

**OVERVALUING ALGORITHMIC ADVICE:  
EVIDENCE FROM A STOCK PRICE  
FORECASTING EXPERIMENT**

Nobuyuki Hanaki  
Bolin Mao  
Tiffany Tsz Kwan Tse  
Wenxin Zhou

Secondly Revised May 2026  
Revised July 2025  
December 2024

The Institute of Social and Economic Research  
The University of Osaka  
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

# Overvaluing Algorithmic Advice: Evidence from a Stock Price Forecasting Experiment

Nobuyuki Hanaki\* Bolin Mao† Tiffany Tsz Kwan Tse‡ Wenxin Zhou§

May 26, 2026

## Abstract

This study investigates willingness to pay (WTP) for stock forecasting advice from algorithms, financial experts, and peers. In two incentivized forecasting experiments, participants purchased advice using an incentive-compatible mechanism and then decided how much to incorporate it into their forecasts. Participants assigned the highest WTP to algorithmic advice and relied on it as heavily as expert advice, despite its forecasting performance being no better than alternative sources. Consequently, participants overpaid for advice, especially algorithmic advice, whose realized benefits were insufficient to offset its cost. A second experiment shows that overpayment persists even after repeated opportunities to revise WTP with detailed feedback on advice quality and realized net benefits. The results suggest that individuals place excessive value on algorithmic advice perceived as sophisticated or credible, even when its realized economic value is limited. These findings highlight the importance of tools and disclosure policies that help individuals better assess the economic value of algorithmic advice.

**Keywords:** algorithms, Becker–DeGroot–Marschak mechanism, experts, financial market, forecasting

**JEL Classification Number:** C90, G1, G4, G17

---

\*Institute of Social and Economic Research, The University of Osaka, Japan and University of Limassol, Cyprus. E-mail: nobuyuki.hanaki@iser.osaka-u.ac.jp, ORCID ID: 0000-0003-3127-690X

†Graduate School of Economics, Hitotsubashi University, Japan. E-mail: meta.bolin.mao@gmail.com, ORCID ID: 0000-0001-7178-0284

‡Corresponding author: Institute of Social and Economic Research, The University of Osaka, Japan. E-mail: tiffany.econ@gmail.com, ORCID ID: 0000-0003-1614-2428

§A non-academic institution. E-mail: haruka\_ivory@outlook.com

**Supplementary Information** The online appendix is provided in a separate file.

**Acknowledgements** We gratefully acknowledge comments and suggestions from John Duffy, Syngjoo Choi, Angela Sutan, Fumio Ohtake, Takeshi Murooka, Peiran Jiao, participants at the East Asian Experimental Behavioral Economics Workshop (Seoul), CREST-CEFM Workshop (Paris), Prediction Markets Workshop (Paris), Workshop on Human-AI interactions (Osaka), and seminars at University of Limassol, Waseda, and Aoyama Gakuin. We thank Satsuki Yamada for her help in conducting the experiment at The University of Osaka.

**Funding** This project is financially supported by the Joint Usage/Research Center at Institute of Social and Economic Research, The University of Osaka, and Japan Society for the Promotion of Science KAKENHI Grant Numbers 18K19954, 20H05631, and 23H00055.

**Availability of data and materials** Data can be made available after publication

**Code availability** The code can be made available after publication.

#### **Declarations**

**Conflict of interest** Not applicable.

**Consent to participate** When registering for the experiment, participants gave their consent to participate.

**Consent for publication** Publication of the manuscript has been approved by all co-authors.

**Ethics approval** The experiment reported in this paper was approved by the Research Ethics Committee at the Institute of Social and Economic Research, The University of Osaka (No. 20231002, No. 20260102).

**Pre-registration** The experiment is pre-registered at [aspredicted.org #150365](https://aspredicted.org/#150365) (<https://aspredicted.org/smzt-dfjq.pdf>) and [#270138](https://aspredicted.org/#270138) (<https://aspredicted.org/sc39e8.pdf>).

# 1 Introduction

The financial advisory industry is undergoing a significant transformation with the rise of algorithmic advisors (OECD, 2019, 2024). Traditional human advisors, while offering personalized services, face challenges such as high fees, behavioral biases (Foerster et al., 2017), conflicts between client profits and employer profits (Hoechle et al., 2018), misconduct (Dimmock et al., 2018; Egan et al., 2019), and reliance on experience-driven decisions (Linnainmaa et al., 2021). In contrast, artificial intelligence (AI) advisors promise lower costs and data-driven recommendations, making them an attractive alternative for investors (D’Acunto et al., 2019). However, despite the increased adoption of algorithms, relatively little is known about the demand for algorithmic advice, how individuals value such advice relative to human expert advice, and whether individuals continue to demand advice even after observing that its cost outweighs its payoff benefits.

The discussions surrounding human-algorithm relationships have often focused on “algorithm aversion” (Dietvorst et al., 2015), wherein individuals underutilize algorithms after learning that they are imperfect, despite their superior performance compared to human peers. This behavioral trend has been observed in numerous studies, such as Burton et al. (2020), Chacon et al. (2022), Gill et al. (2024), and Prahl and Van Swol (2017) and has led to inquiries on how to improve the adoption of algorithm tools (see, for example, Dietvorst et al., 2018; Filiz et al., 2021; OECD, 2019). Scholars have examined whether the degree of algorithm aversion is dependent on the task (Castelo et al., 2019), the domain (Himmelstein and Budescu, 2023), or the incentives that individuals face (Greiner et al., 2025).

Perceived expertise also plays an important role in advice taking. Indi-

viduals are more likely to rely on advice when they perceive the source as competent or knowledgeable. Bonaccio and Dalal (2006) show that people take advice more seriously when the advisor is perceived to have greater expertise. Similarly, Meshi et al. (2012) find that participants place more weight on advice from human experts than from novices in monthly rent estimation tasks. These findings suggest that advice use may depend not only on actual accuracy, but also on perceived competence. In financial forecasting, both algorithms and human experts may be viewed as sophisticated information sources, even when their actual predictive performance is limited.

Recent studies present mixed results regarding preferences for algorithmic advisors versus human advisors in financial decision making. Önkal et al. (2009) find that participants rely less on advice from an algorithm than on advice from a human when forecasting stock prices. Yang et al. (2025) explore the role of human involvement in AI-generated financial advice and its effect on consumer decision making, finding that customers are more likely to follow the advice of human-AI collaboration than pure AI, particularly in the context of high-risk investments. Holzmeister et al. (2023) found that clients prefer to delegate investment decisions to algorithms, followed by experts with aligned incentives, and finally experts compensated with fixed fees. However, among those who are willing to delegate the decisions, the willingness to pay (WTP) to do so is similar across algorithms and experts with varying incentives.<sup>1</sup> Germann and Merkle (2023) suggest that participants prioritize returns but exhibit no strong preferences between human fund managers and investment algorithms when it comes to achieving those

---

<sup>1</sup>In Holzmeister et al. (2023), participants decide whether or not to delegate their investment decisions before submitting their WTP to delegate. WTP is elicited only for those who have decided to delegate.

returns. Logg et al. (2019) find that experienced professionals rely less on algorithmic advice than laypersons do. This suggests that reliance on algorithmic advice may depend on how individuals evaluate the algorithm's expertise relative to their own expertise and to other available information sources.

Most existing studies examining advice-taking behavior, including algorithm aversion and appreciation (Logg et al., 2019), employ the judge-advisor system (JAS) paradigm (Sniezek and Buckley, 1995). In this paradigm, researchers investigate the extent to which participants use the advice they receive for free.<sup>2</sup> In these studies, participants predict variables such as weight, medical conditions, song rankings, and stock prices. After making an initial prediction, they receive advice (predictions made by algorithms, experts, or peers) at no cost and are then asked to either adopt this advice or adjust their predictions accordingly. Reliance on the advice is measured by the frequency of adopting the provided recommendation or by the extent to which the final prediction deviates from the initial one in line with the advice received.

However, in many real-life situations, such as investment decisions and medical consultations, advice is costly. Individuals must therefore decide not only whether to use advice, but also how much it is worth paying for. This distinction is important because a source that is perceived as expert or sophisticated may attract high WTP even when its actual performance does not justify the cost. Thus, rather than solely examining reliance on free advice, it is essential to understand how much individuals are willing

---

<sup>2</sup>Jussupow et al. (2020) summarize 29 publications containing 84 distinct experimental studies related to algorithm aversion. These studies employed the JAS paradigm with free advice. Mahmud et al. (2022) examine 80 empirical studies related to algorithm aversion, while the study by Sutherland et al. (2016) is related to the cost of advice.

to pay for advice from different information sources, how much they use the advice after purchasing it, and whether the benefits of advice are large enough to offset its cost. To the best of our knowledge, with the exception of Holzmeister et al. (2023), few previous studies have investigated participants' WTP for advice from different sources.

To address this gap in the literature, we ask: How much are participants in our stock price forecasting experiment willing to pay for advice in the form of forecasts made by an algorithm, experts, and other participants similar to themselves? Furthermore, to what extent do participants rely on the additional information (forecasts) provided by these sources? Do the benefits of advice justify its cost? How do their willingness to pay and degree of reliance change between when they have no prior experience in the task and after gaining some experience and observing both their own performance and the advisor's performance?<sup>3</sup>

To investigate these questions, we conduct two stock price forecasting experiment akin to those by Bao et al. (2023, 2025) and Tse et al. (2024).<sup>4</sup>

In Experiment 1, participants submit their initial forecasts and have the option of receiving advice in the form of forecasts made by an algorithm (Tse

---

<sup>3</sup>We assume in this experiment that participants do not know the advisor's true performance *ex ante*.

<sup>4</sup>Traditional finance theories, such as the efficient market hypothesis, suggest that prices fully reflect all available information, rendering prediction attempts futile. If participants agree with this hypothesis, they should not be willing to pay anything for advice from any of the information sources we consider. However, not everyone is aware of or agrees with the efficient market hypothesis. For example, the so-called technical analyses rely on charts to determine investment strategies (Lo et al., 2000; Lo and Hasanhodzic, 2011). In addition, existing experimental studies such as Bloomfield and Hales (2002) and Bao et al. (2023, 2025) show that many participants in stock price forecasting experiments consider prices or at least their direction of change, to be predictable even when they are explicitly told that the underlying time series follow a random walk. Thus, we consider investigating how participants value advice from various sources in our stock price forecasting experiment to be a valuable exercise in understanding how an algorithm is perceived compared to human experts.

et al., 2024), experts (Bao et al., 2025), or student participants similar to themselves (Bao et al., 2025) with a fee before finalizing their forecasts.<sup>5</sup> We employ a between-subjects design so that each participant has the possibility of receiving advice from a single source without being aware of the other options. We use the incentive-compatible mechanism proposed by Becker et al. (1964) to elicit participants' WTP for the advice. For each participant, WTP is elicited twice to examine how it evolves with their experience.

Experiment 2 focuses on WTP for advice from the algorithm and experts, and elicits WTP for advice more frequently to better understand the mechanism behind the results we obtained from Experiment 1, including the role played by a detailed description of the information source and limited opportunity for learning.

To the best of our knowledge, our design is unique in that it investigates advice-taking behavior by combining the incentive-compatible method to elicit WTP for advice and the JAS paradigm to study the extent to which participants use the advice for which they have paid. This combined approach allows us to capture a complete picture from valuation to action and thus link the economic evaluation of advice by individuals with their actual usage. In doing so, it offers novel insights into the practical impact of paid advice.

In Experiment 1, participants perceived the advice from the algorithm to be as good as, if not better than, that from experts and showed a high WTP and incorporated it into their final decision when possible. The advice from their peers, on the other hand, was perceived to be the worst with

---

<sup>5</sup>Details of the algorithm can be found in the Online Appendix in Tse et al. (2024). The advice of experts and students are based on the experimental results in Bao et al. (2025).

the lowest WTP among the three, and was relied upon the least. However, the performance of these advice did not match their perceived values. As a result, participants overpaid for advice, especially for algorithmic advice whose quality was the lowest in our experiment. Furthermore, limited feedback provided during the experiment was not sufficient for participants to update their beliefs about advice quality and adjust their WTP to the appropriate level.

While the overvaluation of advice and the absence of appropriate adjustment based on the experience and feedback sound rather alarming, they may have been caused by certain design features of Experiment 1. For example, the detailed description of the algorithm and experts provided to participants in Experiment 1 may have given the impression of high competence and sophistication of these information sources. Furthermore, the feedback was not detailed and frequent enough for participants to learn better.

Experiment 2 was designed to address these concerns. In particular, it aims to investigate the impact of the description of the information sources on initial WTP, together with effects of other design features of Experiment 1, as well as more frequent and detailed feedback on learning about advice quality (and thus, subsequent updating of WTP). Furthermore, it also investigates whether characteristics of the stock-price series influence the WTP for advice.

The results show that detailed descriptions, and thus potential impression of competence and sophistication they may have provided, do not significantly increase WTP for either algorithmic or expert advice. We also found that participants update WTP mainly in response to improvements in forecasting accuracy rather than net benefit gains, i.e., not carefully considering the cost-benefit of receiving these advice. As a result, participants continue

to overpay, particularly for algorithmic advice, even with more frequent and detailed feedback.

These results underscore the need for tools and policies that help individuals evaluate the actual value of advice, rather than relying solely on the perceived expertise or sophistication of the advice source. From a regulatory perspective, transparent disclosures of advice performance, costs, and realized net benefits may help individuals make better decisions about whether advice is worth purchasing. This issue is particularly relevant in high-stakes domains such as finance, where both algorithmic and human expert advice may be perceived as credible and sophisticated, even when the realized benefits are not large enough to justify the cost.

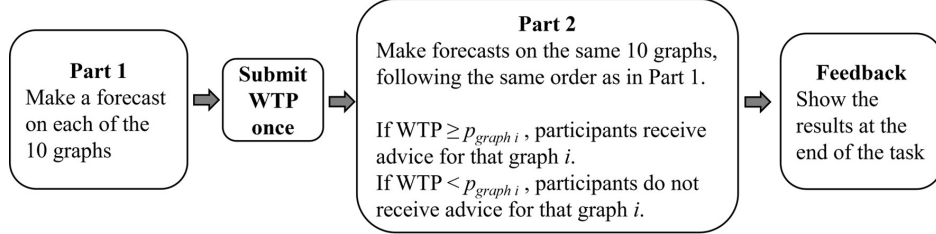
The remainder of the paper is structured as follows: Section 2 presents the design of Experiment 1 and hypotheses; Section 3 summarizes its results, together with a discussion; Section 4 presents the design of Experiment 2; Section 5 presents and discusses its results; and Section 6 concludes the paper.

## **2 Design of Experiment 1**

Experiment 1 is designed to examine how participants value, use, and learn about costly forecasting advice from different information sources. Participants first make forecasts without advice, then decide how much they are willing to pay for advice, and finally determine how much to incorporate the purchased advice into their final forecasts.

We employ three treatments that vary the source of advice, algorithm, experts, and students. In each treatment, there are two main tasks (Tasks

Figure 1: Flow of Task 1 and Task 2



1 and 2) that are identical in structure. Each task consisted of two parts (Parts 1 and 2). Figure 1 illustrates the flow of these tasks.

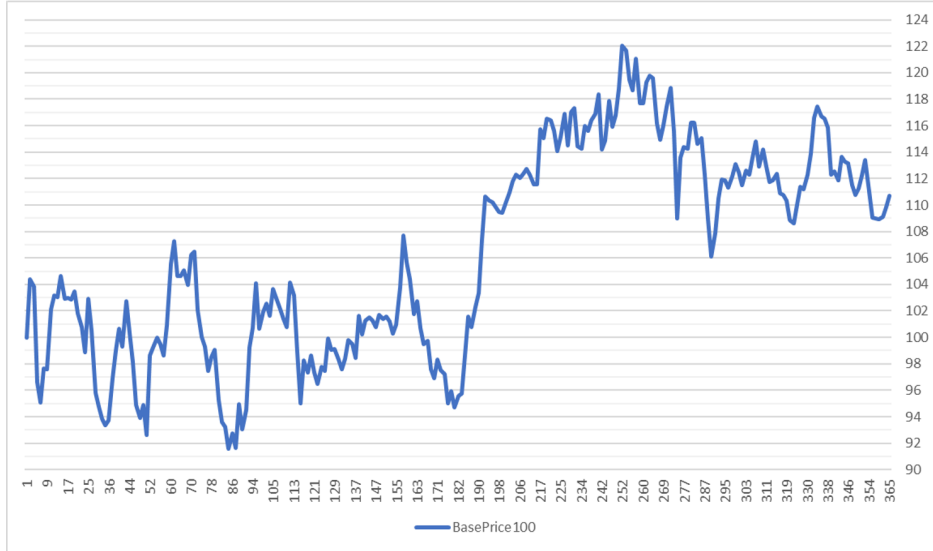
In part 1, participants faced a series of 10 questions. In each question, participants viewed a time-series graph showing 12 months of closing prices for a randomly selected stock from the S&P 500. The starting date for each time series was randomly chosen from a range between January 1, 2008, and June 30, 2018, and participants were not informed of either the stock names or the starting dates. Each time series was normalized so that the starting price equaled 100 (see Figure 2 for an example).

In each question, participants were asked to forecast the stock’s closing price 30 days after the final price shown in the graph within a 40-second time limit.<sup>6</sup> Participants completed forecasts for all 10 questions without receiving any feedback. These 10 questions were presented in the same order across participants. This forecasting task is identical to those in Bao et al. (2023, 2025). Participants did not receive any feedback on their performance before the end of part 2.

After part 1, participants submitted their WTP for advice on a scale from 0.0 to 10.0 in increments of 0.1 following an incentive-compatible BDM

<sup>6</sup>If participants did not submit their forecast within the time limit, the program automatically submitted a value of 0. In the Results section, we excluded data from participants who did not submit within the time limit. Specifically, we excluded 3 WTP observations in Task 1: 1 in each treatment.

Figure 2: Sample graph



mechanism (Becker et al., 1964). They were informed that in part 2, they would once again make forecasts for the same 10 time series graphs in the same order as in part 1.

Let  $WTP_t^i$  denote participant  $i$ 's WTP in Task  $t$ . For each question  $q$  in Part 2, the price of advice,  $p_q^i$ , was independently drawn for each participant from a uniform distribution between 0.1 and 10.0 in increments of 0.1. If  $WTP_t^i \geq p_q^i$ , participant  $i$  received the advice for question  $q$  and paid  $p_q^i$ . Otherwise, no advice was provided and no payment was required.<sup>7</sup> There was a 60-second time limit for submitting a WTP.<sup>8</sup>

<sup>7</sup>Thus, unlike Holzmeister et al. (2023), who elicit WTP only among those who delegate, we elicit WTP for all participants. It is well known that participants have difficulty understanding the BDM mechanism (Cason and Plott, 2014). To encourage participants to truthfully report their WTP, in our experiment, we instruct participants to answer a series of questions to determine their WTP. See slide 14 of the instructions in Online Appendix E.

<sup>8</sup>If participants did not submit their WTP within the time limit, the program will automatically submit a value of 0. In the Results section, we excluded data from participants who did not make submissions within the specified time limit. We excluded a total of 30 initial-forecast observations in Task 1: 9 each in Algorithm and Expert, and 12 in Student; 5 final-forecast observations in Task 1: 3 in Algorithm, 1 each in Expert and Student; 4 initial-forecast observations in Task 2: 3 in Algorithm and 1 in Student; and 2 final-forecast observations in Task 2: 1 each in Algorithm and Expert.

In part 2, participants submitted forecasts on the same set of 10 questions, one by one, as in part 1. For each question, they were informed of the price of the advice,  $p_q^i$ , and whether they receive it. If not, they were only reminded of their forecast from part 1 before submitting their forecast. No performance feedback was provided during part 2.

Upon completing part 2, participants received feedback, including the total points accumulated across both parts 1 and 2, as well as the points awarded for each question in both parts. Additionally, they were reminded of their WTP and the price of advice in each question. See Screen 6 in Online Appendix H for a screenshot after completing Task 1.

Participants were rewarded based on the accuracy of their forecasts. The reward of participant  $i$  for question  $q$ ,  $r_q^i$ , is calculated as follows, where  $(\cdot)^+$  denotes  $\max(\cdot, 0)$ :

$$r_q^i = \left( 20 - \left| \frac{\text{Forecast}_q^i - \text{True value}_q}{\text{True value}_q} \right| \times 100 \right)^+ .$$

If a participant's forecast exactly matched the realized price, they earned 20 points. For every percentage point difference between the forecast and the true value, 1 point was deducted. Participants received no points if their forecast deviated from the true value by more than 20%.<sup>9</sup>

To avoid negative payoffs and standardize the initial endowment, each participant was allocated 10 points per question to cover the cost of advice. The points accrued by participant  $i$  for each question in parts 1 and 2,  $\pi_q^i$ , were  $\pi_q^i = r_q^i + 10$  if they did not receive advice and  $\pi_q^i = r_q^i + 10 - p_q^i$ ,

---

<sup>9</sup>We followed Bao et al. (2025) in using a linear scoring rule, rather than the more commonly used quadratic scoring rule, which penalizes larger deviations more heavily, in order to provide participants in the current experiment with the same incentive scheme faced by the sources of advice.

otherwise.

The final payoff was determined by summing the points earned across the 10 questions from a randomly selected part, which could be any of the following: part 1 of Task 1, part 2 of Task 1, part 1 of Task 2, or part 2 of Task 2. The exchange rate for the experiment was set at 1 point = 6 JPY.

## 2.1 Treatments

We employed a between-subjects design with three treatments that varied the source of advice: algorithm, expert, and student. The algorithm-based advice consisted of forecasts generated by the algorithm used in Tse et al. (2024). Expert advice consisted of the average forecasts submitted by Certified Member Analysts (CMAs) of the Securities Analysts Association of Japan in Bao et al. (2025). Student advice consisted of the average forecasts submitted by students at the University of Osaka in the same study.<sup>10</sup> We denote the algorithm treatment by **Algorithm**, the expert treatment by **Expert**, and the student treatment by **Student**.

Participants in **Algorithm** were informed that forecasts were generated by a machine-learning algorithm trained on historical stock-price data. Participants in **Expert** and **Student** were told that forecasts came from CMAs and students of the same university, respectively, who had previously participated in a similar experiment. Full description provided in the instruction are provided in Table 1.

Participants were not provided with any information about the perfor-

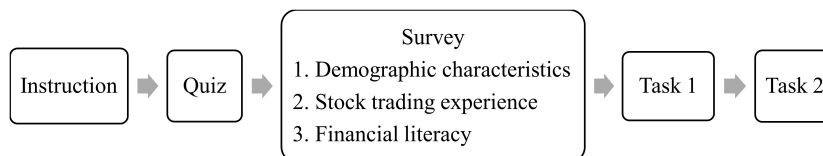
---

<sup>10</sup>Instead of using average forecasts, an alternative design would have been to use forecasts from a randomly selected CMA or student who had previously participated in the experiment. We opted for the current design to reduce the uncertainty associated with variation in individual forecasts, which could otherwise have further reduced participants' willingness to pay for advice from experts and students.

Table 1: Description of Information Sources

<b>Algorithm</b>
<p>The additional information provided here is a stock price forecast submitted, based on the same stock price chart used in this experiment, by an algorithm crafted to predict future stock prices.</p> <p>This algorithm makes a future stock price forecast by learning the historical stock price information, from January 1st, 2000, or from the initial public offering (IPO) day to January 1st, 2020, of 83 target companies ranked at the top in their capital market sectors (i.e., Basic Materials, Consumer Goods, Healthcare, Services, Utilities, Conglomerates, Financial, Industrial Goods, Technology).</p>
<b>Expert</b>
<p>The additional information provided here is the average of stock price predictions submitted by 198 experts (CMAs) who participated in a similar experiment in the past, based on the stock price chart used in this experiment, for stock prices 30 days ahead.</p> <p>CMA stands for Certified Member Analyst of the Securities Analysts Association of Japan. The CMA is a qualification granted to those who have taken the prescribed training courses and passed the examinations based on these courses and fulfilled certain requirements; it is a sign of expertise in the fields of finance and investments. The CMAs' investment valuations are based on the calculation of the corporate value of an investment and the forecasting of its future value. The particularly crucial knowledge required can be broadly categorized into three main areas: evaluation of investment data, decision making on investment policies, and construction and management of portfolios.</p>
<b>Student</b>
<p>The additional information provided here is the average of stock price predictions submitted by 233 students at The University of Osaka who participated in a similar experiment conducted in the past, based on the stock price chart used in this experiment, for stock prices 30 days ahead.</p>

Figure 3: Experiment flow



mance of the advice sources. This design allows us to examine participants' subjective perceptions of different advice sources in Task 1 and how these perceptions evolve after gaining experience in the task.

## 2.2 Procedure

Figure 3 illustrates the flow of the experiment. Initially, participants read the general instructions individually on a computer screen; those instructions outlined the primary tasks and objectives of the experiment. While participants were informed they would have the opportunity to receive advice, the specific source of this advice was not disclosed at this stage. Detailed instructions are available in Online Appendix E.

After reading the instructions, participants completed a comprehension quiz to ensure they understood the experiment. This quiz consisted of five questions which were identical across all three treatments. Feedback on each answer was provided immediately after each question.

After the quiz, participants completed a survey consisting of three sections: demographic characteristics, stock trading experiences, and financial literacy.<sup>11</sup> Upon completing the survey, participants advanced to the main

---

<sup>11</sup>In the demographic section, participants provided information on their gender, academic grade, and whether they were majoring in a science-related field. For stock trading experience, participants indicated any past involvement in stock investments. Those with prior experience were asked about the duration of their activities, the types of stocks they invested in, their returns, the sources of information they consulted during investment decision making, and their risk management approaches in stock investments. In the financial literacy section, participants responded to 12 questions adapted from Fernandes

experimental tasks, Tasks 1 and 2, which followed the same order for all participants.

The experiment was programmed using oTree (Chen et al., 2016) and conducted at the experimental laboratory of the Institute of Social and Economic Research (ISER) at The University of Osaka. A total of 17 sessions took place between October 2023 and June 2024, with each session dedicated to a single treatment. We recruited 311 students from the ISER’s ORSEE database (Greiner, 2015), and each participant attended only one session.<sup>12</sup> See Table B.1 in Online Appendix B for a summary of participants’ characteristics. They received a participation fee of 500 JPY for completing the 90-minute session, with the potential to earn an additional reward of up to 1800 JPY based on their forecasting accuracy.<sup>13</sup> All sessions were conducted by the same experimenter to maintain consistency. Table 2 presents a summary of the sessions.<sup>14</sup>

## 2.3 Hypotheses

Research on advice taking suggests that individuals rely more heavily on advice perceived as competent and knowledgeable (Bonaccio and Dalal, 2006; et al. (2014).

---

<sup>12</sup>Participants in this experiment did not participate in the previous experiment used to generate the student advice.

<sup>13</sup>Participants provided online consent by clicking an approval button prior to registering for the experiment.

<sup>14</sup>The number of sessions differs from our pre-registration because we were only able to recruit a small number of participants in some sessions. Given that this is an individual decision making task, we prioritized reaching the target total number of participants rather than adhering strictly to the pre-registered number of sessions. The sample size was determined based on results from our pilot study, with a target of 95 participants aimed at detecting differences between algorithmic and expert advice. We included data from the pilot sessions in the main analysis, as the sampling frame and experimental methodology were consistent across both. As a robustness check, we conduct the analysis excluding the pilot data. The results are similar to those obtained using the full sample. The detailed results are reported in Online Appendix D.

Table 2: Summary of treatments

Treatments	Algorithm	Expert	Student
No. of participants	106	103	102
No. of sessions	6	5	6
Duration (min)	90	90	90
Avg. payment (JPY)	1884	1907	1921

Meshi et al., 2012). Consistent with this view, both algorithmic and expert advice may be expected to receive higher valuations than advice provided by non-experts such as other students.

At the same time, prior literature provides mixed evidence regarding whether individuals value algorithmic advice more or less than expert advice. Studies on algorithm aversion suggest that individuals may distrust algorithms relative to human experts after observing that algorithms are imperfect (Dietvorst et al., 2015; Jussupow et al., 2020). Consistent with this view, Önkal et al. (2009) find that participants rely less on advice framed as coming from algorithms than on otherwise identical advice framed as coming from human experts in stock-price forecasting tasks. Other studies, however, suggest that algorithms may also be perceived as sophisticated and competent information sources, particularly in data-intensive domains such as financial forecasting (Logg et al., 2019). Recent evidence further suggests that descriptions emphasizing technical sophistication, learning ability, or complexity can increase the perceived competence and acceptance of algorithmic systems (Chacon et al., 2022; Lehmann et al., 2022; Chevrier et al., 2024; Williamson et al., 2025).

