith et al. (1988), which involves dividend payments after each trading period resulting in a decreasing FV, our subjects trade a risky asset with a randomly fluctuating FV, similar to the design 4 of Stöckl et al. (2015). This kind of FV process shares more features with real financial markets since the FV is not deterministic. Specifically, at the end of each period, a cash flow is drawn from a given five-point uniform distribution. No intermediate dividends are paid, however. Instead, the randomly drawn cash flows accumulate. They are paid at the end of the market to owners of the risky asset. In addition to experimental market data, we also collected information on subjects' characteristics. Subjects completed an exit questionnaire, similar to the one proposed by Palan (2015), where they provided demographic information and answered questions on risk aversion.

In a first analysis, we divide the population of subjects (24 sessions with 9 subjects per session) in quartiles with respect to their scores on the NLE task. We find that subjects' earnings in the bottom quartile of NLE accuracy are 16.01% below average while earnings of subjects in the top quartile of NLE accuracy are 11.44% above average. This difference is significant at the 1% level. We then perform a multivariate analysis where we control for differences in trading activity and demographic characteristics (trading volume, order initiation decisions, background education, age, gender, risk aversion, previous experience in experimental markets). Our multivariate analysis confirms that ANS acuity, measured by the accuracy on the NLE task, successfully predicts subjects' earnings.

In a second analysis, we provide additional insights on some of the channels through which high NLE accuracy traders obtain higher earnings. We show that these subjects issue more profitable limit orders than their low NLE accuracy counterparts and execute market orders under better conditions. Our results indicate that, overall, subjects pay a high price for immediacy. On average, market orders take place 4.23% above FV for purchases and 10.96% below FV for sales. The cost of immediacy is particularly important for low NLE accuracy traders since they are ready to pay 6.79% above the FV to acquire the asset and are willing to sell 15.69% below the FV (compared to respectively −0.95% and 1.95% for high NLE accuracy subjects). In addition, we also find that high NLE accuracy traders are faster and better at exploiting price inefficiencies.

A third analysis links mispricing to NLE accuracy. Overall, we find that sessions in which traders score high on the NLE task exhibit lower mispricing. We show that only a few traders with high NLE accuracy are sufficient to reduce mispricing. This result is in line with high NLE accuracy traders being more efficient in dealing with limit orders that significantly deviate from FV.

Our paper makes contributions to several strands of literature. We contribute to the literature on experimental markets by uncovering a new predictor of earnings in market experiments. Previous research on what makes a "good" trader (Corgnet et al., 2018) is based on tests like the Cognitive Reflection Test (CRT) (Frederick, 2005) or Raven Progressive Matrices (RPM). However, recent papers (Skagerlund et al., 2021; Sobkow et al., 2020) show that numerical competencies, and especially approximate numeracy, are strong drivers of decision making. For instance, Sobkow et al. (2020) consider five

[1] See Szkudlarek and Brannon (2017) for a detailed review. The NLE task is widely used because the accuracy achieved during this task correlates with counting abilities (Östergren and Träff, 2013), arithmetic skills (Torbeyns et al., 2015) and standardized test scores (Ashcraft and Moore, 2012).

variables; three measures of numeracy (statistical, subjective and approximate numeracy), one measure of fluid intelligence, and one measure of cognitive reflection (i.e., the score on the standard 7-item CRT). They perform a horse race between models to explain performances in four types of problems: a lottery task where people select among pairs of lotteries, a decision outcome inventory test as in Bruine de Bruin et al. (2007), and two memory tests. The score on the NLE task is the only variable that is significant in predicting decision outcomes in all four types of problems. Moreover, the score on the NLE task is the most significant variable in the four regressions including the five explanatory variables (CRT, fluid intelligence, subjective, statistical, approximate numeracy).[2]

We also contribute to the literature that links mispricing to cognitive abilities. As mentioned above, the distribution of NLE accuracy within a session has an influence on the level of mispricing in this session, even after controlling for average risk aversion and previous experience in economic experiments. Our results resonate with previous experimental studies that demonstrate a link between mispricing and cognitive abilities. Breaban and Noussair (2015) show that deviations from FV correlate negatively with session-average CRT scores. Bosch-Rosa et al. (2018), using three different indicators (CRT, Guessing Game and Race to 60), find lower mispricing in sessions composed of subjects with high cognitive abilities.

2. Experimental design

2.1. Experimental implementation

The experiment was conducted at the LEEM, the computerized laboratory of the University of Montpellier, with the software z-Tree (Fischbacher, 2007). We ran 24 sessions involving a total of 216 subjects (9 subjects in each session), randomly selected from a pool of 5000 volunteers, most of them being students or employees of the University of Montpellier. No subject took part in more than one session. Detailed demographic statistics of the sample are reported in Table A2 in the Appendix.

At the beginning of each session, subjects received written instructions and completed a comprehension questionnaire. The instructions (presented in section E of the Appendix) were read aloud. Subjects were informed that they would then take part in two consecutive markets. They did not receive specific information related to the second market until the end of the first market.

Each session is divided into two distinct markets of 10 periods each. The characteristics of the two consecutive markets are the same with the exception of the cash-flow sequences (described below). At the beginning of each market, subjects are given units of a risky asset and experimental currency. The currency used in the experiment is called ECU. Participants can buy and/or sell as many units of the risky asset as they want, provided they hold enough units of experimental currency to buy or enough units of assets to sell. Short-selling and borrowing are not allowed. Within a given market, holdings of experimental currency and assets are carried over from one period to the next. The order book is visible to all traders at any time.

In each market, subjects trade a single risky asset. The risky asset does not pay intermediate dividends. However, at the end of each period, a cash-flow is drawn from a uniform five-point distribution {0, 0.3, 0.6, 0.9, 1.2}, and publicly announced. At the end of the market, the redemption value is equal to the sum of the ten cash flows. While the cash-flows that are drawn do not have an impact on the cash held by subjects, their revelation induces a change in FV (and thus a change in the traded price of the asset). Since the expected value and variance of each cash-flow are respectively 0.6 and 0.18, $FV = \sum_{t=1}^{10} CF_t$, satisfies $E(FV) = 6$ and $\sigma^2_{FV} = 1.8$. Moreover, the FV fluctuates in the interval [0;12]. Figure 1 illustrates the distribution of redemption values faced by subjects at the start of the market. It is built with 5000 random sequences of ten cash-flows. Adding the terms of each sequence gives a realization of the redemption value. In Figure 1, we have superimposed the histogram of redemption values on the histogram of a gaussian variable with the same expectation and variance. The distribution of the redemption values is close to the gaussian distribution..

As time passes, the FV then fluctuates with respect to the cash-flows already drawn. For instance, if the first cash-flow is equal to $cf_1 = 1.2$, the expected fundamental value increases by 0.6 following the cash-flow revelation (since the expected cash flow is equal to 0.6). More generally, during any period $t \leq 10$, the fundamental value of the asset remains random, equal to

$$FV_t = \sum_{s=1}^{t-1} cf_s + \sum_{u=t}^{T} CF_u \qquad (1)$$

where CF is a uniform random variable with a five-point distribution {0, 0.3, 0.6, 0.9, 1.2} and cf represent past cash-flows (i.e. past realizations of the CF variable). It follows that the redemption value is known with certainty only at the end of period $t = 10$. At this point in time, units of the risky asset are converted into cash with a redemption value equal to the sum of the 10 revealed cash-flows.

Sequences of cash-flows are randomly generated but determined in advance. We consider four different sequences of cash-flows denoted S1 to S4 (see Panel B of Table 1). The sequence S1, in the first market, followed by the sequence S3 in

[2] See Table 2, page 9 in Sobkow et al. (2020). In addition, they find a correlation of 0.34, over more than 500 subjects, between the NLE score and a 7-item CRT score. Although our purpose is not a comparison between the CRT test and the NLE task, we also collect the two scores for 48 subjects in an additional experiment. We obtain as well a correlation of 0.34 (see section C of the Appendix).

