informs
http://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–17
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Forecasting Skills in Experimental Markets: Illusion or Reality?

Brice Corgnet,[a] Cary Deck,[b] Mark DeSantis,[c] David Porter[c]

[a] Emlyon Business School, GATE Lyon Saint-Étienne - Groupe d'Analyse et de Théorie Économique, 69130 Ecully; [b] Department of Economics, Finance, and Legal Studies, The University of Alabama, Tuscaloosa, Alabama 35487; [c] Argyros School of Business and Economics and the Economic Science Institute, Chapman University, Orange, California 92618

Contact: corgnet@em-lyon.com, https://orcid.org/0000-0002-0329-7933 (BC); cdeck@cba.ua.edu, https://orcid.org/0000-0002-4475-1768 (CD); desantis@chapman.edu, https://orcid.org/0000-0002-4771-0419 (MD); dporter@chapman.edu, https://orcid.org/0000-0003-2895-9048 (DP)

Received: February 23, 2021
Revised: May 27, 2021
Accepted: June 3, 2021
Published Online in Articles in Advance: November 5, 2021

https://doi.org/10.1287/mnsc.2021.4160

Abstract. There is an ongoing debate regarding the degree to which a forecaster's ability to draw correct inferences from market signals is real or illusory. This paper attempts to shed light on the debate by examining how personal characteristics do or do not affect forecaster success. Specifically, we investigate the role of fluid intelligence, manipulativeness, and theory of mind on forecast accuracy in experimental asset markets. We find that intelligence improves forecaster performance when market mispricing is low, manipulativeness improves forecaster performance when mispricing is high, and the degree to which theory of mind skills matter depends on both the level of mispricing and how information is displayed. All three of these results are consistent with hypotheses derived from the previous literature. Additionally, we observe that male forecasters outperform female forecasters after controlling for intelligence, manipulativeness, and theory of mind skills as well as risk aversion. Interestingly, we do not find any evidence that forecaster performance improves with experience across markets or within markets.

History: Accepted by Axel Ockenfels, behavioral economics and decision analysis.
Funding: This work was supported by Chapman University and Emlyon Business School.
Supplemental Material: The data files and supplementary material are available at https://doi.org/10.1287/mnsc.2021.4160.

Keywords: forecasting • experimental asset markets • theory of mind • manipulativeness • cognitive skills

1. Introduction

Although fewer and fewer jobs involve direct trading of securities, many tasks in the financial industry require understanding the economic implications of market prices. Forecasting is an essential part of financial analysts' and fund managers' jobs and is key for many professional economists (Elliott and Timmermann 2008). Yet, this practical role of analysts is in contradiction to the idea that markets are informationally efficient (Fama 1965, 1970, 1991; Samuelson 1965). If financial markets are informationally efficient, then no skill beyond basic understanding would be needed to infer the economy's current state from market prices. As a result, any success at forecasting market outcomes would simply be because of luck (Batchelor 1990, Hartzmark 1991, Barber and Odean 2000, Malkiel 2003, Qu et al. 2019), which Kahneman (2011) refers to as an *illusion of skill*.

However, there is strong evidence that actual markets are not perfectly efficient as was shown in both archival (e.g., De Bondt and Thaler 1985, 1987; Lo and MacKinlay 1988; Bernard and Thomas 1990; Cutler et al. 1991; Chopra et al. 1992; Jegadeesh and Titman 1993, 1995; Shleifer 2000; Lo 2019) and experimental studies (e.g., Smith et al. 1988; Biais et al. 2005, Hanson et al. 2006; Veiga and Vorsatz 2010; Page and Siemroth 2017, 2021; Corgnet et al. 2018b, 2020a). In the presence of mispricing, market prices become noisy signals, creating a situation in which accurate forecasting might require *genuine skills*. However, what skills would analysts need to possess to identify market fundamentals accurately in that context? Surprisingly, little is known about the individual characteristics driving forecasting performance. This lack of knowledge could, of course, be because of the illusion of skill (Kahneman 2011). However, the seminal work of Bruguier et al. (2010) (BQB henceforth) finds that social intelligence is an essential skill for predicting directional changes in market prices. More specifically, BQB show that one's capacities to understand others' emotions (as quantitated by Baron-Cohen et al. 1997) and ascribe intentions to apparently random patterns (as quantitated by Heider and Simmel 1944) are key predictors of forecasting performance in an experimental asset market. Collectively, these skills have been referred to as *theory of mind* (Frith and Frith 1999) because they relate to one's inclination to build a mental model of others' behavior. The neurological study of BQB detected an activation in the paracingulate cortex, which is known to activate when inferring others' intentions, for those forecasters who successfully predicted the direction of price changes. BQB also collected some behavioral data showing a positive correlation between theory of mind skills and forecasting performance.

We aim to provide a more comprehensive assessment of the individual drivers of forecasting performance by extending the work of BQB, who only studied 43 participants in their behavioral experiment connecting theory of mind and reasoning test scores with performance in forecasting price changes. Using a larger sample, we assess the relative importance of theory of mind and two other individual characteristics that BQB did not study but have been found to explain financial behavior in archival or experimental studies. These include standard cognitive skills or fluid intelligence (Raven 1936) and the personality trait of manipulativeness (Ashton et al. 2014). To the best of our knowledge, ours is the first study jointly assessing the predictive power of these individual characteristics on forecasting performance.

To accomplish our goal, we showed participants data from previously conducted experimental markets. In those markets, traders bought and sold assets whose fundamental value depended on which of three possible states was randomly selected. Following Plott and Sunder (1982, 1988), traders were endowed with private information, which in aggregate, was sufficient to reveal the true state. Participants in our study were asked to forecast the randomly selected state based upon the observed market activity. This enables participants' individual characteristics to be linked to their forecasting performance in a situation where fundamental value and thus, forecast accuracy are verifiable ex post. Our approach differs from BQB as we asked participants to predict the fundamental value of the asset rather than the direction of price changes. Thus, success in our task exclusively relied on extracting reliable private information from asset prices and market orders, whereas the task in BQB also required the prediction of price movements that might be unrelated to the fundamental value of the asset. It follows that, beyond the study of forecasting in financial markets, our task mimics the situation of corporate decision makers who attempt to extract private information from stock prices to guide their decisions regarding investments, takeovers, and Chief Executive Officer appointments (e.g., Chen et al. 2007, Bakke and Whited 2010, Bond et al. 2010).

Our main findings support the hypothesis that theory of mind skills play a fundamental role in accounting for a person's forecasting success. We extend previous works by putting forth that these skills are relevant when mispricing is high and the graphical display of information is extensive, thus confirming an informal conjecture stated in BQB. Our second main result shows that fluid intelligence (or Intelligence Quotient (IQ)) also contributes to explaining forecasting performance. This finding is in line with the archival study of high-performing forecasters, so-called "superforecasters" (Tetlock and Gardner 2016), who were found to possess high IQ levels. In addition, we also find that the manipulative personality trait (henceforth, manipulativeness) helps explain forecasting performance. This result is consistent with previous research showing a positive link between self-monitoring scores (Snyder and Gangestad 1986), which are closely linked to manipulativeness (Furnham 1989, Osborn et al. 1998), and traders' earnings in double auctions (see Biais et al. 2005).

Our findings that theory of mind skills, as well as standard cognitive skills and manipulativeness, explain forecasting performance support the broader hypothesis that forecasting skills are *real* and not *illusory*.

2. Hypotheses

In this section, we put forth our hypotheses regarding the effects of theory of mind skill, cognitive ability, and manipulativeness on people's ability to make financial forecasts. Our hypotheses contribute to the current literature by establishing distinct mechanisms by which these individual drivers of financial behavior might operate.

2.1. Theory of Mind

Predicting asset prices accurately requires specific skills related to theory of mind, as shown by BQB. In the behavioral setup of BQB, forecasters had to predict the direction of price changes in markets that might or might not be populated by insiders. In BQB, participants' ability to accurately assess the direction of price changes thus crucially hinges upon their capacity to attend and extract traders' private information. In related papers, Corgnet et al. (2018a) and Hefti et al. (2018) have reported a positive effect of theory of mind skills on traders' performance in experimental asset markets. These authors stress that traders with high theory of mind skills perceive market activity as resulting from traders' intentional actions. High theory of mind traders are thus especially attentive to market orders and hence, are more likely to extract valuable insights from observed market activity.

Because the effect of theory of mind skills crucially hinges upon extracting private information from observed market data, these skills will be especially beneficial when mispricing in the market is low. By contrast, in markets with high mispricing, forecasters will not be able to learn much from market data. In the case of asset market bubbles where mispricing is particularly high, De Martino et al. (2013) show that theory of mind skills can be detrimental. Forecasters possessing high theory of mind skills tend to actively use market data and mistakenly infer information about an asset's value from meaningless trends instead of recognizing that the market is being driven by noise. This negative effect of theory of mind is also highlighted by Hefti et al. (2018), who refer to market participants engaging in such behavior as being "semiotic."

As argued by BQB, theory of mind skills closely relate to pattern recognition. Indeed, the measurement tool of Heider and Simmel (1944) involves interpreting the movement of shapes in an animation. Thus, high theory of mind traders tend to be influenced by market movements attributing motivations to what they observe. This tendency may be heightened by the fact that market information is often presented graphically. Bruguier et al. (2010, p. 1722) point to the crucial role of the display of market orders in understanding the impact of theory of mind skills on forecasting performance: "One may wonder whether this success should be attributed to our using a purely graphical interface, where order and trade flows are translated into movement of circles of various sizes and colors." Other researchers have also argued that the display of market activity can prompt pattern recognition (Lo et al. 2000, Lo and Hasanhodzic 2011, De Martino et al. 2013). It follows that the effect of theory of mind skills should depend on the graphical display of the information. When the display of market information is more conducive to recognizing patterns, theory of mind skills will have a greater impact on forecasting performance, although the direction of the impact depends on the noisiness of the market information. Although theory of mind skills are multifaceted (BQB), the crucial feature that underlies our hypotheses relies on a trader's inclination to recognize patterns by attributing intentions to a series of market movements. Thus, pattern recognition skills as measured by the Heider–Simmel (HS; Heider and Simmel 1944) test will be our primary measure of the predictive power of theory of mind skills. More specifically, this measure captures the cognitive dimension of social intelligence, which is distinguished from its affective dimension in standard measures of social intelligence in clinical psychology (e.g., McDonald et al. 2003). The cognitive dimension assesses a person's ability to conduct social inference and attribute intentions to others, whereas the emotional dimension assesses one's ability to recognize emotions.[1] This distinction is also motivated by evidence in neuroscience showing that cognitive and affective dimensions of theory of mind activate distinct brain regions (see, e.g., Singer et al. 2004). This leads to Hypothesis 1.