Our preregistered hypotheses follow the earlier algorithm-aversion literature and predict that participants will value expert advice more highly than algorithmic advice, while both expert and algorithmic advice will be valued

more highly than student advice. Formally, we test the following hypotheses:

**Hypothesis 1.** Participants' WTP for advice is ordered as follows: Expert > Algorithm > Student.

Upon completing Task 1, participants could assess their performance in both parts 1 and 2 of the task. By comparing their initial performance (before accessing advice in part 1) with their performance after adopting advice in part 2, participants could evaluate any improvements attributable to that advice. If participants observed that the advice contributed positively to their performance in Task 1, we anticipate they would be more inclined to increase their WTP for advice in Task 2. Accordingly, we propose the following hypothesis:

**Hypothesis 2.** Participants' WTP for advice in Task 2 increases when they experience improved forecasting performance after adopting advice in Task 1.

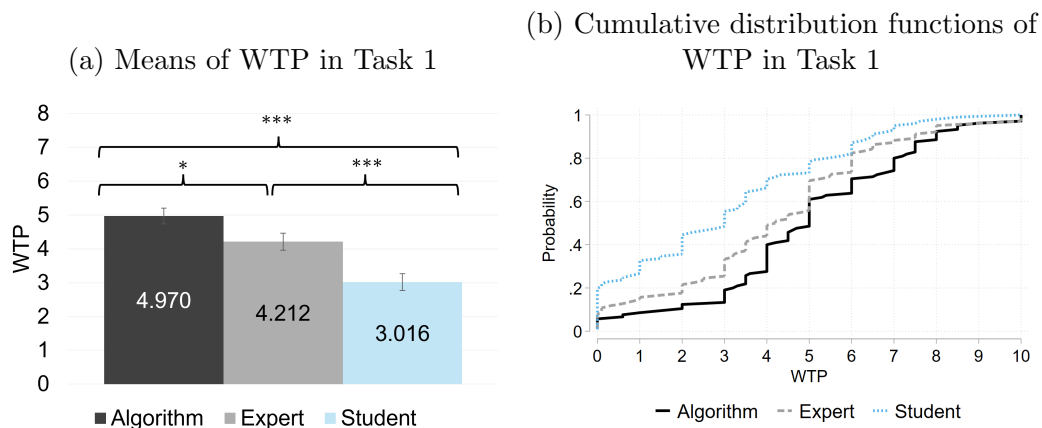
## 3 Results of Experiment 1

We first examine participants' WTP for advice in Task 1, followed by their utilization of advice and the resulting payoff consequences. We then investigate how participants updated their WTP in Task 2 after observing their experiences in Task 1.

### 3.1 WTP for advice in Task 1

When participants submitted their WTP for advice in Task 1, they had not yet received any feedback regarding either their own forecasting performance

Figure 4: Willingness to pay (WTP) in Task 1



*Notes:* We regressed WTP on three treatment dummies using an OLS regression model with robust standard errors clustered at the session level. Pairwise comparisons of estimated dummy coefficients were conducted using F-tests with Bonferroni-adjusted  $p$ -values. Results from two-sided tests are reported. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively.<sup>15</sup>

or the performance of the advice source.

Figure 4 illustrates participants' WTP for advice. Panels (a) and (b) present the means and cumulative distributions of WTPs for the three treatments.

The mean WTPs to receive advice in Task 1, along with their standard deviations, were 4.970 (2.396) for **Algorithm**, 4.212 (2.523) for **Expert**, and 3.016 (2.503) for **Student**. Mean WTP in **Algorithm** was marginally significantly higher than in **Expert** (two-sided test:  $p = 0.095$ ). Mean WTP in both **Algorithm** and **Expert** was significantly higher than in **Student** (one-sided test:  $p < 0.001$  and  $p = 0.007$ , respectively).<sup>16</sup>

Panel (b) of Figure 4 reveals substantial variation in WTP within each

<sup>15</sup>This analysis was partly pre-registered. We additionally cluster robust standard errors at the session level because each session included only one treatment, which may confound treatment effects with session-specific factors. Also, cumulative distribution functions of WTP in Task 1 was not pre-registered.

<sup>16</sup>We also conducted a joint test of equality across the three treatments, rejecting the null hypothesis that  $\text{Algorithm} = \text{Expert} = \text{Student}$  ( $p < 0.001$ ). This indicates that WTP differs significantly across the three treatments.

treatment. Each treatment includes participants with a WTP of 0. The proportion of participants with a WTP of 0 is highest in **Student** (19.80%), more than twice that observed in **Expert** (7.84%) and **Algorithm** (5.71%). The maximum WTP of 10 was observed for both **Algorithm** (3.81%) and **Expert** (3.92%), whereas it was less than 1% in **Student**.<sup>17</sup>

The distribution of WTP in **Student** lies to the left of that in **Expert**, which itself lies to the left of that in **Algorithm**. We conducted a two-sample Kolmogorov–Smirnov test to assess distributional differences in WTP between treatments. The WTP distributions in **Algorithm** and **Expert** were not significantly different ( $p = 0.121$ ), but they were both significantly different from that in **Student** ( $p < 0.001$  and  $p = 0.006$  vs. **Algorithm** and vs. **Expert**, respectively).

**Result 1.** The WTPs for advice are ordered as follows: **Algorithm**  $>^*$  **Expert**  $>^{***}$  **Student**.

Contrary to our preregistered hypothesis based on the algorithm-aversion literature, participants valued algorithmic advice at least as highly, if not better than, as expert advice.

---

<sup>17</sup>We investigated whether personal characteristics correlate with WTP in Task 1 across different information sources, as this measure may reflect participants’ self-assessed competence relative to the advice source. Table B.1 in Online Appendix B presents the results of the OLS regressions. For algorithmic advice and student advice, none of the observable personal characteristics exhibits a significant correlation with WTP in Task 1. By contrast, for expert advice, participants with stock-trading experience have significantly lower WTP ( $p = 0.013$ ), while participants with higher financial literacy scores have significantly higher WTP ( $p = 0.031$ ). Other characteristics are not significantly correlated with WTP.

## 3.2 Utilization of advice

We next examine the extent to which participants incorporated advice into their final forecasts after receiving it.<sup>18</sup> We measure advice utilization using the *shift rate*, a standard measure in the advice-taking literature (Harvey and Fischer, 1997; Snizek et al., 2004). The shift rate captures the extent to which participants move their final forecast toward the advice, relative to their initial forecast. A shift rate of 0 indicates that the participant did not adjust the initial forecast toward the advice, while a shift rate of 1 indicates full adoption of the advice. Our analysis focuses on observations with shift rates between 0 and 1.<sup>19</sup>

The results show that participants relied substantially on both algorithmic and expert advice.<sup>20</sup> The average shift rate was approximately 0.8 in **Algorithm** and **Expert**, and the difference between these two treatments was not statistically significant. By contrast, participants relied less on advice from peers. The average shift rate in **Student** was 0.569 in Task 1, significantly lower than the shift rates in the other two treatments. Thus, participants not only assigned lower WTP to peer advice, but also relied on it less after receiving it.

We also examine whether advice utilization is related to participants' WTP. If participants are more willing to pay for advice that they expect to

---

<sup>18</sup>The analyses reported in this section were not pre-registered. Details on the results about utilization of advice are provided in Online Appendix C.3.

<sup>19</sup>This restriction was not specified in the preregistration. Shift rates below 0 or above 1 capture movements away from, or beyond, the advice and therefore reflect different behavioral responses. We report these cases separately in the Online Appendix C.3. We report the results for the shift rate without imposing the restriction in Online Appendix C.3.1. Approximately 10% of the observations have shift rates outside the range from 0 to 1. The shift rates across treatments in Task 1 show a pattern similar to the results obtained when restricting the shift rate to values between 0 and 1.

<sup>20</sup>We regressed the shift rate on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level.

be accurate, then higher WTP may also be associated with greater reliance on the advice after it is received. However, we find no significant relationship between WTP and the shift rate in any treatment. These results suggest that participants' valuation of advice and their subsequent reliance on it are not strongly linked. Participants may therefore be willing to pay more for advice from certain sources, such as algorithms or experts, without relying more heavily on the advice once it is received.

**Result 2.** Participants were equally likely to incorporate advice in **Algorithm** and **Expert** when they receive it. Reliance on advice was significantly lower in **Student**.

### 3.3 Impacts of advice on performance

So far, we have examined participants' WTP for advice from three different information sources and their use of that advice upon receiving it. We now examine how the use of advice affected participants' forecasting performance and realized payoffs.<sup>21</sup>

The average performance of the advice received by participants differs across the three treatments. Advice received in **Algorithm** performed worse than advice received in **Expert** and **Student**, both in terms of the accuracy error rate and performance points. The performance of advice received is similar between **Expert** and **Student**.<sup>22</sup>

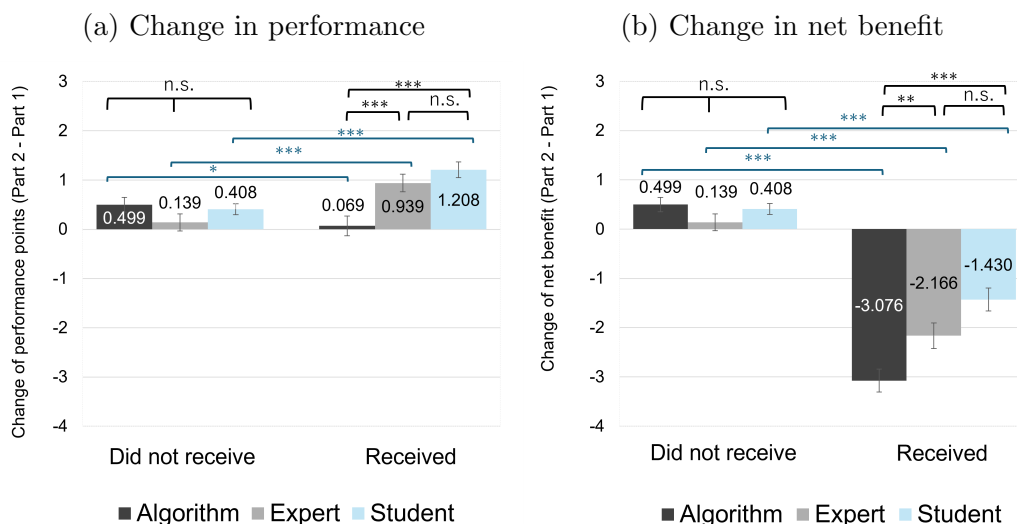
Participants lacked information about the relative performance of the advice sources when forming their valuations. They nevertheless assigned

---

<sup>21</sup>The analyses reported in this section were not pre-registered. We decided to include them based on feedback received during seminars and conferences.

<sup>22</sup>Details on the performance of advice are provided in Online Appendix C.1 and C.2.

Figure 5: Change in performance and net benefit in Task 1



*Notes:* We regressed the change in performance points and net benefits (part 2 – part 1) in task 1 on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level. Pairwise comparisons of the estimated dummy coefficients were performed using F-tests with Bonferroni-adjusted  $p$ -values. We also regressed these changes on an advice-receiving dummy in each treatment using an OLS regression model with robust standard errors clustered at the participant level. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively, while n.s. denotes that the difference is not statistically significant at the 0.1 level.

the highest WTP to the lowest-performing source: algorithmic advice. This may have negatively affected both the accuracy of their final forecasts in part 2 and the payoffs they ultimately received.

Figure 5 displays the average change, from part 1 to part 2, in performance measured by points earned (the left panel) and in payoff measured by net benefit (the right panel) in three treatments.<sup>23</sup> Within each panel, the average change when participants did not receive advice is shown on the left, while when they received advice is shown on the right.

The left panel of Figure 5 shows that when participants did not receive advice, performance improved only slightly from part 1 to part 2, with mean

<sup>23</sup>The results regarding the change in forecasting accuracy measured by the absolute percentage errors are provided in Online Appendix C.6.

gains below one point in all treatments and no significant differences across treatments.

When participants received advice in **Expert** or **Student**, the performance improvement was significantly greater than when they did not receive advice. The mean performance increases were 0.939 points for **Expert** and 1.208 points for **Student**. However, when participants received advice in **Algorithm**, performance improved by an average of only 0.069 points, which was less than the 0.499-point improvement observed when they did not receive it. This difference was marginally significant ( $p = 0.075$ ). These improvements should not be attributed solely to advice, as participants also improved slightly even without receiving advice.

The right panel of Figure 5 shows that despite modest improvements in forecasting performance, the gains were generally too small to offset the cost of advice. As a result, participants who purchased advice experienced negative average changes in net benefits in all treatments. The losses were largest in **Algorithm** ( $-3.076$ ), followed by **Expert** ( $-2.166$ ) and **Student** ( $-1.430$ ).

**Result 3.** Participants overpaid for advice, particularly for algorithmic advice whose performance was the lowest in our experiment.

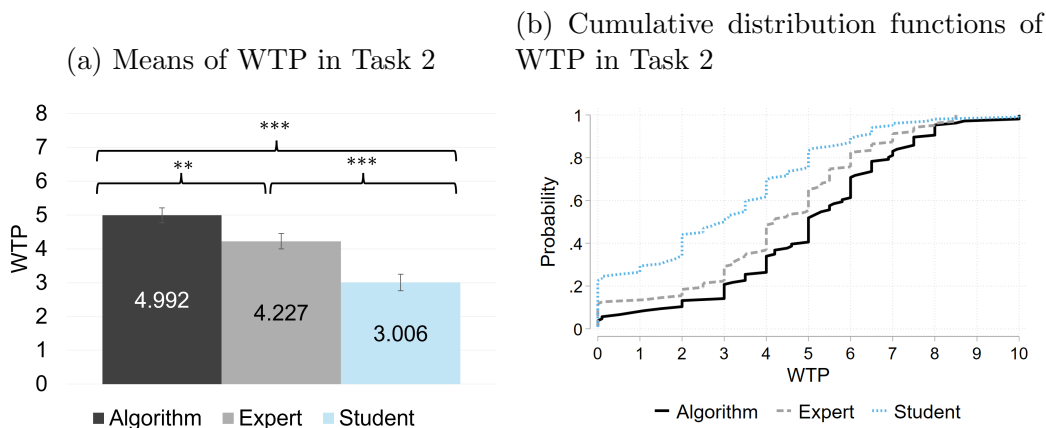
In fact, the maximum WTP required to achieve an average non-decrease in net benefit was 0 for **Algorithm**, 1.942 for **Expert**, and 1.110 for **Student**. These thresholds imply that average WTP exceeded the level required for advice to improve expected net benefits in all treatments.<sup>24</sup>

This overvaluation of advice in Task 1 may reflect participants' initially optimistic beliefs regarding advice quality. How did participants update their

---

<sup>24</sup>Details on the maximum WTP required to achieve an average non-decrease in net benefit are provided in Online Appendix C.4.

Figure 6: WTP in Task 2



*Notes:* We regressed WTP on three treatment dummies using an OLS regression model with robust standard errors clustered at the session level. Pairwise comparisons of the estimated dummy coefficients were performed using F-tests with Bonferroni-adjusted  $p$ -values. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively.<sup>25</sup>

WTP based on observing both their own performance and any improvement in their performance after adopting the advice in Task 1?

### 3.4 Effect of experience

Figure 6 shows WTP for advice in Task 2 across the three treatments. The ordering of both the mean values (Panel a) and the cumulative distributions of WTP (Panel b) in Task 2 closely resembles the patterns observed in Task 1.

Beyond aggregate treatment differences, we are interested in how individual participants updated their WTP between Tasks 1 and 2. Specifically, we aim to understand whether participants' WTP shifted between Tasks 1 and 2 and, if so, how these changes corresponded to their experiences in Task 1. According to Hypothesis 2, a change in WTP between Tasks 1 and 2 should be positively correlated with the improvement in performance resulting from

<sup>25</sup>This analysis was not pre-registered. The cumulative distribution functions of WTP in Task 2 were also not pre-registered.

adopting advice in Task 1.

To test this hypothesis, we run a linear regression where the dependent variable  $\Delta WTP^i = WTP_2^i - WTP_1^i$  is the change in each participant’s WTP between two tasks. Our main explanatory variable is the average improvement in performance obtained from advice in Task 1:  $\Delta\text{Mean performance}^i = \text{avg}(r_{q+10}^i - r_q^i)$ .<sup>26</sup>

Additionally, in place of  $\Delta\text{Mean performance}^i$ , we consider  $\Delta\text{Net benefit}^i$  as an alternative independent variable. This variable represents the increase in the payoff from part 1 to part 2 in Task 1, net of the price paid for advice.<sup>27</sup>

Table 3 presents a summary of the regression results, with a participant serving as the unit of observation. We restricted the analysis to participants who received advice at least once. Models (1) to (3) consider  $\Delta\text{Mean performance}$  in Task 1 and the number of pieces of advice received in Task 1 as independent variables, while models (4) to (6) consider  $\Delta\text{Net benefit}$  in Task 1 as the independent variable.<sup>28</sup>

Regarding the effect of  $\Delta\text{Mean performance}$  in Task 1, the estimated coefficients are positive and significant only for **Algorithm**. These results suggest that participants responded differently to performance improvements depending on the advice source.

From an economic perspective, however, the key question is whether the improvement in forecasting performance is large enough to justify the cost of advice. This aspect is examined in models (4) through (6), where  $\Delta\text{Net ben-}$

---

<sup>26</sup>Recall that questions are ordered so that participants make forecasts based on the same time series forecast in question  $q$  of part 1 and question  $q + 10$  of part 2. The average is taken over those  $q$  such that participant  $i$  received the advice in  $q + 10$ .

<sup>27</sup> $\Delta\text{Net benefit}^i = \text{avg}(\pi_{q+10}^i - \pi_q^i) = \text{avg}(r_{q+10}^i - r_q^i - p_{q+10}^i)$ , where the average is taken over those  $q$  such that participant  $i$  received the advice in  $q + 10$ .

<sup>28</sup>The analyses of models (1) to (6) were partly pre-registered, which we additionally control for the number of pieces of advice received in Task 1 and cluster robust standard errors at the session level.

Table 3: Changes in WTP from Task 1 to Task 2

VARIABLES	(1) Algorithm $\Delta$ WTP	(2) Expert $\Delta$ WTP	(3) Student $\Delta$ WTP	(4) Algorithm $\Delta$ WTP	(5) Expert $\Delta$ WTP	(6) Student $\Delta$ WTP
$\Delta$ Mean performance in Task 1	0.286** (0.110)	0.120 (0.110)	-0.117 (0.086)			
$\Delta$ Net benefit in Task 1				0.305* (0.127)	0.262*** (0.051)	0.080 (0.064)
Number of received advice in Task 1	-0.396*** (0.039)	-0.576*** (0.124)	-0.313** (0.110)	-0.294*** (0.043)	-0.476*** (0.092)	-0.296** (0.098)
Constant	2.094*** (0.165)	2.478** (0.595)	1.176** (0.356)	2.389*** (0.223)	2.548*** (0.392)	1.057** (0.365)
Obs.	97	84	73	97	84	73
Cluster at session level	6	5	6	6	5	6
$R^2$	0.312	0.388	0.206	0.346	0.442	0.202

Notes: a) In models 1–3, we regressed  $\Delta$ WTP on  $\Delta$ Mean performance from graphs with advice using an OLS regression model with robust standard errors clustered at session level.  $\Delta$ WTP is calculated as WTP in Task 2 minus WTP in Task 1.  $\Delta$ Mean performance is calculated as the mean performance from graphs with advice in part 2 of Task 1 minus the mean performance from graphs with advice in part 1 of Task 1. Performance ranges from 0 to 20, based on the points obtained in each question. In models 4–6, we regressed  $\Delta$ WTP on  $\Delta$ Net benefit from graphs with advice using an OLS regression model, with robust standard errors.  $\Delta$ Net benefit is calculated as the mean net benefit from graphs with advice in part 2 of Task 1 minus the mean net benefit from graphs with advice in part 1 of Task 1. The net benefit in each graph in part 1 equals the rewards in each graph plus 10 points. The net benefit in each graph in part 2 equals the rewards in each graph plus 10 points minus the price of advice. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively.

b) The unit of observation is the participant who received at least one piece of advice in Task 1. The total number of observations is the number of participants who received at least one piece of advice in Task 1 in each treatment.

c) Numbers in parentheses represent robust standard error clustered at session level.

efit in Task 1, which accounts for the cost of advice, serves as the independent variable.

The estimated coefficients for  $\Delta$ Net benefit in Task 1 are positive and significant for **Algorithm** and **Expert**, but not for **Student**. In addition, in models (1) to (6), as the number of pieces of advice received in Task 1 increases, WTP decreases significantly in all treatments.

These findings indicate that participants increased their WTP for advice in **Algorithm** when they observed an improvement in performance as well as net benefit. While they increased their WTP when net benefit increased in **Expert**, their WTP did not respond to either in **Student**.

**Result 4.** Participants increased their WTP for advice after experiencing improved performance from adopting it only in **Algorithm**. In **Expert**, participants increased their WTP only when advice improved net benefits.

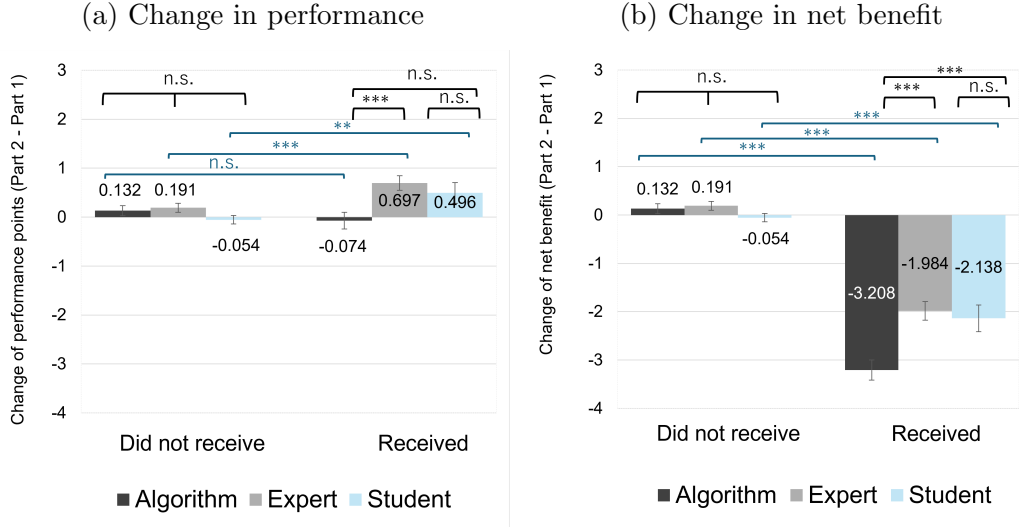
Thus, Hypothesis 2 is partially supported. While participants updated their WTPs in response to experienced changes in performance and payoffs. However, as we observe in Figure 7, they continued to overpay for advice, particularly algorithmic advice, in Task 2.<sup>29</sup>

---

<sup>29</sup>In addition, another insightful finding can also be observed in Table 3. Based on the regression results, and suppose that the other variables ( $\Delta$ Mean performance in Task 1, or  $\Delta$ Net benefit in Task 1) are equal to zero, we calculate the predicted number of received advices in Task 1 that would make  $\Delta$ WTP equal to zero. Specifically, this value is obtained by dividing the absolute value of the constant term by the coefficient on the number of received advices in Task 1. The resulting values are as follows: **Algorithm** = 5.29, **Expert** = 4.30, and **Student** = 3.76 in models (1)-(3), and **Algorithm** = 8.13, **Expert** = 5.35, and **Student** = 3.57 in models (4)-(6). In both cases, the ordering is: **Algorithm** > **Expert** > **Student**. This suggests that, holding changes in performance or net benefits constant, participants' demand for advice is least easily satisfied when using algorithmic advisors, while it is most easily satisfied when using student advisors.

<sup>30</sup>This analysis was not pre-registered

Figure 7: Change in performance and net benefit in Task 2



*Notes:* We regressed the change in performance points and net benefits (part 2 – part 1) in task 2 on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level. Pairwise comparisons of the estimated dummy coefficients were performed using F-tests with Bonferroni-adjusted  $p$ -values. We also regressed these changes on an advice-receiving dummy in each treatment using an OLS regression model with robust standard errors clustered at the participant level. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively, while n.s. denotes that the difference is not statistically significant at the 0.1 level.<sup>30</sup>

### 3.5 Discussion on Experiment 1

Participants assigned the highest WTP to algorithmic advice and relied on it as heavily as expert advice, despite its forecasting performance being no better than the alternatives. As a result, participants overpaid for advice, particularly for algorithmic advice. Moreover, the feedback provided after Task 1 was not sufficient for participants to fully adjust their beliefs about advice quality and revise their WTP accordingly. We discuss three possible explanations for these findings.

First, the high initial WTP for algorithmic and expert advice may reflect participants' perceptions of these sources as sophisticated and competent, reinforced by the detailed descriptions provided in the instructions (Lehmann

et al., 2022; Chevrier et al., 2024; Williamson et al., 2025). Similarly, stating that an algorithm “learns” may increase its acceptance because it is perceived as more competent (Chacon et al., 2022).<sup>31</sup>

Second, participants may have had difficulty accurately evaluating the economic value of advice in a noisy forecasting environment. In Experiment 1, WTP was elicited only twice, and the feedback did not clearly separate participants’ own forecasting ability, advice quality, and the net benefits from purchasing advice. As a result, participants may have struggled to determine whether observed improvements in performance were sufficiently large to justify the cost of advice.

Third, learning opportunities were endogenous to participants’ initial WTP. Participants with high WTP purchased advice more frequently and therefore had more opportunities to observe its performance, whereas participants with very low WTP observed relatively little information about advice quality. This endogenous exposure may have slowed belief updating and contributed to persistent misvaluation.

Experiment 2 is designed to investigate these mechanisms more directly. In particular, it examines whether simplifying the descriptions of the advice sources reduces initial WTP and whether more frequent and detailed feedback improves participants’ ability to evaluate the economic value of advice.

## 4 Design of Experiment 2

We employ a 2×2 between-subjects design in Experiment 2. In one dimension, we vary the source of advice: algorithm vs experts. In another

---

<sup>31</sup>Participants’ WTP may also have been affected by curiosity, prior beliefs about advice performance, overconfidence, and the house-money effect.

dimension, we vary the description of these sources: with a detailed description (as in Experiment 1) vs without description in which the source was described simply as “algorithm” or “human expert,” without additional details. The four treatments are therefore: Algorithm–No Description (Alg-No D), Algorithm–Description (Alg-D), Expert–No Description (Exp-No D), and Expert–Description (Exp-D).

All treatments used the same set of 20 stock-price time series as in Experiment 1. However, unlike Experiment 1, in each period, WTP for advice is elicited, and detailed feedback is provided regardless of whether the participant purchased the advice.<sup>32</sup> This feature controls for the effects of curiosity and differential learning opportunities that may have influenced WTP in Experiment 1. It also provides equal learning opportunities to all the participants regardless of their WTP. Furthermore, to reduce potential house-money effects, participants performed a real-effort task before the forecasting task in order to earn the endowment used to purchase advice. The repeated elicitation of WTP together with detailed feedback allows us to investigate whether additional learning opportunities improve participants’ ability to evaluate the economic value of advice.

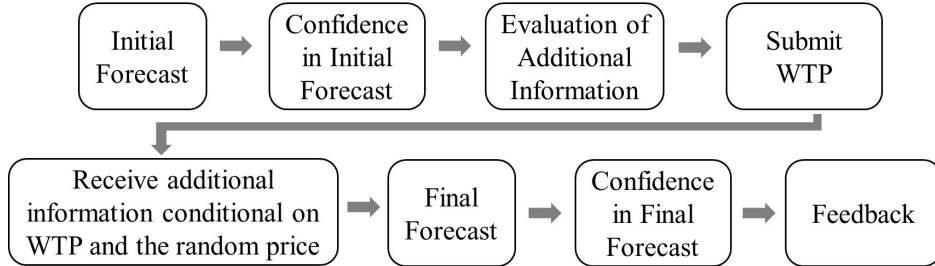
Figure 8 illustrates the sequence of decisions in each round. Participants first viewed a stock-price graph and submitted an initial forecast. They then reported confidence in that forecast and evaluated the expected usefulness of the additional information.<sup>33</sup> Next, WTP for advice was elicited using

---

<sup>32</sup>As the feedback, participants are informed of (1) WTP they submitted, (2) the price of the advice, (3) their initial and final forecasts, (4) the advice, (5) the realized stock price, (6) the forecast errors of and the reward points associated with the initial and final forecast as well as the advice, (7) the final rewards based on the initial and final forecasts, (8) the change in points if they had relied entirely on the advice, and (9) the change in net benefit from the initial to the final forecast.