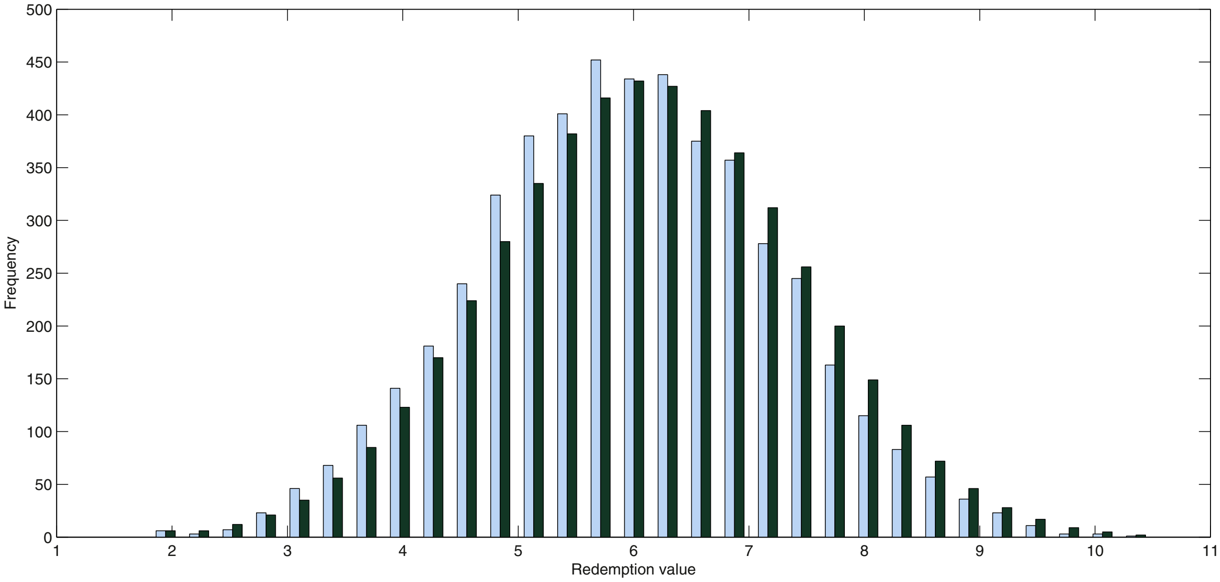

Fig. 1. Distribution of redemption value vs. normal distribution. This figure shows a sample of 5000 redemption values (light blue bars), compared to what would be expected from a normal distribution (dark blue bars).

Table 1
Experimental design.

Panel A: Sessions

Session	FV magnitude	Stakes	CF sequences	Units of assets			Units of cash			Exchange rate
				P1	P2	P3	P1	P2	P3	(1 Ecu in Euros)
1–3	Large	Low	S1 - S3	3	6	9	984	768	552	0.0167
4–6	Large	Low	S2 - S4	3	6	9	984	768	552	0.0167
7–9	Large	High	S1 - S3	3	6	9	984	768	552	0.0333
10–12	Large	High	S2 - S4	3	6	9	984	768	552	0.0333
13–15	Small	Low	S1 - S3	3	6	9	82	64	46	0.2
16–18	Small	Low	S2 - S4	3	6	9	82	64	46	0.2
19–21	Small	High	S1 - S3	3	6	9	82	64	46	0.4
21–24	Small	High	S2 - S4	3	6	9	82	64	46	0.4

Panel B: Time series of cash-flows

Periods	1	2	3	4	5	6	7	8	9	10
Basic sequence 1 (S1)	0.6	0	0.3	0.9	0.6	1.2	0.9	0.3	0.3	0.9
Basic sequence 2 (S2)	0.6	1.2	0.9	0.3	0.6	0	0.3	0.9	0.9	0.3
Mirrored sequence 1 (S3)	0.9	0.6	0	0.6	0.6	1.2	0.6	0	0.9	0.6
Mirrored sequence 2 (S4)	0.3	0.6	1.2	0.6	0.6	0	0.6	1.2	0.3	0.6

This table presents the main characteristics of the market design. Panel A shows the structure of sessions according to the different treatments in terms of FV magnitude, size of stakes, cash-flow sequences, and portfolio composition (assets plus cash). For example, 12 sessions are performed with a large FV, among which 6 correspond to low stakes among which 3 use the cash-flow sequences S1 and S3. Panel B shows the four sequences of cash-flows used in the Small FV treatment. In High FV treatments, the sequences are multiplied by 12.

the second market were used in half of the sessions. The other sessions used sequence S2 for the first market and sequence S4 for the second market. Sequences S2 and S4 are “mirrored” versions of S1 and S3 (with respect to the unconditional fundamental value). Using mirrored sequences allows us to control for the FV trend, which is known to influence mispricing (Stöckl et al., 2015). Gillette et al. (1999) and Kirchler (2009) show that markets with predominantly decreasing (increasing) FV tend to exhibit overvaluation (undervaluation).

We also introduce alternative treatments to assess the robustness of our results to changes in FV magnitude (measured in ECUs) and/or earnings magnitude (measured in Euros).[3] First, in half the sessions, we multiply random cash-flows by 12. The

[3] See Camerer and Hogarth (1999) for a discussion of financial incentives in economic experiments. See Kocher et al. (2017) for the importance of stake size in experimental asset markets.

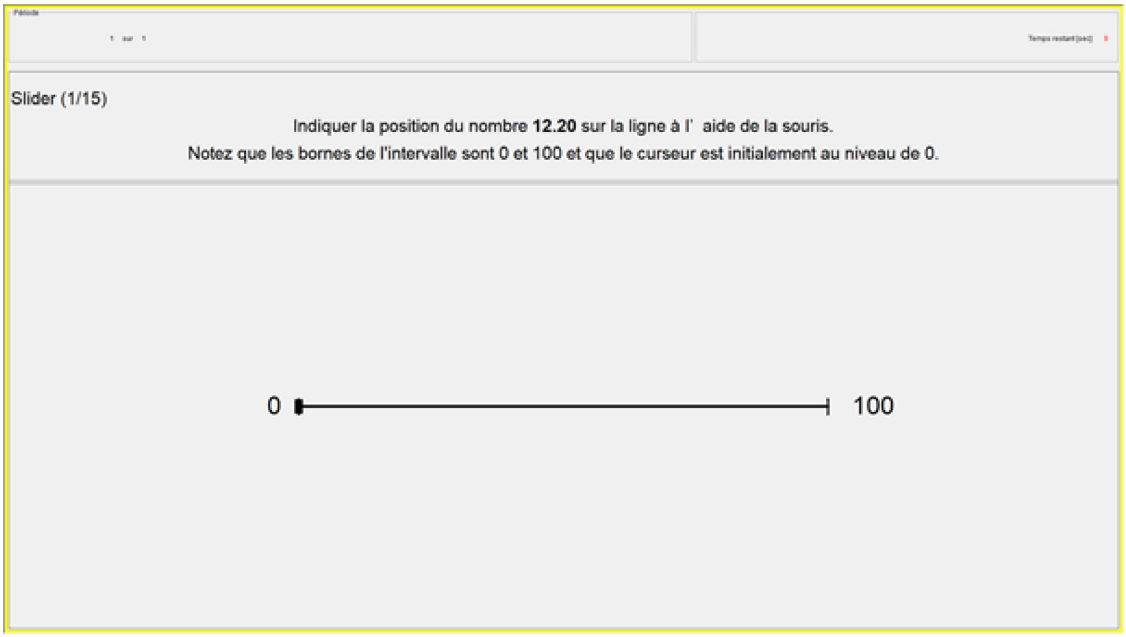

Fig. 2. Ilustration of the NLE task. Figure 2 provides an illustration of the NLE task for the first trial. The target number is 12.20 and the slider is initially positioned at 0. The translation of the text in the figure is: "Indicate the position of number 12.20 on the line using the mouse. Note that the boundaries of the interval are 0 and 100 and that the slider is initially at 0.".

uniform distribution of cash-flows becomes {0, 3.6, 7.2, 10.8, 14.4}. As a result, the unconditional expected FV is 72 at the beginning of the market. Cash endowments are adjusted accordingly to keep the cash/asset ratio constant (Caginalp et al., 1998; 2001) show that this ratio can influence market prices). Second, we introduce two exchange rates ECUs/Euros. Half the sessions involve a low exchange rate and the other half a high exchange rate. On average, subjects earned 20 (40) euros in the low (high) exchange rate treatment. The experimental design is provided in Table 1. In summary, we implement a 2×2 factorial design: two exchange rates (low and high) combined with two FV magnitudes (low and high). We run six sessions in each cell. Since FV magnitude and exchange rate had no influence on our results, we pooled all the sessions together and scaled our measure of earnings accordingly. Detailed results, which control for these two treatments are available in section D.1 of the Appendix.

2.2. The NLE task

The design of the NLE task subjects have to complete is adapted from Siegler and Opfer (2003), Siegler and Booth (2004) and Booth and Siegler (2006). Participants complete 15 trials. Figure 2 provides an illustration for a given trial. In each trial, subjects are facing an empty number line on the screen. The left end is marked with number 0 and the right end is marked with number 100. A real number between 0 and 100 is then shown on the subject's screen for 10 seconds. Participants are asked to locate this number on the empty number line using a slider initially positioned at the left end (i.e., 0). In half of the 24 sessions, subjects are instructed that their accuracy is rewarded at the end of the 15 trials. Specifically, one trial is drawn at random and yields a reward (in €) calculated as:

$$EARNINGS_{NLE} = Max(3 - 0.05 \times |Estimate - Value|; 0) \tag{2}$$

where *Estimate* corresponds to the position of the slider on the number line and *Value* is the target number. In the remaining 12 sessions, the NLE task is not incentivized. A Mann-Whitney test (p-value=0.2241) indicates that NLE accuracy is not significantly different in the two treatments (see Table A3 in the Appendix for more details).