Hypothesis 1 (Theory of Mind and Forecasting Performance).

a. *Theory of mind skills* (*as measured with pattern recognition*) *will enhance forecasting performance when mispricing is low and information is presented in a manner conducive to pattern recognition*.

b. *Theory of mind skills* (*as measured with pattern recognition*) *will hurt forecasting performance when mispricing is high and information is presented in a manner conducive to pattern recognition*.

2.2. Cognitive Ability

We focus on fluid intelligence (or IQ) as the relevant measure of cognitive ability because it consistently has been found to explain financial performance. IQ, which measures one's capacity to perform abstract mental calculations (Mackintosh and Mackintosh 2011), has been shown to relate to stock market participation and successful investment decisions (Kezdi and Willis 2003, Cole and Shastry 2009, Christelis et al. 2010, Grinblatt et al. 2011, Benjamin et al. 2013). Using a unique database of adult Finnish men, Grinblatt et al. (2012) found that high-IQ people exhibited better market timing than their low-IQ counterparts, thus being more likely to buy winning stocks and sell losing stocks. In experimental asset markets with private information similar to those used in the current study, Corgnet et al. (2018a) show that high-IQ traders earned significantly more than low-IQ traders.[2] Higher IQ test scores have also been correlated with fewer Bayesian updating errors (Charness et al. 2018, Corgnet et al. 2018a) and higher strategic levels of reasoning (Civelli and Deck 2018).[3] We thus predict that greater fluid intelligence should help forecasters infer the true asset value. Because standard cognitive skills' positive impact on forecasting performance hinges upon extracting information from market data, the benefit of stronger cognitive skills is greater when mispricing is low. By contrast, when mispricing is high, the benefit of stronger cognitive skills is minimal because the market information is noisy.

At first, the prediction that cognitive skills matter might seem at odds with the finding of BQB that participants' scores on a reasoning task did not significantly explain forecasting performance. However, our forecasting task differs from BQB as they asked participants to predict the direction of price changes in the market. Instead, we ask participants to identify the fundamental asset value based on market data. Furthermore, the reasoning task used in BQB is a measure not of IQ but rather, of cognitive reflection (or inhibitory control; Diamond 2013).[4] The null result of BQB might be because of the fact that cognitive reflection is an imperfect proxy of cognitive ability with correlation coefficients around 0.30 (see, e.g., Toplak et al. 2011, Corgnet et al. 2018a). Our primary measure of cognitive ability is thus fluid intelligence.

In contrast to theory of mind, we do not have an a priori reason to expect that the way information is displayed will influence the effect of standard cognitive skills on forecasting performance. This leads to Hypothesis 2.

Hypothesis 2 (Cognitive Ability and Forecasting Performance).

Cognitive ability will enhance forecasting performance when mispricing is low but will not affect forecasting performance when mispricing is high.

2.3. Manipulativeness

The finance literature has generally ignored personality traits despite recent work showing their relevance in explaining individual economic success (see Borghans et al. 2008; Barrick and Mount 2009; Almlund et al. 2011; Heckman and Kautz 2012; Corgnet et al. 2015, 2020b). The experimental economics literature has also shown correlations between personality traits and strategic sophistication. For example, Gill and Prowse (2016) show that agreeable and emotionally stable people tend to be more strategically sophisticated in beauty contest games and play strategies closer to the Nash equilibrium.

Only a few papers have studied the impact of personality traits on traders' earnings and strategies in experimental markets. One such study, Biais et al. (2005), shows that self-monitoring explains traders' performance.[5] More specifically, Biais et al. (2005, p. 298) claim that people who score high on self-monitoring "may assume that other market participants are also behaving strategically and trying to manipulate the market as they do. Accordingly, high self-monitors should be less likely to take market prices at face value and will reason about the signals and strategies that generated them." Even when traders are not explicitly incentivized to manipulate prices (as in Hanson et al. 2006, Deck et al. 2013), they can place orders that contradict their private information so as to distort prices and thus, trade advantageously (Biais et al. 2005). Self-monitoring skills should be especially critical in markets in which mispricing is high and attempts to distort prices are likely to be successful. By definition, markets with low mispricing have not experienced successful manipulation attempts, and as such, self-monitoring is unlikely to help forecasters. By contrast with theory of mind and standard cognitive skills, self-monitoring will lead traders to downplay the role of prices as valuable signals of the true asset value. Traders who are high in self-monitoring will thus be more manipulative and more likely to perceive other traders as trying to manipulate market prices (Biais et al. 2005). Following the original work of Snyder and Gangestad (1986), self-monitoring has been explicitly integrated in personality theory thanks to the development of the HEXACO scale (Lee and Ashton 2004), which includes honesty as the sixth fundamental trait of personality in addition to the big five (see, e.g., John et al. 1991). Also, the honesty trait includes a facet of personality that assesses whether a person is manipulative or not. Manipulativeness is measured with items, such as "I would pretend to like someone just to get that person to do favors for me," which are very closely related to the self-monitoring items, such as "I may deceive people by being friendly when I really dislike them." These two scales are both conceptually similar and empirically correlated (Tseëlon 1992, Grieve 2011). Because the self-monitoring scale was not developed within a theory of personality, it is not orthogonal to the five fundamental traits of personality. In particular, self-monitoring scores correlate with extraversion (see Furnham 1989, Osborn et al. 1998). Because of these psychometric concerns, we measure manipulativeness and extraversion separately using the HEXACO scale instead of the self-monitoring scale.

Finally, like cognitive skills, we have no a priori reason to believe that the manner in which information is presented will interact with manipulativeness.

Hypothesis 3 (Manipulativeness and Forecasting Performance).

Manipulativeness will enhance forecasting performance when mispricing is high but will not affect forecasting performance when mispricing is low.

3. Experimental Design

We test the mechanisms identified in the previous section in a laboratory setting in which we can manipulate mispricing levels in markets and the graphical display of information available to forecasters. The experiment was conducted in three phases: instructions, forecasting task, and questionnaire.

3.1. Instruction Phase

The full set of instructions is available in the supplementary material. The participants were recruited from a pool of individuals who had previously traded in the same type of experimental asset markets used in the forecasting task. These previous markets followed the design of Plott and Sunder (1982, 1988).[6] In phase 1, the participants were reminded that they had previously participated in a market experiment in which a single asset was traded for five minutes. This asset paid a liquidating dividend of either 50, 240, or 490 at the end of the market period with probabilities commonly known among the 12 traders in the market (these parameters are originally from Plott and Sunder 1988). Private signals of the form "Not 50," "Not 240," and "Not 490" were available to the traders in these markets. In Plott and Sunder (1982)-style markets, some traders were *fully informed*, meaning they were given two of these signals at the beginning of the period so that they knew the true asset value with certainty. The other traders were *uninformed* in that they did not receive private signals. In Plott and Sunder (1988)-style markets, all traders were *partially informed*, meaning that each received exactly one private signal; however, half received one signal, and the other half received the other possible signal. Thus, in aggregate, the participants had complete information.

Participants were given a reminder of how the double-auction trading market worked as well as the basic private signal and asset value structures. They completed two unpaid practice markets lasting five minutes. In the first practice market, participants were informed that all traders would be *partially informed*. In the second practice market, they were informed that one-half of the traders would be *fully informed*. As in the historic markets, these markets were conducted with groups of 12 participants. At the conclusion of the second practice market, participants were informed that they would be shown past experimental sessions and be asked to make inferences about what they observed. An incentivized three-question comprehension quiz was then administered ($0.50 per correct answer). Upon completion of the quiz, a monitor read the solutions to the participants and publicly answered any questions.

3.2. Forecasting Task

In phase 2, participants were placed in the role of a forecaster observing data from a previous financial market experiment (market data taken from Corgnet et al. 2018a, 2020a).[7] The forecasters' main task was to predict the true value of the asset being traded in the observed markets.

Of the available data from Corgnet et al. (2018a, 2020a), 12 sessions were randomly selected: 4 (of 10) of the sessions in which there were 12 *partially informed* traders, 4 (of 10) of the sessions in which there were 6 *fully informed* traders, and 4 (of 5) of the sessions in which there were only 2 *fully informed* traders.[8] Five market periods were then randomly drawn from each of these sessions. These five market periods from a single historic session were grouped together and treated as a single "block." The ordering of the blocks (five periods from a historic session) and the ordering of the markets within a block were randomized for each participant. That is, each participant observed a (different) random sequence of the 12 historic sessions listed. Moreover, each participant observed a (different) random sequence of the five markets within each historic session. That said, all participants observed data from the same 60 market periods. This randomization enables us to identify if forecasting performance improves with the forecaster's overall experience, the forecaster's experience with a given set of traders, or the experience of the traders in the market being observed.

Participants forecast the true asset value for each market they observed by assigning a probability, $Prob_v$, to each of the three possible asset values: $v \in V := \{50, 240, 490\}$.[9] Upon conclusion of a block, participants were asked to forecast two additional characteristics of the original session from which the five (just observed) markets were drawn. First, participants assigned probabilities to each of the three possible information structures: 12 *partially informed*, 6 *fully informed*, and 2 *fully informed*. Second, participants assigned probabilities to the number of participants in the original session who had high cognitive reflection test (CRT) scores (Frederick 2005).[10] In addition to allowing us to examine forecasters' beliefs about the cognitive sophistication of the group of traders they just observed, this served to reinforce to the forecaster that each block of five markets was from a distinct group of traders.