<sup>33</sup>Participants’ confidence in the initial forecast is evaluated based on a five-point scale from “1. not at all confident” to “5. very confident.” Similarly they indicate how they

Figure 8: Flow of steps in each round



the same BDM mechanism as in Experiment 1. Conditional on submitted WTP and the randomly drawn advice price, participants either received the advice and submitted a final forecast or proceeded without advice. Finally, participants reported confidence in their final forecast and received detailed feedback. Unlike Experiment 1, no time limits were imposed.

The payment rule was designed to make both the initial and final forecasts payoff relevant. As in Experiment 1, participants earned 20 points if their forecast exactly matched the realized price, with one point deducted for each percentage-point deviation from the true value.

At the end of the experiment, one of the 20 forecasting rounds was randomly selected for payment, together with either the initial or final forecast from that round. Participants received the 10-point endowment earned in the slider task and the payoff associated with the selected forecast. If the final forecast was selected and advice had been purchased, the advice cost was deducted. The exchange rate was 1 point = 60 JPY.

---

expect the accuracy of their initial forecast to compare with the accuracy of the forecast based on the additional information, using a five-point scale from “1. much less accurate” to “5. much more accurate.”

Table 4: Summary of treatments in Experiment 2

Treatments	Alg - No D	Alg - D	Exp - No D	Exp - D
No. of participants	50	51	50	50
Duration (min)	90	90	90	90
Avg. payment (JPY)	2443	2456	2477	2413

## 4.1 Procedure

The experiment consisted of five parts. Participants first read the instructions and completed a comprehension quiz. Unlike in Experiment 1, participants had to answer every quiz question correctly before proceeding to the next step. They then answered a survey on demographic characteristics, stock-trading experience, and financial literacy. In the third part, to reduce the house-money effect, participants completed a slider task designed to let them earn the endowment used to purchase advice.<sup>34</sup> In the fourth part, participants completed the stock-price forecasting task. Finally, participants completed a post-experiment survey.

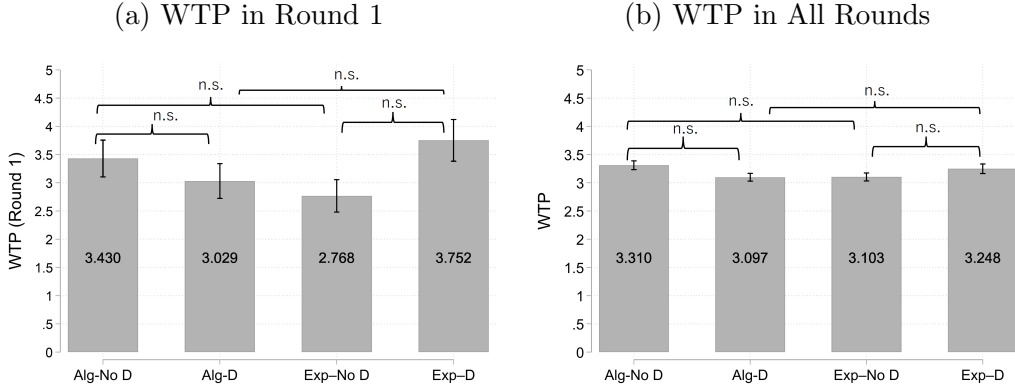
Furthermore, to examine whether characteristics of the stock-price series affected WTP, we organized the 20 time series into blocks based on statistical and technical features following Bao et al. (2023). Using k-means clustering, we grouped the time series into four clusters and constructed 4 blocks of 5 rounds containing a balanced mix of stock-price patterns across blocks.

We conducted 11 laboratory sessions at the University of Osaka in February 2026 involving 201 subjects. See Table 4 for a summary.<sup>35</sup>

<sup>34</sup>Specifically, participants completed a standard slider task (Gill and Prowse, 2012), earning 0.5 points for each completed slider and up to 10 points in total.

<sup>35</sup>Descriptive statistics of participants' characteristics are summarized in Table K.1 in Online Appendix K.

Figure 9: WTP in Round 1 and Across All Rounds in Experiment 2



*Notes:* We regressed the WTP in round 1 (in all rounds) on four treatment dummies by an OLS regression model with robust standard errors (clustered at participant level). Pairwise comparisons of estimated dummy coefficients are done by F-test with Bonferroni-adjusted  $p$ -values. Results from two-sided tests are reported. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively. See Online Appendix J.1 for details.

## 5 Results of Experiment 2

### 5.1 Initial and subsequent WTP and the role of descriptions

We first examine whether detailed descriptions of the advice sources increased participants' WTP.<sup>36</sup>

Figure 9 shows WTP for advice in Round 1 and across all rounds in Experiment 2. Detailed descriptions of the information sources did not increase WTP for either algorithmic or expert advice, either in Round 1 or across all rounds.

The average WTPs in Round 1 were 3.029 in Alg-D and 3.752 in Exp-D. These values are lower than those in Task 1 of Experiment 1, where the average WTP were 4.970 and 4.212 for algorithmic and expert advice,

<sup>36</sup>All reported results in the Results of Experiment 2 section were not pre-registered, except for the analysis of WTP in Round 1. The pre-registered analysis is presented in Online Appendices L.3 and L.4.

respectively. The lower WTP observed in Experiment 2 may reflect the reduction of curiosity-driven demand and house-money effects, both of which are more tightly controlled for in Experiment 2 than in Experiment 1.

**Result 5.** Detailed descriptions do not increase WTP for either algorithmic or expert advice.

## 5.2 Utilization of advice

In Experiment 1, participants valued algorithmic advice as highly as expert advice and relied on both sources to a similar extent after receiving the advice. We next examine whether detailed descriptions and richer feedback influenced participants' reliance on purchased advice. As in Experiment 1, reliance on advice is measured using the shift rate. We again restrict the analysis to observations with shift rates between 0 and 1.

The average shift rate, across 20 rounds, was 0.617, 0.589, 0.707, and 0.771 in Alg-No D, Alg-D, Exp-No D, and Exp-D, respectively.<sup>37</sup> Shift rates for expert advice are significantly higher than those for algorithmic advice in both the description and no-description treatments ( $p < 0.001$  and  $p = 0.025$ , respectively). In contrast, detailed descriptions do not significantly affect reliance on either source of advice.

Thus, participants relied more heavily on expert advice than on algorithmic advice after purchasing it, despite assigning similar WTP to the two sources.

**Result 6.** Participants relied more on expert advice than algorithmic advice upon purchasing it.

---

<sup>37</sup>See Online Appendix J.2 for details.

Because detailed descriptions do not significantly affect either WTP or advice utilization, we pool the treatments with and without descriptions in the remaining analyses while controlling for the presence of descriptions through a treatment dummy.

### 5.3 Learning and updating under repeated feedback

How did repeated feedback affect participants' updating of WTP and shift rates over time? Table 5 reports regressions examining the determinants of WTP across rounds, while Table 6 reports corresponding regressions for shift rates. Some specifications additionally include characteristics of the stock-price series to account for variation in the underlying forecasting environment.<sup>38</sup>

Table 5 shows the results of random effect panel regressions investigating the relationship between WTP and (i) confidence in initial forecasts (Columns 1 to 6), (ii) experiences in performance improvements with and without receiving advice (Columns 3 to 6), and (iii) time series characteristics (Columns 5 and 6).

The negative and significant coefficients on "Round" indicate that WTP for both types of advice declines over time. The negative and significant coefficients on the variables related to beliefs about forecast accuracy and confidence indicate that participants with stronger beliefs in the accuracy of their own initial forecasts and greater confidence in those forecasts exhibit significantly lower WTP.

In terms of learning from experience, greater performance improvements from purchased advice in previous rounds significantly increase WTP for

---

<sup>38</sup>Following Bao et al. (2023), we consider five characteristics of the time series: the autocorrelation coefficient, amplitude, variance, MACD reversals, and reversals.

algorithmic advice only. In addition, stronger performance improvements achieved without purchasing advice significantly reduce WTP for algorithmic advice only.

These relationships hold after controlling for the characteristics of the time series. Among the five time-series characteristics considered, greater amplitude increases WTP for expert advice, whereas higher variance reduces it. A higher autocorrelation coefficient increases WTP for algorithmic advice, whereas more MACD reversals reduce it. In addition, after controlling for past experience and time-series characteristics, providing a description of algorithmic advice has a marginally significant negative effect on WTP.

Table 6 reports random-effects panel regressions examining the relationship between shift rates and WTP, beliefs about forecast accuracy, confidence in initial forecasts, changes in confidence after receiving advice, and past experiences with and without purchased advice.

WTP is positively and significantly associated with shift rates for algorithmic advice, but not for expert advice. After controlling for WTP, the positive and marginally significant coefficients on prior performance improvements with purchased advice indicate that participants rely more heavily on advice when they previously experienced forecasting improvements from using it, regardless of the advice source. By contrast, prior performance improvements obtained without purchasing advice, which negatively affected WTP for algorithmic advice, are not significantly related to shift rates.

In addition, both the number of previously purchased advice observations and confidence in initial forecasts are negatively associated with shift rates for algorithmic advice. No comparable relationship is observed for expert advice.

Table 5: The effects of beliefs about accuracy and confidence in initial forecasts, past experience, and stock price characteristics on WTP

Variables	(1) Algorithm WTP	(2) Expert WTP	(3) Algorithm WTP	(4) Expert WTP	(5) Algorithm WTP	(6) Expert WTP
Description dummy	-0.260 (0.367)	0.089 (0.414)	-0.602* (0.342)	0.019 (0.411)	-0.600* (0.341)	0.019 (0.411)
Round	-0.041*** (0.011)	-0.031*** (0.010)	-0.052*** (0.016)	-0.037** (0.015)	-0.053*** (0.016)	-0.043*** (0.015)
Confidence in initial forecasts						
Evaluation of own accuracy in the initial forecast	-0.366*** (0.089)	-0.423*** (0.086)	-0.328*** (0.095)	-0.394*** (0.080)	-0.326*** (0.095)	-0.393*** (0.081)
Confidence in the initial forecast	-0.324*** (0.054)	-0.382*** (0.072)	-0.344*** (0.052)	-0.405*** (0.078)	-0.338*** (0.053)	-0.395*** (0.077)
Experience						
Mean prior performance change w/ purchased advice			0.161*** (0.050)	0.089 (0.056)	0.168*** (0.051)	0.089 (0.055)
Mean prior performance change w/o purchased advice			-0.233** (0.097)	-0.086 (0.153)	-0.232** (0.097)	-0.077 (0.155)
Number of advice received			0.004 (0.052)	0.010 (0.034)	0.009 (0.052)	0.011 (0.034)
Time Series Characteristics						
Autocorrelation coefficient					0.831* (0.487)	0.043 (0.416)
Amplitude					-0.004 (0.005)	0.016*** (0.005)
Variance					0.000 (0.001)	-0.002** (0.001)
MACD reversals					-0.025** (0.012)	0.006 (0.011)
Reversals					-0.001 (0.004)	0.001 (0.005)
Constant	5.398*** (0.364)	5.304*** (0.367)	5.689*** (0.408)	5.390*** (0.403)	5.724*** (0.622)	4.750*** (0.649)
Observations	2,020	2,000	1,590	1,475	1,590	1,475
Cluster at participant level	101	100	96	95	96	95
Wald $\chi^2$	91.649	83.976	120.428	74.971	132.736	85.540
Prob > $\chi^2$	0.000	0.000	0.000	0.000	0.000	0.000

Notes: a) We regress WTP in all rounds on the treatment dummy, evaluation of own accuracy in the initial forecast, confidence in the initial forecast and mean prior performance change wth purchased advice, mean prior performance change without purchased advice, number of received advice from previous rounds, using a random effect panel data OLS regression model with standard errors clustered at participant level. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.10, 0.05, and 0.01 levels, respectively.  
b) The unit of observation is each subject's WTP in round 1. The total number of observations is the number of subjects in all treatments.  
c) Numbers in parentheses are standard errors clustered at participant level.

Table 6: The effects of WTP, beliefs about accuracy and confidence in initial forecasts, change in confidence after receiving the advice, and past experience with and without receiving advices on Shift Rate

Variables	(1)		(2)		(3)		(4)		(5)		(6)	
	Algorithm SHIFT	Expert SHIFT	Algorithm SHIFT	Expert SHIFT	Algorithm SHIFT	Expert SHIFT	Algorithm SHIFT	Expert SHIFT	Algorithm SHIFT	Expert SHIFT	Algorithm SHIFT	Expert SHIFT
WTP	0.031*** (0.012)	0.006 (0.009)	0.059*** (0.021)	0.008 (0.010)	0.057*** (0.020)	0.007 (0.011)						
Round number	-0.006 (0.005)	-0.006*** (0.002)	0.024 (0.015)	-0.003 (0.006)	0.022 (0.014)	-0.004 (0.006)						
Description dummy	-0.017 (0.073)	-0.019 (0.046)	-0.081 (0.064)	-0.019 (0.055)	-0.069 (0.060)	-0.020 (0.055)						
Experiences												
Mean prior performance change with purchased advice			0.064* (0.035)	0.019* (0.010)	0.059* (0.035)	0.016* (0.010)						
Mean prior performance change without purchased advice			0.190 (0.152)	-0.005 (0.015)	0.189 (0.151)	-0.002 (0.015)						
Number of advice received			-0.040*** (0.015)	-0.001 (0.012)	-0.036** (0.014)	0.000 (0.012)						
Confidence												
Evaluation of own accuracy in the initial forecast			0.037 (0.043)	-0.000 (0.025)	0.033 (0.041)	0.001 (0.025)						
Confidence in the initial forecast			-0.109*** (0.039)	-0.014 (0.019)	-0.083** (0.037)	0.012 (0.023)						
Change in confidence												
Constant	0.522*** (0.134)	0.810*** (0.068)	0.376 (0.235)	0.769*** (0.110)	0.299 (0.256)	0.689*** (0.112)						
Observations	604	569	478	445	478	445						
Clustered at participant level	97	94	91	85	91	85						
Wald $\chi^2$	16.097	9.083	44.855	9.176	51.458	16.578						
Prob > $\chi^2$	0.001	0.028	0.000	0.328	0.000	0.056						

Notes: a) We use data for graphs for which participants received advice. We also restrict the sample to observations where the shift rate is between 0 and 1. We regress shift rate on the WTP in each round using a random effect panel data OLS model, with standard errors clustered at the participant level. WTP indicates the submitted WTP in each round. In Models (3) and (4), we add control variables, including the mean performance change from all previous rounds with purchased advice, the mean performance change from all previous rounds without purchased advice, the number of pieces of advice received in previous rounds, the evaluation of the accuracy of the initial forecast, confidence in the initial forecast. In Models (5) and (6), we add a control variable: the change in confidence from the initial forecast to the final forecast. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively.

b) The unit of observation is a round in which a participant received advice. The total number of observations is the number of SHIFT observations across all such rounds in the algorithm and expert treatments, including both the description and no-description conditions.

c) Numbers in parentheses are standard errors clustered at participant level.

Overall, repeated performance feedback led participants to update their WTP for algorithmic advice, but not for expert advice. It also affected participants' reliance on both algorithmic and expert advice over time. In contrast, repeated net-benefit feedback did not lead participants to update either their WTP or their advice utilization.<sup>39</sup>

## 5.4 Impacts of advice on performance

Did the richer feedback and repeated updating opportunities in Experiment 2 improve participants' ability to evaluate advice appropriately?

Figure 10 shows the average changes in performance points and net benefits from the initial to the final forecast, separately by whether participants received advice, for Rounds 1–10 and Rounds 11–20.

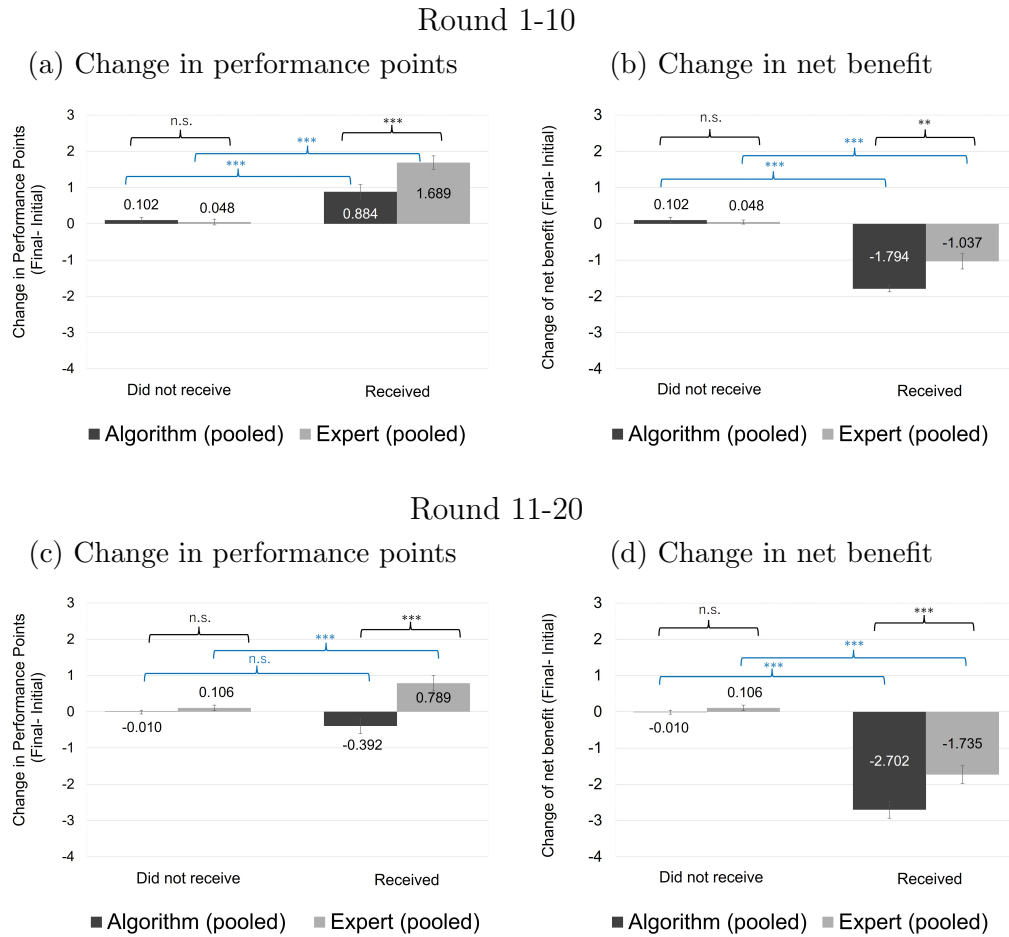
In the first ten rounds, receiving either algorithmic or expert advice significantly improved forecasting performance, with the improvement in latter being significantly larger. However, regardless of the advice source, participants' net benefits decreased on average, with significantly larger losses for algorithmic advice than for expert advice.

In the second ten rounds, as participants gained experience, the performance improvements generated by both types of advice became significantly smaller ( $p < 0.01$  for both algorithmic and expert advice). For algorithmic advice, the improvement was no longer statistically different from zero. This also led to a further decline in net benefits from receiving advice. This reduction in the net benefit from advice reflects improvements in participants' initial forecasts over time. Among all participants, average performance points from the initial forecast increased from 12.667 to 14.490 in the algorithm

---

<sup>39</sup>Detailed results on net benefit are reported in Online Appendix J.4.

Figure 10: Change in performance points and net benefit (Top) from Round 1 to 10 and (Bottom) from Round 11 to 20



*Notes:* We regressed the change in performance points and the change in net benefit from (Top) round 1 to 10 and (Bottom) round 11 to 20, both measured as final forecast minus initial forecast, on an algorithm dummy using OLS regressions with robust standard errors clustered at the participant level. All regression models control for a description dummy, which equals one if participants received a detailed description of the additional information and zero otherwise. Pairwise comparisons of the estimated coefficients were conducted using F-tests. We also estimated, within each pooled treatment, regressions of the change in performance points and the change in net benefit on an advice-receiving dummy, again using robust standard errors clustered at the participant level. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively, while n.s. indicates that the difference is not statistically significant at the 0.1 level. See Online Appendix J.3 for details.

treatments, and from 12.414 to 14.598 in the expert treatments. These increases of approximately 2 points between the first and last 10 rounds are statistically significant ( $p < 0.01$  for both algorithmic and expert advice).

These findings are broadly consistent with those from Experiment 1: although advice can improve forecasting performance, the resulting gains in accuracy are often insufficient to offset the cost of purchasing advice.

**Result 7.** Even after repeated opportunities for learning and updating, participants continued to assign valuations that exceeded the realized economic value of the advice.

## 6 Conclusion

Using two incentivized forecasting experiments, we investigated participants' willingness to pay (WTP) for advice from algorithms, financial experts, and peers; the extent to which they used this advice when making final forecasts; how advice affected forecasting performance and net benefits; and how WTP and reliance changed with experience.

In Experiment 1, participants assigned the highest WTP to algorithmic advice and relied on it as heavily as expert advice, despite its forecasting performance not being superior to that of alternative sources. Peer advice received the lowest WTP and was relied upon the least.

However, participants overpaid for advice, particularly algorithmic advice, whose realized benefits were insufficient to offset its cost. Moreover, the limited feedback provided in Experiment 1 was insufficient for participants to fully revise their beliefs about advice performance and adjust their WTP accordingly.

Three possible mechanisms may explain the results of Experiment 1. First, participants may have perceived algorithmic and expert advice as sophisticated and competent, an impression potentially reinforced by the detailed descriptions of these sources. Second, participants may have had difficulty accurately evaluating the value of advice in a noisy forecasting environment with limited feedback. Third, learning opportunities were endogenous, as they depended on participants' initial WTP.

Experiment 2 was designed to investigate these mechanisms. In particular, it examined whether withholding detailed descriptions of algorithms and experts reduced WTP, and whether more frequent and detailed feedback, provided independently of participants' WTP, improved their ability to evaluate the value of advice.

The results show that detailed descriptions do not significantly affect participants' WTP. Detailed feedback helped participants adjust their WTP for algorithmic advice, as well as their utilization of both algorithmic and expert advice, based on their experiences with and without advice. Participants' unaided initial forecasts improved over time, and their WTP declined as they gained experience. However, as in Experiment 1, participants continued to overpay for advice, particularly algorithmic advice.

Given participants' tendency to overvalue advice from sources perceived as sophisticated despite mixed performance, particularly algorithmic advice, regulatory bodies could establish standards for transparent disclosure of advice performance, especially in finance and other high-stakes domains. Clear, accessible information on the accuracy and limitations of algorithm-driven predictions may help mitigate overreliance and encourage better-informed decision-making.

Policies that foster algorithm literacy—through educational programs or certification standards in financial literacy and algorithmic competency—could help consumers and professionals critically evaluate when to rely on algorithmic or human expertise. Such initiatives could help protect consumer interests, improve decision quality, and support the balanced integration of algorithms across various sectors.

Our experiment does not involve conflicts of interest on the part of the advisors. However, advisors often provide guidance that can affect their own benefit. Therefore, a fruitful direction for future research is to study how the potential presence of a conflict of interest between advisor and advisee affects the latter’s WTP for and reliance on advice, depending on the type of advisor.

Finally, another fruitful direction for future research is to investigate whether our findings are specific to the task employed in this study or instead generalize more broadly, and, if so, which characteristics of tasks (Castelo et al., 2019) or domains (Himmelstein and Budescu, 2023) determine the applicability of these findings.

## References

- BAO, T., B. CORGNET, N. HANAKI, K. OKADA, Y. E. RIYANTO, AND J. ZHU (2025): “Financial forecasting in the lab and the field: Qualified professionals vs. smart students,” *Journal of Behavioral and Experimental Finance*, 46, 101051, <https://doi.org/10.1016/j.jbef.2025.101051>.
- BAO, T., B. CORGNET, N. HANAKI, Y. E. RIYANTO, AND J. ZHU (2023): “Predicting the unpredictable: New experimental evidence on forecasting

- random walks,” *Journal of Economic Dynamics and Control*, 146, 104571, <https://doi.org/10.1016/j.jedc.2022.104571>.
- BECKER, G. M., M. H. DEGROOT, AND J. MASCHAK (1964): “Measuring utility by a single-response sequential method,” *Behavioral Science*, 9, 226–232, <https://doi.org/10.1002/bs.3830090304>.
- BLOOMFIELD, R., AND J. HALES (2002): “Predicting the next step of a random walk: experimental evidence of regime-shifting beliefs,” *Journal of Financial Economics*, 65, 397–414, [https://doi.org/10.1016/S0304-405X\(02\)00147-2](https://doi.org/10.1016/S0304-405X(02)00147-2).
- BONACCIO, S., AND R. S. DALAL (2006): “Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences,” *Organizational Behavior and Human Decision Processes*, 101, 127–151, <https://doi.org/10.1016/j.obhdp.2006.07.001>.
- BURTON, J. W., M.-K. STEIN, AND T. B. JENSEN (2020): “A systematic review of algorithm aversion in augmented decision making,” *Journal of Behavioral Decision Making*, 33, 220–239, <https://doi.org/10.1002/bdm.2155>.
- CASON, T. N., AND C. R. PLOTT (2014): “Misconceptions and Game Form Recognition: Challenges to Theories of Revealed Preference and Framing,” *Journal of Political Economy*, 122, 1235–1270, <https://doi.org/10.1086/677254>.
- CASTELO, N., M. W. BOS, AND D. R. LEHMANN (2019): “Task-dependent algorithm aversion,” *Journal of Marketing Research*, 56, 809–825, <https://doi.org/10.1177/0022243719851788>.

- CHACON, A., E. E. KAUSEL, AND T. REYES (2022): “A longitudinal approach for understanding algorithm use,” *Journal of Behavioral Decision Making*, 35, e2275, <https://doi.org/10.1002/bdm.2275>.
- CHEN, D. L., M. SCHONGER, AND C. WICKENS (2016): “oTree—An open-source platform for laboratory, online, and field experiments,” *Journal of Behavioral and Experimental Finance*, 9, 88–97, <https://doi.org/10.1016/j.jbef.2015.12.001>.
- CHEVRIER, M., B. CORGNET, E. GUERCI, AND J. ROSAZ (2024): “Algorithm credulity: Human and algorithmic advice in prediction experiments,” *Available at SSRN 4828701*.
- D’ACUNTO, F., N. PRABHALA, AND A. G. ROSSI (2019): “The promises and pitfalls of robo-advising,” *The Review of Financial Studies*, 32, 1983–2020, <https://doi.org/10.1093/rfs/hhz014>.
- DIETVORST, B. J., J. P. SIMMONS, AND C. MASSEY (2015): “Algorithm aversion: people erroneously avoid algorithms after seeing them err.,” *Journal of Experimental Psychology: General*, 144, 114, <https://doi.org/10.1037/xge0000033>.
- (2018): “Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them,” *Management Science*, 64, 1155–1170, <https://doi.org/10.1287/mnsc.2016.2643>.
- DIMMOCK, S. G., W. C. GERKEN, AND N. P. GRAHAM (2018): “Is fraud contagious? Coworker influence on misconduct by financial advisors,” *The Journal of Finance*, 73, 1417–1450, <https://doi.org/10.1111/jofi.12613>.