3. Measures and control variables

This section presents the different variables used in our analysis. A summary can be found in Table A1 in the Appendix.

3.1. Dependent variables

Trading performance measure

In a given session, we measure subjects' trading performance as the average of their trading performance over the two consecutive markets. For a given market, the performance of subject i is the ratio of her final wealth to the average final wealth of the nine subjects involved in the same market. Formally, for a given market $k = 1, 2$, the trading performance of subject i writes:

$$PERF_i^k = \frac{Wealth_i^k}{\overline{Wealth^k}} - 1 \tag{3}$$

where $Wealth_i^k$ is the wealth in ECUs of subject i for market k and $\overline{Wealth^k}$ is the average wealth in ECUs of the 9 subjects participating in market k.

$PERF_i^k$ therefore measures the relative earnings of subject i in market k. This measure is relevant for our purpose because the aggregate profit in a given market is 0. Moreover, this relative measure allows us to compare the performance across sessions with different exchange rates. The overall performance of a given subject is obtained by taking the arithmetic average of the performance achieved over the two consecutive markets.

Mispricing at the session level

Following Stöckl et al. (2010), we quantify mispricing by the absolute price deviation with respect to the fundamental value of the risky asset. The Relative Absolute Deviation (RAD) is defined as

$$RAD = \frac{1}{N} \sum_{t=1}^{N} \frac{\left|\overline{P_t} - FV_t\right|}{\overline{FV}} \tag{4}$$

where t denotes the period number in a given market and N is the total number of periods ($N = 10$ in our experiment). $\overline{P_t}$ is the period-t average transaction price and FV_t is the beginning of period-t fundamental value. $\overline{FV}$ is the average fundamental value over the N periods.

3.2. Independent variables

Accuracy on the NLE task

Following Siegler and Booth (2004) and Booth and Siegler (2006), we base our measure of accuracy on the NLE task on the Percent Absolute Error (PAE). For each trial, we calculate the Percent Absolute Error (PAE) as:

$$PAE_{i,j} = \left|\frac{Estimate_{i,j} - Value_j}{100}\right| \tag{5}$$

where $Value_j$ is the number that subjects are asked to locate on the number line for trial j. $Estimate_{i,j}$ is the number selected with the slider by subject i. The denominator corresponds to the scale of estimates. It is equal to 100 since our number line is delimited by 0 on the left and 100 on the right. Our measure of NLE accuracy writes

$$NLEacc_{i,j} = 1 - PAE_{i,j}. \tag{6}$$

For a given session and a given subject i, we set the overall NLE accuracy $NLEacc_i$ equal to the average $NLEacc$ for trials 2 to 15.[4]

Control variables

Our analysis also takes into account demographic characteristics. Our information on subjects comes from a questionnaire they answered to at the end of the experiment (see Palan (2015) for a prototype of the exit questionnaire). The complete questionnaire and descriptive statistics of the answers can be found in section B.1 in the Appendix. In short, subjects were asked the following information:

1) age and gender;
2) attitude toward risk on several dimensions (Dohmen et al., 2011): (a) in general, (b) while driving, (c) when making financial investments, (d) when practicing sport, (e) in the workplace, (f) regarding healthcare, and, (g) when trusting others.
3) whether they are currently a student, what is their curriculum major, which degree they are studying for and whether they ever attended a Finance course.
4) whether they had already participated in economic experiments.

Since the seven variables that account for attitude toward risk are not independent (see Table A6 in the Appendix), we summarize risk attitudes by the first principal component of a Principal Component Analysis (PCA) including these seven dimensions of risk. We refer to this variable as *Risk attributes*. A high value for *Risk attributes* corresponds to low risk aversion. The variance of this first component represents about 35% of the trace of the covariance matrix.[5]

[4] A large proportion of subjects (42.76%) did not manage to complete the first trial on time (see Table A3 in section B.3 of the Appendix for trial-by-trial answer rates). As a consequence, we decided to exclude the first trial when computing *NLEacc*. However, the results based on all 15 trials are unchanged and can be found in the Appendix (see Tables A9 to A12 in section D.2).

[5] In a robustness test, we include the second principal component (Table A15 in the Appendix) in the main regression. We also consider instead the seven risk dimensions directly (Table A16 in the Appendix). Results are unchanged when these alternative approaches are considered.

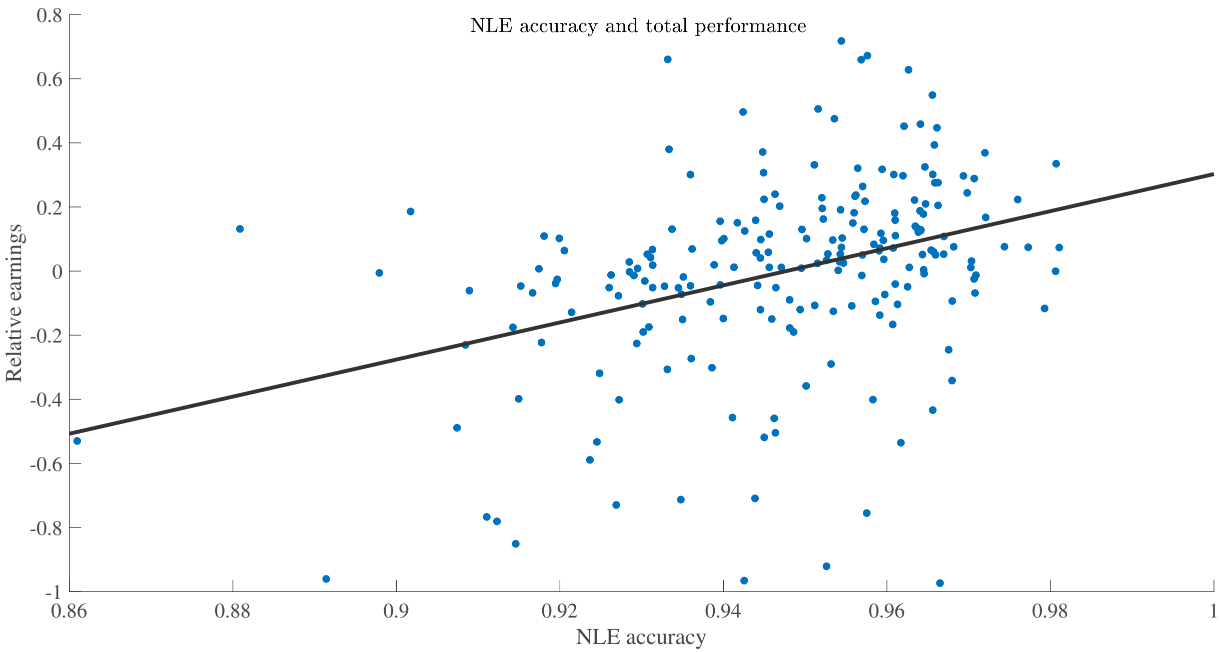

Fig. 3. NLE accuracy and financial performance. The figure shows how NLE accuracy (horizontal axis) and relative earnings (vertical axis) are distributed in a two-dimensional space. The least-squares line (in black) shows the positive relationship between the two variables.

In addition to demographic characteristics, we also consider trading dynamics. The literature on individual investors initiated by Odean (1998) shows that the trading performance of retail investors partially depends on their trading dynamics (Barber and Odean, 2000) through explicit and implicit trading costs. To account for a potential link between trading dynamics and performance, we introduce two control variables. First, we use the volume of trades, measured by the logarithm of the number of trades. Second, we consider the proportion of trades that are limit orders. In a continuous double-auction, subjects can either trade against outstanding orders (i.e. market orders) or enter new orders in the order book (i.e. limit orders). These two types of orders imply different implicit transaction costs (bid-ask spreads) since the former represents a demand for immediacy while the latter is a liquidity supply. These two variables, *Log number of trades* and *Proportion of limit orders* incorporate different dimensions of trading dynamics, as indicated by the low correlation ($\rho = 0.0094$) between them.

4. Results

4.1. NLE accuracy and earnings

Figure 3 provides a scatter plot of subjects' average relative earnings versus NLE accuracy. The regression line appears in black. A quick examination of the graph reveals a positive relationship between NLE accuracy and relative earnings (i.e., the regression line has a positive slope). Overall, NLE accuracy ranges mainly between 0.91 and 0.98, as shown on the histogram of NLE accuracy reported in Figure A3 of the Appendix. Few subjects seem to perform particularly poorly on the NLE task and could potentially be qualified as outliers. However, the relationship between NLE accuracy and performance is unchanged if we discard these observations. The correlation is equal to 0.3676 (and highly significant) when we consider the whole sample (see Table A4 in the Appendix). It is equal to 0.3696 (respectively 0.3590) when we exclude the two (five) subjects with the worst performance on the NLE task.