A quadratic scoring rule was used for each elicited belief (60 forecasts of true asset value, 12 forecasts of information structure, and 12 forecasts of CRT level). The payment per question was determined via the following equation:

$$\frac{1}{2}\left[1 + 2 \times Prob_{correct} - \sum_{v \in V} Prob_v^2\right],$$

where $Prob_{correct}$ is the probability the forecaster assigned to the true asset value. Thus, the payment per question could range between $0.00 for a guess that placed all probability on an incorrect value and $1.00 for a guess that placed all probability on the correct value, whereas placing equal weight on all values would result in a payoff of $0.67.[11] Participants were informed of the correct answer after submitting a forecast. Thus, they received immediate feedback regarding the accuracy of their forecast.

3.3. Questionnaire

In phase 3, participants completed a series of six surveys: (1) Heider–Simmel test, (2) cognitive ability test, (3) bomb risk elicitation task (BRET), (4) eye-gaze test, (5) HEXACO personality test, and (6) demographic questions.

The HS test is commonly used to assess theory of mind skills related to pattern recognition (Bossaerts et al. 2019). It was operationalized in a manner similar to BQB. Participants watched a video in five-second intervals of three geometric objects: a circle and two triangles. After each interval, the video was paused for 10 seconds, and participants were asked to forecast whether, after the next 5-second interval, the large triangle was going to be closer to, farther from, or at the same distance from the small triangle. Participants were paid $1 for each correct answer and incurred a $0.25 penalty for not responding within the 10 seconds.

Participants made a maximum of 14 guesses in this task. Our measure of theory of mind skill (HS score) is the number of correct responses the participant makes on this task.

The cognitive ability test was adapted from Civelli and Deck (2018). In this test, participants were shown a three-by-three table of images, with the image in the lower right corner missing. Participants were asked to select the image (from a given set of images) that logically completes the sequence similarly to the Raven (1936) test.[12] Participants were given six minutes to answer 12 such questions. Each correct answer was worth $0.50. Our measure of cognitive ability is the number of correct responses the participant makes on this task.

The BRET is adapted from Crosetto and Filippin (2013).[13] In this test, participants were shown a seven-by-seven square grid of boxes. A bomb was randomly placed behind 1 of the 49 boxes. Participants were instructed to select a number between 1 and 49, indicating the number of boxes they wished to collect (collection occurred left to right and then top to bottom). If the bomb was behind one of the collected boxes, then the participant earned $0.00 for this task. If the bomb was not behind one of the collected boxes, then the participant earned $0.10 for each collected box. Participants were given two minutes to complete this task. Even though we do not have a specific hypothesis regarding the effect of risk attitudes on forecasting performance, we included it as a control. Risk attitudes have been shown to relate to financial behavior in experimental markets, accounting for lower bids on risky assets (e.g., Fellner and Maciejovsky 2007, Breaban and Noussair 2015). Our measure of risk attitudes is based on the number of boxes selected by the participant. Rather than using the raw number, we rely on the implied level of relative risk aversion because this is a more standard measure and a nonlinear transformation of the number of boxes selected.[14]

The eye-gaze test is adapted from Baron-Cohen et al. (1997) and is another task commonly used to assess theory of mind skills. Participants were shown sets of eyes and instructed to select the word (from a list of four) that best described what the person in the image was thinking or feeling. Participants were given six minutes to complete this task, which was not incentivized.[15] We collected this additional theory of mind skills measure following BQB. However, as it is not a measure of pattern recognition, we do not expect that this aspect of theory of mind would affect forecasting performance, although BQB do report that it has some influence. Because this is not our main measure of theory of mind skills, we only collected the short version of the test that uses 10 instead of 36 questions (Olderbak et al. 2015). The eye-gaze test analysis is exploratory and relegated to the appendix (see Section A.4 in the appendix).

The HEXACO personality test asks participants to decide on a scale of one (strongly disagree) to five (strongly agree) how much they agreed with each of several statements.[16] This survey assesses six major dimensions of personality (honesty-humility, emotionality, extraversion, agreeableness, conscientiousness, and openness to experience). Participants were given 10 minutes to complete the 60 questions. Although the literature does not suggest any hypotheses regarding the predictive power for personality traits beyond the manipulativeness facet, we decided to measure all traits. This is motivated by the methodological concern of conducting the HEXACO test in its standard form rather than isolating a subset of questions.

Demographic questions consisted of the following three items. (1) What is your gender? (2) Do you regularly look at stock prices? (3) In what school are you enrolled?

3.4. Treatments

The design exogenously manipulates the level of mispricing by altering the structure of information, which we analyze within subject. It also varies the graphical display of information presented to forecasters, which we analyze between subject. With either 31 or 32 forecasters per graphical display each making 20 forecasts for each of three information structures, we observe a total of 7,500 forecasts for an average of 625 forecasts per combination of graphical display and information structure.

3.4.1. Mispricing. To assess mispricing in a market, we calculate the mean absolute deviation (MAD) of market prices and the true asset value. Mispricing was exogenously manipulated in each experimental session by having participants make forecasts under three possible information structures: 12 *partially informed*, 6 *fully informed*, and 2 *fully informed*. Corgnet et al. (2020a) show that prices typically track fundamentals in markets where insiders are one-half of the traders (six *fully informed*). In contrast, mispricing is often high in the 12 *partially informed* and 2 *fully informed* markets.[17] Using markets with different information structures assures considerable heterogeneity of MAD across all markets. In the subsequent analysis, we conduct a median split of the randomly selected markets based on MAD to identify low- and high-mispricing markets. The average MAD in the low-mispricing markets is 47.89, whereas the average MAD in the high-mispricing markets is 211.85. At a conceptual level, we are distinguishing between low and high levels of mispricing as reflecting cases in which private information is either successfully incorporated into prices or not. Thus, markets with low MAD are ones in which the rational expectations model forecasts prices relatively well, whereas markets with high MAD are ones in which the rational expectations model does not forecast prices well.

3.4.2. Graphical Display. The second design dimension corresponds to the nature of the graphical display shown to the forecasters. Although forecasters were always shown a time series of market prices in a chart, there were four variations based upon combinations of two distinct features (see Figures S1–S4 in the supplementary material). The order book, which contains bids and asks (see Offers to Sell and Offers to Buy on the right side of Figure S1 in the supplementary material), was either visible or not visible. Further, the market charts were displayed either statically or dynamically. When the display was dynamic, the participants observed a time-compressed 30-second video replay of the market that originally lasted for five minutes. In these videos, transactions appeared sequentially over the 30-second window. When the display was static, for 30 seconds the participants were shown the final image from the market replay video. Although forecasters could assign probabilities for each possible true asset value at any point during the 30-second observation window, they could not submit the probabilities until the 30 seconds had elapsed.

In total, we had four different graphical displays. We refer to the case where the image was static and the order book was not displayed as the *Basic Graphical Display*. This was the treatment in which the least amount of information was displayed. *Extensive Graphical Display* was dynamic and displayed an updated order book throughout the replay, and thus, it was the treatment in which the most information was graphically displayed. We refer to the case where the image was dynamic and the order book was not displayed and the case in which the image was static and the closing order book was displayed jointly as the *Intermediate Graphical Display*. These two cases provided more graphical information than the *Basic Graphical Display* and less than the *Extensive Graphical Display*. A priori, we cannot hypothesize whether the graphical information about the book or the dynamic display of transaction prices would matter the most in explaining the impact of theory of mind skills on forecasting performance. Yet, we conducted two treatments at the *Intermediate Graphical Display* in order to implement a full two-by-two factorial design. In addition, this allowed us to test, a posteriori, whether these two intermediate levels differ in explaining the effect of theory of mind skills on forecasting performance.[18]

3.5. Procedures

A total of 125 participants completed this study in six sessions. The average earnings of the participants were \$49.85.[19] This includes average payments of \$0.66 for the comprehension quiz in phase 1, \$33.95 for phase 2 (forecasting task), and \$8.24 for phase 3 (questionnaire).[20] In addition, participants were given a \$7 participation payment for the two-hour study.

Each experimental session was conducted with either 12 or 24 participants, and each participant was randomly assigned to one of the four graphical display formats.[21] This ensures that the observations are balanced across treatments, and it reduces any impact of session effects on treatment effects.

4. Results

We start by providing Table 1, which provides a correlation matrix for our main predictors of forecasting performance: theory of mind skills (as measured with HS), cognitive ability, and manipulativeness.

To the best of our knowledge, correlations between these variables have not been previously reported in the literature. We find it interesting that there is no substantial correlation between these factors. Further, we do not observe significant correlation between gender and any of the characteristics listed in Table 1 (i.e., no p-values < 0.10). Our measure of risk aversion is marginally correlated with cognitive ability (correlation coefficient = −0.160 with p-value = 0.074), although it is not significantly correlated with manipulativeness, HS, or gender. As a result, these variables can be used simultaneously in our analysis without introducing collinearity issues.[22] We also note that the mean score for each of the three characteristics listed in Table 1 did not differ across graphical display levels, which is the only dimension of the experimental design that was administered between subjects.[23]

To examine the effects of individual characteristics, we define the *Price Forecasting Error* as the difference (in absolute terms) between the predicted expected value of the asset based on the probabilities reported by the forecaster and the true asset value. We test our three hypotheses by regressing price forecasting error against our three predictors: *HS*, *Cognitive Ability*, and *Manipulativeness*. We use linear panel regressions with robust standard errors. We control for gender (*Male* dummy variable takes the value of one for males and zero otherwise), the true asset value in the observed market (*True Asset Value* is either 50, 240, or 490), and *Risk Aversion*. We also include two additional variables, *Original Round* and *Times Observed*. The former corresponds to the observed market's round number

Table 1. Correlations (p-Values) Between Individual Characteristics

	HS	Cognitive ability
Cognitive ability	−0.039 (0.663)	
Manipulativeness	−0.034 (0.707)	0.159 (0.077)

(1–17) within the original session and thus, is a measure of the experience of the traders in the market that is being observed. The latter indicates how many times the forecaster has observed that specific group of traders (one through five). These variables are included because forecasters might perform better when using market data generated by groups of traders who have been together longer and forecasters' predictions might improve after they have familiarized themselves with a specific group of traders. We also utilize session and time fixed effects. The former are included to account for any effect that might be session specific such as the time of day or the check-in process, whereas the latter are included to control for learning as forecasters' predictions might improve over time. We also include random effects for participants.[24]

Our hypotheses suggest that each of the three main characteristics interacts with the level of mispricing in the market. So, we interact each variable with the *High Mispricing* dummy variable, which takes the value of one if the MAD value of the observed market is greater than the median MAD value across all 60 historical markets and takes the value of zero otherwise. Hypothesis 1 further suggests that the effect of theory of mind skills depends on the manner in which information is presented graphically. Thus, we also include interactions of the *HS* and *High Mispricing* with dummy variables for *Intermediate Graphical Display* and *Extensive Graphical Display*. Results from this regression are presented in Table 2. Because the dependent variable is an error, negative coefficients indicate marginal improvements in forecasting performance.