- EGAN, M., G. MATVOS, AND A. SERU (2019): “The market for financial adviser misconduct,” *Journal of Political Economy*, 127, 233–295, <https://doi.org/10.1086/700735>.
- FERNANDES, D., J. G. L. JR., AND R. G. NETEMEYER (2014): “Financial literacy, financial education, and downstream financial behaviors,” *Management Science*, 60, 1861–1883, <https://doi.org/10.1287/mnsc.2013.1849>.
- FILIZ, I., J. R. JUDEK, M. LORENZ, AND M. SPIWOKS (2021): “Reducing algorithm aversion through experience,” *Journal of Behavioral and Experimental Finance*, 31, 100524, <https://doi.org/10.1016/j.jbef.2021.100524>.
- FOERSTER, S., J. T. LINNAINMAA, B. T. MELZER, AND A. PREVITERO (2017): “Retail financial advice: does one size fit all?” *The Journal of Finance*, 72, 1441–1482, <https://doi.org/10.1111/jofi.12514>.
- GERMANN, M., AND C. MERKLE (2023): “Algorithm aversion in delegated investing,” *Journal of Business Economics*, 93, 1691–1727, <https://doi.org/10.1007/s11573-022-01121-9>.
- GILL, A., R. M. GILLENKIRCH, J. ORTNER, AND L. VELTHUIS (2024): “Dynamics of Reliance on Algorithmic Advice,” *Journal of Behavioral Decision Making*, 37, e2414, <https://doi.org/10.1002/bdm.2414>.
- GILL, D., AND V. PROWSE (2012): “A structural analysis of disappointment aversion in a real effort competition,” *American Economic Review*, 102, 469–503.
- GREINER, B. (2015): “Subject pool recruitment procedures: Organizing

- experiments with ORSEE,” *Journal of the Economic Science Association*, 1, 114–125, <https://doi.org/10.1007/s40881-015-0004-4>.
- GREINER, B., P. GRÜNWARD, T. LINDNER, G. LINTNER, AND M. WIERNSPERGER (2025): “Incentives, framing, and reliance on algorithmic advice: An experimental study,” *Management Science*, <https://doi.org/10.1287/mnsc.2022.02777>.
- HARVEY, N., AND I. FISCHER (1997): “Taking advice: Accepting help, improving judgment, and sharing responsibility,” *Organizational behavior and human decision processes*, 70, 117–133, <https://doi.org/10.1006/obhd.1997.2697>.
- HIMMELSTEIN, M., AND D. V. BUDESCU (2023): “Preference for human or algorithmic forecasting advice does not predict if and how it is used,” *Journal of Behavioral Decision Making*, 36, e2285, <https://doi.org/10.1002/bdm.2285>.
- HOECHLE, D., S. RUENZI, N. SCHAUB, AND M. SCHMID (2018): “Financial advice and bank profits,” *The Review of Financial Studies*, 31, 4447–4492, <https://doi.org/10.1093/rfs/hhy046>.
- HOLZMEISTER, F., M. HOLMÉN, M. KIRCHLER, M. STEFAN, AND E. WENGSTRÖM (2023): “Delegation decisions in finance,” *Management Science*, 69, 4828–4844, <https://doi.org/10.1287/mnsc.2022.4555>.
- JUSSUPOW, E., I. BENBASAT, AND A. HEINZL (2020): “Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion,” in *In Proceedings of the 28th European Conference on Information Systems (ECIS)*, [https://aisel.aisnet.org/ecis2020\\_rp/168](https://aisel.aisnet.org/ecis2020_rp/168).

- LEHMANN, C. A., C. B. HAUBITZ, A. FÜGENER, AND U. W. THONEMANN (2022): “The risk of algorithm transparency: How algorithm complexity drives the effects on the use of advice,” *Production and Operations Management*, 31, 3419–3434.
- LINNAINMAA, J. T., B. T. MELZER, AND A. PREVITERO (2021): “The misguided beliefs of financial advisors,” *The Journal of Finance*, 76, 587–621, <https://doi.org/10.1111/jofi.12995>.
- LO, A. W., AND J. HASANHODZIC (2011): *The Evolution of Technical Analysis: Financial Prediction from Babylonian Tablets to Bloomberg Terminals* Volume 139: John Wiley & Sons.
- LO, A. W., H. MAMAYSKY, AND J. WANG (2000): “Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation,” *Journal of Finance*, 55, 1705–1765, <https://doi.org/10.1111/0022-1082.00265>.
- LOGG, J. M., J. A. MINSON, AND D. A. MOORE (2019): “Algorithm appreciation: People prefer algorithmic to human judgment,” *Organizational Behavior and Human Decision Processes*, 151, 90–103, <https://doi.org/10.1016/j.obhdp.2018.12.005>.
- MAHMUD, H., A. N. ISLAM, S. I. AHMED, AND K. SMOLANDER (2022): “What influences algorithmic decision-making? A systematic literature review on algorithm aversion,” *Technological Forecasting and Social Change*, 175, 121390, <https://doi.org/10.1016/j.techfore.2021.121390>.
- MESHI, D., G. BIELE, C. W. KORN, AND H. R. HEEKEREN (2012): “How expert advice influences decision making,” *PLoS One*, 7, e49748.

- OECD (2019): “Artificial Intelligence in Society,” *OECD Publishing, Paris*, <https://doi.org/10.1787/eedfee77-en>.
- (2024): “Regulatory approaches to Artificial Intelligence in finance,” *OECD Publishing, Paris*, <https://doi.org/10.1787/f1498c02-en>.
- ÖNKAL, D., P. GOODWIN, M. THOMSON, S. GÖNÜL, AND A. POLLOCK (2009): “The relative influence of advice from human experts and statistical methods on forecast adjustments,” *Journal of Behavioral Decision Making*, 22, 390–409.
- PRAHL, A., AND L. VAN SWOL (2017): “Understanding algorithm aversion: When is advice from automation discounted?” *Journal of Forecasting*, 36, 691–702, <https://doi.org/10.1002/for.2464>.
- SNIEZEK, J. A., AND T. BUCKLEY (1995): “Cueing and cognitive conflict in judge-advisor decision making,” *Organizational behavior and human decision processes*, 62, 159–174.
- SNIEZEK, J. A., G. E. SCHRAH, AND R. S. DALAL (2004): “Improving judgement with prepaid expert advice,” *Journal of Behavioral Decision Making*, 17, 173–190, <https://doi.org/10.1002/bdm.468>.
- SUTHERLAND, S. C., C. HARTEVELD, AND M. E. YOUNG (2016): “Effects of the advisor and environment on requesting and complying with automated advice,” *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 6, 1–36, <https://doi.org/10.1145/2905370>.
- TSE, T. T. K., N. HANAKI, AND B. MAO (2024): “Beware the performance of an algorithm before relying on it: Evidence from a stock price forecasting

experiment,” *Journal of Economic Psychology*, 102, 102727, <https://doi.org/10.1016/j.joep.2024.102727>.

WILLIAMSON, K., R. KIZILCEC, S. FATH, AND N. HEFFERNAN (2025): “Algorithm appreciation in education: Educators prefer complex over simple algorithms,” in *Proceedings of the 15th International Learning Analytics and Knowledge Conference*, 527–536.

YANG, C. L., K. BAUER, X. LI, AND O. HINZ (2025): “My Advisor, Her AI and Me: Evidence from a Field Experiment on Human-AI Collaboration and Investment Decisions,” *arXiv preprint arXiv:2506.03707*, <https://doi.org/10.48550/arXiv.2506.03707>.

# Online Appendix to “Overvaluing Algorithmic Advice: Evidence from a Stock Price Forecasting Experiment”

## Contents

<b>A Results on which figures and text are based in manuscript for Experiment 1</b>	<b>3</b>
<b>B Analyses of experimental results conditional on personal characteristics in Experiment 1</b>	<b>9</b>
<b>C Other results in Experiment 1</b>	<b>19</b>
C.1 Performance of advice . . . . .	19
C.2 Performance of advice received by participants . . . . .	21
C.3 Utilization of Advice: Shift rate . . . . .	23
C.3.1 Shift rate without restriction . . . . .	26
C.4 Maximum WTP to increase net benefit . . . . .	28
C.5 Potential drivers of Variation in WTP across information sources . . . . .	28
C.6 Change of absolute percentage error (APE) . . . . .	30
C.7 WTP vs. Net benefit in Part 2 . . . . .	31
C.8 WTP vs. Change of performance (points) . . . . .	32
C.9 Changes in Performance and Net Benefit among Participants Who Received Advice Frequently and Infrequently . . . . .	32
C.10 Comparing Constant Terms in Table 3 . . . . .	36
<b>D Results without Pilot Sessions in Experiment 1</b>	<b>38</b>
<b>E Experimental Instruction in Experiment 1 (English Translation)</b>	<b>45</b>
<b>F Quiz in Experiment 1 (English Translation)</b>	<b>55</b>
<b>G Questionnaire in Experiment 1 and 2 (English Translation)</b>	<b>56</b>
G.1 Survey on demographic characteristics . . . . .	56
G.2 Survey on stock trading experience . . . . .	57
G.3 Financial literacy test . . . . .	58

<b>H</b>	<b>Screenshots of Tasks 1 and 2 in Experiment 1 (English Translation)</b>	<b>62</b>
<b>I</b>	<b>Stock-price time series graphs</b>	<b>69</b>
<b>J</b>	<b>Results on which figures and text are based in manuscript for Experiment 2</b>	<b>72</b>
J.1	WTP . . . . .	72
J.2	Utilization of advice . . . . .	73
J.3	Change in performance points and net benefit between the first 10 rounds and the last 10 rounds . . . . .	75
J.4	Factors influence WTP and reliance on advice . . . . .	78
<b>K</b>	<b>Analyses of experimental results conditional on personal characteristics in Experiment 2</b>	<b>80</b>
<b>L</b>	<b>Other Results in Experiment 2</b>	<b>90</b>
L.1	WTP trend across rounds . . . . .	90
L.2	WTP across Blocks . . . . .	91
L.3	The Effects of Changes in Performance Points and Net Benefit on WTP	93
L.4	The Characteristics of Stock Price . . . . .	98
L.5	Shift rate without restriction . . . . .	98
<b>M</b>	<b>Experimental Instruction in Experiment 2 (English Translation)</b>	<b>101</b>
<b>N</b>	<b>Screenshots in Task 1 in Experiment 2 (English Translation)</b>	<b>111</b>
<b>O</b>	<b>Screenshots in Task 2 in Experiment 2 (English Translation)</b>	<b>112</b>
<b>P</b>	<b>Post-experiment survey in Experiment 2</b>	<b>122</b>
<b>Q</b>	<b>Quiz in Experiment 2</b>	<b>126</b>
<b>R</b>	<b>Clustering the stock price into blocks</b>	<b>127</b>

## A Results on which figures and text are based in manuscript for Experiment 1

Table A1: Predicted WTP in Task 1 and Task 2 (for Figure 4(a))

VARIABLES	(1) WTP Task 1	(2) WTP Task 2
Baseline = Algorithm treatment dummy		
Expert treatment dummy	-0.759** (0.322)	-0.765** (0.3229)
Student treatment dummy	-1.954*** (0.296)	-1.987*** (0.166)
Constant	4.970*** (0.171)	4.992*** (0.084)
Observations	308	311
Cluster at session level	17	17
R-squared	0.097	0.110
Treatments	Predicted WTP Task 1	Predicted WTP Task 2
Algorithm	4.970 (0.171)	4.992 (0.084)
Experts	4.212 (0.273)	4.227 (0.213)
Students	3.016 (0.242)	3.006 (0.144)
One-sided test	Prob > F	Prob > F
Expert > Algorithm#	0.953	0.994
Algorithm > Student #	0.000	0.000
Expert > Student #	0.007	0.000
Two-sided test	Prob > F	Prob > F
Algorithm = Experts = Students	0.000	0.000
Algorithm = Expert #	0.095	0.012
Algorithm = Student #	0.000	0.000
Expert = Student #	0.014	0.001

*Notes:* (a) We regressed the WTP on three treatment dummies by OLS regression model with robust standard errors clustered at session level. (b) The unit of observation is the number of participants. The total number of observations is the number of participants in all treatments. (c) The robust standard errors are in parentheses. (d) # indicates Bonferroni-adjusted p-values. (e) 3 observations were removed in Task 1 because 3 participants did not submit the WTP within the time limit.

Table A2: Change of performance (points) (for Figures 5(a) and 7(a))

VARIABLES	Advice received	(1)	(2)	(3)	(4)
		$\Delta$ Perf. Task 1 No	$\Delta$ Perf. Task 1 Yes	$\Delta$ Perf. Task 2 No	$\Delta$ Perf. Task 2 Yes
Baseline = Algorithm treatment dummy					
Expert treatment dummy		-0.360 (0.227)	0.870*** (0.269)	0.059 (0.139)	0.770*** (0.225)
Student treatment dummy		-0.091 (0.184)	1.139*** (0.259)	-0.186 (0.136)	0.570** (0.271)
Constant		0.499*** (0.147)	0.069 (0.202)	0.132 (0.103)	-0.074 (0.170)
Observations		1,850	1,225	1,850	1,254
Clusters		298	254	305	261
R-squared		0.002	0.015	0.002	0.010
Treatments		Predicted $\Delta$ Perf. Task 1	Predicted $\Delta$ Perf. Task 2	Predicted $\Delta$ Perf. Task 1	Predicted $\Delta$ Perf. Task 2
Algorithm		0.499 (0.147)	0.069 (0.202)	0.132 (0.103)	-0.074 (0.170)
Experts		0.139 (0.172)	0.939 (0.178)	0.191 (0.093)	0.697 (0.147)
Students		0.408 (0.110)	1.208 (0.162)	-0.054 (0.088)	0.496 (0.211)
		Prob > F	Prob > F	Prob > F	Prob > F
Algorithm = Experts = Students		0.265	0.000	0.137	0.003
Algorithm = Expert #			0.004		0.002
Algorithm = Student #			0.000		0.109
Expert = Student #			0.793		1.000

*Notes:* (a) We regressed the change in performance points (Part 2 - Part 1) ( $\Delta$ Perf.) on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level. (b) The unit of observation is the forecast from each participant in each round. The total number of observations is the total number of graphs with/without advice received in each task in all treatments. (c) The robust standard errors are in parentheses. (d) # indicates Bonferroni-adjusted p-values. (e) Due to no submissions of predictions within the time limit, 14 observations were removed in Model (1), 16 observations in Model (2), 2 observations in Model (3), and 2 observations in Model (4).

Table A3: Change in performance (points) when comparing with and without receiving advice (for Figures 5(a) and 7(a))

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	$\Delta$ Perf.	$\Delta$ Perf.	$\Delta$ Perf.	$\Delta$ Perf.	$\Delta$ Perf.	$\Delta$ Perf.
	Task 1	Task 1	Task 1	Task 2	Task 2	Task 2
	Algorithm	Expert	Student	Algorithm	Expert	Student
Advice received dummy	-0.430*	0.799***	0.801***	-0.206	0.506***	0.550**
	(0.239)	(0.241)	(0.184)	(0.208)	(0.161)	(0.232)
Constant	0.499***	0.139	0.408***	0.132	0.191**	-0.054
	(0.148)	(0.173)	(0.110)	(0.104)	(0.094)	(0.088)
Observations	1,048	1,020	1,007	1,056	1,029	1,019
Clusters	106	103	102	106	103	102
R-squared	0.003	0.012	0.014	0.001	0.009	0.00

*Notes:* (a) We regressed the change in performance points (Part 2 - Part 1) ( $\Delta$ Perf.) on advice received dummy in each task and each treatment using an OLS regression model with robust standard errors clustered at the participant level. The advice received dummy equals 1 if the participant receives advice for that graph, and 0 otherwise. (b) The unit of observation is the forecast from each participant in each round. The total number of observations is the total number of forecasts with/without advice received in each task in all treatments. (c) The robust standard errors are in parentheses. (d) Due to no submissions of predictions within the time limit, 12 observations were removed in Model (1), 10 observations in Model (2), 13 observations in Model (3), 4 observations in Model (4), 1 observations in Model (5) and 1 observations in Model (6).

Table A4: Change of net benefit (points) (for Figures 5(b) and 7(b))

VARIABLES	Advice received	(1)	(2)	(3)	(4)
		$\Delta$ NB. Task 1 No	$\Delta$ NB. Task 1 Yes	$\Delta$ NB. Task 2 No	$\Delta$ NB. Task 2 Yes
Baseline = Algorithm treatment dummy					
Expert treatment dummy		-0.360 (0.227)	0.910*** (0.349)	0.059 (0.139)	1.224*** (0.283)
Student treatment dummy		-0.091 (0.184)	1.645*** (0.329)	-0.186 (0.136)	1.070*** (0.346)
Constant		0.499*** (0.147)	-3.076*** (0.235)	0.132 (0.103)	-3.208*** (0.209)
Observations		1,850	1,225	1,850	1,254
Clusters		298	254	305	261
R-squared		0.002	0.022	0.002	0.020
Treatments		Predicted $\Delta$ NB. Task 1	Predicted $\Delta$ NB. Task 2	Predicted $\Delta$ NB. Task 1	Predicted $\Delta$ NB. Task 2
Algorithm		0.499 (0.147)	-3.076 (0.235)	0.132 (0.103)	-3.208 (0.209)
Experts		0.139 (0.172)	-2.166 (0.258)	0.191 (0.093)	-1.984 (0.191)
Students		0.408 (0.110)	-1.430 (0.231)	-0.054 (0.088)	-2.138 (0.277)
		Prob > F	Prob > F	Prob > F	Prob > F
Algorithm = Experts = Students		0.265	0.000	0.137	0.000
Algorithm = Expert #			0.029		0.000
Algorithm = Student #			0.000		0.007
Expert = Student #			0.104		1.000

*Notes:* (a) We regressed the change in net benefit (Part 2 - Part 1) ( $\Delta$ NB.) on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level. (b) The unit of observation is the forecast from each participant in each round. The total number of observations is the total number of graphs with/without advice received in each task in all treatments. (c) The robust standard errors are in parentheses. (d) # indicates Bonferroni-adjusted p-values. (e) Due to no submissions of predictions within the time limit, 17 observations were removed in Model (1), 18 observations in Model (2), 4 observations in Model (3), and 2 observations in Model (4).

Table A5: Change in net benefit (points) when comparing with and without receiving advice (for Figures 5(b) and 7(b))

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	$\Delta$ NB.	$\Delta$ NB.	$\Delta$ NB.	$\Delta$ NB.	$\Delta$ NB.	$\Delta$ NB.
	Task 1 Algorithm	Task 1 Expert	Task 1 Student	Task 2 Algorithm	Task 2 Expert	Task 2 Student
Advice received dummy	-3.575*** (0.273)	-2.305*** (0.310)	-1.838*** (0.252)	-3.340*** (0.242)	-2.175*** (0.195)	-2.084*** (0.293)
Constant	0.499*** (0.148)	0.139 (0.173)	0.408*** (0.110)	0.132 (0.104)	0.191** (0.094)	-0.054 (0.088)
Observations	1,048	1,020	1,007	1,056	1,029	1,019
Clusters	106	103	102	106	103	102
R-squared	0.153	0.081	0.065	0.184	0.117	0.104

*Notes:* (a) We regressed the change in net benefit (Part 2 - Part 1) ( $\Delta$ NB.) on advice received dummy in each task and each treatment using an OLS regression model with robust standard errors clustered at the participant level. The advice received dummy equals 1 if the participant receives advice for that graph, and 0 otherwise. (b) The unit of observation is the forecast from each participant in each round. The total number of observations is the total number of forecasts with/without advice received in each task in all treatments. (c) The robust standard errors are in parentheses. (d) Due to no submissions of predictions within the time limit, 12 observations were removed in Model (1), 10 observations in Model (2), 13 observations in Model (3), 4 observations in Model (4), 1 observations in Model (5) and 1 observations in Model (6).

Table A6: Summary of variables

VARIABLES	Algorithm		Expert		Student	
	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2
Average WTP	4.970 (0.234)	4.992 (0.224)	4.212 (0.250)	4.269 (0.224)	3.016 (0.249)	3.026 (0.244)
Frequency of WTP=0 (%)	5.714 (2.276)	3.774 (1.860)	7.843 (2.675)	11.650 (3.177)	19.802 (3.985)	22.549 (4.158)
Frequency of WTP=10 (%)	3.810 (1.877)	2.830 (1.618)	3.922 (1.931)	0	0.990 (0.990)	1.961 (1.380)
Average price of paid advice	3.146 (0.087)	3.141 (0.091)	3.104 (0.106)	2.680 (0.088)	2.625 (0.120)	2.634 (0.105)
Frequency of receiving advice (%) when WTP≠0	54.040 (1.585)	49.412 (1.566)	43.191 (1.616)	47.692 (1.657)	37.284 (1.700)	40.253 (1.746)
Performance of advice (APE)	0.073 (0.021)	0.058 (0.013)	0.062 (0.020)	0.047 (0.011)	0.060 (0.020)	0.051 (0.009)
Performance of advice (points)	12.870 (1.975)	14.157 (1.268)	14.044 (1.740)	15.323 (1.124)	14.379 (1.687)	14.852 (0.936)
Performance (points) in Part 1	12.954 (0.184)	14.314 (0.131)	13.048 (0.189)	14.397 (0.137)	12.992 (0.193)	14.368 (0.133)
Performance (points) in Part 2	13.200 (0.181)	14.342 (0.123)	13.498 (0.177)	14.798 (0.120)	13.624 (0.182)	14.481 (0.126)
Improvement of performance (points) with advice	0.069 (0.207)	-0.074 (0.178)	0.939 (0.183)	0.697 (0.151)	1.208 (0.182)	0.496 (0.178)
Improvement of performance (points) without advice	0.499 (0.139)	0.132 (0.098)	0.139 (0.142)	0.191 (0.094)	0.408 (0.112)	-0.054 (0.092)
Improvement of net benefit (points) with advice	-3.076 (0.219)	-3.208 (0.200)	-2.166 (0.208)	-1.984 (0.176)	-1.430 (0.210)	-2.138 (0.201)
Improvement of net benefit (points) without advice	0.499 (0.139)	0.132 (0.098)	0.139 (0.142)	0.191 (0.094)	0.408 (0.112)	-0.054 (0.092)

*Notes:* Standard errors are in parentheses.

## B Analyses of experimental results conditional on personal characteristics in Experiment 1

Table B1: Summary of personal characteristics

Treatments	Algorithm	Expert	Student	<i>p</i> -value
Male (%)	70.755 (4.439)	62.136 (4.480)	60.606 (4.936)	0.245
Undergraduate (%)	51.887 (4.876)	56.311 (4.911)	65.686 (4.724)	0.116
Science major (%)	74.528 (4.252)	67.961 (4.620)	68.627 (4.617)	0.507
Stock trading experience (%)	20.755 (3.958)	11.650 (3.177)	8.824 (2.822)	0.048
Financial literacy score (out of 12)	7.981 (0.215)	8.000 (0.205)	8.127 (0.216)	0.872
Score of comprehension quiz	4.491 (0.074)	4.379 (0.092)	4.431 (0.079)	0.629
No. of participants	106	103	102	

*Notes:* a) Each personal characteristic was regressed on treatment dummies using an OLS regression model, with robust standard errors. The estimated dummy coefficients were compared using an F-test, with results presented as *p*-values.

b) Numbers in parentheses represent standard errors.

c) In *Student*, 3 participants refused to answer their gender, so the observation for *Male (%)* is based on 99 participants.

Table B1 summarizes the frequency of male participants, undergraduate grade level, science majors, stock trading experience, financial literacy scores, and the score in comprehension quiz.

Most participants in all treatments displayed a solid understanding of the experimental rules, with comprehension levels showing no significant variation across treatments. The mean financial literacy score was 7.981, 8.000, and 8.127 in *Algorithm*, *Expert*, and *Student*, respectively, with no statistically significant differences observed across treatments.

The characteristics of participants are well balanced across three treatments except for stock trading experience. In *Algorithm*, 20.755% of participants had stock trading experience, compared to 11.650% and 8.824% in *Expert* and *Student*, respectively.<sup>1</sup> To address this, we analyzed the experimental results conditional on personal characteristics.

We investigated whether personal characteristics correlate with WTP in Task 1, as

<sup>1</sup>Stock trading experience was not balanced across the three treatments, possibly due to a lack of proper randomization. Since participants in the lab experiment were allowed to choose their session dates, this may have contributed to the imperfect randomization across treatments, as each session was conducted using only one treatment condition.

this measure may reflect participants' perceptions of their own skills relative to the advice source. Table B2 presents these results. We excluded data from subjects who did not make submissions within the specified time limit. Models (1) through (3) use WTP in Task 1 as the dependent variable.

Table B2: Personal characteristics and WTP in Task 1

VARIABLES	(1) Algorithm WTP <sub>Task1</sub>	(2) Expert WTP <sub>Task1</sub>	(3) Student WTP <sub>Task1</sub>
Male dummy	-0.211 (0.599)	-0.340 (0.541)	-0.909 (0.582)
Undergraduate student dummy	0.139 (0.469)	0.962* (0.558)	-0.393 (0.598)
Science major dummy	0.517 (0.628)	0.203 (0.617)	-0.002 (0.584)
Stock trading experience dummy	0.424 (0.694)	-1.877** (0.743)	0.626 (1.208)
Financial literacy score	-0.094 (0.120)	0.288** (0.132)	0.007 (0.154)
Constant	5.326*** (1.075)	1.641 (1.290)	3.686*** (1.386)
Obs.	105	102	98
$R^2$	0.016	0.104	0.042

*Notes:* a) We regressed  $WTP$  in Task 1 on personal characteristics using an OLS regression model, with robust standard errors. The male dummy equals 1 for male participants and 0 otherwise. The undergraduate student dummy equals 1 for undergraduate students and 0 otherwise. The science major dummy equals 1 for science majors and 0 otherwise. The stock trading experience dummy equals 1 for participants with stock trading experience and 0 otherwise. The financial literacy score ranges from 0 to 12. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively.

b) The unit of observation is the participant. The total number of observations corresponds to the number of participants in each treatment.

c) Numbers in parentheses represent standard errors.

d) In each treatment, one observation was removed in Task 1 because a participant did not submit their WTP within the time limit.

For algorithmic advice and student advice, none of the observable personal characteristics show a significant correlation with WTP in Task 1. In contrast, for expert advice, participants who are undergraduate students exhibit a higher WTP, with a statistically significant result ( $p = 0.088$ ). Participants with stock trading experience exhibit a lower WTP, with a statistically significant result ( $p = 0.034$ ); and participants with higher financial literacy score exhibit a higher WTP, with a statistically significant result ( $p = 0.031$ ).

Overall, most personal characteristics do not significantly correlate with WTP for advice from any source. While this study focused on a limited set of personal characteristics, future research could expand this exploration to additional attributes to assess their potential impact on WTP for advice from various information sources.

Table B3: Predicted WTP in Task 1 and Task 2 conditional on personal characteristics

VARIABLES	(1) WTP Task 1	(2) WTP Task 2
<i>Baseline = Algorithm treatment dummy</i>		
Expert treatment dummy	−0.824** (0.352)	−0.878*** (0.317)
Student treatment dummy	−2.090*** (0.361)	−2.084*** (0.329)
Male dummy	−0.381 (0.333)	−0.604* (0.307)
Undergraduate student dummy	0.189 (0.313)	−0.215 (0.287)
Science major dummy	0.006 (0.352)	−0.545* (0.319)
Stock trading experience dummy	−0.221 (0.462)	−0.384 (0.414)
Financial literacy score	0.050 (0.078)	−0.030 (0.070)
Constant	4.786*** (0.723)	6.259*** (0.616)
Observations	305	308
R-squared	0.106	0.143
Treatments	Predicted WTP Task 1	Predicted WTP Task 2
Algorithm	5.025 (0.241)	5.063 (0.221)
Expert	4.201 (0.248)	4.185 (0.225)
Student	2.936 (0.258)	2.980 (0.243)
	Prob > F	Prob > F
Algorithm = Expert = Student	0.000	0.000
Algorithm = Expert #	0.060	0.018
Algorithm = Student #	0.000	0.000
Expert = Student #	0.001	0.001

*Notes:* (a) We regressed the WTP on three treatment dummies by OLS regression model with robust standard errors, controlling personal characteristics. (b) The unit of observation is the number of participants. The total number of observations is the number of participants in all treatments. (c) The robust standard errors are in parentheses. (d) # indicates Bonferroni-adjusted p-values. (e) 3 observations were removed in Task 1 because 3 participants did not submit the WTP within the time limit.