Table 2 compares the average relative earnings of subjects per quartile of NLE accuracy. The first column averages financial performance over the two markets, columns 2 and 3 provide separate results for markets 1 and 2. The performance of a subject in the top quartile of NLE accuracy (*NLEacc*) is 11.44% above average while the performance of a subject in the bottom quartile is 16.01% below average. A Wilcoxon-Mann-Whitney test indicates that the difference of performance between the top and bottom quartiles of *NLEacc* (0.2745) is significantly different from 0 at the 1% level ($z = 5.3708$). Results are qualitatively unchanged when we consider first markets and second markets separately. The difference between the top and the bottom quartile remains highly significant with 0.3387 for first markets ($z = 4.9570$) and 0.2104 for second markets ($z = 4.6853$). Similar results using the median instead of the mean are presented in section D.3 of the Appendix.

To test the link between NLE accuracy and trading performance in a multidimensional setting, we run the following regression:

$$PERF_i = \alpha + \beta NLEacc_i + \gamma \mathrm{CONTROLS}_i + \epsilon_i \tag{7}$$

Table 2
Earnings and quartiles of NLE accuracy.

	All markets	First markets	Second markets
Bottom quartile (Smallest NLEacc)	−0.1601	−0.1877	−0.1326
Quartile 2	−0.0297	−0.0662	0.0067
Quartile 3	0.0698	0.0875	0.0522
Top quartile (Highest NLEacc)	0.1144	0.1510	0.0778
Difference Top-Bottom	0.2745***	0.3387***	0.2104***
WMW (z value)	5.3708	4.9570	4.6853

This table shows the average (relative) earnings in subsets defined by quartiles of NLE accuracy. The first column averages the individual performance across markets, columns 2 and 3 provide separately the same statistics for markets 1 and 2. The last row gives the Wilcoxon-Mann-Whitney (WMW) statistic for the difference of performance between top and bottom quartiles. ***, **, and * indicate statistical significance at the 1%, 5%, and 10%, respectively.

Table 3
NLE accuracy and earnings: a multivariate analysis.

	Overall earnings		Earnings in first market		Earnings in second market	
NLE accuracy	5.7865***	3.7644***	7.2697***	5.4964***	4.3034***	2.7017**
	(1.0102)	(1.2287)	(1.2802)	(1.6930)	(1.0097)	(1.0807)
Proportion of limit orders		0.0040***		0.0051***		0.0027***
		(0.0008)		(0.0011)		(0.0007)
Log number of trades		−0.1244***		−0.1196***		−0.0875***
		(0.0186)		(0.0263)		(0.0207)
Log age		−0.4520***		−0.5002***		−0.4152***
		(0.1352)		(0.1533)		(0.1327)
Risk attributes		−0.0122		−0.0240		−0.0074
		(0.0129)		(0.0168)		(0.0130)
Male dummy		0.0641*		0.0924**		0.0480
		(0.0342)		(0.0436)		(0.0359)
Previous participation dummy		0.0636**		0.0495		0.0702*
		(0.0267)		(0.0295)		(0.0361)
Finance educated dummy		−0.0172		0.0107		−0.0533*
		(0.0284)		(0.0481)		(0.0280)
Intercept	−5.4839***	−1.9587	−6.8915***	−3.6246*	−4.0762***	−1.1927
	(0.9573)	(1.3518)	(1.2132)	(1.8044)	(0.9569)	(1.1242)
N	212	212	212	212	212	212
R^2	0.1351	0.3952	0.1331	0.3523	0.0796	0.2885

This table presents the results of the regression of (relative) earnings on NLE accuracy (*NLEacc*), controlling for individual characteristics (number of trades, proportion of limit orders, previous participation in economic experiments, risk attributes, age, gender, background in finance). In the two first columns, we average relative earnings over the two successive markets. Columns 3–4 and 5–6 provide separate results for first markets and second markets. Robust standard errors (reported in parentheses) are clustered at the session level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10%, respectively.

where $PERF_i$ is the average performance over the two consecutive markets of subject i and $NLEacc_i$ is the accuracy of subject i on the NLE task averaged over trials 2 to 15. Finally, $\mathrm{CONTROLS}_i$ is the vector of control variables for subject i defined in Section 3.2.[6] Since the performance of a given subject depends on the performance of other subjects in the same session, we cluster standard errors by session.

The results, in Table 3, show that NLE accuracy has a statistically significant impact on subjects' trading performance.[7] Traders who exhibit greater (smaller) NLE accuracy tend to outperform (underperform) their peers. In the first column, we present the results of the regression without control variables. The NLE accuracy coefficient is equal to 5.7865 and is significant at the 1% level. The R-squared is equal to 0.1351. When controls are added (in column 2), the coefficient takes the value 3.7644. Actively trading has a negative and significant influence on performance as shown by the negative coefficient of the log number of trades. This finding is in line with Odean (1998) and Barber and Odean (2000) who show that individual investors damage their performance by trading. We also find that the proportion of limit orders influences positively trading performance, consistent with the idea that buying and selling through market orders bear immediacy/liquidity costs. The other significant control variables are the logarithm of subject's age and whether the subject had previous experience in experimental markets. Results are similar when focusing on the performance on first and second markets separately. Acquiring experience within the experiment seems to slightly mitigate the link between NLE accuracy and performance. A similar trend is found for the main control variables: the magnitude of the coefficients decreases in second markets, though they

[6] Descriptive statistics on control variables and univariate comparisons can be found in Tables A4 and A5 in the Appendix

[7] Two subjects did not complete the NLE task (in sessions 16 and 17). In addition, two subjects deliberately failed the task and were thus removed from the analysis (in sessions 6 and 18). This yields a total of 212 observations.

Table 4
Order dynamics.

Panel A: All orders

	Executed limit orders				Market orders				
	Number of orders	Number of traders	P/FV on buy orders	P/FV on sell orders	Number of orders	Number of traders	P/FV on buy orders	P/FV on sell orders	Time to hit orders
All traders	6953	211	0.8904	1.0423	6875	209	1.0445	0.8884	9.1455
Top NLE accuracy	2393	71	0.8681	1.0927	1804	71	0.9905	0.9805	7.3678
Average NLE accuracy	2490	72	0.8655	1.0367	1997	70	1.0671	0.8971	9.2997
Bottom NLE accuracy	2070	68	0.9513	0.9969	3074	68	1.0679	0.8431	10.0886
WMW z-value			−3.9886***	8.6510***			−8.0280***	8.9330***	−11.0033***

Panel B: Unfavorable limit orders

	Executed limit orders				Market orders				
	Number of orders	Number of traders	P/FV on buy orders	P/FV on sell orders	Number of orders	Number of traders	P/FV on buy orders	P/FV on sell orders	Time to hit orders
All traders	2523	193	1.2244	0.8216	2522	199	0.8233	1.2173	6.3573
Top NLE accuracy	722	64	1.1543	0.8517	890	65	0.7965	1.2764	5.3666
Average NLE accuracy	896	63	1.2236	0.8333	776	68	0.8221	1.2165	6.4082
Bottom NLE accuracy	905	66	1.2951	0.7882	856	66	0.8537	1.1627	7.3412
WMW z-value			−4.2623***	4.6488***			−3.9575***	2.7391***	−5.8885***

Panel C: Extreme limit orders

	Executed limit orders				Market orders				
	Number of orders	Number of traders	P/FV on buy orders	P/FV on sell orders	Number of orders	Number of traders	P/FV on buy orders	P/FV on sell orders	Time to hit orders
All traders	276	84	2.2139	0.5358	271	101	0.5410	2.1266	6.2345
Top NLE accuracy	57	18	1.7415	0.5844	103	37	0.5022	2.5640	4.4689
Average NLE accuracy	105	29	2.1755	0.5862	87	35	0.5071	1.9628	6.9918
Bottom NLE accuracy	114	37	2.5426	0.4647	81	29	0.6285	1.8084	7.6664
WMW z-value			−1.9192*	1.9220*			−2.8700***	0.5864	−2.0132**

This table reports statistics on the number of executed orders, the average ratios of buying and selling prices divided by the fundamental value of the risky asset (denoted P/FV) and the time-to-execution. Panel A aggregates all orders. Panel B contains the subset of unfavorable orders, defined by a ratio P/FV higher (lower) than 1 for buy (sell) orders. Panel C concerns Extreme orders, that is buy orders for which the price is above the maximum redemption value or sell orders for which the price is below the minimum redemption value. Each Panel contains five rows. The first row concerns all orders in the Panel. Rows 2 to 4 refer to orders posted or executed by different subsets of subjects (defined with respect to NLE accuracy). The last row gives the Wilcoxon-Mann-Whitney (WMW) statistic for the difference between top and bottom NLE accuracy categories for different variables (P/FV or Time-to-execution). The left part (Columns 2 to 5) of the table provides statistics on limit orders entered in the order book (orders posted). The right part gives the same information for market orders (orders executed against limit orders). The last column (Time-to-execution) gives the average duration of orders in the book before they are executed. ***, **, and * indicate statistical significance at the 1%, 5%, and 10%, respectively.

remain statistically significant. Overall, we find evidence that NLE accuracy correlates with performance in experimental markets. Our results hold when additional controls, including initial endowments, and additional information on subjects, are added to the regression (See Table A8 in the Appendix).