Before assessing our hypotheses, we note that, if the regression reported in Table 2 is repeated without any interaction terms, the coefficient on *HS* is negative and significant (p-value = 0.032) indicating that, in aggregate, theory of mind skills enhance forecasting performance. Further, the coefficient on *Cognitive Ability* is found to be negative and marginally significant (p-value = 0.0996), whereas the coefficient on *Manipulativeness* is found to be insignificant (p-value = 0.470). These results are available in Table A.1 in the appendix.[25] However, our hypotheses suggest that such an approach to evaluating the importance of these characteristics is incomplete. Thus, we test our three hypotheses and present the findings in a series of results.

In line with Hypothesis 1, we show that theory of mind has explanatory power for forecasting performance that interacts with mispricing levels and the graphical display. In addition, cognitive ability and manipulativeness also explain forecasting performance as predicted in Hypotheses 2 and 3.

Result 1(a). In line with Hypothesis 1(a), we show that theory of mind skills (as measured with pattern recognition) enhance forecasting performance when mispricing is low and the graphical display of information is extensive.

Support. From the statistical tests reported in Table 2, when mispricing is low, the coefficient for HS score is not significant (which is the relevant test for *Basic Graphical Display*; p-value = 0.129). For *Extensive Graphical Display*, the relevant test is $HS + HS \times Extensive\ Graphical\ Display = 0$, which is negative and statically significant (p-value = 0.042). Interestingly, for *Intermediate Graphical Display*, the effect of *HS* when mispricing is low is $HS + HS \times Intermediate\ Graphical\ Display$, which is negative and significantly different from zero (p-value = 0.022). Our findings confirm Hypothesis 1(a) and suggest that the graphical display does not have to be extensive for theory of mind skills to play a critical role. □

Result 1(b). In line with Hypothesis 1(b), we show that theory of mind skills (as measured with pattern recognition) reduce forecasting performance when mispricing is high and the graphical display is extensive.

Support. From the statistical tests reported in Table 2, at the *Basic Graphical Display* when mispricing is high, the relevant test for the HS score is $HS + HS \times High\ Mispricing$, which is not significant (p-value = 0.172). However, for *Extensive Graphical Display*, the test is $HS + HS \times High\ Mispricing + HS \times Extensive\ Graphical\ Display + HS \times High\ Mispricing \times Extensive\ Graphical\ Display$, which is positive and marginally significant (p-value = 0.064). For *Intermediate Graphical Display*, the effect of *HS* when mispricing is high is $HS + HS \times High\ Mispricing + HS \times Intermediate\ Graphical\ Display + HS \times High\ Mispricing \times Intermediate\ Graphical\ Display$. This test is not significant (p-value = 0.628). Our findings are consistent with Hypothesis 1(b), suggesting that the graphical display needs to be sufficiently conducive to recognizing patterns for theory of mind skills to be detrimental when mispricing is high. □

Result 2. In line with Hypothesis 2, cognitive ability enhances forecasting performance when mispricing is low.

Support. Table 2 indicates that the coefficient for *Cognitive Ability* is negative and statistically significant (p-value = 0.038), which is the relevant test when mispricing is low. When mispricing is high, the effect of cognitive ability is captured by $Cognitive\ Ability + Cognitive\ Ability \times High\ Mispricing$, which is found to not be statistically different from zero (p-value = 0.815). □

Result 3. Manipulativeness enhances forecasting performance when mispricing is high.

Table 2. Analysis of Price Forecasting Error

	Price forecasting error
HS	−3.557 (2.343)
HS × *High Mispricing*	0.019 (2.795)
HS × *Intermediate Graphical Display*	−0.910 (3.015)
HS × *High Mispricing* × *Intermediate Graphical Display*	3.103 (4.527)
HS × *Extensive Graphical Display*	−2.065 (3.605)
HS × *High Mispricing* × *Extensive Graphical Display*	12.134*** (4.693)
Manipulativeness	2.912* (1.739)
Manipulativeness × *High Mispricing*	−8.060*** (2.847)
Cognitive Ability	−3.250** (1.566)
Cognitive Ability × *High Mispricing*	2.891 (2.095)
High Mispricing	138.335*** (2.107)
Male	−5.599** (2.496)
Risk Aversion	0.009 (1.462)
Original Round	−0.006 (0.189)
Times Observed	1.471 (2.382)
True Asset Value	−0.010 (0.007)
Constant	56.334*** (8.685)
Observations	7,494
Significance of Wald Tests on coefficients[a]	
HS + *HS* × *Extensive Graphical Display*= 0	0.042
HS + *HS* × *Intermediate Graphical Display* = 0	0.022
HS + *HS* × *High Mispricing* = 0	0.172
HS + *HS* × *High Mispricing* + *HS* × *Extensive Graphical Display* + *HS* × *High Mispricing* × *Extensive Graphical Display* = 0	0.064
HS + *HS* × *High Mispricing* + *HS* × *Intermediate Graphical Display* + *HS* × *High Mispricing* × *Intermediate Graphical Display* = 0	0.628
Cognitive Ability + *Cognitive Ability* × *High Mispricing* = 0	0.815
Manipulativeness + *Manipulativeness* × *High Mispricing* = 0	0.011

Notes. Robust standard errors in parentheses with session and time fixed effects are included as well as random effects for each participant. The Heider–Simmel (*HS*), *Manipulativeness*, and *Cognitive Ability* variables correspond to participants' standardized scores on each of these tests. The number of observations is equal to the number of forecasters (125) multiplied by the number of forecasts (60) made by each individual. Six observations were lost because of a computer error. For space considerations, we do not report the time and session fixed effect estimates.

[a]To assess the impact and significance of the interaction terms, we present results from Wald tests of coefficient equality. *p*-values are reported.

*Significance at the 0.1 level using the two-tailed test; **Significance at the 0.05 level using the two-tailed test; ***significance at the 0.01 level using the two-tailed test.

Support. In line with Hypothesis 3, when mispricing is high, the effect of manipulativeness is captured by *Manipulativeness* + *Manipulativeness* × *High Mispricing*, which is found to be negative and statistically different from zero (*p*-value = 0.011). Table 2 may also indicate that manipulativeness reduces forecasting performance when mispricing is low because the coefficient for *Manipulativeness* is positive and marginally statistically significant (*p*-value = 0.094), which is the relevant test when mispricing is low. □

Although we conduct multiple tests in support of Results 1–3, each of these tests is based on a priori independent hypotheses. Because we expect each of these tests to be significant, we are not in a case in which we expect an unspecified number of tests within a large subset to be significant. Still, one can apply a Holm–Bonferroni correction (Holm 1979) to our four main statistical tests: *HS* + *HS* × *Extensive Graphical Display* = 0; *HS* + *HS* × *High Mispricing* + *HS* × *Extensive Graphical Display* + *HS* × *High Mispricing* × *Extensive Graphical Display* = 0; *Cognitive Ability* = 0; and *Manipulativeness* + *Manipulativeness* × *High Mispricing* = 0. Before applying such a correction, we note that the relationships specified in our hypotheses are all one sided, whereas the *p*-values reported in Table 2 and discussed are two sided in keeping with convention. Thus, we apply the correction to the one-sided *p*-values. Although only the test of *Manipulativeness* + *Manipulativeness* × *High Mispricing* = 0 remains significant when applying the Holm–Bonferroni correction using $\alpha = 0.05$, all four tests remain significant when using $\alpha = 0.06$. We further note that the probability of all four tests yielding the predicted patterns by chance if none of the relationships are real is only 6.25×10^{-6}.

The regression results in Table 2 also suggest several other interesting findings. However, we do not report these as formal results because they are not based on a priori hypotheses and thus, should be viewed as exploratory. First, males make significantly better forecasts than females after controlling for the other characteristics considered. Second, risk attitudes do not appear to affect forecasting performance. Finally, experience does not appear to improve performance. This is true for the trader's experience (coefficient on *Original Round* is not statistically different from zero), the experience of the forecaster with the traders (coefficient on *Times Observed* is not statically different from zero), and the overall experience of the forecaster.[26]

Although we do not have any hypotheses regarding the interaction of the graphical display with either cognitive ability or manipulativeness, for completeness we also consider a specification with those interaction terms included and allowed to vary with mispricing. The results of that specification are presented in Table A.1 in the appendix. In the full specification, the coefficients are qualitatively similar to those shown in Table 2. Further, the p-value for the F test that all interaction terms involving cognitive ability or manipulativeness and the graphical display in the full specification are jointly zero is 0.384.

5. Discussion

This paper reports the results of an experiment in which participants observe asset markets from prior experiments and forecast the underlying fundamental value of the asset being traded. In addition to the forecasting task, we use standard tests to measure the forecasters' theory of mind skills, cognitive ability, and personality traits.