Table B4: Predicted shift rate in Task 1 and Task 2 conditional on personal characteristics

VARIABLES	(1) Shift rate Task 1	(2) Shift rate Task 2
Baseline = Algorithm treatment dummy		
Expert treatment dummy	0.041 (0.031)	-0.029 (0.034)
Student treatment dummy	-0.235*** (0.039)	-0.328*** (0.035)
Male dummy	-0.038 (0.031)	-0.025 (0.033)
Undergraduate student dummy	0.026 (0.031)	0.036 (0.029)
Science major dummy	0.052 (0.037)	0.004 (0.032)
Stock trading experience dummy	0.089** (0.040)	0.051 (0.038)
Financial literacy score	-0.010 (0.008)	0.004 (0.008)
Constant	0.847*** (0.082)	0.776*** (0.066)
Observations	1,095	1,117
Clusters	249	257
R-squared	0.138	0.186
Treatments	Predicted shift rate Task 1	Predicted shift rate Task 2
Algorithm	0.807 (0.022)	0.820 (0.022)
Experts	0.848 (0.022)	0.791 (0.025)
Students	0.572 (0.032)	0.492 (0.026)
	Prob > F	Prob > F
Algorithm = Experts = Students	0.000	0.000
Algorithm = Expert #	0.579	1.000
Algorithm = Student #	0.000	0.000
Expert = Student #	0.000	0.000

*Notes:* (a) We used the data of the graphs which participants received the advice. We also used the data which shift rate is between 0 and 1. We regressed the shift rate on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level, controlling personal characteristics. (b) The unit of observation is the graph with received advice of each participant. The total number of observations is the number of graphs with received advice in each task in all treatments. (c) The robust standard errors are in parentheses. (d) # indicates Bonferroni-adjusted p-values. (e) 8 observations were removed in Task 1 and 2 observations were removed in Task 2 due to no submissions of predictions within the time limit.

Table B5: Change of performance (points) conditional on personal characteristics

VARIABLES	Advice received	(1)	(2)	(3)	(4)
		$\Delta$ Perf. Task 1 No	$\Delta$ Perf. Task 1 Yes	$\Delta$ Perf. Task 2 No	$\Delta$ Perf. Task 2 Yes
Baseline = Algorithm treatment dummy					
Expert treatment dummy		-0.357 (0.236)	0.821*** (0.267)	0.057 (0.141)	0.802*** (0.231)
Student treatment dummy		-0.117 (0.195)	1.037*** (0.264)	-0.151 (0.131)	0.583** (0.275)
Male dummy		0.004 (0.199)	0.050 (0.250)	0.094 (0.119)	-0.222 (0.233)
Undergraduate student dummy		0.011 (0.175)	0.252 (0.224)	-0.052 (0.108)	-0.163 (0.238)
Science major dummy		-0.112 (0.207)	-0.330 (0.279)	-0.156 (0.136)	-0.227 (0.240)
Stock trading experience dummy		-0.037 (0.301)	0.090 (0.339)	0.102 (0.183)	0.355 (0.323)
Financial literacy score		0.031 (0.039)	0.019 (0.053)	-0.041 (0.031)	-0.003 (0.051)
Constant		0.327 (0.425)	-0.013 (0.509)	0.519* (0.286)	0.266 (0.463)
Observations		1,834	1,212	1,828	1,246
Clusters		296	252	302	259
R-squared		0.003	0.017	0.003	0.014
Treatments		Predicted $\Delta$ Perf. Task 1	Predicted $\Delta$ Perf. Task 2	Predicted $\Delta$ Perf. Task 1	Predicted $\Delta$ Perf. Task 2
Algorithm		0.502 (0.152)	0.096 (0.201)	0.124 (0.103)	-0.088 (0.174)
Experts		0.145 (0.171)	0.917 (0.176)	0.181 (0.093)	0.714 (0.148)
Students		0.385 (0.117)	1.133 (0.163)	-0.028 (0.084)	0.495 (0.212)
		Prob > F	Prob > F	Prob > F	Prob > F
Algorithm = Experts = Students		0.302	0.000	0.206	0.002
Algorithm = Expert #			0.007		0.002
Algorithm = Student #			0.000		0.108
Expert = Student #			1.000		1.000

Notes: (a) We regressed the change in performance points (Part 2 - Part 1) ( $\Delta$ Perf.) on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level, controlling personal characteristics. (b) The unit of observation is the forecast from each participant in each round. The total number of observations is the total number of graphs with/without advice received in each task in all treatments. (c) The robust standard errors are in parentheses. (d) # indicates Bonferroni-adjusted p-values. (e) Due to no submissions of predictions within the time limit, 17 observations were removed in Model (1), 17 observations in Model (2), 4 observations in Model (3), and 2 observations in Model (4).

Table B6: Change in performance (points) when comparing with and without receiving advice conditional on personal characteristics

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	$\Delta$ Perf.	$\Delta$ Perf.	$\Delta$ Perf.	$\Delta$ Perf.	$\Delta$ Perf.	$\Delta$ Perf.
	Task 1 Algorithm	Task 1 Expert	Task 1 Student	Task 2 Algorithm	Task 2 Expert	Task 2 Student
Advice received dummy	-0.438* (0.251)	0.712*** (0.226)	0.744*** (0.171)	-0.241 (0.212)	0.541*** (0.159)	0.493** (0.234)
Male dummy	0.051 (0.298)	0.077 (0.300)	-0.130 (0.230)	0.098 (0.213)	-0.037 (0.213)	-0.171 (0.205)
Undergraduate student dummy	0.515** (0.254)	0.266 (0.249)	-0.572** (0.220)	-0.289 (0.196)	0.123 (0.192)	-0.153 (0.218)
Science major dummy	-0.283 (0.346)	-0.356 (0.345)	0.236 (0.203)	-0.355 (0.219)	-0.132 (0.237)	-0.314 (0.204)
Stock trading experience dummy	0.446 (0.311)	-0.430 (0.492)	-0.215 (0.382)	-0.054 (0.248)	0.774** (0.384)	-0.144 (0.303)
Financial literacy score	-0.024 (0.065)	0.044 (0.059)	0.038 (0.044)	-0.054 (0.042)	-0.069 (0.055)	0.057 (0.053)
Constant	0.514 (0.663)	-0.079 (0.603)	0.409 (0.427)	0.938** (0.387)	0.683 (0.528)	-0.059 (0.490)
Observations	1,048	1,020	978	1,056	1,029	989
Clusters	106	103	99	106	103	99
R-squared	0.010	0.019	0.023	0.005	0.017	0.014

*Notes:* (a) We regressed the change in performance points (Part 2 - Part 1) ( $\Delta$ Perf.) on advice received dummy in each task and each treatment using an OLS regression model with robust standard errors clustered at the participant level, controlling personal characteristics. The advice received dummy equals 1 if the participant receives advice for that graph, and 0 otherwise. (b) The unit of observation is the forecast from each participant in each round. The total number of observations is the total number of forecasts with/without advice received in each task in all treatments. (c) The robust standard errors are in parentheses. (d) Due to no submissions of predictions within the time limit, 12 observations were removed in Model (1), 10 observations in Model (2), 12 observations in Model (3), 4 observations in Model (4), 1 observations in Model (5) and 1 observations in Model (6).

Table B7: Change of net benefit (points) conditional on personal characteristics

VARIABLES	Advice received	(1)	(2)	(3)	(4)
		$\Delta$ NB. Task 1 No	$\Delta$ NB. Task 1 Yes	$\Delta$ NB. Task 2 No	$\Delta$ NB. Task 2 Yes
Baseline = Algorithm treatment dummy					
Expert treatment dummy		-0.357 (0.236)	0.836** (0.351)	0.057 (0.141)	1.253*** (0.289)
Student treatment dummy		-0.117 (0.195)	1.517*** (0.361)	-0.151 (0.131)	1.070*** (0.349)
Male dummy		0.004 (0.199)	-0.255 (0.338)	0.094 (0.119)	-0.168 (0.293)
Undergraduate student dummy		0.011 (0.175)	0.438 (0.300)	-0.052 (0.108)	0.166 (0.300)
Science major dummy		-0.112 (0.207)	-0.235 (0.382)	-0.156 (0.136)	-0.027 (0.313)
Stock trading experience dummy		-0.037 (0.301)	0.059 (0.484)	0.102 (0.183)	0.381 (0.424)
Financial literacy score		0.031 (0.039)	0.003 (0.069)	-0.041 (0.031)	-0.027 (0.068)
Constant		0.327 (0.425)	-2.991*** (0.616)	0.519* (0.286)	-3.030*** (0.624)
Observations		1,834	1,212	1,828	1,246
Clusters		296	252	302	259
R-squared		0.003	0.026	0.003	0.022
Treatments		Predicted $\Delta$ NB. Task 1	Predicted $\Delta$ NB. Task 2	Predicted $\Delta$ NB. Task 1	Predicted $\Delta$ NB. Task 2
Algorithm		0.502 (0.152)	-3.027 (0.234)	0.124 (0.103)	-3.221 (0.213)
Experts		0.145 (0.170)	-2.190 (0.256)	0.181 (0.093)	-1.973 (0.192)
Students		0.385 (0.117)	-1.509 (0.257)	-0.028 (0.084)	-2.151 (0.279)
		Prob > F	Prob > F	Prob > F	Prob > F
Algorithm = Experts = Students		0.302	0.000	0.206	0.000
Algorithm = Expert #			0.053		0.000
Algorithm = Student #			0.000		0.007
Expert = Student #			0.171		1.000

*Notes:* (a) We regressed the change in net benefit (Part 2 - Part 1) ( $\Delta$ NB.) on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level, controlling personal characteristics. (b) The unit of observation is the forecast from each participant in each round. The total number of observations is the total number of graphs with/without advice received in each task in all treatments. (c) The robust standard errors are in parentheses. (d) # indicates Bonferroni-adjusted p-values. (e) Due to no submissions of predictions within the time limit, 17 observations were removed in Model (1), 17 observations in Model (2), 4 observations in Model (3), and 2 observations in Model (4).

Table B8: Change in net benefit (points) when comparing with and without receiving advice conditional on personal characteristics

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	$\Delta$ NB. Task 1 Algorithm	$\Delta$ NB. Task 1 Expert	$\Delta$ NB. Task 1 Student	$\Delta$ NB. Task 2 Algorithm	$\Delta$ NB. Task 2 Expert	$\Delta$ NB. Task 2 Student
Advice received dummy	-3.580*** (0.285)	-2.375*** (0.292)	-1.846*** (0.256)	-3.366*** (0.245)	-2.129*** (0.195)	-2.114*** (0.295)
Male dummy	0.070 (0.339)	-0.234 (0.318)	-0.156 (0.263)	0.147 (0.244)	-0.033 (0.256)	-0.080 (0.246)
Undergraduate student dummy	0.562** (0.273)	0.286 (0.296)	-0.487** (0.227)	-0.230 (0.216)	0.168 (0.230)	0.110 (0.285)
Science major dummy	-0.409 (0.381)	-0.307 (0.398)	0.341 (0.254)	-0.415 (0.275)	-0.087 (0.275)	-0.165 (0.250)
Stock trading experience dummy	0.457 (0.362)	-0.210 (0.535)	-0.682 (0.522)	-0.267 (0.313)	0.930** (0.420)	0.099 (0.360)
Financial literacy score	-0.030 (0.070)	0.055 (0.067)	0.027 (0.050)	-0.042 (0.049)	-0.074 (0.069)	0.012 (0.069)
Constant	0.615 (0.735)	-0.054 (0.694)	0.425 (0.468)	0.856** (0.429)	0.643 (0.640)	-0.060 (0.642)
Observations	1,048	1,020	978	1,056	1,029	989
Clusters	106	103	99	106	103	99
R-squared	0.160	0.087	0.075	0.187	0.126	0.105

*Notes:* (a) We regressed the change in net benefit (Part 2 - Part 1) ( $\Delta$ NB.) on advice received dummy in each task and each treatment using an OLS regression model with robust standard errors clustered at the participant level, controlling personal characteristics. The advice received dummy equals 1 if the participant receives advice for that graph, and 0 otherwise. (b) The unit of observation is the forecast from each participant in each round. The total number of observations is the total number of forecasts with/without advice received in each task in all treatments. (c) The robust standard errors are in parentheses. (d) Due to no submissions of predictions within the time limit, 12 observations were removed in Model (1), 10 observations in Model (2), 12 observations in Model (3), 4 observations in Model (4), 1 observations in Model (5) and 1 observations in Model (6).

Table B9: Changes in WTP from Task 1 to Task 2 conditional on personal characteristics

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Algorithm $\Delta$ WTP	Expert $\Delta$ WTP	Student $\Delta$ WTP	Algorithm $\Delta$ WTP	Expert $\Delta$ WTP	Student $\Delta$ WTP
$\Delta$ Mean performance in Task 1	0.297** (0.113)	0.199*** (0.074)	-0.028 (0.111)			
Average cost per advice paid in Task 1	-0.807*** (0.152)	-1.149*** (0.234)	-0.599*** (0.210)			
$\Delta$ Net benefit in Task 1				0.421*** (0.103)	0.451*** (0.119)	0.230* (0.122)
Male dummy	-0.185 (0.359)	-0.075 (0.366)	0.021 (0.425)	-0.207 (0.356)	-0.312 (0.395)	0.249 (0.464)
Undergraduate student dummy	-0.720** (0.344)	-0.016 (0.449)	-0.749* (0.444)	-0.824** (0.371)	0.043 (0.588)	-0.591 (0.465)
Science major dummy	-0.580 (0.442)	-0.386 (0.515)	-0.441 (0.447)	-0.546 (0.460)	-0.080 (0.620)	-0.520 (0.485)
Stock trading experience dummy	-0.207 (0.508)	-0.027 (0.455)	-1.225 (0.977)	-0.381 (0.495)	0.051 (0.402)	-1.656 (1.153)
Financial literacy score	-0.029 (0.077)	-0.040 (0.104)	-0.110 (0.094)	0.006 (0.080)	-0.033 (0.113)	-0.139 (0.106)
Constant	3.384*** (0.934)	3.347*** (1.221)	2.814** (1.265)	2.108** (0.891)	1.038 (1.221)	1.882 (1.153)
Obs.	97	84	71	97	84	71
$R^2$	0.352	0.448	0.324	0.278	0.246	0.193

Notes: a) In models 1-3, we regressed  $\Delta$ WTP on  $\Delta$ Mean performance from graphs with advice using an OLS regression model, controlling for the average cost per advice paid in Task 1 and personal characteristics, with robust standard errors.  $\Delta$ WTP is calculated as WTP in Task 2 minus WTP in Task 1.  $\Delta$ Mean performance is calculated as the mean performance from graphs with advice in Part 2 of Task 1 minus the mean performance from graphs with advice in Part 1 of Task 1. Performance ranges from 0 to 20, based on the points obtained in each question. In models 4-6, we regressed  $\Delta$ WTP on  $\Delta$ Net benefit from graphs with advice using an OLS regression model controlling personal characteristics, with robust standard errors.  $\Delta$ Net benefit is calculated as the mean net benefit from graphs with advice in Part 2 of Task 1 minus the mean net benefit from graphs with advice in Part 1 of Task 1. The net benefit in each graph in Part 1 equals the rewards in each graph plus 10 points. The net benefit in each graph in Part 2 equals the rewards in each graph plus 10 points minus the price of advice. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively.

b) The unit of observation is the participant who received at least one piece of advice in Task 1. The total number of observations is the number of participants who received at least one piece of advice in Task 1 in each treatment.

c) Numbers in parentheses represent standard errors.

Table B10: Price of Advice and Shift Rate conditional on personal characteristics

VARIABLES	(1) Algorithm SHIFT	(2) Expert SHIFT	(3) Student SHIFT
Task 2 dummy	0.007 (0.021)	-0.049* (0.026)	-0.072** (0.033)
WTP	0.004 (0.008)	0.000 (0.011)	0.009 (0.013)
Male dummy	-0.035 (0.044)	-0.051 (0.046)	0.012 (0.049)
Undergraduate student dummy	0.030 (0.038)	0.064 (0.045)	-0.006 (0.049)
Science major dummy	0.028 (0.046)	0.024 (0.054)	0.038 (0.045)
Stock trading experience dummy	0.091* (0.048)	0.054 (0.054)	-0.007 (0.073)
Financial literacy score	-0.009 (0.010)	-0.001 (0.011)	0.004 (0.012)
Constant	0.829*** (0.103)	0.817*** (0.125)	0.461*** (0.120)
Obs.	929	779	504
Clusters	101	93	78
$R^2$	0.025	0.026	0.024

*Notes:* a) We used the data of the graphs which participants received the advice. We also used the data which shift rate is between 0 and 1. We regressed shift rate on price of advice in each graph, using an OLS regression model, controlling for the Task 2 dummy, WTP and personal characteristics, with robust standard errors clustered at participant level. Price indicates the random price of each graphs in each task. Task 2 dummy equals to 1 for Task 2 and 0 for Task 1. WTP indicates the submitted WTP in each task. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively.

b) The unit of observation is the graph with received advice of each participant. The total number of observations is the number of graphs with received advice in Task 1 and Task 2 in each treatment.

c) Numbers in parentheses represent standard errors.

d) Due to no submissions of predictions within the time limit, 8 observations were removed in Model (1), 4 observations in Model (2), 2 observations in Model (3), 8 observations in Model (4), 4 observations in Model (5) and 2 observations in Model (6).

## C Other results in Experiment 1

### C.1 Performance of advice

Figure C1 illustrates the average performance of the three information sources, along with freely available information; that is, the last price shown in the graph. Panel (a) displays the mean absolute percentage error (APE), calculated over 10 forecasts in each task from each source. The mean APE for a given source  $s$  is defined as

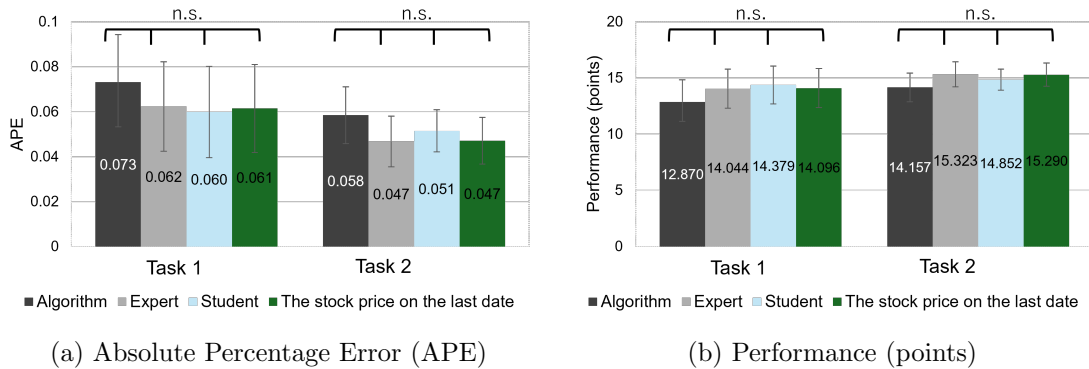
$$\text{Mean APE}^s = \frac{1}{10} \sum_{g=1}^{10} \left| \frac{\text{Advice}_g^s - \text{True value}_g}{\text{True value}_g} \right|,$$

where  $\text{Advice}_g^s$  is the advice provided by source  $s$  for graph  $g$  and  $\text{True value}_g$  denotes the actual target value for graph  $g$ . The mean APE is calculated by averaging the 10 graphs in each task. As this measure represents the deviation of the advice from the true value, a higher APE indicates lower performance.

Panel (b) illustrates the average performance in terms of points participants could have earned if they had fully adopted the advice. Specifically, the mean performance points for source  $s$  are defined as follows, where  $(\cdot)^+$  denotes  $\max(\cdot, 0)$ :

$$\text{Mean performance points}^s = \frac{1}{10} \sum_{g=1}^{10} \left( 20 - \left| \frac{\text{Advice}_g^s - \text{True value}_g}{\text{True value}_g} \right| \times 100 \right)^+$$

Thus, a higher mean performance point indicates better performance. This metric reflects the potential points participants could have gained by aligning their forecasts fully with the provided advice.



*Notes:* We regressed (a) the APE of advice and (b) the performance of advice in points on four dummy variables (three treatment dummies and last price dummy) using an OLS regression model with robust standard errors. We compared the estimated dummy coefficients using an F-test, with results illustrated by  $p$ -values. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively, and n.s. denotes that the difference is not statistically significant at the 0.1 level.

Figure C1: Performance of advice

As is evident from Figure C1, although the performance—as measured by both mean APE and mean performance points—does not significantly differ across the three information sources in either task, it is worst in **Algorithm**. The performance is similar in **Expert** and **Student** across both tasks. Even these two better-performing sources, however, are not any better than the last observed price, which is freely available.

Consequently, participants unknowingly expressed the highest WTP for advice from the lowest-performing source. This tendency likely has adverse implications, both in terms of the accuracy of their final forecasts (submitted in part 2) and in the payoffs they ultimately received.

Table C1: Performance of advice in Task 1 and Task 2 (for Figure 8)

VARIABLES	(1)	(2)	(3)	(4)
	APE Task 1	APE Task 2	Points Task 1	Points Task 2
Baseline = Algorithm treatment dummy				
Expert treatment dummy	-0.011 (0.029)	-0.012 (0.017)	1.174 (2.632)	1.166 (1.695)
Student treatment dummy	-0.013 (0.029)	-0.007 (0.016)	1.509 (2.597)	0.695 (1.576)
Closing price dummy	-0.012 (0.029)	-0.011 (0.016)	1.226 (2.628)	1.133 (1.638)
Constant	0.073*** (0.021)	0.058*** (0.013)	12.870*** (1.975)	14.157*** (1.268)
Observations	40	40	40	40
R-squared	0.007	0.020	0.012	0.020
Treatments	Predicted APE Task 1	Predicted APE Task 2	Predicted Points Task 1	Predicted Points Task 2
Algorithm	0.073 (0.021)	0.058 (0.013)	12.870 (1.975)	14.157 (1.268)
Experts	0.062 (0.020)	0.047 (0.011)	14.044 (1.740)	15.323 (1.124)
Students	0.060 (0.020)	0.051 (0.009)	14.379 (1.687)	14.852 (0.936)
Closing prices	0.061 (0.020)	0.047 (0.010)	14.096 (1.733)	15.290 (1.037)
	Prob > F	Prob > F	Prob > F	Prob > F
Algorithm = Experts = Students = Closing prices	0.968	0.891	0.945	0.891

*Notes:* (a) We regressed the APE of advice / the performance of advice in points on three treatment dummies using an OLS regression model with robust standard errors. (b) The unit of observation is the prediction of a graph from each information source. The total number of observations is the total number of graphs in all treatments. (c) The robust standard errors are in parentheses.

## C.2 Performance of advice received by participants

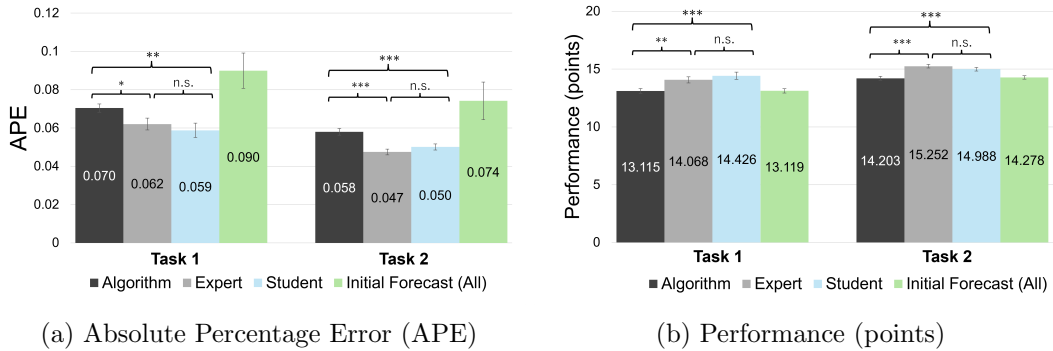


Figure C2: Performance of advice received by participants

*Notes:* We calculated the average performance of the advice received by participants in each task across the three treatments, with one observation in Task 1 and one observation in Task 2 for each participant. We regressed (a) the APE of received advice and (b) the performance of received advice in points on three treatment dummy variable using an OLS regression model with robust standard errors. Pairwise comparisons of the estimated dummy coefficients were performed using an F-test with Bonferroni-adjusted p-values. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively, and n.s. denotes that the difference is not statistically significant at the 0.1 level.

Figure C2 illustrates the average performance of the advice received by participants across the three treatments. In both tasks, the average performance of the received advice, measured by APE and performance points, is worse in the **Algorithm** treatment than in the **Expert** and **Student** treatments, while performance is similar between the **Expert** and **Student** treatments.

Participants' initial forecasts in the rounds in which they received advice perform worse than the received advice across all treatments. To compare the performance of received advice with participants' own initial forecasts, we examine the difference between the performance of the advice received and the performance of participants' initial forecasts. The detailed results are shown in Figure C3.

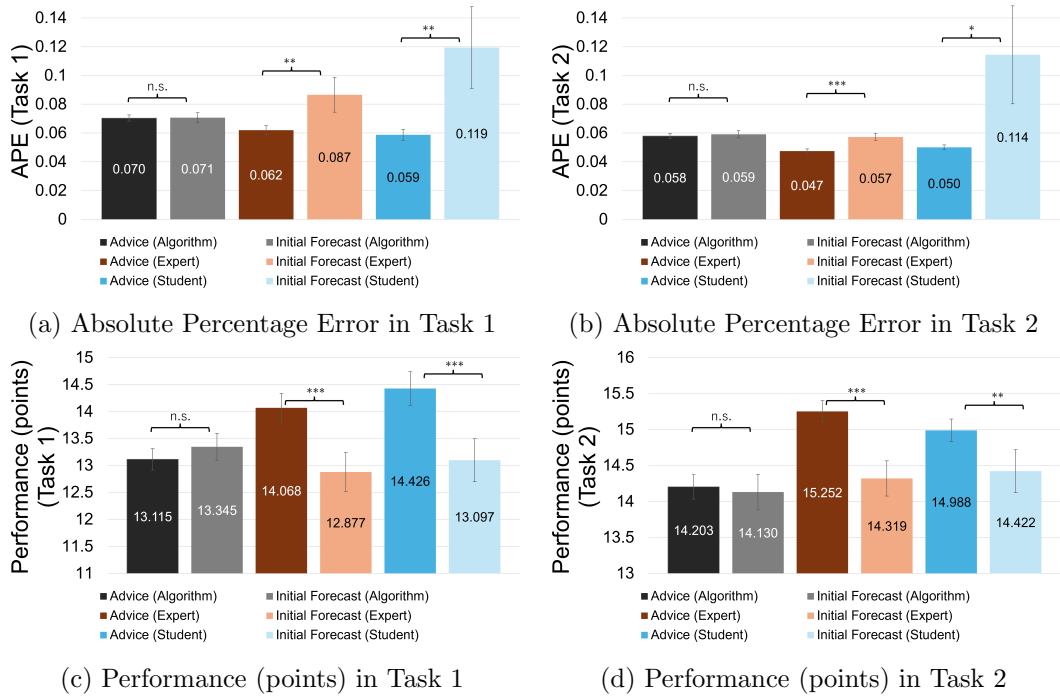
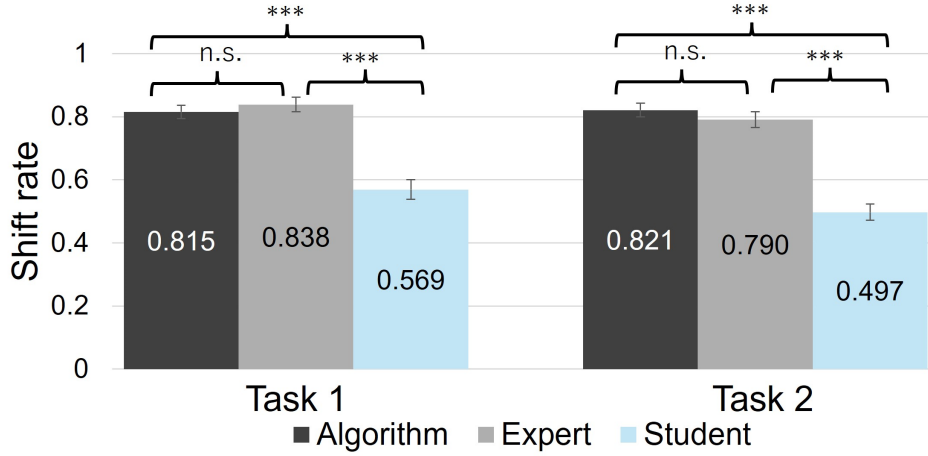


Figure C3: Performance of Advice Received vs. Participants' Initial Forecasts

*Notes:* For each participant, we calculated the average performance of the advice received in each task, measured by APE and performance points. We also calculated the average performance of participants' initial forecasts in the rounds in which they received advice, using the same two measures. Finally, for each participant, we computed the difference between the average performance of the received advice and the average performance of participants' initial forecasts. We estimate separate OLS regressions of these difference measures on a constant for each treatment and task, using robust standard errors. Specifically, we examine (a) the difference in APE between received advice and initial forecasts and (b) the difference in performance points between received advice and initial forecasts. The constant term represents the mean difference within each treatment-task cell. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively, and n.s. indicates that the difference is not statistically significant at the 0.1 level.