4.2. NLE accuracy and order dynamics

To gain a better understanding on the influence of NLE accuracy on earnings magnitude, we focus on order dynamics. Our objective is to understand how the trading behavior of high NLE accuracy traders allows them to capture higher earnings, compared to low NLE accuracy traders. To do so, we analyze both order types: those entered in the order book (limit orders) and those executed against orders already standing in the book (market orders). The results of our analysis are presented in Table 4. Table 4 contains three panels, which relate to different sets of orders. Panel A contains all the orders in the order book. Panel B includes only unfavorable orders. Unfavorable orders must be understood from the perspective of the subject that submits the order. Buy (sell) orders are classified as unfavourable if they are issued at a price above (below) the FV. Finally, Panel C contains only extreme orders, that is, buy (sell) orders issued at a price greater (lower) than the maximum (minimum) final FV.[8] Extreme orders are not compatible with the absence of arbitrage opportunities since executing against such orders provides a profit at no risk.

[8] For example, let us assume that the two first cash-flows are 0.3 and 0.6. There are eight cash-flows not already drawn, each with a maximum value of 1.2. Thus, the maximum final FV (i.e., the redemption value) is $8 \times 1.2 + 0.6 + 0.3 = 10.5$. A buying order at a price of 11 during the third period of the market would be considered an extreme order.

Each panel is divided into two parts. Columns 2 to 5 provide statistics on limit orders that were eventually executed. Columns 6 to 10 refer to market orders (orders that execute against standing limit orders). The four first rows in each panel correspond to different sets of subjects. The first row aggregates all subjects. The second row (i.e., Top NLE accuracy) corresponds to the three subjects, in each session, that exhibit the highest NLE accuracy (T1 subjects, hereafter). The third row (i.e., Average NLE accuracy) provides information about subjects ranked 4th to 6th in NLE accuracy. Finally, the fourth row (i.e., Bottom NLE accuracy) corresponds to subjects with the lowest NLE accuracy (T3 subjects, hereafter). The last row gives the Wilcoxon-Mann-Whitney statistic for the difference of performance between top and bottom NLE accuracy categories.

For each type of order and each set of subjects, we provide the following statistics: 1) the total number of orders, 2) the number of distinct traders who issue these orders, 3) the ratio price to fundamental value (P/FV) for buy orders and, 4) the same ratio for sell orders. In addition, for market orders, we also indicate the average time it takes for market orders to execute against a standing limit order (i.e., or equivalently, the average duration of limit orders that are eventually executed).

Overall, the results in Panel A indicate that traders pay a high price for immediacy and get rewarded for providing liquidity. The average price of executed limit buy orders amounts to 89.04% of FV. This figure indicates, in turn, that subjects are willing to sell at a steep discount when issuing market orders. A similar pattern is found for sell limit orders, which take place at 4.23% above FV. The first driver of trading performance of high NLE accuracy subjects directly relates to this balance between liquidity demand and supply. T1 subjects post more limit orders in the book (2393) than T3 subjects do (2070). On the contrary, T1 subjects' demand for liquidity is lower (1804 market orders by T1 subjects compared to 3074 by T3 subjects). These differences in liquidity supply and demand are also revealed by the relative prices at which T1 and T3 subjects trade. T1 subjects post buy (sell) limit orders at lower (higher) price than T3 subjects. Indeed, while buy (sell) limit orders are approximately 13% (9%) below (above) the FV for T1 traders, T3 subjects post orders close to the FV (-5% for buy orders and about 0% for sell orders). We find similar results when looking at market orders. T1 subjects execute against standing orders that are, on average, at prices close to the FV. In contrast, T3 subjects' buy orders exceed FV by about 7%. Their sell orders are typically completed below FV (approximately -16%). These differences between T1 and T3 subjects are highly significant, both for executed limit orders and market orders. The z-values of the Wilcoxon-Mann-Whitney tests are -3.9886 (-8.0280) for buy limit (market) orders and 8.6510 (8.9330) for sell limit (market) orders. The comparative advantage of high NLE accuracy traders comes both from their role as liquidity providers and from being faster at executing against standing orders. Indeed, the last column of Table 4 shows that, on average, T1 subjects execute their market orders against standing limit orders 7.37 seconds after these limit orders were entered in the book. For T3 subjects, the average time-to-execution is 10.09 seconds. The WMW z-value for the difference is -11.00.

Panels B and C provide further insights on the better order dynamics of high NLE accuracy traders (who post limit orders at better prices and are quicker at executing market orders against unfavorable limit orders). In Panel B, we find that T1 subjects are less likely to post unfavorable limit orders. There are 722 unfavorable limit orders posted by T1 subjects compared to 905 orders posted by T3 subjects. In addition, when posting unfavorable limit orders, T1 subjects' orders take place at prices closer to the FV than T3 subjects. Considering unfavorable limit orders, T1 subjects buy about 15% above the FV ($P/FV = 1.1543$) and sell about 15% below ($P/FV = 0.8517$). The corresponding figures for T3 subjects are 29.51% above the FV for purchases and 21.18% below for sales. These differences are highly significant (the WMW z-values are above 4 in absolute value). These figures resonate with the findings of Linnainmaa (2010) who shows that retail investors' losses can mainly be explained by a sub-optimal use of limit orders. In a market experiment, Corgnet et al. (2018) find that a trader's proportion of favorable limit orders relates positively and significantly to trader earnings. The analysis of market orders (that execute against unfavorable limit orders) provides additional elements on trading behavior. We find that T1 subjects exploit the most unfavorable limit orders by buying 20.35% below FV and selling 27.64% above FV. The corresponding figures for T3 subjects are 14.63% below the FV for purchases and 16.27% for sales. The differences between the two sets of subjects are significant at the 1% level on both sides (purchases and sales). Moreover, the last column shows that T1 subjects are faster to exploit unfavorable orders: the time-to-execution is 5.37 seconds for T1 subjects versus 7.34 seconds for T3 subjects (WMW z-value = -5.8885). These results echo the findings of Grinblatt et al. (2012) who show that high-IQ investors exhibit superior market timing and trade execution.

Section B.2 in the Appendix (Figures A1 and A2) shows the time-series of prices and fundamental values for the 24 sessions. It appears on those graphs that some trades are executed at prices that largely deviate from the FV. As a consequence, Panel C of Table 4, focuses on extreme orders (limit orders that generate a sure loss at the end of the experimental market). We observe that extreme limit orders entered in the book by T1 subjects represent 2.38% (57/2393) of their limit orders while this figure is 5.51% for T3 subjects. There are also less orders posted by T1 subjects (and less traders posting such orders) compared to T3 subjects. Moreover, when posting extreme orders, T3 subjects are more extreme than T1 subjects. Regarding market orders, we find the same results as for unfavorable orders. First, T1 subjects are quicker at arbitraging extreme orders (4.47 seconds compared to 7.67 for T3 subjects). Second, they exploit the most profitable arbitrage opportunities by buying on average at about half the FV (compared to $P/FV = 0.6285$ for T3 subjects) and selling at 2.56 times the FV (1.81 for T3 subjects).

Table 5
Spearman correlation between RAD and measures of NLE accuracy.

	Correlation with RAD		
	All markets	First markets	Second markets
Maximum *NLEacc*	−0.4670**	−0.5113**	−0.3435
Average of two largest *NLEacc*	−0.5591***	−0.5800***	−0.4522**
Average of three largest *NLEacc*	−0.5096**	−0.5400***	−0.4026*
Average of four largest *NLEacc*	−0.5870***	−0.5904***	−0.4948**
Average of five largest *NLEacc*	−0.5104**	−0.4609**	−0.4774**
Average of six largest *NLEacc*	−0.4904**	−0.4348**	−0.4296**
Average of seven largest *NLEacc*	−0.4609**	−0.3991*	−0.3983*
Average of eight largest *NLEacc*	−0.3852*	−0.3557*	−0.3278
Average *NLEacc*	−0.3635*	−0.3426	−0.3548*

This table provides Spearman rank-correlation coefficients of the Relative Absolute Deviation (RAD) with given measures of NLE accuracy (*NLEacc*). Each correlation coefficient is computed with 24 observations (i.e., the number of sessions). We consider partial averages of NLE accuracy aggregated at the session level. The first measure is the maximum NLE accuracy in each session. The second measure averages the NLE accuracy of the two subjects with the largest NLE accuracy. The following measures are partial averages (in decreasing order of NLE accuracy) over 3, 4,... 9 subjects. ***, **, and * indicate statistical significance at the 1%, 5%, and 10%, respectively.