Our behavioral results identify that certain characteristics enhance forecasting ability and thus, that forecasting success is not simply luck. Our research design lays out and tests precise mechanisms, relating to the level of mispricing in the market and the graphical display of information, through which each set of skills operates. This allows us to shed light on previous literature results, such as the apparent contradictory effect of theory of mind skills in different market contexts.

In particular, we have shown that manipulativeness enhances forecasting performance when markets exhibit higher levels of mispricing, whereas cognitive ability does not. However, when mispricing in the market is low, the opposite pattern emerges as cognitive ability is beneficial and manipulativeness is not. These findings suggest that manipulative traders might perform relatively well in markets with high levels of mispricing in comparison with those who are not manipulative. If this was the case, then markets with high levels of mispricing may be more likely to be populated by manipulative traders and covered by manipulative analysts, thus reducing further the informational efficiency of these markets. By contrast, because those with high cognitive ability perform relatively better than those with low cognitive ability when mispricing is low, these markets may be more likely to be populated by high-cognitive ability traders and covered by similar analysts, which would enhance the informational efficiency of these markets.

We have also shown the effect of theory of mind skills to be more nuanced than those of manipulativeness and cognitive ability. That is, theory of mind skills hinder forecasting performance when mispricing is high and the graphical display of market activity is extensive. However, when market mispricing is low and the graphical display of market activity is extensive, theory of minds skills enhance forecasting performance. These findings help to reconcile previous results in the literature showing that theory of mind skills can be both beneficial and detrimental (Bruguier et al. 2010, Martino et al. 2013, Corgnet et al. 2018a, Hefti et al. 2018). Our findings corroborate the informal conjecture in BQB that the graphical display of market information is critical in understanding the impact of theory of mind skills on forecasting performance. Specifically, we identify an interaction effect between the level of mispricing and the graphical display of market information. Even though our results indicate that forecasting skill is not illusory, more research is needed in this area to identify other skills that provide a real benefit for forecasters. Further, more research as to how the features of the environment impact the relevance of individual skills for forecasting success is warranted. We hope that this paper helps to spark that effort.

Acknowledgments

Brice Corgnet acknowledges that this research has benefited from the financial support of IDEXLYON from Université de Lyon (project INDEPTH) within the Program Investissements d'Avenir (ANR-16-IDEX-0005) and was performed within the framework of the LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon, within the program Investissements d'Avenir (ANR-11-IDEX-007) operated by the French National Research Agency.

Appendix. Additional Analyses

This appendix provides additional analysis. Section A.1 augments the analysis in the main text by examining the interactions of graphical display with cognitive ability and manipulativeness. Section A.2 reconducts the analyses in the main text using closing prices to compute MAD. Section A.3 evaluates how well participants forecast other aspects of the markets they have been observing. Section A.4 contrasts the eye-gaze test for theory of mind with the Heider–Simmel test for theory of mind.

A.1. Interaction of Graphical Display with Cognitive Ability and Manipulativeness

Table A.1 reports two sets of regression results similar to Table 2. Specification (1) identifies the main effect of each of the three characteristics of interest. Specification (2) includes all interaction terms between *Graphical Display* and *Cognitive Ability* as well as between *Graphical Display* and *Manipulativeness*. The Wald test that every term included

Table A.1. Analysis of Price Forecasting Error with Additional Interaction Terms

	Price forecasting error	
	(1)	(2)
HS	−2.482** (1.156)	−3.514 (2.228)
HS × *High Mispricing*		0.263 (2.542)
HS × *Intermediate Graphical Display*		−1.002 (2.945)
HS × *High Mispricing* × *Intermediate Graphical Display*		2.817 (4.255)
HS × *Extensive Graphical Display*		−1.849 (3.390)
HS × *High Mispricing* × *Extensive Graphical Display*		11.764*** (4.199)
Manipulativeness	−0.886 (1.227)	2.997 (3.784)
Manipulativeness × *High Mispricing*		−5.656 (5.538)
Manipulativeness × *Intermediate Graphical Display*		1.418 (4.304)
Manipulativeness × *High Mispricing* × *Intermediate Graphical Display*		−4.805 (6.972)
Manipulativeness × *Extensive Graphical Display*		−3.363 (4.911)
Manipulativeness × *High Mispricing* × *Extensive Graphical Display*		0.197 (6.741)
Cognitive Ability	−1.853* (1.125)	−1.295 (3.992)
Cognitive Ability × *High Mispricing*		0.487 (4.379)
Cognitive Ability × *Intermediate Graphical Display*		−3.010 (4.497)
Cognitive Ability × *High Mispricing* × *Intermediate Graphical Display*		0.776 (5.406)
Cognitive Ability × *Extensive Graphical Display*		−1.604 (4.817)
Cognitive Ability × *High Mispricing* × *Extensive Graphical Display*		7.097 (5.306)
High Mispricing	138.003*** (2.248)	138.000*** (2.070)
Male	−5.332** (2.546)	−5.671** (2.481)
Risk Aversion	−0.051 (1.441)	−0.064 (1.416)
Original Round	−0.007 (0.189)	−0.005 (0.190)
Times Observed	1.331 (2.395)	1.544 (2.393)
True Asset Value	−0.010 (0.007)	−0.0104 (0.007)
Constant	56.63*** (8.737)	56.562*** (8.670)
Observations	7,494	7,494
Significance of joint test that certain coefficients equal zero[a]		0.384

Notes. Robust standard errors are in parentheses with session and time fixed effects included as well as random effects for each participant. The Heider–Simmel (*HS*), *Manipulativeness*, and *Cognitive Ability* variables correspond to participants' standardized scores on each of these tests. The number of observations is equal to the number of forecasters (125) multiplied by the number of forecasts (60) made by each individual. Six observations were lost because of a computer error. For space considerations, we do not report the time and session fixed effect estimates.

[a]We use a Wald test to jointly test that the following variables equal zero: *Cognitive Ability* × *Intermediate Graphical Display*, *Cognitive Ability* × *Extensive Graphical Display*, *Cognitive Ability* × *High Mispricing* × *Intermediate Graphical Display*, *Cognitive Ability* × *High Mispricing* × *Extensive Graphical Display*, *Manipulativeness* × *Intermediate Graphical Display*, *Manipulativeness* × *Extensive Graphical Display*, *Manipulativeness* × *High Mispricing* × *Intermediate Graphical Display*, and *Manipulativeness* × *High Mispricing* × *Extensive Graphical Display*. *p*-values are reported.

**Significance at the 0.1 level using the two-tailed test; **significance at the 0.05 level using the two-tailed test; ***significance at the 0.01 level using the two-tailed test.

in specification (2) of Table A.2 but not Table 2 is jointly zero has a p-value of 0.384. This suggests that either the effects of intelligence and manipulativeness do not depend upon how the information is presented or those effects are too small for us to identify given our sample size.

A.2. Analysis Based upon Closing Prices and Not-240 Instead of High Mispricing

Table A.2 replicates the analysis in Table 2 with MAD based upon the last three transactions in a market period. Additionally, the second and third columns of Table A.2 report the results of a similar analysis restricted to markets where the true asset value is 240 and to markets where the true asset value is 50 or 490, respectively, as mispricing tends to be lower in markets where the true asset value is 240. When the true asset value is 240 (not 240), 95.83% (77.78%) of the markets are classified as low-mispricing (high-mispricing) markets. Table A.3 replicates the analysis in Table 2 with the *High Mispricing* dummy variable replaced by the *Not-240* dummy variable, which takes the value of one if the true asset value in a market is not 240.

A.3. Predictions of CRT Composition of Markets and Distribution of Private Information

After observing a block of five markets and making forecasts of each market's true asset value, forecasters were then asked two questions regarding the block of markets. These questions were intended to give participants a break during the experiment and to make it clear they were transitioning from one block of past markets to the next. Forecasters were asked to predict the number of traders who correctly answered at least four of seven CRT questions correctly. Forecasters assigned weights to the following categories: 0–3 people answered at least four questions correctly, 4–8 people answered at least four questions correctly, and 9–12 people answered at least four questions correctly. Forecasters were also asked to guess the correct structure of private information in the

Table A.2. Analysis of Price Forecasting Error Based on Closing Prices

	Price forecasting error: All	Price forecasting error: True value = 240	Price forecasting error: True value = 50 or 490
HS	−4.831* (2.585)	−3.176 (2.605)	−9.178* (5.053)
HS × *High Mispricing*	2.688 (2.627)	−3.490 (9.336)	6.715* (3.675)
HS × *Intermediate Graphical Display*	0.494 (3.169)	1.680 (4.098)	0.182 (7.716)
HS × *High Mispricing* × *Intermediate Graphical Display*	0.166 (3.843)	2.971 (10.37)	0.374 (5.939)
HS × *Extensive Graphical Display*	−0.148 (4.248)	−2.825 (4.198)	8.194 (10.28)
HS × *High Mispricing* × *Extensive Graphical Display*	8.109* (4.712)	4.687 (11.03)	0.860 (8.878)
Manipulativeness	1.522 (1.808)	3.488 (2.195)	−2.020 (3.506)
Manipulativeness × *High Mispricing*	−5.160* (2.653)	4.300 (3.894)	−3.326 (3.373)
Cognitive Ability	−3.907** (1.694)	−2.930 (2.024)	−5.267 (4.499)
Cognitive Ability × *High Mispricing*	4.209** (2.092)	0.920 (3.769)	5.950 (4.047)
High Mispricing	134.4*** (1.894)	15.30*** (3.734)	143.4*** (3.296)
Male	−5.571** (2.500)	−0.872 (4.163)	−8.293* (4.372)
Risk Aversion	0.0367 (1.466)	0.998 (3.494)	−0.429 (2.857)
Original Round	0.957*** (0.189)	−1.144*** (0.287)	−1.781*** (0.238)
Times Observed	1.089 (2.470)	9.677*** (2.912)	−3.277 (3.065)
True Asset Value[a]	0.0222*** (0.00740)	—	0.0265*** (0.00732)
Constant	39.73*** (9.273)	18.192* (9.715)	93.15*** (12.00)
Observations	7,494	2,996	4,498
Significance of Wald Tests on coefficients[b]			
HS + *HS* × *Extensive Graphical Display*= 0	0.138	0.082	0.913
HS + *HS* × *Intermediate Graphical Display* = 0	0.019	0.631	0.116
HS + *HS* × *High Mispricing* = 0	0.336	0.457	0.397
HS + *HS* × *High Mispricing* + *HS* × *Extensive Graphical Display* + *HS* × *High Mispricing* × *Extensive Graphical Display* = 0	0.049	0.422	0.065
HS + *HS* × *High Mispricing* + *HS* × *Intermediate Graphical Display* + *HS* × *High Mispricing* × *Intermediate Graphical Display* = 0	0.535	0.745	0.503
Cognitive Ability + *Cognitive Ability* × *High Mispricing* = 0	0.828	0.628	0.700
Manipulativeness + *Manipulativeness* × *High Mispricing* = 0	0.046	0.063	0.010

Notes. Robust standard errors are in parentheses with session and time fixed effects included as well as random effects for each participant. The Heider–Simmel (*HS*), *Manipulativeness*, and *Cognitive Ability* variables correspond to participants' standardized scores on each of these tests. The number of observations is equal to the number of forecasters (125) multiplied by the number of forecasts (60) made by each individual. Six observations were lost because of a computer error. For space considerations, we do not report the time and session fixed effect estimates. If a market period had fewer than three transactions, then all of the transactions are used in the calculation of the MAD.