### C.3 Utilization of Advice: Shift rate



*Notes:* We regressed the shift rate on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level. Pair-wise comparisons of the estimated dummy coefficients are done by an F-test with Bonferroni-adjusted  $p$ -values. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively, and n.s. means that the difference is not statistically significant at the 0.1 level.

Figure C4: Mean shift rate

Let us now examine how the degree of advice use varies across different information sources.<sup>2</sup> We quantify the degree of utilization using the *shift rate* which is a common measure in the literature on information use (Harvey and Fischer, 1997; Sniezek et al., 2004).<sup>3</sup> For each participant  $i$  and each graph  $g$ , the shift rate is defined as

$$\text{shift rate}_g^i = \frac{\text{Final forecast}_g^i - \text{Initial forecast}_g^i}{\text{Advice}_g^i - \text{Initial forecast}_g^i},$$

where *Initial forecast* $_g^i$  and *Final forecast* $_g^i$  are the forecasts submitted by participant  $i$  for graph  $g$  in parts 1 and 2, respectively, and *Advice* $_g^i$  is the advice provided from the information source for graph  $g$ .

A shift rate of 0 indicates that participant  $i$  did not alter their forecast between parts 1 and 2, regardless of the advice received. Conversely, a shift rate of 1 signifies that participant  $i$  fully incorporated the advice, aligning their part 2 forecast exactly with the advice. Although shift rates can be less than 0 (indicating that the forecast in part 2 moved in the opposite direction of the advice) or greater than 1 (indicating that the forecast in part 2 exceeded the advice received), our focus is on cases where the shift

<sup>2</sup>The analyses reported in this subsection were not pre-registered. We decided to include them based on feedback received during seminars and conferences.

<sup>3</sup>The shift rate in this paper is calculated in the same way as the weight of advice in the JAS paradigm, as used in previous studies (see, for example, the literature summarized in Bailey et al. (2023); Bonaccio and Dalal (2006)).

Table C2: Price of Advice and Shift Rate

VARIABLES	(1)	(2)	(3)
	Algorithm SHIFT	Expert SHIFT	Student SHIFT
WTP	0.007 (0.009)	-0.004 (0.012)	0.008 (0.013)
Task 2 dummy	0.006 (0.021)	-0.049* (0.027)	-0.071** (0.033)
Constant	0.772*** (0.055)	0.861*** (0.065)	0.528*** (0.064)
Obs.	929	779	522
Clusters	101	93	80
$R^2$	0.003	0.008	0.016

*Notes:* a) We use data for graphs for which participants received advice. We also restrict the sample to observations where the shift rate is between 0 and 1. We regress shift rate on the price of advice for each graph using an OLS model, controlling for the Task 2 dummy and WTP, with robust standard errors clustered at the participant level. Task 2 dummy equals 1 for Task 2 and 0 for Task 1. WTP indicates the submitted WTP in each task. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively.

b) The unit of observation is a graph for which advice was received by a participant. The total number of observations is the number of such graphs in Task 1 and Task 2 in each treatment.

c) Numbers in parentheses are standard errors.

rate falls between 0 and 1 inclusive.

Figure C4 displays the mean shift rates for the three treatments in Task 1 (left) and Task 2 (right). For both *Algorithm* and *Expert*, the mean shift rates are approximately 0.8 across both tasks, with no statistically significant difference between the two treatments. However, the mean shift rates for *Student* are significantly lower, at 0.569 and 0.497 in Tasks 1 and 2, respectively. This suggests that participants are not only willing to pay significantly less for advice from their peers but also tend to incorporate this advice less when they receive it.

We now turn to investigating whether the shift rate is influenced by the prices participants paid or their WTPs and whether these relationships vary by the source of advice. Table C2 presents the results of linear regressions where the shift rate is the dependent variable and the price paid for advice is an independent variable. In Models (4) through (6), WTP is also included as an independent variable.

Given the information source, the estimated coefficients of the price paid are not significantly different from zero in any of the models shown in Table C2. The insignificance of the estimated coefficients is also true for WTP in Models (4) through (6). Furthermore, the shift rate is significantly lower in Task 2 for both *Expert* and *Student*, as indicated by the significant negative coefficients for the Task 2 dummy in Models (2), (3), (5), and (6). Thus, controlling for the information source, neither a higher price paid nor a greater WTP leads to increased use of a given piece of advice.

By observing the shift rate, we can understand how participants rely on different

sources of advice and how they form expectations about the accuracy of that advice. We expect participants to be more likely to rely on advice when they anticipate it to be more accurate. Accordingly, we also expect that participants would be willing to pay more for advice they believe to be more accurate. Interestingly, in Models (4), (5), and (6), we observed no significant correlation between WTP and shift rate. Additionally, we found no correlation between changes in shift rate and changes in WTP from part 1 to part 2 in any treatments ( $p = 0.126$  for Algorithm,  $p = 0.153$  for Expert, and  $p = 0.355$  for Student).<sup>4</sup>

Table C3: Predicted shift rate in Task 1 and Task 2 (for Figure 7)

VARIABLES	(1)	(2)
	Shift rate Task 1	Shift rate Task 2
Baseline = Algorithm treatment dummy		
Expert treatment dummy	0.023 (0.031)	-0.031 (0.033)
Student treatment dummy	-0.246*** (0.038)	-0.324*** (0.033)
Constant	0.815*** (0.021)	0.821*** (0.022)
Observations	1,106	1,124
Clusters	251	259
R-squared	0.124	0.178
Treatments	Predicted shift rate Task 1	Predicted shift rate Task 2
Algorithm	0.815 (0.021)	0.821 (0.022)
Experts	0.838 (0.023)	0.790 (0.025)
Students	0.569 (0.031)	0.497 (0.026)
	Prob > F	Prob > F
Algorithm = Experts = Students	0.000	0.000
Algorithm = Expert #	1.000	1.000
Algorithm = Student #	0.000	0.000
Expert = Student #	0.000	0.000

*Notes:* (a) We used the data of the graphs which participants received the advice. We also used the data which shift rate is between 0 and 1. We regressed the shift rate on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level. (b) The unit of observation is the graph with received advice of each participant. The total number of observations is the number of graphs with received advice in each task in all treatments. (c) The robust standard errors are in parentheses. (d) # indicates Bonferroni-adjusted p-values. (e) 16 observations were removed in Task 1 and 2 observations were removed in Task 2 due to no submissions of predictions within the time limit.

<sup>4</sup>This analysis focused on cases where the shift rate is between 0 and 1. We regressed  $\Delta$ Shift rate on  $\Delta$ WTP using an OLS regression model with robust standard errors. Here,  $\Delta$ Shift rate is calculated as the mean shift rate for graphs with advice in Task 2 minus the mean shift rate for graphs with advice in Task 1, and  $\Delta$ WTP is calculated as WTP in Task 2 minus WTP in Task 1.

Table C4: Change in shift rate (Part 2 - Part 1)  
vs. Changes in WTP (Part 2 - Part 1)

	(1)	(2)	(3)
	$\Delta\text{Shift}$	$\Delta\text{Shift}$	$\Delta\text{Shift}$
VARIABLES	Algorithm	Expert	Student
$\Delta\text{WTP}$	0.019 (0.012)	-0.028 (0.019)	0.019 (0.020)
Constant	-0.004 (0.020)	-0.028 (0.021)	-0.059 (0.039)
Observations	93	78	65
R-squared	0.033	0.061	0.006

*Notes:* (a) We regressed the change in shift rate (Part 2 - Part 1) ( $\Delta\text{Shift}$ ) on change in WTP (Part 2 - Part 1) ( $\Delta\text{WTP}$ ) in each treatment using an OLS regression model with robust standard errors.  $\Delta\text{Shift}$  is calculated as Mean shift rate in Task 2 minus mean shift rate in Task 1. We only include data in which the shift rate ranges between 0 and 1. (b) The unit of observation is the participant who received at least one piece of advice in Task 1 and Task 2, respectively. The total number of observations is the number of participants who received at least one piece of advice in Task 1 and Task 2 in each treatment, respectively. (c) The robust standard errors are in parentheses.

### C.3.1 Shift rate without restriction

In the manuscript, our analysis focuses on observations with shift rates between 0 and 1. Shift rates below 0 or above 1 capture movements away from the advice, or beyond it, and therefore reflect different behavioral responses.

In this section, we present results without imposing the restriction that the shift rate lies between 0 and 1. That is, we include all observations, including those with shift rates below 0 or above 1. However, we exclude observations with extremely large submitted predictions. Specifically, in Task 1, 9 initial-forecast observations and 5 final-forecast observations are greater than 700; in Task 2, 4 initial-forecast observations and 5 final-forecast observations are greater than 700. Because these extreme predictions generate unusually large shift rates, we treat observations with shift rates greater than 60 or less than  $-60$  as outliers and exclude them from the analysis. This excludes 4 observations in total: 2 in the **Algorithm** treatment (2 in Task 2), 1 in the **Expert** treatment (1 in Task 1), and 1 in the **Student** treatment (1 in Task 1).

After excluding outliers, the number of observations with shift rates greater than 1 is 90 in the **Algorithm** treatment (43 in Task 1 and 47 in Task 2), 56 in the **Expert** treatment (31 in Task 1 and 25 in Task 2), and 83 in the **Student** treatment (38 in Task 1 and 45 in Task 2). The number of observations with shift rates below 0 is 8 in the **Algorithm** treatment (4 in Task 1 and 4 in Task 2) and 8 in the **Student** treatment (1 in Task 1 and 7 in Task 2). The number of observations with shift rates between 0 and 1 is 929 in the **Algorithm** treatment (480 in Task 1 and 449 in Task 2), 779 in the **Expert** treatment (370 in Task 1 and 409 in Task 2), and 522 in the **Student** treatment (256 in Task 1 and 266 in Task 2).

Table C5: Predicted shift rate without restriction in Task 1 and Task 2

VARIABLES	(1)	(2)
	Shift rate Task 1	Shift rate Task 2
Baseline = Algorithm treatment dummy		
Expert treatment dummy	0.022 (0.030)	0.051 (0.080)
Student treatment dummy	-0.172*** (0.037)	-0.052 (0.193)
Constant	0.835*** (0.020)	0.760*** (0.076)
Observations	1,223	1,252
Clusters	254	261
R-squared	0.054	0.000
Treatments	Predicted shift rate Task 1	Predicted shift rate Task 2
Algorithm	0.835 (0.020)	0.760 (0.076)
Experts	0.857 (0.022)	0.811 (0.025)
Students	0.663 (0.031)	0.708 (0.177)
	Prob > F	Prob > F
Algorithm = Experts = Students	0.000	0.701
Algorithm = Expert #	1.000	
Algorithm = Student #	0.000	
Expert = Student #	0.000	

*Notes:* (a) We used the data of the graphs which participants received the advice. We also used the data which shift rate is between  $-60$  and  $60$ . We regressed the shift rate on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level. (b) The unit of observation is the graph with received advice of each participant. The total number of observations is the number of graphs with received advice in each task in all treatments. (c) The robust standard errors are in parentheses. (d) # indicates Bonferroni-adjusted p-values. (e) 16 observations were removed in Task 1 and 2 observations were removed in Task 2 due to no submissions of predictions within the time limit. We removed 4 outlier observations with extremely small or large shift rates, which were driven by extreme submitted predictions.

## C.4 Maximum WTP to increase net benefit

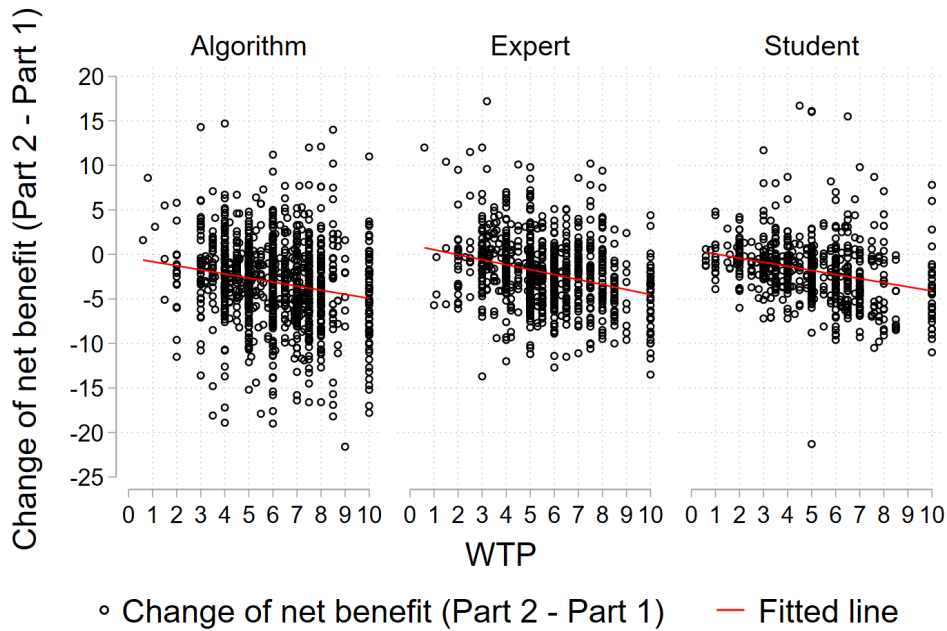


Figure C5: WTP vs. Change of net benefit in points (part 2 – part 1)

Figure C5 shows a scatter plot of WTP and corresponding changes in net benefits, as measured in points (part 2 – part 1), across treatments. Data from both Tasks 1 and 2 are incorporated. The red line represents the linear fit of the change in net benefit on WTP. To estimate this relationship, we regressed the change in net benefits on WTP using an OLS regression model with robust standard errors clustered at the participant level. From this model, we determined the WTP value at which the average change in net benefit equals zero.

The maximum WTP required to achieve an average non-decrease in net benefit was 0 for Algorithm, 1.942 for Expert, and 1.110 for Student. This result suggests that participants, on average, overpaid for advice from all information sources, as their average WTP exceeded the threshold required to increase their net benefits.

## C.5 Potential drivers of Variation in WTP across information sources

In Task 1, when participants had no prior information about the performance of the advice, the average WTP was approximately 5, 4, and 3 in Algorithm, Expert, and Student, respectively. These differences in WTP across treatments may reflect participants' perceptions of the performance of each advice source, as well as their personal inherent preferences regarding advice from different sources. Unfortunately, our experiment does not allow us to separately quantify the effect of these two possible reasons.

After gaining experience, participants who received a positive payoff from purchasing

advice in Task 1 increased their WTP in Task 2 compared to Task 1. After controlling for the average change in net benefit between the two parts in Task 1, we find that WTP increased, on average, by 1.192 in *Algorithm*, 0.594 in *Expert*, and 0.252 in *Student*. Although the differences in the increase in WTP across the three treatments are not statistically significant ( $p = 0.227$ ), it does suggest that the advice was appreciated more in *Algorithm* than in *Expert* and *Student* after gaining similar positive experience relying on it.<sup>5</sup>

We also observe seemingly irrational behavior. Despite experiencing a decrease in payoff from purchasing advice in Task 1, some participants increased their WTP in Task 2. The proportion of participants who raised their WTP after observing the negative net benefit from advice in Task 1 was 30%, 22%, and 18% in *Algorithm*, *Expert*, and *Student*. Although the differences across the three treatments are not statistically significant ( $p = 0.102$ ), a larger proportion of participants increased their WTP for advice in *Algorithm* even after incurring a loss, relative to those who received advice in *Expert* or *Student*.<sup>6</sup> These patterns suggest that participants tend to prefer advice from an algorithm over advice from human beings, regardless of actual performance.

---

<sup>5</sup>Using data from participants who gained a positive payoff from purchasing advice in Task 1, we regressed the change in WTP ( $\Delta$ WTP) on three treatment dummies using an OLS regression model with robust standard errors, controlling the average change in net benefit between two parts after receiving the advice. Comparisons of the estimated dummy coefficients were performed using F-tests.

<sup>6</sup>An irrational dummy variable was constructed, taking a value of 1 for participants who increased their WTP in Task 2 relative to Task 1, despite observing a negative net benefit from advice in Task 1, and 0 otherwise. We regressed the irrational dummy on three treatment dummies using a logit regression model with robust standard errors. Comparisons of the estimated dummy coefficients were performed using F-tests.

## C.6 Change of absolute percentage error (APE)

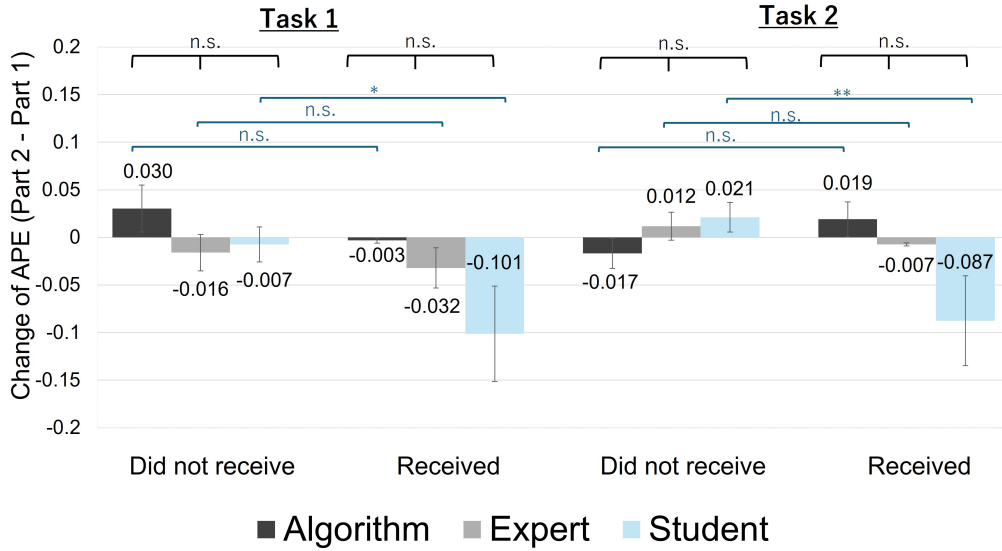


Figure C6: Change of APE

*Notes:* Outliers (change of APE  $\leq -20$  or  $\geq 20$ ) are removed in this analysis, resulting in the removal of three observations. We regressed the change in APE (Part 2 - Part 1) on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level. Pairwise comparisons of the estimated dummy coefficients were performed using an F-test with Bonferroni-adjusted p-values. We also regressed the change in performance points (Part 2 - Part 1) on an advice-receiving dummy in each treatment using an OLS regression model with robust standard errors clustered at the participant level. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively, while n.s. denotes that the difference is not statistically significant at the 0.1 level.

Figure C6 illustrates the average change in absolute percentage error (APE) from Part 1 to Part 2 in Task 1 (left) and Task 2 (right) for each of the three treatments. In both tasks, the performance improvement from Part 1 to Part 2 among participants who received advice is better than that of participants who did not receive advice (except for those in the algorithm treatment in Task 2). These results are significant for advice from algorithms and students in Task 1, and for advice from students in Task 2.

## C.7 WTP vs. Net benefit in Part 2

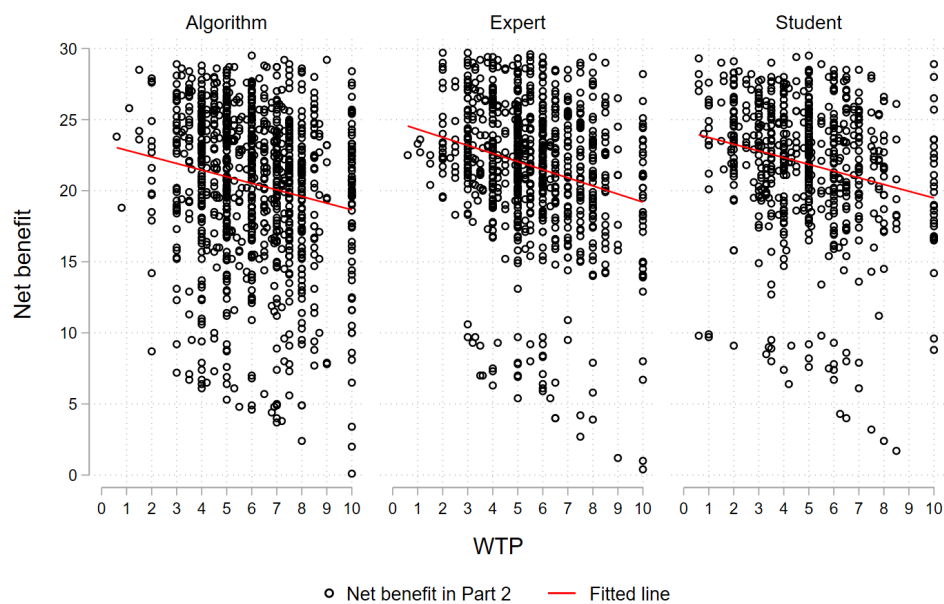


Figure C7: WTP vs. Change of performance (points)

Figure C7 presents a scatter plot of WTP values and their corresponding net benefits in Part 2 for both tasks, measured in points (Part 2 - Part 1), across treatments, incorporating data from both Task 1 and Task 2. The red line represents the linear fit of net benefits in Part 2 (points) on WTP. Across all treatments, net benefits in Part 2 decrease as WTP increases, after paying and receiving for the advice..

## C.8 WTP vs. Change of performance (points)



Figure C8: WTP vs. Change of performance (points)

Figure C8 presents a scatter plot of WTP values and their corresponding changes in performance, measured in points (Part 2 - Part 1), across treatments, incorporating data from both Task 1 and Task 2. The red line represents the linear fit of the change in performance (points) on WTP. To estimate this relationship, we regressed the change in performance (points) on WTP using an OLS regression model with robust standard errors clustered at the participant level. Based on this model, we identified the WTP value at which the change in performance (points) equals zero.

The minimum WTP required to achieve an increase in performance (points) was 6.094 for the algorithm. For both the expert and student treatments, all observed WTP values resulted in an increase in performance (points).

## C.9 Changes in Performance and Net Benefit among Participants Who Received Advice Frequently and Infrequently

Among participants who received at least one piece of advice in each task, the average number of pieces of advice received was 4.427 in Task 1 and 4.440 in Task 2. Based on these averages, we classify participants who received five or more pieces of advice as frequent advice recipients, and those who received fewer than five pieces of advice as infrequent advice recipients.

For rounds in which participants received advice, the WTP of infrequent advice recipients is lower than that of frequent advice recipients in Task 1 across all treatments,

with the differences statistically significant at the 1% level. In Task 1, the WTP of infrequent advice recipients is 4.043 in *Algorithm*, 3.745 in *Expert*, and 3.266 in *Student*. The WTP of frequent advice recipients is 6.752 in *Algorithm*, 6.632 in *Expert*, and 6.109 in *Student*.

In Task 2, the WTP of infrequent advice recipients is also lower than that of frequent advice recipients across all treatments, with the differences statistically significant at the 5% level. The WTP of infrequent advice recipients is 5.623 in *Algorithm*, 5.000 in *Expert*, and 4.317 in *Student*. The WTP of frequent advice recipients is 6.417 in *Algorithm*, 5.960 in *Expert*, and 5.793 in *Student*.

Table C6: Change of performance (points) between frequent and infrequent advice recipients

VARIABLES	Advice received	(1)	(2)	(3)	(4)
		$\Delta$ Perf. Task 1 Yes Infrequent	$\Delta$ Perf. Task 1 Yes Frequent	$\Delta$ Perf. Task 2 Yes Infrequent	$\Delta$ Perf. Task 2 Yes Frequent
Baseline = Algorithm treatment dummy					
Expert treatment dummy		0.886*	0.838**	0.834**	0.705**
		(0.477)	(0.322)	(0.334)	(0.305)
Student treatment dummy		0.585	1.375***	0.644*	0.431
		(0.426)	(0.322)	(0.378)	(0.382)
Constant		0.378	-0.032	-0.038	-0.101
		(0.352)	(0.243)	(0.249)	(0.233)
Observations		352	873	619	635
Clusters		127	127	140	121
R-squared		0.010	0.018	0.010	0.008
Treatments					
		Predicted $\Delta$ Perf. Task 1	Predicted $\Delta$ Perf. Task 2	Predicted $\Delta$ Perf. Task 1	Predicted $\Delta$ Perf. Task 2
Algorithm		0.378	-0.032	-0.038	-0.101
		(0.352)	(0.243)	(0.249)	(0.233)
Experts		1.264	0.805	0.796	0.604
		(0.323)	(0.212)	(0.223)	(0.196)
Students		0.964	1.342	0.606	0.330
		(0.240)	(0.211)	(0.284)	(0.303)
		Prob > F	Prob > F	Prob > F	Prob > F
Algorithm = Experts = Students		0.175	0.000	0.041	0.073
Algorithm = Expert #			0.031	0.041	0.067
Algorithm = Student #			0.000	0.272	0.785
Expert = Student #			0.225	1.000	1.000

*Notes:* (a) We regressed the change in performance points (Part 2 - Part 1) ( $\Delta$ Perf.) on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level. (b) The unit of observation is the forecast from each participant in each round. The total number of observations is the total number of graphs with advice received in each task from infrequent or frequent advice recipients in all treatments. (c) The robust standard errors are in parentheses. (d) # indicates Bonferroni-adjusted p-values.

Table C7: Change in performance (points) when comparing frequent and infrequent advice recipient

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	$\Delta$ Perf.	$\Delta$ Perf.	$\Delta$ Perf.	$\Delta$ Perf.	$\Delta$ Perf.	$\Delta$ Perf.
	Task 1	Task 1	Task 1	Task 2	Task 2	Task 2
	Algorithm	Expert	Student	Algorithm	Expert	Student
Frequent advice receipt dummy	-0.411	-0.459	0.379	-0.064	-0.192	-0.276
	(0.427)	(0.386)	(0.321)	(0.342)	(0.298)	(0.417)
Constant	0.378	1.264***	0.964***	-0.038	0.796***	0.606**
	(0.351)	(0.323)	(0.241)	(0.249)	(0.224)	(0.285)
Observations	527	402	296	502	434	318
Clusters	97	84	73	97	90	74
R-squared	0.001	0.003	0.003	0.000	0.001	0.002

*Notes:* (a) We regressed the change in performance points (Part 2 - Part 1) ( $\Delta$ Perf.) on frequent advice recipient dummy in each task and each treatment using an OLS regression model with robust standard errors clustered at the participant level. The frequent advice receipt dummy equals 1 if the participant receives advice 5 pieces or more, and 0 otherwise. (b) The unit of observation is the forecast from each participant in each round. The total number of observations is the total number of forecasts with advice received in each task in all treatments. (c) The robust standard errors are in parentheses.

Table C8: Change of net benefit (points) between frequent and infrequent advice recipients

VARIABLES	Advice received	(1)	(2)	(3)	(4)
		$\Delta$ NB. Task 1 Yes Infrequent	$\Delta$ NB. Task 1 Yes Frequent	$\Delta$ NB. Task 2 Yes Infrequent	$\Delta$ NB. Task 2 Yes Frequent
Baseline = Algorithm treatment dummy					
Expert treatment dummy		0.855*	0.820**	1.355***	1.060***
		(0.515)	(0.406)	(0.409)	(0.382)
Student treatment dummy		1.089**	1.649***	1.231***	0.626
		(0.445)	(0.410)	(0.457)	(0.484)
Constant		-1.719***	-3.520***	-2.932***	-3.421***
		(0.370)	(0.272)	(0.285)	(0.296)
Observations		352	873	619	635
Clusters		127	127	140	121
R-squared		0.015	0.020	0.024	0.014
Treatments					
		Predicted $\Delta$ NB. Task 1	Predicted $\Delta$ NB. Task 2	Predicted $\Delta$ NB. Task 1	Predicted $\Delta$ NB. Task 2
Algorithm		-1.719	-3.520	-2.932	-3.421
		(0.370)	(0.272)	(0.285)	(0.296)
Experts		-0.864	-2.700	-1.577	-2.361
		(0.359)	(0.302)	(0.293)	(0.242)
Students		-0.630	-1.870	-1.701	-2.795
		(0.248)	(0.307)	(0.357)	(0.383)
		Prob > F	Prob > F	Prob > F	Prob > F
Algorithm = Experts = Students		0.052	0.001	0.002	0.024
Algorithm = Expert #		0.299	0.137	0.004	0.019
Algorithm = Student #		0.048	0.000	0.024	0.596
Expert = Student #		1.000	0.168	1.000	1.000

Notes: (a) We regressed the change in net benefit (Part 2 - Part 1) ( $\Delta$ NB.) on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level. (b) The unit of observation is the forecast from each participant in each round. The total number of observations is the total number of graphs with advice received in each task from frequent advice recipient or infrequent advice recipient in all treatments. (c) The robust standard errors are in parentheses. (d) # indicates Bonferroni-adjusted p-values.