To sum up, Table 4 illustrates the channels through which high NLE accuracy subjects obtain superior earnings.[9] First, acting as liquidity providers, these subjects manage to enter limit orders in the order book at better conditions than low NLE accuracy subjects. Second, high NLE accuracy subjects are faster to react to potential opportunities.

4.3. NLE accuracy and mispricing

Previous research on market experiments show that market efficiency improves when rules and procedures are easily understood (Huber and Kirchler, 2012; Kirchler et al., 2012; Lei and Vesely, 2009). Cognitive abilities are thus likely to influence mispricing: subjects with higher cognitive abilities should have a better understanding of experiments. Breaban and Noussair (2015) and Bosch-Rosa et al. (2018) find that the magnitude of price deviations from FV is inversely proportional to the session average CRT score. Charness and Neugebauer (2019) show, in an experiment on the Modigliani-Miller theorem, that (cross-asset) mispricing is smaller in sessions populated by high CRT subjects. Additional studies evidence that the presence of subjects with high cognitive abilities decreases price volatility (Cueva and Rustichini, 2015) and improves price discovery (Noussair et al., 2016).

We now investigate whether the distribution of NLE accuracy across subjects in a session influences mispricing. Given our findings on order dynamics in the previous section, we expect that the presence of high NLE accuracy subjects in a session mitigates mispricing. However, given the distribution of initial portfolios across traders and, since short sales and borrowing are not allowed, a single subject with a high NLE accuracy would not be able, alone, to exploit all trading opportunities. In a given market, several high NLE accuracy subjects may be necessary to efficiently reduce mispricing. We therefore define several measures of session-level NLE accuracy. We consider partial averages of NLE accuracy scores ranked in a decreasing order. Our first measure is defined as the maximum NLE accuracy score in a given session. Our second measure is the average of the two largest NLE accuracy scores and so on.

Table 5 provides the Spearman rank-correlation coefficients of the RAD[10] with the partial averages of NLE accuracy. Overall, there is a significant negative correlation between session-level NLE accuracy and RAD. When taking the average NLE accuracy, the correlation is equal to −0.3635. However, taking the average NLE accuracy is likely not the best choice if one wants to investigate the relationship between Approximate Number System (ANS) acuity and mispricing. Indeed, we expect price discrepancies to be exploited mainly by subjects who exhibit higher NLE accuracy. Considering correlations between mispricing and the average NLE accuracy of the few most accurate subjects in each session, yields large and significant negative coefficients. The correlation between RAD and partial average NLE accuracy is equal to −0.4670 when only the NLE accuracy of the highest scorer is considered. It increases in magnitude to −0.5870 when we measure NLE accuracy with the average of the four largest NLE accuracy scores. Mispricing has been shown to decrease when subjects gain experience (Boening et al., 1993; Dufwenberg et al., 2005; Haruvy et al., 2007; Huber et al., 2016; King et al., 1993; King, 1991; Smith et al., 1988). However, a more recent paper by Kopányi-Peuker and Weber (2021) shows that mispricing is not eliminated when the market is repeated several times with the same subjects. Our results indicate that acquiring experience within

[9] Table 4 provides averages. An analysis that uses the median instead of the mean is available in section D.3 of the Appendix.
[10] A session RAD is the average of the RADs of the two markets in the session.

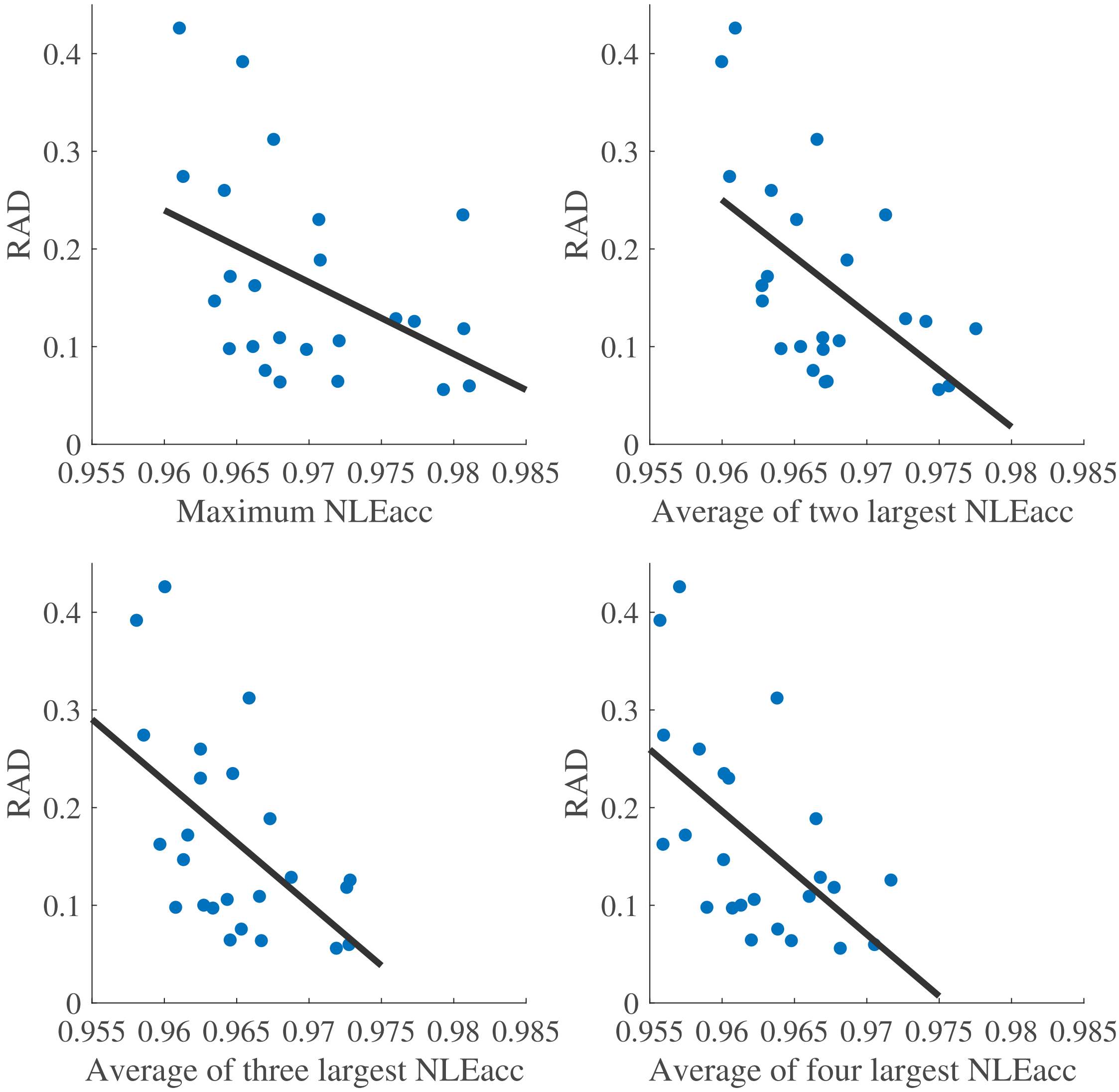

Fig. 4. Session level NLE and mispricing. The four graphs show the relationship between the session-level NLE accuracy (measured either by the NLE accuracy of the best performing subject of the session, or by different partial averages of NLE accuracies, namely the two, three and four best performing subjects on the NLE task). The least-squares lines (in black) shows the negative relationships between the different variables.

the experiment slightly mitigates the impact of NLE accuracy on mispricing. Indeed, the correlation coefficients are slightly smaller in absolute value (although significant) in second markets.

Figure 4 presents scatter plots of partial averages of NLE accuracy versus RAD. We focus on the measures of session-level NLE accuracy for which the correlation with RAD is the highest. As a result of the negative correlation between session-level NLE accuracy and RAD, we find a negative slope for the least-squares line (in black). The graphs show that when session-level NLE accuracy is low, RAD takes either small or large values. In contrast, when session-level NLE accuracy is high, RAD takes only small values. In other words, having subjects with high NLE accuracy in a session is a sufficient condition to reduce mispricing but not a necessary one.

In Table 6, we regress RAD on the different partial averages of NLE accuracy, controlling for variables that are either significant at the individual level in Table 3 or linked to mispricing in the literature (i.e., trading volume, previous experience and risk attributes). Five models appear in Table 6, depending on how session-level NLE accuracy is assessed. In Model 1, we consider the average NLE accuracy over all subjects in a session. Models 2 to 5 use partial averages of NLE accuracy. Session-level NLE accuracy coefficients are significant regardless of the measure used in the regression. R-squared values range from 0.3860 for average NLE accuracy to 0.5138 for the partial average NLE accuracy of the four most accurate subjects. Overall, our analyses confirm the hypothesis that high NLE accuracy subjects help mitigate mispricing. Regarding control variables, we find a negative coefficient for the *Log number of trades*, which shows that mispricing is a decreasing function of the

Table 6
NLE accuracy at the session level and mispricing.