[a]Because the second specification only includes data from a single true asset value (240), the *True Asset Value* term is omitted.

[b]To assess the impact and significance of the interaction terms, we present results from Wald tests of coefficient equality. *p*-values are reported.

*Significance at the 0.1 level using the two-tailed test; **significance at the 0.05 level using the two-tailed test; ***significance at the 0.01 level using the two-tailed test.

five markets' observed block by assigning weights to 12 *partially informed*, 2 *fully informed*, and 6 *fully informed*. In total, forecasters made 12 guesses for each of these questions. Because these predictions are for categorical variables, we cannot rely on forecasting error as we did for asset value. Therefore, for each forecaster we calculate the average of the weights assigned to the correct number of high-CRT participants and the private information structure, respectively. We use ordinary least squares (OLS) regressions where these variables, *Average Weight on Correct Information Structure* and *Average Weight on Correct CRT Category*, are regressed against our primary measures of interest: Heider–Simmel (*HS*), *Cognitive Ability*, and *Manipulativeness*. We also control for the graphical display, gender, and level of risk aversion (via an estimated CRRA parameter).

The regression analysis presented in Table A.4 indicates that theory of mind as measured by the Heider–Simmel test helps forecasters identify the information structure of the market being observed. None of the considered

Table A.3. Analysis of Price Forecasting Using *Not-240* Instead of High Mispricing (the *Not-240* Dummy Variable Takes the Value of One if the *True Asset Value* in a Market Is Not 240)

	Price forecasting error
HS	−3.263 (2.681)
HS × *Not-240*	−0.458 (4.329)
HS × *Intermediate Graphical Display*	0.926 (4.064)
HS × *Not-240* × *Intermediate Graphical Display*	−0.597 (7.119)
HS × *Extensive Graphical Display*	−3.425 (4.248)
HS × *Not-240* × *Extensive Graphical Display*	12.040* (7.296)
Manipulativeness	3.655* (2.210)
Manipulativeness × *Not-240*	−7.728** (3.514)
Cognitive Ability	−3.354* (1.938)
Cognitive Ability × *Not-240*	2.512 (3.339)
Not-240	113.450*** (3.119)
Male	−5.611** (2.498)
Risk Aversion	0.042 (1.467)
Original Round	−1.627*** (0.174)
Times Observed	0.098 (2.874)
Constant	66.730*** (10.380)
Observations	7,494
Significance of Wald Tests on coefficients[a]	
HS + *HS* × *Extensive Graphical Display* = 0	0.045
HS + *HS* × *Intermediate Graphical Display* = 0	0.454
HS + *HS* × *Not-240* = 0	0.229
HS + *HS* × *Not-240* + *HS* × *Extensive Graphical Display* + *HS* × *Not-240* × *Extensive Graphical Display* = 0	0.242
HS + *HS* × *Not-240* + *HS* × *Intermediate Graphical Display* + *HS* × *Not-240* × *Intermediate Graphical Display* = 0	0.292
Cognitive Ability + *Cognitive Ability* × *Not-240* = 0	0.679
Manipulativeness + *Manipulativeness* × *Not-240* = 0	0.048

Notes. Robust standard errors are in parentheses with session and time fixed effects included as well as random effects for each participant. The Heider–Simmel (*HS*), *Manipulativeness*, and *Cognitive Ability* variables correspond to participants' standardized scores on each of these tests. The number of observations is equal to the number of forecasters (125) multiplied by the number of forecasts (60) made by each individual. Six observations were lost because of a computer error. For space considerations, we do not report the time and session fixed effect estimates.

[a]To assess the impact and significance of the interaction terms, we present results from Wald tests of coefficient equality. *p*-values are reported.

*Significance at the 0.1 level using the two-tailed test; **significance at the 0.05 level using the two-tailed test; ***significance at the 0.01 level using the two-tailed test.

characteristics help forecasters predict the number of participants in the market who had high CRT scores.

A.4. Comparisons of Eye Gaze and HS

Theory of mind is a broad concept referring to one's ability to understand what others are thinking. As such, there are multiple ways that it can be measured that may be relevant to forecasting performance. The HS test involves recognizing intention behind relative movement and as such, should matter when interpreting market trends (see, for example, BQB). By contrast, the eye-gaze test (Baron-Cohen et al. 1997) involves identifying emotions from pictures of people's eyes, and as such, it is less clear why that skill would be beneficial for interpreting market data. However, as both are identified as measures of theory of mind, we examine the relationship between these two measures for our participants. Olderbak

Table A.4. Average Weight Placed on the Correct Information Structure and Number of High-CRT Traders

	Average weight on correct information structure (1)	Average weight on correct CRT category (2)
HS	2.127** (0.995)	0.0927 (0.568)
Cognitive Ability	−1.211 (1.021)	0.959 (0.583)
Manipulativeness	−1.902* (1.003)	−0.585 (0.573)
Intermediate Graphical Display	2.309 (2.451)	0.855 (1.399)
Extensive Graphical Display	5.618** (2.829)	−2.156 (1.615)
Male	1.442 (2.055)	1.002 (1.173)
Risk Aversion	1.661 (1.442)	−0.234 (0.824)
Constant	35.85*** (2.179)	40.22*** (1.244)
Observations	125	125
Adjusted R^2	0.076	0.014

Note. Standard errors are in parentheses.

*Significance at the 0.1 level using the two-tailed test; **significance at the 0.05 level using the two-tailed test; ***significance at the 0.01 level using the two-tailed test.

Table A.5. Correlations (p-Values) Between Measures of Theory of Mind

	HS	Eye-gaze 10
Eye-gaze 10	−0.021 (0.815)	
Eye-gaze 36	−0.094 (0.297)	0.260 (0.004)

et al. (2015) have argued that a 10-question version of the eye-gaze test is as valid as the 36-question version of Baron-Cohen et al. (1997). As all of our participants had previously participated in a study in which the 36-question test had been applied, we also compare scores on the 10-question test with prior scores on the 36-question test in order to verify the robustness of the results of Olderbak et al. (2015). Table A.5 indicates that the 10- and 36-question versions of the eye-gaze test provide positively and significantly correlated information about a respondent's ability to perceive emotions, although the magnitude of the correlation coefficient is moderate. Interestingly, respondents' scores from these eye-gaze tests are not significantly correlated with their HS scores. We replicated Table 2, substituting eye-gaze test scores for Heider–Simmel scores. Let *Eye Gaze 10* (*Eye Gaze 36*) represent the participant's standardized score (i.e., the number of correct responses) on the 10-question (36-question) version of the eye gaze test. Only 1 (*Eye Gaze 10* × *High Mispricing* × *Intermediate Graphical Display*) of the 12 coefficients (*Eye Gaze Z*, *Eye Gaze Z* × *High Mispricing*, *Eye Gaze Z* × *Intermediate Graphical Display*, *Eye Gaze Z* × *High Mispricing* × *Intermediate Graphical Display*, *Eye Gaze Z* × *Extensive Graphical Display*, and *Eye Gaze Z* × *High Mispricing* × *Extensive Graphical Display* for Z = 10 or 36) for this alternative measure of theory mind was significantly different from zero, whereas the other results remain qualitatively unchanged. That 1 of 12 coefficients turns out to be significant is unsurprising given the number of exploratory tests. We thus conclude that Hypothesis 1 does not hold when replacing Heider–Simmel scores with eye-gaze scores.

Endnotes

[1] Because previous research (e.g., Bruguier et al. 2010, Corgnet et al. 2018a, Hefti et al. 2018) has also elicited the emotional dimension of social intelligence using the eye-gaze test (Baron-Cohen et al. 1997) as a predictor of forecasting and trading performance in experimental markets, we also collected this measure.

[2] Similar results are found by Hefti et al. (2018) using the canonical design of Smith et al. (1988) for exploring price bubbles in asset markets.

[3] High-IQ people have also been shown to induce greater truthful revelation in second price auctions (Lee et al. 2020).

[4] Although cognitive reflection scores correlate positively with IQ scores, both constructs are distinct because they relate to different executive functions. IQ can be seen as a measure of working memory, whereas cognitive reflection relates to inhibitory control (Stanovich 2011, 2016).

[5] Snyder and Gangestad (1986) refer to self-monitoring as a stable personality trait, which is defined as one's inclination to attend social cues.

[6] We recruited participants who were experienced in trading to ensure they would understand the financial environment in which they had to make forecasts.

[7] Because of a recruitment error, 6 of the 125 participants in the current study actually participated in those markets. Similar results are obtained if we drop these participants from the analysis.

[8] The parentheses report the number of historic markets of the given type that were conducted by Corgnet et al. (2018a, 2020a). In each of these markets, there were a total of 12 traders.