Table C9: Change in net benefit (points) when comparing infrequent and frequent advice recipient

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	$\Delta$ NB. Task 1 Algorithm	$\Delta$ NB. Task 1 Expert	$\Delta$ NB. Task 1 Student	$\Delta$ NB. Task 2 Algorithm	$\Delta$ NB. Task 2 Expert	$\Delta$ NB. Task 2 Student
Frequent advice recipient dummy	-1.800*** (0.459)	-1.836*** (0.470)	-1.240*** (0.396)	-0.489 (0.411)	-0.784** (0.381)	-1.094** (0.526)
Constant	-1.719*** (0.369)	-0.864** (0.359)	-0.630** (0.249)	-2.932*** (0.285)	-1.577*** (0.294)	-1.701*** (0.359)
Observations	527	402	296	502	434	318
Clusters	97	84	73	97	90	74
R-squared	0.024	0.040	0.027	0.003	0.011	0.022

*Notes:* (a) We regressed the change in net benefit (Part 2 - Part 1) ( $\Delta$ NB.) on advice received dummy in each task and each treatment using an OLS regression model with robust standard errors clustered at the participant level. The frequent advice receipt dummy equals 1 if the participant receives advice 5 pieces or more, and 0 otherwise. (b) The unit of observation is the forecast from each participant in each round. The total number of observations is the total number of forecasts with advice received in each task in all treatments. (c) The robust standard errors are in parentheses.

### C.10 Comparing Constant Terms in Table 3

In Table 3 of the manuscript, models (1) through (3) regress the change in WTP from Task 1 to Task 2 on the performance difference associated with the advice received in Task 1 and the number of pieces of advice received. These models are estimated using the subsample of participants who received at least one piece of advice in Task 1. The estimated constant terms are positive and statistically significant across all three treatment groups. Since the dependent variable is  $WTP_{T_2} - WTP_{T_1}$ , these constants indicate that the predicted change in WTP is positive when the included covariates are set to zero. Specifically, the constants capture the predicted change in WTP for participants with zero change in mean performance after receiving advice in Task 1 and zero pieces of advice received.

The magnitude of the constant term is largest for **Expert**, followed by **Algorithm** and **Student**. Pairwise comparisons of the constants across treatment-specific models indicate a statistically significant difference between **Algorithm** and **Student** ( $p = 0.035$ ), but no statistically significant difference between **Expert** and **Algorithm** ( $p = 1.000$ ) or between **Expert** and **Student** ( $p = 0.121$ ).<sup>7</sup>

In models (4) through (6), the dependent variable remains the change in WTP from Task 1 to Task 2, but the key explanatory variable is the change in net benefit after receiving advice in Task 1. These models are also estimated using the subsample of participants who received at least one piece of advice in Task 1. The estimated constant terms are again positive and statistically significant across all three treatment groups. Formally, the constants represent the predicted change in WTP when the change in net

<sup>7</sup>Comparisons of the constant terms across models were performed using F-tests, with standard errors clustered at the session level. Bonferroni-adjusted  $p$ -values are reported for the pairwise comparisons.

benefit after receiving advice in Task 1 and the number of pieces of advice are both equal to zero.

The magnitude of the constant term is largest for **Expert**, followed by **Algorithm** and **Student**. Pairwise comparisons indicate statistically significant differences between **Expert** and **Student** ( $p = 0.007$ ) and between **Algorithm** and **Student** ( $p = 0.002$ ). By contrast, the difference between **Algorithm** and **Expert** is not statistically significant ( $p = 1.000$ ).<sup>8</sup>

These results suggest that, after controlling for the change in net benefit or performance points from advice and the number of pieces of advice received, participants' baseline change in WTP is similar for expert and algorithmic advice, and both are higher than for student advice. This pattern is similar to the WTP observed in Task 1, before participants had learned about the performance of each advice source. This similarity suggests that participants' initial perceptions of the advice sources may continue to influence their WTP, even after they have gained experience with the advice.

---

<sup>8</sup>Comparisons of the constant terms across models were performed using F-tests after seemingly unrelated estimation, with standard errors clustered at the session level. Bonferroni-adjusted  $p$ -values are reported for the pairwise comparisons.

## D Results without Pilot Sessions in Experiment 1

Table D1: Predicted WTP in Task 1 and Task 2 without pilot sessions

VARIABLES	(1)	(2)
	WTP Task 1	WTP Task 2
Baseline = Algorithm treatment dummy		
Expert treatment dummy	-0.720*	-0.952***
	(0.393)	(0.230)
Student treatment dummy	-2.049***	-2.005***
	(0.318)	(0.182)
Constant	4.935***	5.052***
	(0.200)	(0.076)
Observations	255	258
Cluster at session level	14	14
R-squared	0.109	0.114
Treatments	Predicted WTP Task 1	Predicted WTP Task 2
Algorithm	4.935	5.052
	(0.200)	(0.076)
Experts	4.216	4.100
	(0.338)	(0.217)
Students	2.886	3.047
	(0.247)	(0.165)
One-sided test	Prob > F	Prob > F
Expert > Algorithm#	0.865	0.998
Algorithm > Student #	0.000	0.000
Expert > Student #	0.011	0.003
Two-sided test	Prob > F	Prob > F
Algorithm = Experts = Students	0.000	0.000
Algorithm = Expert #	0.270	0.004
Algorithm = Student #	0.000	0.000
Expert = Student #	0.022	0.006

*Notes:* (a) We regressed the WTP on three treatment dummies by OLS regression model with robust standard errors clustered at session level. (b) The unit of observation is the number of participants. The total number of observations is the number of participants in all treatments. (c) The robust standard errors are in parentheses. (d) # indicates Bonferroni-adjusted p-values. (e) Pilot sessions were removed. 3 observations were removed in Task 1 because 3 participants did not submit the WTP within the time limit.

Table D2: Changes in WTP from Task 1 to Task 2

VARIABLES	(1) Algorithm $\Delta$ WTP	(2) Expert $\Delta$ WTP	(3) Student $\Delta$ WTP	(4) Algorithm $\Delta$ WTP	(5) Expert $\Delta$ WTP	(6) Student $\Delta$ WTP
$\Delta$ Mean performance in Task 1	0.261* (0.120)	0.121 (0.134)	-0.121 (0.106)			
$\Delta$ Net benefit in Task 1				0.284 (0.141)	0.298*** (0.044)	0.058 (0.091)
Number of received advice in Task 1	-0.366*** (0.036)	-0.614** (0.132)	-0.273 (0.135)	-0.268*** (0.036)	-0.486** (0.107)	-0.262* (0.113)
Constant	1.954*** (0.105)	2.558** (0.685)	1.181** (0.414)	2.211*** (0.151)	2.546** (0.464)	1.048* (0.410)
Obs.	82	69	59	82	69	59
Cluster at session level	5	4	5	5	4	5
$R^2$	0.265	0.429	0.195	0.299	0.496	0.186

*Notes:* a) In models 1–3, we regressed  $\Delta$ WTP on  $\Delta$ Mean performance from graphs with advice using an OLS regression model with robust standard errors clustered at session level.  $\Delta$ WTP is calculated as WTP in Task 2 minus WTP in Task 1.  $\Delta$ Mean performance is calculated as the mean performance from graphs with advice in part 2 of Task 1 minus the mean performance from graphs with advice in part 1 of Task 1. Performance ranges from 0 to 20, based on the points obtained in each question. In models 4–6, we regressed  $\Delta$ WTP on  $\Delta$ Net benefit from graphs with advice using an OLS regression model, with robust standard errors.  $\Delta$ Net benefit is calculated as the mean net benefit from graphs with advice in part 2 of Task 1 minus the mean net benefit from graphs with advice in part 1 of Task 1. The net benefit in each graph in part 1 equals the rewards in each graph plus 10 points minus the price of advice. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively.

b) The unit of observation is the participant who received at least one piece of advice in Task 1. The total number of observations is the number of participants who received at least one piece of advice in Task 1 in each treatment.

c) Numbers in parentheses represent robust standard error clustered at session level.

Table D3: Change of performance without pilot sessions (points)

VARIABLES	Advice received	(1)	(2)	(3)	(4)
		$\Delta$ Perf. Task 1 No	$\Delta$ Perf. Task 1 Yes	$\Delta$ Perf. Task 2 No	$\Delta$ Perf. Task 2 Yes
Baseline = Algorithm treatment dummy					
Expert treatment dummy		-0.424* (0.251)	0.643** (0.297)	0.007 (0.157)	0.596** (0.250)
Student treatment dummy		-0.155 (0.204)	1.088*** (0.297)	-0.194 (0.148)	0.461 (0.281)
Constant		0.533*** (0.163)	0.141 (0.229)	0.174 (0.118)	0.006 (0.183)
Observations		1,549	1,001	1,529	1,045
Clusters		247	210	252	218
R-squared		0.003	0.012	0.002	0.006
Treatments		Predicted $\Delta$ Perf. Task 1	Predicted $\Delta$ Perf. Task 2	Predicted $\Delta$ Perf. Task 1	Predicted $\Delta$ Perf. Task 2
Algorithm		0.533 (0.163)	0.141 (0.229)	0.174 (0.118)	0.006 (0.183)
Experts		0.109 (0.191)	0.784 (0.189)	0.181 (0.104)	0.602 (0.170)
Students		0.378 (0.123)	1.230 (0.189)	-0.021 (0.090)	0.467 (0.213)
		Prob > F	Prob > F	Prob > F	Prob > F
Algorithm = Experts = Students		0.240	0.002	0.253	0.052
Algorithm = Expert #			0.095		0.054
Algorithm = Student #			0.001		0.305
Expert = Student #			0.292		1.000

Notes: (a) We regressed the change in performance points (Part 2 - Part 1) ( $\Delta$ Perf.) on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level. (b) The unit of observation is the forecast from each participant in each round. The total number of observations is the total number of graphs with/without advice received in each task in all treatments. (c) The robust standard errors are in parentheses. (d) # indicates Bonferroni-adjusted p-values. (e) Pilot sessions were removed. Due to no submissions of predictions within the time limit, 12 observations were removed in Model (1), 13 observations in Model (2), 2 observations in Model (3), and 2 observations in Model (4).

Table D4: Change in performance (points) when comparing with and without receiving advice without pilot sessions

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	$\Delta$ Perf.	$\Delta$ Perf.	$\Delta$ Perf.	$\Delta$ Perf.	$\Delta$ Perf.	$\Delta$ Perf.
	Task 1	Task 1	Task 1	Task 2	Task 2	Task 2
	Algorithm	Expert	Student	Algorithm	Expert	Student
Advice received dummy	-0.392 (0.271)	0.675** (0.261)	0.851*** (0.208)	-0.167 (0.229)	0.421** (0.183)	0.488** (0.236)
Constant	0.533*** (0.163)	0.109 (0.192)	0.378*** (0.123)	0.174 (0.118)	0.181* (0.105)	-0.021 (0.091)
Observations	879	832	839	886	839	849
Clusters	89	84	85	89	84	85
R-squared	0.002	0.009	0.015	0.001	0.006	0.008

*Notes:* (a) We regressed the change in performance points (Part 2 - Part 1) ( $\Delta$ Perf.) on advice received dummy in each task and each treatment using an OLS regression model with robust standard errors clustered at the participant level. The advice received dummy equals 1 if the participant receives advice for that graph, and 0 otherwise. (b) The unit of observation is the forecast from each participant in each round. The total number of observations is the total number of forecasts with/without advice received in each task in all treatments. (c) The robust standard errors are in parentheses. (d) Pilot sessions were removed. Due to no submissions of predictions within the time limit, 11 observations were removed in Model (1), 8 observations in Model (2), 11 observations in Model (3), 4 observations in Model (4), 1 observations in Model (5) and 1 observations in Model (6).

Table D5: Change of net benefit (points) without pilot sessions

VARIABLES	Advice received	(1)	(2)	(3)	(4)
		$\Delta$ NB. Task 1 No	$\Delta$ NB. Task 1 Yes	$\Delta$ NB. Task 2 No	$\Delta$ NB. Task 2 Yes
Baseline = Algorithm treatment dummy					
Expert treatment dummy		-0.424*	0.621	0.007	1.095***
		(0.251)	(0.377)	(0.157)	(0.312)
Student treatment dummy		-0.155	1.635***	-0.194	0.961**
		(0.204)	(0.366)	(0.148)	(0.382)
Constant		0.533***	-2.985***	0.174	-3.121***
		(0.163)	(0.256)	(0.118)	(0.234)
Observations		1,549	1,001	1,529	1,045
Clusters		247	210	252	218
R-squared		0.003	0.020	0.002	0.016
Treatments		Predicted $\Delta$ NB. Task 1	Predicted $\Delta$ NB. Task 2	Predicted $\Delta$ NB. Task 1	Predicted $\Delta$ NB. Task 2
Algorithm		0.533	-2.985	0.174	-3.121
		(0.163)	(0.256)	(0.118)	(0.234)
Experts		0.109	-2.364	0.181	-2.027
		(0.191)	(0.277)	(0.104)	(0.206)
Students		0.378	-1.350	-0.021	-2.160
		(0.123)	(0.262)	(0.090)	(0.302)
		Prob > F	Prob > F	Prob > F	Prob > F
Algorithm = Experts = Students		0.240	0.000	0.253	0.002
Algorithm = Expert #			0.304		0.002
Algorithm = Student #			0.000		0.038
Expert = Student #			0.025		1.000

Notes: (a) We regressed the change in net benefit (Part 2 - Part 1) ( $\Delta$ NB.) on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level. (b) The unit of observation is the forecast from each participant in each round. The total number of observations is the total number of graphs with/without advice received in each task in all treatments. (c) The robust standard errors are in parentheses. (d) # indicates Bonferroni-adjusted p-values. (e) Due to no submissions of predictions within the time limit, 12 observations were removed in Model (1), 13 observations in Model (2), 2 observations in Model (3), and 2 observations in Model (4).

Table D6: Change in net benefit (points) when comparing with and without receiving advice without pilot sessions

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	$\Delta$ NB.	$\Delta$ NB.	$\Delta$ NB.	$\Delta$ NB.	$\Delta$ NB.	$\Delta$ NB.
	Task 1 Algorithm	Task 1 Expert	Task 1 Student	Task 2 Algorithm	Task 2 Expert	Task 2 Student
Advice received dummy	-3.518*** (0.301)	-2.472*** (0.336)	-1.728*** (0.284)	-3.295*** (0.273)	-2.207*** (0.210)	-2.139*** (0.318)
Constant	0.533*** (0.163)	0.109 (0.192)	0.378*** (0.123)	0.174 (0.118)	0.181* (0.105)	-0.021 (0.091)
Observations	879	832	839	886	839	849
Clusters	89	84	85	89	84	85
R-squared	0.152	0.093	0.054	0.174	0.126	0.114

*Notes:* (a) We regressed the change in net benefit (Part 2 - Part 1) ( $\Delta$ NB.) on advice received dummy in each task and each treatment using an OLS regression model with robust standard errors clustered at the participant level. The advice received dummy equals 1 if the participant receives advice for that graph, and 0 otherwise. (b) The unit of observation is the forecast from each participant in each round. The total number of observations is the total number of forecasts with/without advice received in each task in all treatments. (c) The robust standard errors are in parentheses. (d) Due to no submissions of predictions within the time limit, 11 observations were removed in Model (1), 8 observations in Model (2), 11 observations in Model (3), 4 observations in Model (4), 1 observations in Model (5) and 1 observations in Model (6).

Table D7: Predicted shift rate in Task 1 and Task 2 without pilot sessions

VARIABLES	(1)	(2)
	Shift rate Task 1	Shift rate Task 2
Baseline = Algorithm treatment dummy		
Expert treatment dummy	0.043 (0.034)	-0.042 (0.037)
Student treatment dummy	-0.266*** (0.044)	-0.326*** (0.036)
Constant	0.811*** (0.024)	0.822*** (0.023)
Observations	905	930
Clusters	207	216
R-squared	0.147	0.179
Treatments	Predicted shift rate Task 1	Predicted shift rate Task 2
Algorithm	0.811 (0.024)	0.822 (0.023)
Experts	0.853 (0.023)	0.780 (0.029)
Students	0.545 (0.037)	0.496 (0.028)
	Prob > F	Prob > F
Algorithm = Experts = Students	0.000	0.000
Algorithm = Expert #	0.613	0.751
Algorithm = Student #	0.000	0.000
Expert = Student #	0.000	0.000

*Notes:* Notes: (a) We used the data of the graphs which participants received the advice. We also used the data which shift rate is between 0 and 1. We regressed the shift rate on three treatment dummies using an OLS regression model with robust standard errors clustered at the participant level. (b) The unit of observation is the graph with received advice of each participant. The total number of observations is the number of graphs with received advice in each task in all treatments. (c) The robust standard errors are in parentheses. (d) # indicates Bonferroni-adjusted p-values. (e) 13 observations were removed in Task 1 and 2 observations were removed in Task 2 due to no submissions of predictions within the time limit.

## E Experimental Instruction in Experiment 1 (English Translation)

# Experimental Instruction

1

## Welcome to the experiment

- We will now begin the experiment. In addition to the participation fee of 500 yen for taking part in this experiment, you will also receive reward based on the choices you make during the experiment.
- You are not allowed to communicate with other participants during the experiment. Additionally, please turn off all electronic devices such as smartphones and iPads during the experiment.
- If you have any questions during the experiment, please raise your hand. We will answer them individually.

2

## **Experiment details**

The experiment consists of three parts:

Part 1: Questionnaire

Part 2: Experiment Task 1

Part 3: Experiment Task 2

The total duration of the experiment is approximately one hour.

3

## **Experiment details: Questionnaire**

- In the questionnaire, you will be asked mainly about your experience with stock investments and your knowledge of financial literacy to assess your knowledge and experience related to financial investment.
- Your responses will not affect your reward, but please answer honestly according to your personal situation.
- The estimated time for completing the questionnaire is 10 minutes.
- If the time limit is exceeded, you will be automatically moved to the next step.
- Once you have completed the questionnaire, click the 'Next' button to begin Task 1.

4

## Experiment details: Tasks 1 & 2

- In Tasks 1 and 2, you will be asked to **predict future stock prices based on past stock price information**. Your reward will depend on how accurate your predictions are, as will be explained later.
- In each question, you will be shown a graph displaying the closing prices of a particular stock over a 12-month period. You will then be asked to predict the stock price 30 days after the last day shown on the graph.
- The period for each graph is randomly selected from the time between January 1, 2008, and June 30, 2018, and will cover a 12-month period starting from that randomly selected date.
- The stock will be randomly chosen from the S&P 500.
- The stock name and the exact period will not be displayed.

5

## Experiment details: Tasks 1 & 2

- Tasks 1 and 2 will follow the same procedure.
- Each task is divided into two parts: the first half and the second half.

Task 1	Task 2	Steps
First half	First half	After reading the graph of a certain stock, make a prediction.
		Submit your willingness to pay for additional information.
Second half	Second half	After reading the graph and additional information, make a prediction.

Same 10 graphs      Same 10 graphs      Total 20 graphs

- More details will be provided later.

6

## Experiment details : The first half of Tasks 1 & 2

- In the first half, as mentioned earlier, you will read a graph displaying the closing prices of a particular stock over a 12-month period and predict the stock price 30 days after the last day shown on the graph.
- All graphs are normalized so that the stock price on the first day of the graph is set to 100.
- Please submit your predicted value with up to three decimal places.
- There are 10 questions in total.
- **The time limit for each question is 40 seconds.**
- In the next slide, a sample question will be displayed.

7

### Sample Question

Please read the below graph displaying the closing prices of a particular stock over a 12-month period and predict the stock price 30 days after the last day shown on the graph. The closing price on the last day is 110.693.



Your prediction:

Note: Please submit your prediction up to three decimal places.

Next

8

## Experiment details : The second half of Tasks 1 & 2

- In the second half, the same 10 graphs will be displayed in the same order, and you will make predictions again. The time limit for each question is 40 seconds.
- However, in the second half, you may have the opportunity to obtain additional information to assist with your predictions.
- Whether you can obtain the additional information depends on the amount you submit between the first and second halves as **your willingness to pay** (WTP, the maximum amount you are willing to pay to purchase the additional information) and the **randomly determined price** ( $p$ ) of the additional information.
- This will be explained in more detail later.

9

## Experiment details : The second half of Tasks 1 & 2

- **Participants who obtained additional information:** they can refer to both the graph and the additional information to predict the stock price again.
- **Participants who did not obtain additional information:** they will predict the stock price again using only the graph, just as in the first half.
- As in the first half, please submit your predicted value with up to three decimal places.
- There are 10 graphs in total.

10

## Obtaining additional information

- Before the second half begins, an explanation will be provided on how the availability of additional information is determined, based on the submitted willingness to pay (WTP) and the price (p) of the additional information, which is randomly set for each question.
- In this experiment, the WTP that can be submitted is limited to numbers **from 0.0 to 10.0, with one decimal place.**
- The price of the additional information (p) ranges **from 0.1 to 10.0 and is randomly determined** for each of the 10 questions in the second half. All prices are chosen with the same probability.

11

## Purchase of additional information

- In a given graph, if the WTP is greater than or equal to the price of the additional information (p) ( $WTP \geq p$ ):
  - You will pay p and obtain the additional information.
- Conversely, if the WTP is less than p ( $WTP < p$ ):
  - You will not obtain the additional information, and no payment will occur.
- The WTP in all 10 questions is the same and is submitted only once before the second half begins.
- The price of the additional information (p) is determined randomly for each of the 10 questions in the second half.

12

## Purchase of additional information

- Example 1

- For instance, suppose you submit a WTP of 10.0.
- In this case, since the price of the additional information determined randomly for each graph is at most 10.0, you will obtain additional information for all graphs and pay the price of the additional information for each graph.

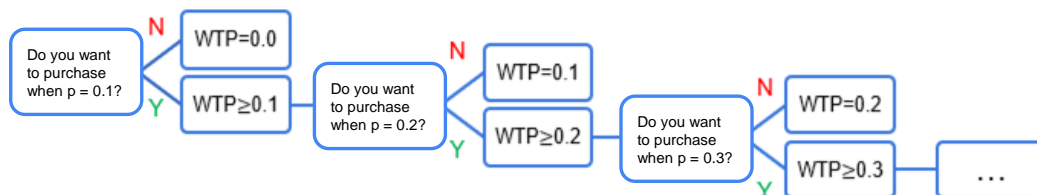
- Example 2

- For instance, suppose you submit a WTP of 0.0.
- In this case, since the minimum price of the additional information determined randomly for each graph is 0.1, you will not obtain additional information for any graph, and no payment for the additional information will occur either.

13

## How to determine the willingness to pay (WTP)

When determining the WTP, first consider the following questions:



Continue this until your answer changes from 'yes' to 'no'.

The price at which your answer changes is your WTP.

14

## Endowment for purchasing additional information

- For Task 1 and Task 2, you are provided with 100 points endowment.
- Out of these 100 points, any points not used for purchasing additional information will be added to your rewards at a rate of 1 point = 6 yen.
- Example 1: If you submit a WTP of 0.0 and do not purchase any additional information, these 100 points will be added to your rewards.
- Example 2: If you submit a WTP of 10.0 and obtain additional information for each of the 10 problems, spending a total of 90 points, the remaining 10 points will be added to your rewards.

15

## Example of Purchasing Additional Information and Point Calculation (WTP = 5.0)

Question	Price of additional information(p)	Can the additional information be purchased?	Remaining points
Q1	3.5	Yes	$100 - 3.5 = 96.5$
Q2	6.2	No	96.5
Q3	0.3	Yes	$96.5 - 0.3 = 96.2$
Q4	7.3	No	96.2
Q5	8.2	No	96.2
Q6	5.0	Yes	$96.2 - 5.0 = 91.2$
Q7	3.8	Yes	$91.2 - 3.8 = 87.4$
Q8	5.1	No	87.4
Q9	5.0	Yes	$87.4 - 5.0 = 82.4$
Q10	9.2	No	<b>82.4</b>

16

## Calculation of rewards

- In addition to a participation fee of JPY 500, you can earn additional rewards based on your performance in Tasks 1 and 2.
- One of the four parts—Task 1 first half, Task 1 second half, Task 2 first half, and Task 2 second half—will be randomly selected, and additional rewards will be determined based on the accuracy of the 10 predictions you submitted.
- Please refer to the table below for the calculation method of the additional rewards for each prediction:

Degree of deviation from the true value	Points
±0%	20
±0% ~ ±20%	20 -  Degree of deviation × 100
±20%~	0

- The total points gained from the 10 questions is used to calculate the additional reward.
- The points earned are converted to Japanese yen at a rate of 1 point = 6 yen. 17

## Example of point calculation

For example, suppose the true value for a particular question 118.890.

- If your prediction is 130.123 :

The degree of deviation from the true value is

$$\left| \frac{130.123 - 118.890}{118.890} \right| \approx 0.09448 \text{ (9.448\%)}$$

You get  $20 - |0.09448 \times 100| \approx 10.552$  points.

- If your prediction is 90.002 :

The degree of deviation from the true value is

$$\left| \frac{90.002 - 118.890}{118.890} \right| \approx 0.243 \text{ (24.3\%)}$$

You get no points.

18

## Summary regarding rewards

In today's experiment, in addition to a participation fee of 500 yen, the total rewards paid to you will include:

- The amount based on the points not used for purchasing additional information, from the 100 points provided for purchasing additional information in Task 1 and Task 2.
- The amount based on the points earned according to the accuracy of the 10 predictions in one of the four parts—Task 1 first half, Task 1 second half, Task 2 first half, Task 2 second half—randomly selected for rewards.

This total amount will be paid to you as total rewards.

19

This concludes the explanation of the experiment.  
Please answer the quiz to confirm whether you have correctly understood the content of the explanation.

20

## F Quiz in Experiment 1 (English Translation)

1. What can you do in this experiment?
  - 1 View the stock price graph of a company you specify
  - 2 Submit your predictions
  - 3 Decide the part used for calculating additional rewards
  - 4 Choose the source of additional information you want to obtain

[Correct answer: 2]

2. In which of the following cases do you obtain additional information?

Case	Willingness to pay (WTP)	Price of additional information (p)
1	10	5
2	4	6
3	8	8
4	3	9

- 1 Case 1 and Case 3
- 2 Case 2 and Case 4
- 3 Case 1 and Case 4
- 4 Case 2 and Case 3

[Correct answer: 1]

3. Which of the following is true about the willingness to pay (WTP)?
  - 1 It is possible to submit the WTP 10 times per task.
  - 2 It is possible to submit 12.
  - 3 It is not possible to submit 5.5.
  - 4 It is the maximum number of points you are willing to pay to acquire additional information.

[Correct answer: 4]

4. Which of the following is true about the price of additional information (p)?
  - 1  $p = 12$  is possible.
  - 2 There is only one per task.
  - 3 It is determined randomly for each question in the second half of the task.
  - 4 You can set it yourself.

[Correct answer: 3]

5. Please select the correct option regarding the purchase of additional information.

- 1 If  $WTP \geq p$ , you decide whether to buy it yourself.
- 2 If  $WTP < p$ , you cannot purchase, but you will still pay the points.
- 3 If  $WTP = p$ , you cannot purchase.
- 4 If  $WTP \geq p$ , you automatically pay  $p$  points.