	Model 1	Model 2	Model 3	Model 4	Model 5
Average *NLEacc*	−6.8581***				
	(2.1451)				
Maximum *NLEacc*		−6.7616**			
		(3.2018)			
Average of 2 largest *NLEacc*			−10.9201***		
			(2.7421)		
Average of 3 largest *NLEacc*				−12.3936***	
				(3.1579)	
Average of 4 largest *NLEacc*					−12.1250***
					(3.3711)
Log number of trades	−0.1337***	−0.1193**	−0.1116**	−0.1138***	−0.1102***
	(0.0434)	(0.0452)	(0.0411)	(0.0380)	(0.0363)
Proportion of experienced subjects	−0.0274	−0.0217	0.0035	0.0246	0.0509
	(0.1864)	(0.1622)	(0.1457)	(0.1534)	(0.1585)
Average risk attributes	0.0661**	0.0550**	0.0579**	0.0632**	0.0600**
	(0.0293)	(0.0256)	(0.0232)	(0.0247)	(0.0243)
Intercept	7.4363***	7.4112**	11.3519***	12.7431***	12.4136***
	(2.0923)	(3.2664)	(2.7837)	(3.1519)	(3.3361)
N	24	24	24	24	24
R^2	0.3860	0.4062	0.5004	0.5064	0.5138

This table presents the results of the regression of the Relative Absolute Deviation (*RAD*) on different measures of aggregate NLE accuracy (*NLEacc*). The dependent variable is the average of the RAD over the two markets of a given session. For a given market, $RAD = \frac{1}{N}\sum_{t=1}^{N}\frac{|\overline{P_t}-FV_t|}{\overline{FV}}$ where t denotes the period number and N is the total number of periods ($N = 10$ in our experiment). $\overline{P_t}$ is the period-t average transaction price and FV_t is the beginning of period-t fundamental value. Finally, $\overline{FV}$ is the average fundamental value over the N periods. The main independent variable is a function of subjects' NLE accuracy. In model 1, we consider the average NLE accuracy across all subjects in a session. Models 2 to 5 use partial averages of the best scorers on the NLE task (i.e., highest *NLEacc*). The control variables are the logarithm of the trading volume, the proportion of subjects that already participated in an economic experiment and the risk attributes. The risk attributes variable is the first principal component of the Principal Component Analysis of risk attributes defined in the exit questionnaire. ***, **, and * indicate statistical significance at the 1%, 5%, and 10%, respectively.

number of trades.[11] The *Proportion of experienced subjects* is not significant, in contrast with previous studies that show that the presence of experienced traders act as a price stabilizer and helps reduce mispricing (Akiyama et al., 2014; Huber et al., 2016). Our definition of experience, however, is slightly different from the one used in these previous studies since we consider the experience in all types of economic experiments and not specifically market experiments. Finally, the coefficient of *Average risk attributes* is positive and significant at the 5% level. Mispricing is larger when the subjects in a session are more risk tolerant. A similar result is found by Breaban and Noussair (2015) and Eckel and Füllbrunn (2015).

5. Conclusion

Our analysis shows that ANS acuity is a strong driver of trading performance. Our results, obtained in a controlled trading environment, suggest that number perception and approximate numeracy is linked to the ability to successfully implement profitable trading strategies. Overall, we find that traders who exhibit high ANS acuity tend to buy (sell) the risky asset at lower (higher) prices than their low ANS acuity counterparts. In addition, they react faster to favorable trading opportunities. Our findings also show that high ANS acuity subjects improve market efficiency. Sessions populated with few acute subjects typically exhibit lower mispricing.

Our paper provides novel insights into the determinants of trading performance. We show that ANS acuity, measured with a quick and simple test (the NLE task), explains a large proportion of trading profit differences across subjects in a standard market experiment. In this regard, our article also contributes to the debate on cognitive skills measurement in experimental markets. Beyond the correlation analyses mentioned in this paper, further research is warranted on the relationship between ANS acuity and popular measures of cognitive skills such as CRT tests and Raven matrices.

[11] The debate on the link between market efficiency and trading volume can be traced back to Milgrom and Stokey (1982). These authors show that no trade takes place in a speculative market because the willingness to trade signals private information and adverse selection. But Glosten and Milgrom (1985) show that trades take place when liquidity traders participate to the market. In De Long et al. (1990), noise traders can bid up the price of already over-priced assets. In such a framework, the arbitrageurs sometimes inflate the starting bubble and this 'positive-feedback trading' drives prices further from fundamentals. In the finance literature, linking trading volume to market efficiency (and therefore mispricing) is difficult for two reasons. First, the expected fundamental value of assets is unknown on financial markets. Second, part of the trading volume is devoted to hedging or risk-sharing. In experimental markets, however, trades are only speculative. We find that more trading means less mispricing. Our result is in line with Brown and Yang (2017) who study the horse race market in the United Kingdom. The horse race market is similar to our experimental framework. While we progressively reveal information to subjects, bettors accumulate information before and during the race, with a fundamental value publicly known at the end of the race.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.jedc.2021.104293.