[9] In addition to the market activity, forecasters were given the same commonly known probabilities (35%, 45%, and 20%) as the traders about the asset being worth 50, 240, or 490.

[10] Corgnet et al. (2018a, 2020a) administered a seven-question CRT survey to participants in the original sessions. Our forecasters were asked to predict how many of the original traders answered at least four of the seven CRT questions correctly: 0–4, 5–8, and 9–12. Our forecasters were provided with the CRT questions prior to making this prediction. Similar to IQ, CRT is a measure of cognitive sophistication (Oechssler et al. 2009; Hoppe and Kusterer 2011; Toplak et al. 2011; Corgnet et al. 2018a, 2021), and many experimental papers have shown the predictive power of CRT scores on traders' performance (see Breaban and Noussair 2015; Corgnet et al. 2015, 2020a; Noussair et al. 2016; Hanaki et al. 2017; Duchenne et al. 2019; Schneider and Porter 2020).

[11] Although one might expect that more risk-averse forecasters would place more equal weight on each outcome than less risk-averse forecasters, we do not observe a significant correlation between the average maximum weight placed on an outcome and our measure of risk attitudes, as described. Specifically, the observed correlation is 0.006 (p-value = 0.949). Analyses of forecasts of information structure and CRT level are provided in Section A.3 in the appendix.

[12] Although our measure of cognitive ability taken from Civelli and Deck (2018) and our measure of theory of mind taken from Heider and Simmel (1944) both rely on the search for patterns, the former examines the ability to find objective logical relationships and relates to fluid intelligence (Mackintosh and Mackintosh 2011), whereas the latter examines the ability to infer intentions from dynamic relationships and relates to theory of mind (Frith and Frith 1999).

[13] We opted for this risk elicitation tool because it provides a granular partition of risk attitudes while only requiring a single decision. Tasks that involve more questions can yield data inconsistent with standard models of risk or impose structure on the decision maker.

[14] Formally, we assume that each participant's risk preferences are captured by the constant relative risk aversion model and then, take the midpoint of the range of parameter values that would be consistent with the observed choice. A few participants selected only one box, which is a level of risk aversion described by Holt and Laury (2002) as "stay in bed." For these participants, we assume the risk parameter is one, which is consistent with the choice of one box.

[15] Although it is standard not to incentivize this test (Baron-Cohen et al. 1997, Corgnet et al. 2018a), BQB paid for performance.

[16] See http://hexaco.org/hexaco-inventory for more details regarding the HEXACO personality test.

[17] In related experiments, Bossaerts et al. (2014) show that mispricing decreases as the number of fully informed traders in the market increases.

[18] Ultimately, our analysis does not reveal substantial differences between the two cases that comprise *Intermediate Graphical Display*. Specifically, the specification reported in Table 2 can be modified to include separate variables for the two intermediate graphical display treatments: *Intermediate Graphical Display A* and *Intermediate Graphical Display B*. The p-value when testing *HS* × *Intermediate Graphical Display A* equals *HS* × *Intermediate Graphical Display B* is 0.762, and the p-value when testing *HS* × *High Mispricing* ×

Intermediate Graphical Display A equals *HS* × *High Mispricing* × *Intermediate Graphical Display B* is 0.506.

[19] The exchange rate was one experimental dollar equals U.S. $1 in sessions 1 and 2 (12 and 24 participants). It was one experimental dollar equals U.S. $0.50 in sessions 3–6 (24 participants in sessions 3–5 and 18 in session 6). This exchange rate adjustment was made to keep average earnings in the range dictated by laboratory policy. In our analyses, we include session fixed effects, which control for differences in exchange rates as well as other idiosyncratic session-specific shocks.

[20] On average, participants answered 2.04 of 3 comprehension quiz questions correctly.

[21] When possible, the number of participants assigned to each graphic level was balanced. However, one session was run with 18 participants, in which case the assignment was not fully balanced. For the practice market periods, the participants in this session were divided into two groups of nine. Also, in session 5, 1 of the 24 participants exited halfway through phase 2 of the experiment, so we had to drop this observation from the analyses.

[22] Using an OLS regression with the specification considered in Table 2, we report a mean variance inflation factor of 2.15 across all variables, which is substantially lower than the standard cutoff values used for spotting multicollinearity concerns (see, e.g., Gordon 2015, Hair et al. 2018).

[23] Specifically, regression analysis yields p-values of 0.715, 0.451, and 0.905 when testing that mean scores are the same across graphical display treatments for HS, cognitive ability, and manipulativeness, respectively.

[24] A Hausman test fails to reject the adequacy of random effects (p-value > 0.500).

[25] Section A.2 in the appendix provides analysis similar to that in Table 2 but with MAD based on the last three trades in a market (see Table A.2). We also provide an analysis in which we replace the *High Mispricing* dummy variable with the *Not-240* dummy variable, which takes the value of one if the true asset value in a market is not 240 (see Table A.3). The results are generally consistent with those in Table 2.

[26] Only four of the estimated time fixed effects in Table 2 are significantly different from zero (p-value < 0.05), which is similar to the three that one would expect to show significance by chance given the number of effects estimated. The significant effects occurred in periods 16, 36, 44, and 57.

References

Almlund M, Duckworth AL, Heckman JJ, Kautz T (2011) Personality psychology and economics. Hanushek E, Machin S, Woessman L, eds. *Handbook of the Economics of Education*, vol. 4 (Elsevier, Amsterdam), 1–181.

Ashton MC, Lee K, de Vries RE (2014) The HEXACO Honesty-Humility, Agreeableness, and Emotionality Factors: A review of research and theory. *Personality Soc. Psych. Rev.* 18:139–152.

Bakke TE, Whited TM (2010) Which firms follow the market? An analysis of corporate investment decisions. *Rev. Financial Stud.* 23(5):1941–1980.

Barber BM, Odean T (2000) Trading is hazardous to your wealth: The common stock investment performance of individual investors. *J. Finance* 55(2):773–806.

Baron-Cohen S, Jolliffe T, Mortimore C, Robertson M (1997) Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger syndrome. *J. Child Psych. Psychiatry* 38(7):813–822.

Barrick MR, Mount MK (2009) Select on conscientiousness and emotional stability. Locke E, ed. *Handbook of Principles of Organizational Behavior* (John Wiley and Sons, Ltd., Malden, MA), 19–40.

Batchelor RA (1990) All forecasters are equal. *J. Bus. Econom. Statist.* 8(1):143–144.

Benjamin DJ, Brown SA, Shapiro JM (2013) Who is 'behavioral'? Cognitive ability and anomalous preferences. *J. Eur. Econom. Assoc.* 11(6):1231–1255.

Bernard VL, Thomas JK (1990) Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. *J. Accounting Econom.* 13(4):305–340.

Biais B, Hilton D, Mazurier K, Pouget S (2005) Judgemental overconfidence, self-monitoring, and trading performance in an experimental financial market. *Rev. Econom. Stud.* 72(2):287–312.

Bond P, Goldstein I, Prescott ES (2010) Market-based corrective actions. *Rev. Financial Stud.* 23(2):781–820.

Borghans L, Duckworth AL, Heckman JJ, Ter Weel B (2008) The economics and psychology of personality traits. *J. Human Resources* 43(4):972–1059.

Bossaerts P, Frydman C, Ledyard J (2014) The speed of information revelation and eventual price quality in markets with insiders: Comparing two theories. *Rev. Finance* 18(1):1–22.

Bossaerts P, Suzuki S, O'Doherty JP (2019) Perception of intentionality in investor attitudes toward financial risks. *J. Behav. Experiment. Finance* 23:189–197.

Breaban A, Noussair CN (2015) Trader characteristics and fundamental value trajectories in an asset market experiment. *J. Behav. Experiment. Finance* 8:1–17.

Bruguier AJ, Quartz SR, Bossaerts P (2010) Exploring the nature of "trader intuition." *J. Finance* 65(5):1703–1723.

Charness G, Rustichini A, Van de Ven J (2018) Self-confidence and strategic behavior. *Experiment. Econom.* 21(1):72–98.

Chen Q, Goldstein I, Jiang W (2007) Price informativeness and investment sensitivity to stock price. *Rev. Financial Stud.* 20(3):619–650.

Chopra N, Lakonishok J, Ritter JR (1992) Measuring abnormal performance: Do stocks overreact? *J. Financial Econom.* 31(2):235–268.

Christelis D, Jappelli T, Padula M (2010) Cognitive abilities and portfolio choice. *Eur. Econom. Rev.* 54(1):18–38.

Civelli A, Deck C (2018) A flexible and customizable method for assessing cognitive abilities. *Rev. Behav. Econom.* 5(2):123–147.

Cole SA, Shastry GK (2009) Smart money: The effect of education, cognitive ability, and financial literacy on financial market participation. Harvard Business School Finance Working Paper No. 09-071, Harvard Business School, Boston.

Cole SA, Paulson A, Shastry GK (2014) Smart money? The effect of education on financial outcomes. *Rev. Financial Stud*. 27(7):2022–2051.

Corgnet B, DeSantis M, Porter D (2018a) What makes a good Trader? On the role of intuition and reflection on trading performance. *J. Finance* 73(3):1113–1137.

Corgnet B, DeSantis M, Porter D (2020a) The distribution of information and the price efficiency of markets. *J. Econom. Dynam. Control* 110:103671.

Corgnet B, DeSantis M, Porter D (2021) Information aggregation and the cognitive make-up of market participants. Zimmermann K, ed. *Eur. Econom. Rev*. 133:103667.

Corgnet B, Gaechter S, Hernán González R (2020b) Personality, emotional intelligence, and rationality. Zimmermann K, ed. *Handbook of Labor, Human Resources and Population Economics* (Springer, Cham, Switzerland), 1–28.

Corgnet B, Hernan-Gonzalez RH, Mateo R (2015) Cognitive reflection and the diligent worker: An experimental study of millennials. *PLoS One* 10(11):e0141243.

Corgnet B, Deck C, DeSantis M, Porter D (2018b) Information (non) aggregation in markets with costly signal acquisition. *J. Econom. Behav. Organ.* 154:286–320.

Crosetto P, Filippin A (2013) The "bomb" risk elicitation task. *Experiment. Econom.* 47(1):31–65.

Cutler DM, Poterba JM, Summers LH (1991) Speculative dynamics. *Rev. Econom. Stud.* 58(3):529–546.

De Bondt WF, Thaler R (1985) Does the stock market overreact? *J. Finance* 40(3):793–805.

De Bondt WF, Thaler RH (1987) Further evidence on investor overreaction and stock market seasonality. *J. Finance* 42(3):557–581.

Deck C, Lin S, Porter D (2013) Affecting policy by manipulating prediction markets: Experimental evidence. *J. Econom. Behav. Organ.* 85(1):48–62.

De Martino B, O'Doherty JP, 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.

Diamond A (2013) Executive functions. *Annual Rev. Psych.* 64: 135–168.

Duchêne S, Guerci E, Hanaki N, Noussair CN (2019) The effect of short selling and borrowing on market prices and traders' behavior. *J. Econom. Dynam. Control* 107:103734.

Elliott G, Timmermann A (2008) Economic forecasting. *J. Econom. Literature* 46(1):3–56.

Fama EF (1965) The behavior of stock-market prices. *J. Bus.* 38(1): 34–105.

Fama E (1970) Efficient capital markets: A review of theory and empirical work. *J. Finance* 25(2):383–417.

Fama E (1991) Efficient capital markets. II. *J. Finance* 46(5): 1575–1617.

Fellner, Maciejovsky (2007) Risk attitude and market behavior: Evidence from experimental asset markets. *J. Econom. Psych.* 28(3):338–350.

Frederick S (2005) Cognitive reflection and decision making. *J. Econom. Perspect.* 19(4):25–42.

Frith CD, Frith U (1999) Interacting minds–a biological basis. *Science* 286(5445):1692–1695.

Furnham A (1989) Personality correlates of self-monitoring: The relationship between extraversion, neuroticism, type A behaviour and Snyder's self-monitoring construct. *Personality Individual Differences* 10(1):35–42.

Gill D, Prowse V (2016) Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis. *J. Political Econom.* 124(6):1619–1676.

Gordon RA (2015) *Regression Analysis for the Social Sciences* (Routledge, New York).

Grieve R (2011) Mirror: The role of self-monitoring and sincerity in emotional manipulation. *Personality Individual Differences* 51(8): 981–985.

Grinblatt M, Keloharju M, Linnainmaa J (2011) IQ and stock market participation. *J. Finance* 66(6):2121–2164.

Grinblatt M, Keloharju M, Linnainmaa JT (2012) IQ, trading behavior, and performance. *J. Financial Econom.* 104(2):339–362.

Hair J, Black W, Babin B, Anderson R (2018) *Multivariate Data Analysis*, 8th ed. (Cengage Learning EMEA, Delhi, India).

Hanaki N, Akiyama E, Funaki Y, Ishikawa R (2017) Diversity in cognitive ability enlarges mispricing in experimental asset markets. GREDEG working paper 2017-08, Université Côte d'Azur, Nice, France.

Hanson R, Oprea R, Porter D (2006) Information aggregation and manipulation in an experimental market. *J. Econom. Behav. Organ.* 60(4):449–459.

Hartzmark ML (1991) Luck vs. forecast ability: Determinants of trader performance in futures markets. *J. Bus.* 64(1):49–74.

Heckman JJ, Kautz T (2012) Hard evidence on soft skills. *Labour Econom.* 19(4):451–464.

Hefti A, Heinke S, Schneider F (2018) Mental capabilities, heterogeneous trading patterns and performance in an experimental asset market. University of Zurich, Department of Economics Working Paper No. 234, University of Zurich, Zurich, Switzerland.

Heider F, Simmel M (1944) An experimental study of apparent behavior. *Amer. J. Psych.* 57:243–259.

Holm S (1979) A simple sequentially rejective multiple test procedure. *Scandinavian J. Statist.* 6(2):65–70.

Holt CA, Laury SK (2002) Risk aversion and incentive effects. *Amer. Econom. Rev.* 92(5):1644–1655.

Hoppe EI, Kusterer DJ (2011) Behavioral biases and cognitive reflection. *Econom. Lett.* 110(2):97–100.

Jegadeesh N, Titman S (1993) Returns to buying winners and selling losers: Implications for stock market efficiency. *J. Finance* 48(1): 65–91.

Jegadeesh N, Titman S (1995) Overreaction, delayed reaction, and contrarian profits. *Rev. Financial Stud.* 8(4):973–993.

John OP, Donahue EM, Kentle RL (1991) *The Big Five Inventory - Versions 4a and 54* (University of California, Berkeley, Institute of Personality and Social Research, Berkeley).

Kahneman D (2011) *Thinking, Fast and Slow* (Macmillan, New York).

Kezdi G, Willis RJ (2003) Who becomes a stockholder? Expectations, subjective uncertainty, and asset allocation. Michigan Retirement Research Center Research Paper No. WP 39, Michigan Retirement Research Center, Ann Arbor.

Lee K, Ashton MC (2004) Psychometric properties of the HEXACO personality inventory. *Multivariate Behav. Res.* 39(2):329–358.

Lee JY, Nayga R, Deck C, Drichoutis AC (2020) Cognitive ability and bidding behavior in second price auctions: An experimental study. *Amer. J. Agricultural Econom.* 102(5):1494–1510.

Lo AW (2019) *Adaptive Markets: Financial Evolution at the Speed of Thought* (Princeton University Press, Princeton, NJ).

Lo AW, Hasanhodzic J (2011) *The Evolution of Technical Analysis: Financial Prediction from Babylonian Tablets to Bloomberg Terminals* (John Wiley & Sons, Hoboken, NJ).

Lo AW, MacKinlay AC (1988) Stock market prices do not follow random walks: Evidence from a simple specification test. *Rev. Financial Stud.* 1(1):41–66.

Lo AW, Mamaysky H, Wang J (2000) Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *J. Finance* 55(4):1705–1765.

Mackintosh N, Mackintosh NJ (2011) *IQ and Human Intelligence* (Oxford University Press, Oxford, United Kingdom).

Malkiel BG (2003) The efficient market hypothesis and its critics. *J. Econom. Perspect.* 17(1):59–82.

McDonald S, Flanagan S, Rollins J, Kinch J (2003) TASIT: A new clinical tool for assessing social perception after traumatic brain injury. *J. Head Trauma Rehabilitation* 18(3):219–238.

Noussair C, Tucker S, Xu Y (2016) Futures markets, cognitive ability, and mispricing in experimental asset markets. *J. Econom. Behav. Organ.* 130:166–179.

Oechssler J, Roider A, Schmitz PW (2009) Cognitive abilities and behavioral biases. *J. Econom. Behav. Organ.* 72(1):147–152.

Olderbak S, Wilhelm O, Olaru G, Geiger M, Brenneman MW, Roberts RD (2015) A psychometric analysis of the reading the mind in the eyes test: Toward a brief form for research and applied settings. *Frontiers Psych.* 6:1503.

Osborn SM, Field HS, Veres JG (1998) Introversion-extraversion, self-monitoring, and applicant performance in a situational panel interview: A field study. *J. Bus. Psych.* 13(2):143–156.

Page L, Siemroth C (2017) An experimental analysis of information acquisition in prediction markets. *Games Econom. Behav.* 101:354–378.

Page L, Siemroth C (2021) How much information is incorporated into financial asset prices? Experimental evidence. *Rev. Financial Stud.* 34(9):4412–4449.

Plott CR, Sunder S (1982) Efficiency of experimental security markets with insider information: An application of rational-expectations models. *J. Political Econom.* 90(4):663–698.

Plott CR, Sunder S (1988) Rational expectations and the aggregation of diverse information in laboratory security markets. *Econometrica* 56(5):1085–1118.

Qu R, Timmermann A, Zhu Y (2019) Do any economists have superior forecasting skills? Working paper, Indiana University, Bloomington.

Raven JC (1936) Mental tests used in genetic studies: The performance of related individuals on tests mainly educative and mainly reproductive. MSc thesis, University of London, London.

Samuelson PA (1965) Rational theory of warrant pricing. *Indust. Management Rev.* 6(2):13–39.

Schneider M, Porter D (2020) Effects of experience, choice architecture, and cognitive reflection in strategyproof mechanisms. *J. Econom. Behav. Organ.* 171:361–377.

Shleifer A (2000) *Inefficient Markets: An Introduction to Behavioural Finance* (OUP, Oxford, United Kingdom).

Singer T, Seymour B, O'Doherty J, Kaube H, Dolan RJ, Frith CD (2004) Empathy for pain involves the affective but not sensory components of pain. *Science* 303(5661):1157–1162.

Smith VL, Suchanek GL, Williams AW (1988) Bubbles, crashes, and endogenous expectations in experimental spot asset markets. *Econometrica* 56(5):1119–1151.

Snyder M, Gangestad S (1986) On the nature of self-monitoring: Matters of assessment, matters of validity. *J. Personality Soc. Psych.* 51(1):125–139.

Stanovich K (2011) *Rationality and the Reflective Mind* (Oxford University Press, Oxford, United Kingdom).

Stanovich K (2016) *The Rationality Quotient: Toward a Test of Rational Thinking* (MIT Press, Cambridge, MA).

Tetlock PE, Gardner D (2016) *Superforecasting: The Art and Science of Prediction* (Broadway Books, New York).

Toplak M, West R, Stanovich K (2011) The Cognitive Reflection Test as a predictor of performance on heuristics and biases tasks. *Memory Cognition* 39(7):1275–1289.

Tseëlon E (1992) Is the presented self sincere? Goffman, impression management and the postmodern self. *Theory Culture Soc.* 9(2):115–128.

Veiga H, Vorsatz M (2010) Information aggregation in experimental asset markets in the presence of a manipulator. *Experiment. Econom.* 13(4):379–398.