References

Agrillo, C., Dadda, M., Serena, G., Bisazza, A., 2008. Do fish count? Spontaneous discrimination of quantity in female mosquitofish. Anim. Cogn. 11 (3), 495–503.
Agrillo, C., Dadda, M., Serena, G., Bisazza, A., 2009. Use of number by fish. PLoS ONE 4 (3), e4786.
Akiyama, E., Hanaki, N., Ishikawa, R., 2014. How do experienced traders respond to inflows of inexperienced traders? An experimental analysis. J. Econ. Dyn. Control 45, 1–18.
Ashcraft, M.H., Moore, A.M., 2012. Cognitive processes of numerical estimation in children. J. Exp. Child Psychol. 111 (2), 246–267.
Barber, B.M., Odean, T., 2000. Trading is hazardous to your wealth: The common stock investment performance of individual investors. J. Finance 55 (2), 773–806.
Boening, M.V.V., Williams, A.W., LaMaster, S., 1993. Price bubbles and crashes in experimental call markets. Econ. Lett. 41 (2), 179–185.
Booth, J.L., Siegler, R.S., 2006. Developmental and individual differences in pure numerical estimation. Dev. Psychol. 42 (1), 189–201.
Bosch-Rosa, C., Meissner, T., Bosch-Domènech, A., 2018. Cognitive bubbles. Exp. Econ. 21, 132–153.
Breaban, A., Noussair, C.N., 2015. Trader characteristics and fundamental value trajectories in an asset market experiment. J. Behav. Exp. Finance 8, 1–17.
Brown, A., Yang, F., 2017. The role of speculative trade in market efficiency: Evidence from a betting exchange. Rev. Financ. 21 (2), 583–603.
Bruine de Bruin, W., Parker, A., Fischhoff, B., 2007. Individual differences in adult decision-making competence. J. Pers. Soc. Psychol. 92 (5), 938–956.
Caginalp, G., Porter, D.P., Smith, V.L., 1998. Initial cash/asset ratio and asset prices: An experimental study. Proc. Natl. Acad. Sci. 95 (2), 756–761.
Caginalp, G., Porter, D.P., Smith, V.L., 2001. Financial bubbles: Excess cash, momentum, and incomplete information. J. Psychol. Financ. Mark. 2 (2), 80–99.
Camerer, C.F., Hogarth, R.M., 1999. The effects of financial incentives in experiments: A review and capital-labor-production framework. J. Risk Uncertain. 19 (1), 7–42.
Charness, G., Neugebauer, T., 2019. A test of the Modigliani-Miller invariance theorem and arbitrage in experimental asset markets. J. Finance 74 (1), 493–529.
Chen, Q., Li, J., 2014. Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. Acta Psychol. 148, 163–172.
Corgnet, B., Desantis, M., Porter, D., 2018. What makes a good trader? On the role of intuition and reflection on trader performance. J. Finance 73 (3), 1113–1137.
Cueva, C., Rustichini, A., 2015. Is financial instability male-driven? Gender and cognitive skills in experimental asset markets. J. Econ. Behav. Organ. 119, 330–344.
De Long, J.B., Shleifer, A., Summers, L.H., Waldmann, R.J., 1990. Noise trader risk in financial markets. J. Polit. Economy 98 (4), 703–738.
Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G.G., 2011. Individual risk attitudes: Measurement, determinants, and behavioral consequences. J. Eur. Econ. Assoc. 9 (3), 522–550.
Dufwenberg, M., Lindqvist, T., Moore, E., 2005. Bubbles and experience: An experiment. Am. Econ. Rev. 95 (5), 1731–1737.
Eckel, C.C., Füllbrunn, S.C., 2015. Thar SHE blows? Gender, competition, and bubbles in experimental asset markets. Am. Econ. Rev. 105 (2), 906–920.
Fazio, L.K., Bailey, D.H., Thompson, C.A., Siegler, R.S., 2014. Relations of different types of numerical magnitude representations to each other and to mathematics achievement. J. Exp. Child Psychol. 123, 53–72.
Feigenson, L., Dehaene, S., Spelke, E., 2004. Core systems of number. Trends Cogn. Sci. 8 (7), 307–314.
Fischbacher, U., 2007. z-tree: Zurich toolbox for ready-made economic experiments. Exp. Econ. 10 (2), 171–178.
Frederick, S., 2005. Cognitive reflection and decision making. J. Econ. Perspect. 19 (4), 25–42.
Gallistel, C.R., Gelman, R., 1992. Preverbal and verbal counting and computation. Cognition 44 (1–2), 43–74.
Garland, A., Low, J., Burns, K.C., 2012. Large quantity discrimination by North Island robins (Petroica longipes). Anim. Cogn. 15 (6), 1129–1140.
Gillette, A.B., Stevens, D.E., Watts, S.G., Williams, A.W., 1999. Price and volume reactions to public information releases: An experimental approach incorporating traders' subjective beliefs. Contemp. Account. Res. 16 (3), 437–479.
Glosten, L.R., Milgrom, P.R., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. J. Financ. Econ. 14 (1), 71–100.
Grinblatt, M., Keloharju, M., Linnainmaa, J.T., 2012. IQ, trading behavior, and performance. J. Financ. Econ. 104 (2), 339–362.
Halberda, J., Mazzocco, M.M., Feigenson, L., 2008. Individual differences in non-verbal number acuity correlate with maths achievement. Nature 455 (7213), 665–668.
Haruvy, E., Lahav, Y., Noussair, C.N., 2007. Traders' expectations in asset markets: Experimental evidence. Am. Econ. Rev. 97 (5), 1901–1920.
Huber, J., Kirchler, M., 2012. The impact of instructions and procedure on reducing confusion and bubbles in experimental asset markets. Exp. Econ. 15 (1), 89–105.
Huber, J., Kirchler, M., Stöckl, T., 2016. The influence of investment experience on market prices: Laboratory evidence. Exp. Econ. 19 (2), 394–411.
Hyde, D.C., Spelke, E.S., 2011. Neural signatures of number processing in human infants: Evidence for two core systems underlying numerical cognition. Dev. Sci. 14 (2), 360–371.
Izard, V., Sann, C., Spelke, E.S., Streri, A., 2009. Newborn infants perceive abstract numbers. Proc. Natl. Acad. Sci. 106 (25), 10382–10385.
Kahneman, D., 2011. Thinking, Fast and Slow. New York: Farrar, Straus and Giroux.
King, R., Smith, V., Williams, A., Van Boening, M., 1993. The robustness of bubbles and crashes in experimental stock markets. In: Non Linear Dynamics and Evolutionary Economics. Oxford University Press, pp. 183–200.
King, R.R., 1991. Private information acquisition in experimental markets prone to bubble and crash. J. Financ. Res. 14 (3), 197–206.
Kirchler, M., 2009. Underreaction to fundamental information and asymmetry in mispricing between bullish and bearish markets: An experimental study. J. Econ. Dyn. Control 33 (2), 491–506.
Kirchler, M., Huber, J., Stöckl, T., 2012. Thar she bursts: Reducing confusion reduces bubbles. Am. Econ. Rev. 102 (2), 865–883.
Kocher, M.G., Martinsson, P., Schindler, D., 2017. Overpricing and stake size: On the robustness of results from experimental asset markets. Econ. Lett. 154, 101–104.
Kopányi-Peuker, A., Weber, M., 2021. Experience does not eliminate bubbles: Experimental evidence. Rev. Financ. Stud. 34 (9), 4450–4485.
Lei, V., Vesely, F., 2009. Market efficiency: Evidence from a no-bubble asset market experiment. Pacific Econ. Rev. 14 (2), 246–258.
Libertus, M.E., Brannon, E.M., 2009. Behavioral and neural basis of number sense in infancy. Curr. Dir. Psychol. Sci. 18 (6), 346–351.
Libertus, M.E., Brannon, E.M., 2010. Stable individual differences in number discrimination in infancy. Dev. Sci. 13 (6), 900–906.
Linnainmaa, J.T., 2010. Do limit orders alter inferences about investor performance and behavior? J. Finance 65 (4), 1473–1506.
Lipton, J.S., Spelke, E.S., 2003. Origins of number sense. Psychol. Sci. 14 (5), 396–401.
Markowitz, H., 1952. Portfolio selection. J. Finance 7 (1), 77–91.
Mehlis, M., Thünken, T., Bakker, T.C.M., Frommen, J.G., 2015. Quantification acuity in spontaneous shoaling decisions of three-spined sticklebacks. Anim. Cogn. 18 (5), 1125–1131.
Milgrom, P., Stokey, N., 1982. Information, trade and common knowledge. J. Econ. Theory 26 (1), 17–27.
Nieder, A., Dehaene, S., 2009. Representation of number in the brain. Annu. Rev. Neurosci. 32 (1), 185–208.

Noussair, C.N., Tucker, S., Xu, Y., 2016. Futures markets, cognitive ability, and mispricing in experimental asset markets. J. Econ. Behav. Organ. 130, 166–179.
Odean, T., 1998. Are investors reluctant to realize their losses? J. Finance 53, 1775–1798.
Odic, D., Starr, A., 2018. An introduction to the approximate number system. Child Dev. Perspect. 12 (4), 223–229.
Östergren, R., Träff, U., 2013. Early number knowledge and cognitive ability affect early arithmetic ability. J. Exp. Child Psychol. 115 (3), 405–421.
Palan, S., 2015. GIMS - software for asset market experiments. J. Behav. Exp. Finance 5, 1–14.
Pepperberg, I.M., 2006. Grey parrot numerical competence: A review. Anim. Cogn. 9 (4), 377–391.
Perdue, B.M., Talbot, C.F., Stone, A.M., Beran, M.J., 2012. Putting the elephant back in the herd: Elephant relative quantity judgments match those of other species. Anim. Cogn. 15 (5), 955–961.
Peters, E., Bjalkebring, P., 2015. Multiple numeric competencies: When a number is not just a number. J. Pers. Soc. Psychol. 108 (5), 802–822.
Peters, E., Slovic, P., Västfjäll, D., Mertz, C., 2008. Intuitive numbers guide decisions. Judgm. Decis. Mak. 3, 619–635.
Schley, D.R., Peters, E., 2014. Assessing "economic value": Symbolic-number mappings predict risky and riskless valuations. Psychol. Sci. 25 (3), 753–761.
Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., De Smedt, B., 2017. Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. Dev. Sci. 20 (3), e12372.
Schneider, M., Merz, S., Stricker, J., De Smedt, B., Torbeyns, J., Verschaffel, L., Luwel, K., 2018. Associations of number line estimation with mathematical competence: A meta-analysis. Child Dev. 89 (5), 1467–1484.
Sharpe, W.F., 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. J. Finance 19 (3), 425–442.
Siegler, R.S., Booth, J.L., 2004. Development of numerical estimation in young children. Child Dev. 75 (2), 428–444.
Siegler, R.S., Opfer, J.E., 2003. The development of numerical estimation: Evidence for multiple representations of numerical quantity. Psychol. Sci. 14, 237–243.
Skagerlund, K., Forsblad, M., Tinghög, G., Västfjäll, D., 2021. Decision-making competence and cognitive abilities: Which abilities matter? J. Behav. Decis. Mak. 35 (1), e2242.
Smith, V., Suchanek, G., Williams, A., 1988. Bubbles, crashes, and endogenous expectations in experimental spot asset markets. Econometrica 56 (5), 1119–1151.
Sobkow, A., Olszewska, A., Traczyk, J., 2020. Multiple numeric competencies predict decision outcomes beyond fluid intelligence and cognitive reflection. Intelligence 80, 101452.
Stöckl, T., Huber, J., Kirchler, M., 2010. Bubble measures in experimental asset markets. Exp. Econ. 13 (3), 284–298.
Stöckl, T., Huber, J., Kirchler, M., 2015. Multi-period experimental asset markets with distinct fundamental value regimes. Exp. Econ. 18 (2), 314–334.
Szkudlarek, E., Brannon, E.M., 2017. Does the approximate number system serve as a foundation for symbolic mathematics? Lang. Learn. Dev. 13 (2), 171–190.
Torbeyns, J., Schneider, M., Xin, Z., Siegler, R.S., 2015. Bridging the gap: Fraction understanding is central to mathematics achievement in students from three different continents. Learn. Instr. 37, 5–13.
Xu, F., Spelke, E.S., 2000. Large number discrimination in 6-month-old infants. Cognition 74 (1), B1–B11.
Xu, F., Spelke, E.S., Goddard, S., 2005. Number sense in human infants. Dev. Sci. 8 (1), 88–101.