[Correct answer: 4]

## **G Questionnaire in Experiment 1 and 2 (English Translation)**

### **G.1 Survey on demographic characteristics**

1. Gender:

- Female
- Male
- Other
- Do not want to answer

2. Grade level:

- Undergraduate Year 1
- Undergraduate Year 2
- Undergraduate Year 3
- Undergraduate Year 4
- Master Year 1
- Master Year 2
- Other

3. Major:

- Humanities
- Sciences

## G.2 Survey on stock trading experience

d) Do you have experience in stock investment? Yes (Go to Q2 - Q6) No

- How much experience did you have with stock investments during your student days?
- Less than six months
  - Six months to one year
  - One to three years
  - More than three years
2. What is the main type of stock you target in your investments?
- Large-cap stocks
  - Small-cap stocks
  - Technology stocks
  - Consumer goods stocks
  - Financial stocks
  - Other
3. What is the performance of your investments?
- Stable profits
  - Stable losses
  - Balanced between profits and losses
  - Uncertain with no clear trend
4. What are the main sources of information you use when investing in stocks?
- Financial news and media coverage
  - Corporate financial reports and fundamental analysis
  - Technical analysis and chart patterns
  - Social media and forum discussions
  - Professional research reports
  - Other
5. What is your perception of the importance of risk management in stock investing?
- Extremely important, and I actively employ risk management strategies.
  - Important, but there may be room for improvement.
  - Not very emphasized, focusing more on profits.
  - I do not understand the concept of risk management at all.

### G.3 Financial literacy test

The financial literacy scores were measured by following Fernandes et al. (2014). The 12 questions were as follows.

1. Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, would you be able to buy:
  - 1 More than today with the money in this account.
  - 2 Exactly the same as today with the money in this account.
  - 3 Less than today with the money in this account.
  - 4 Don't know.
  - 5 Refuse to answer.

[Correct answer: 3]

2. Do you think that the following statement is true or false? "Bonds are normally riskier than stocks."
  - 1 True.
  - 2 False.
  - 3 Don't know.
  - 4 Refuse to answer.

[Correct answer: 2]

3. Considering a long time period (for example 10 or 20 years), which asset described below normally gives the highest return?
  - 1 Savings accounts.
  - 2 Stocks.
  - 3 Bonds.
  - 4 Don't know.
  - 5 Refuse to answer.

[Correct answer:2]

4. Normally, which asset described below displays the highest fluctuations over time?
  - 1 Savings accounts.
  - 2 Stocks.
  - 3 Bonds.

4 Don't know.

5 Refuse to answer.

[Correct answer:2]

5. When an investor spreads his money among different assets, does the risk of losing a lot of money:

1 Increase.

2 Decrease.

3 Stay the same.

4 Don't know.

5 Refuse to answer.

[Correct answer:2]

6. Do you think that the following statement is true or false? "If you were to invest \$1000 in a stock mutual fund, it would be possible to have less than \$1000 when you withdraw your money?"

1 True.

2 False.

3 Don't know.

4 Refuse to answer.

[Correct answer: 1]

7. Do you think that the following statement is true or false? "A stock mutual fund combines the money of many investors to buy a variety of stocks."

1 True.

2 False.

3 Don't know.

4 Refuse to answer.

[Correct answer: 1]

8. Do you think that the following statement is true or false? "A 15-year mortgage typically requires higher monthly payments than a 30-year mortgage, but the total interest paid over the life of the loan will be less."

1 True.

- 2 False.
- 3 Don't know.
- 4 Refuse to answer.

[Correct answer: 1]

9. Suppose you had \$100 in a savings account and the interest rate is 20% per year and you never withdraw money or interest payments. After 5 years, how much would you have on this account in total?

- 1 More than \$200.
- 2 Exactly \$200.
- 3 Less than \$200.
- 4 Don't know.
- 5 Refuse to answer.

[Correct answer: 1]

10. Which of the following statements is correct?

- 1 Once one invests in a mutual fund, one cannot withdraw the money in the first year.
- 2 Mutual funds can invest in several assets, for example invest in both stocks and bonds.
- 3 Mutual funds pay a guaranteed rate of return which depends on their past performance.
- 4 None of the above
- 5 Don't know.
- 6 Refuse to answer.

[Correct answer: 2]

11. Which of the following statements is correct? If somebody buys a bond of firm B:

- 1 He owns a part of firm B.
- 2 He has lent money to firm B.
- 3 He is liable for firm B's debts.
- 4 None of the above
- 5 Don't know.
- 6 Refuse to answer.

[Correct answer: 2]

12. Suppose you owe \$3,000 on your credit card. You pay a minimum payment of \$30 each month. At an Annual Percentage Rate of 12% (or 1% per month), how many years would it take to eliminate your credit card debt if you made no additional new charges?

- 1 Less than 5 years.
- 2 Between 5 and 10 years.
- 3 Between 10 and 15 years.
- 4 Cannot be fully repaid.
- 5 Don't know.
- 6 Refuse to answer.

[Correct answer: 4]

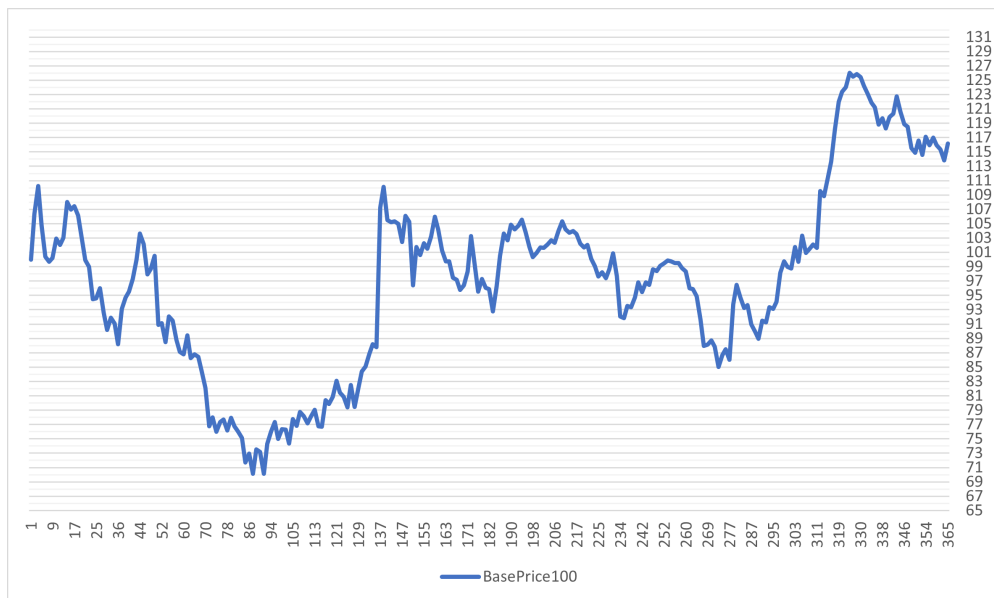
## H Screenshots of Tasks 1 and 2 in Experiment 1 (English Translation)

[Screen 1] There are 10 questions in Part 1 of both Task 1 and Task 2. Below is an example of Question 1 from Part 1 of Task 1.

### Question 1

The remaining time: 0:40

Please read the below graph displaying the closing prices of a particular stock over a 12-month period and predict the stock price 30 days after the last day shown on the graph. The closing price on the last day is 116.183.



Your prediction:

Note: Please submit your prediction up to three decimal places.

Next

[Screen 2]

## Question 1

The remaining time: 0:10

Your prediction has been successfully sent.

Your prediction for Question 1 is 118.3.

Next

[Screen 3] (shown in algorithm treatment)

## Information source and Willingness to pay (WTP)

The remaining time: 0:60

The additional information provided here is a stock price forecast submitted, based on the same stock price chart used in this experiment, by [an algorithm](#) crafted to predict future stock prices.

This algorithm makes a future stock price forecast by learning the historical stock price information, from January 1st, 2000, or from the initial public offering (IPO) day to January 1st, 2020, of 83 target companies ranked at the top in their capital market sectors (i.e., Basic Materials, Consumer Goods, Healthcare, Services, Utilities, Conglomerates, Financial, Industrial Goods, Technology).

Now your endowment is 100 points.

How much are you willing to pay, up to a maximum of 10.0, for the additional information?

Note: WTP submissions are **once** per task only.

Submissions can range from 0 to 10, up to one decimal place.

Next

[Screen 3] (shown in expert treatment)

## Information source and Willingness to pay (WTP)

The remaining time: 0:60

The additional information provided here is the average of stock price predictions submitted by 198 [experts \(CMAs\)](#) who participated in a similar experiment in the past, based on the stock price chart used in this experiment, for the stock prices 30 days ahead.

CMA stands for Certified Member Analyst of the Securities Analysts Association of Japan. The CMA is a qualification granted to those who have taken the prescribed training courses and passed the examinations based on these courses and fulfilled certain requirements; it is a sign of a expertise in the fields of finance and investments. The CMAs' investment valuations are based on the calculation of the corporate value of an investment and the forecasting of its future value. The particularly crucial knowledge required can be broadly categorized into three main areas: evaluation of investment data, decision making on investment policies, and construction and management of portfolios.

[Now your endowment is 100 points.](#)

How much are you willing to pay, up to a maximum of 10.0, for the additional information?

Note: WTP submissions are **once** per task only.

Submissions can range from 0 to 10, up to one decimal place.

Next

[Screen 3] (shown in student treatment)

## Information source and Willingness to pay (WTP)

The remaining time: 0:60

The additional information provided here is the average of stock price predictions submitted by 233 [students at the University of Osaka](#) who participated in a similar experiment conducted in the past, based on the stock price chart used in this experiment, for stock prices 30 days ahead.

Now your endowment is 100 points.

How much are you willing to pay, up to a maximum of 10.0, for the additional information?

Note: WTP submissions are **once** per task only.

Submissions can range from 0 to 10, up to one decimal place.

Next

[Screen 4] (shown in the case  $WTP \geq p$ )

### Notice

The remaining time: 0:15

The Price of additional information is 1.1.

Your willingness to pay is 4.0.

Because  $WTP \geq p$ , you can obtain additional information in the next question.

Your current points is 98.9 points.

Next

[Screen 4] (shown in the case  $WTP < p$ )

### Notice

The remaining time: 0:15

The Price of additional information is 8.0.

Your willingness to pay is 4.0.

Because  $WTP < p$ , you cannot obtain additional information in the next question.

Your current points is 100.0 points.

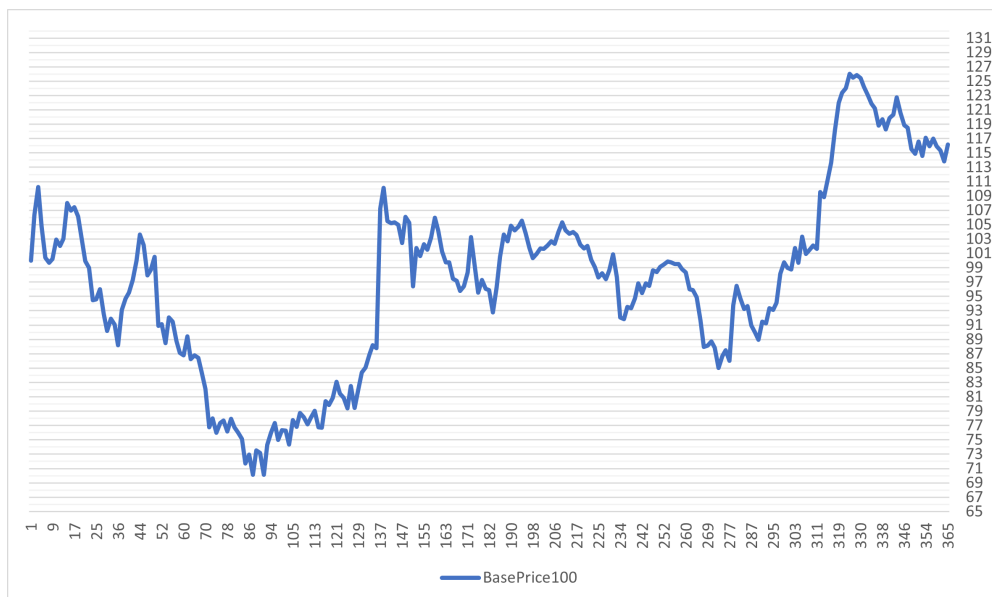
Next

[Screen 5] There are 10 questions in Part 2 of both Task 1 and Task 2. The order of questions in Part 2 is the same as in Part 1 in each task. Below is an example of Question 11 from Part 2 of Task 1.

### Question 11

The remaining time: 0:40

Please read the below graph displaying the closing prices of a particular stock over a 12-month period and predict the stock price 30 days after the last day shown on the graph. The closing price on the last day is 116.183.



Additional information: The prediction value from the algorithm is 124.502. (shown in algorithm treatment and in the case  $WTP \geq p$ )

Additional information: The prediction value from the experts is 115.984. (shown in

expert treatment and in the case  $WTP \geq p$ )

Additional information: The prediction value from the Osaka University students is 113.969. (shown in student treatment and in the case  $WTP \geq p$ )

Your prediction in first half is 118.3.

Your prediction:

Note: Please submit your prediction up to three decimal places.

Next

[Screen 6] The feedback is shown at the end of Task 1 and Task 2. The following is an example of the feedback given at the end of Task 1.

### Completion of Task 1 and Notification of Results

Thank you for your participation in Task 1.

We will inform you of the results of Task 1.

Total points in Part 1:  $132.1 + 100 = 232.1$  points

Total points in Part 2:  $137.8 + 89.1 = 226.9$  points

Please refer to the table below for the details of the gained points.

[Click the next button to start Task 2.](#)

Part 1

Questions	Your forecast	Points
1	118.3	10.1
2	140	8.4
3	113	16.8
4	144	15.9
5	100	12
6	143	0
7	133	19.7
8	101	18.6
9	107	17.1
10	123	13.5

Part 2

Questions	WTP	p	Your forecast	Points
11	4.0	1.1	124	4.8
12	4.0	9	140	8.4
13	4.0	5.7	112	17.7
14	4.0	2.3	140	18.8
15	4.0	6.2	107	18.5
16	4.0	9.4	144	0
17	4.0	9.4	133	19.7
18	4.0	5.1	101	18.6
19	4.0	3.8	104	14.4
20	4.0	3.7	119	16.9

Next

[Screen 7]

## The end of experiment and reward announcement

Thank you for participating in the experiment today.  
Please wait on this screen until everyone has finished.

We will inform you about the rewards.

Task 1:

Points in the first half: 232.1 points

Points in the second half: 226.9 points

Task 2:

Points in the first half: 2252.6 points

Points in the second half: 235.5 points

The selected part for reward calculation is first half in Task 1.

The reward you will receive is calculated as  $232.1 \text{ points} \times 6 \text{ JPY/point} + \text{JPY } 500 = \text{JPY } 1892.6$ .

# I Stock-price time series graphs

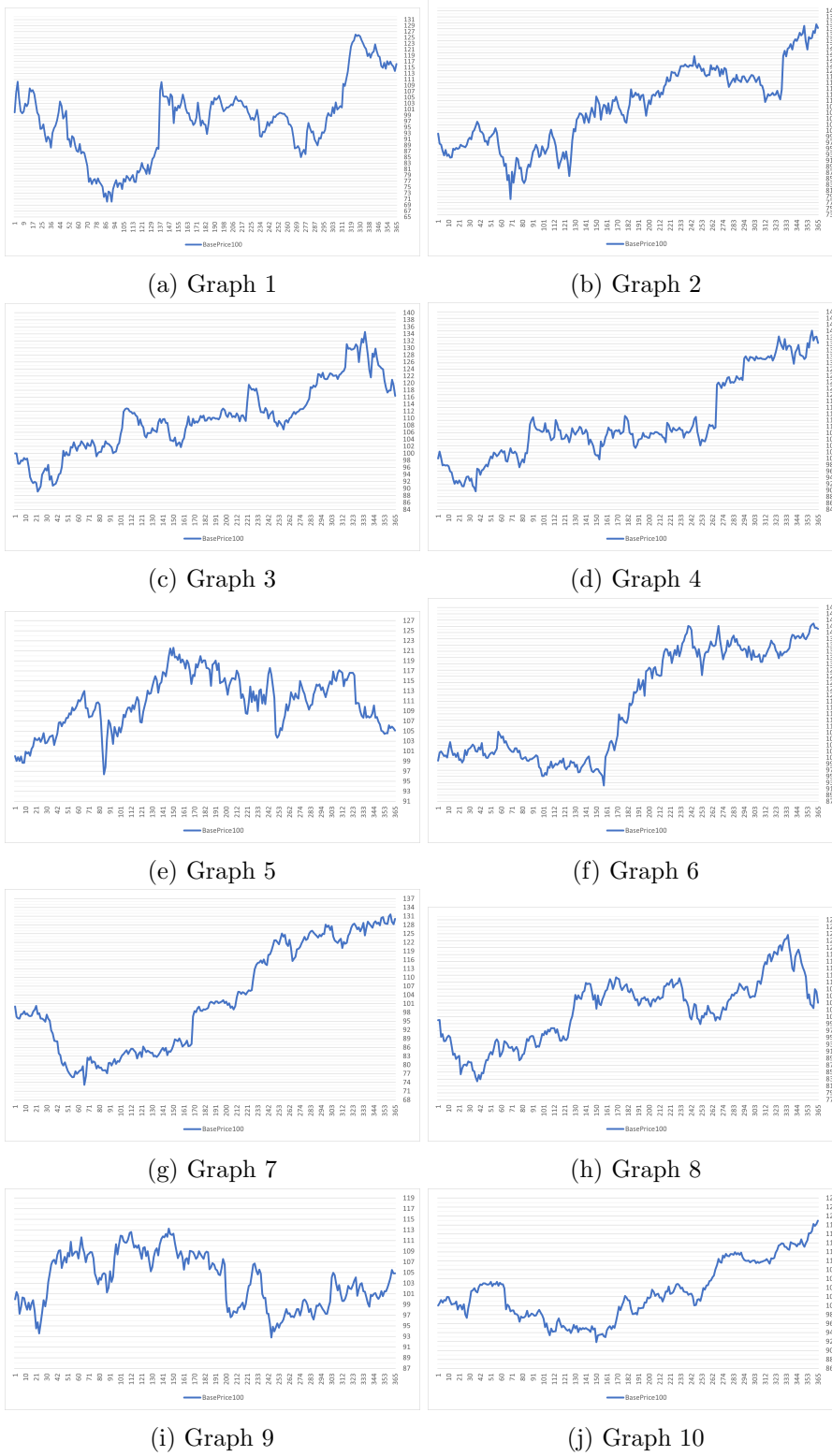
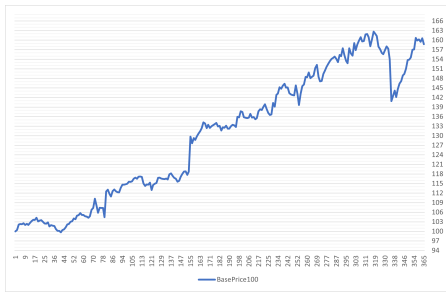
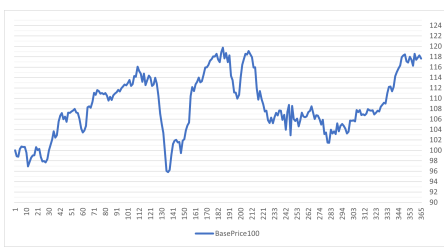


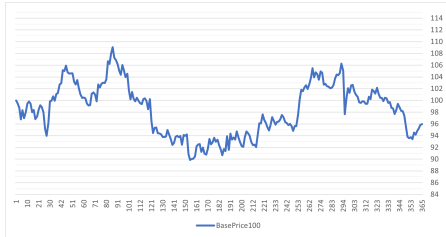
Figure I9: 10 graphs in Task 1



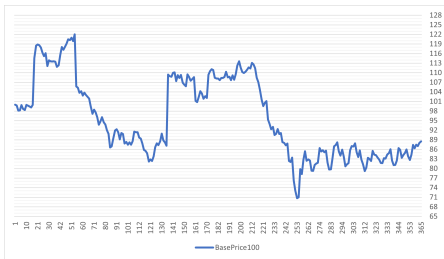
(a) Graph 11



(b) Graph 12



(c) Graph 13



(d) Graph 14



(e) Graph 15



(f) Graph 16



(g) Graph 17



(h) Graph 18



(i) Graph 19



(j) Graph 20

Figure I10: 10 graphs in Task 2

Table II: Stock Data with algorithm, expert, and student advice

Graph	Stock Ticker	Period (Start June)	Date from	Closing Price at Last Date (Normalized)	Closing Price One Month Later (Normalized)	Algorithm Advice	Expert Advice	Student Advice
1	STX	2010-2011		116.183	107.676	124.502	115.984	113.969
2	LKQ	2011-2012		135.314	125.498	136.293	134.380	133.211
3	UPS	2009-2010		116.381	109.475	118.215	116.255	115.061
4	DLTR	2009-2010		136.241	138.273	139.922	137.484	135.960
5	SBUX	2015-2016		105.113	108.656	105.449	105.973	105.031
6	FOXA	2017-2018		142.146	183.223	143.164	141.515	139.942
7	TSN	2012-2013		130.098	132.576	130.824	130.918	130.610
8	COP	2009-2010		105.073	102.479	107.670	105.877	105.305
9	F	2013-2014		104.847	110.204	104.032	104.407	103.783
10	FISV	2016-2017		118.940	115.456	118.870	122.045	120.121
11	EL	2017-2018		158.742	151.572	162.390	157.382	157.686
12	ADM	2014-2015		117.637	108.027	119.225	113.552	114.655
13	NKE	2016-2017		95.962	96.288	97.026	96.434	96.796
14	TRIP	2015-2016		88.537	83.505	92.265	91.306	90.399
15	WM	2013-2014		106.869	106.940	109.501	107.334	110.741
16	MLM	2016-2017		118.546	117.107	122.650	116.721	118.174
17	HES	2013-2014		91.097	97.096	93.155	89.991	90.655
18	SIVB	2016-2017		130.597	136.070	134.396	130.488	129.888
19	SPGI	2008-2009		78.589	73.714	77.623	78.882	79.430
20	FMC	2014-2015		73.448	67.383	75.437	73.581	73.534

## J Results on which figures and text are based in manuscript for Experiment 2

### J.1 WTP

Table J1: Predicted WTP in Round 1 and All Rounds (for Figure 9)

VARIABLES	(1) WTP Round 1	(2) WTP All Rounds
Baseline: Alg-No D dummy		
Alg-D dummy	-0.401 (0.448)	-0.213 (0.381)
Exp-No D dummy	-0.662 (0.434)	-0.207 (0.386)
Exp-D dummy	0.322 (0.493)	-0.062 (0.437)
Constant	3.430*** (0.326)	3.310*** (0.283)
Observations	201	4,020
Cluster at participant level	/	201
R-squared	0.027	0.001
Treatments	Predicted WTP Task 1	Predicted WTP Task 2
Alg-No D	3.430 (0.326)	3.310 (0.283)
Alg-D	3.029 (0.308)	3.097 (0.255)
Exp-No D	2.768 (0.287)	3.104 (0.262)
Exp-D	3.753 (0.370)	3.248 (0.333)
Two-sided test	Prob > F	Prob > F
Alg-No D = Alg-D = Exp-No D = Exp-D	0.156	0.931
Alg-No D = Alg-D #	1.000	1.000
Exp-No D = Exp-D #	0.148	1.000
Alg-No D = Exp-No D#	0.516	1.000
Alg-D = Exp-D#	0.539	1.000

*Notes:* (a) We regressed WTP in Round 1 and across all rounds on treatment dummies using OLS regressions. Model (1) uses robust standard errors, while Model (2) uses robust standard errors clustered at the participant level. Pairwise comparisons of the estimated treatment coefficients were conducted using F-tests with Bonferroni-adjusted  $p$ -values. Results from two-sided tests are reported. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively. (b) The unit of observation is the participant in Model (1) and the participant-round observation in Model (2). The total number of observations equals the number of participants in Round 1 for Model (1) and the number of participant-round observations across all rounds for Model (2). (c) Standard errors are reported in parentheses. Model (1) reports robust standard errors, while Model (2) reports robust standard errors clustered at the participant level. (d) # indicates Bonferroni-adjusted  $p$ -values.

## J.2 Utilization of advice

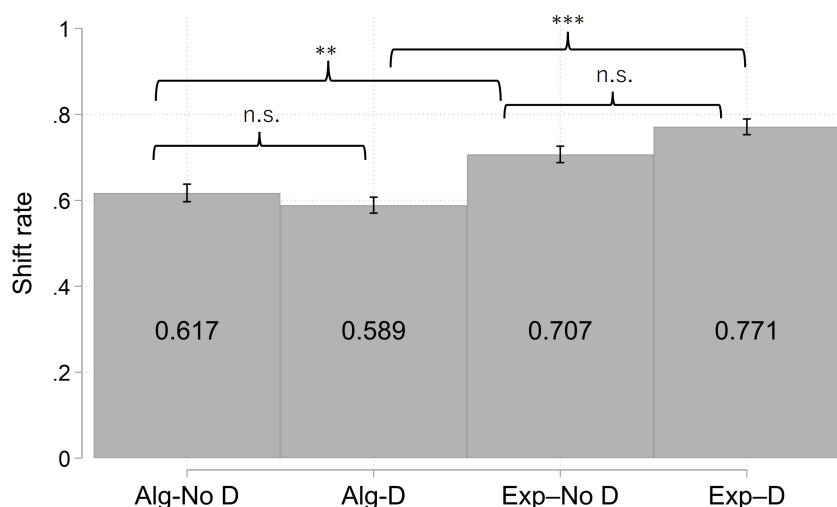


Figure J11: Shift rate in Experiment 2

*Notes:* We consider only observations for which shift rate between 0 and 1. We regress shift rate on treatment dummies, using a random-effects panel-data OLS regression model with standard errors clustered at the participant level, controlling for round number. Pairwise comparisons of estimated dummy coefficients are done by F-test with Bonferroni-adjusted  $p$ -values. Results from two-sided tests are reported. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively.

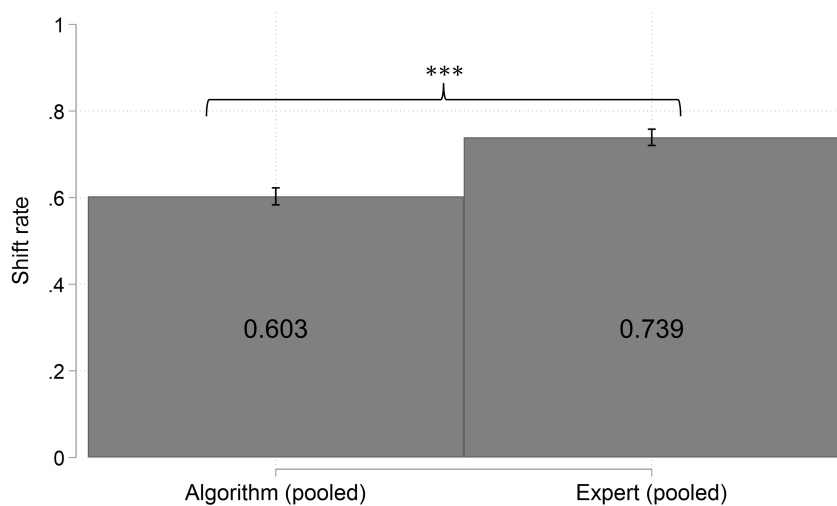


Figure J12: Shift rate between algorithm and expert in Experiment 2

*Notes:* We consider only observations for which shift rate between 0 and 1. We regress shift rate on pooled treatment dummies, using a random-effects panel-data OLS regression model with standard errors clustered at the participant level, controlling for round number and description dummy. Pairwise comparisons of estimated dummy coefficients are done by F-test. Results from two-sided tests are reported. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively.

Table J2: Predicted Shift Rate

VARIABLES	(1) Shift	(2) Shift
Baseline: Alg-No D dummy		
Alg-D dummy	-0.014 (0.040)	
Exp-No D dummy	0.113*** (0.041)	
Exp-D dummy	0.157*** (0.040)	
Algorithm dummy		-0.142*** (0.029)
Description dummy		0.014 (0.029)
Round	-0.010*** (0.001)	-0.010*** (0.001)
Constant	0.699*** (0.029)	0.827*** (0.028)
Observations	1,103	1,103
Cluster at participant level	190	190
Wald $\chi^2$	68.197	67.721
Prob > chi2	0.000	0.000
Two-sided test	Prob > F	Prob > F
Alg-No D = Alg-D = Exp-No D = Exp-D	0.000	
Alg-No D = Alg-D #	1.000	
Exp-No D = Exp-D #	1.000	
Alg-No D = Exp-No D #	0.025	
Alg-D = Exp-D #	0.000	

*Notes:* (a) In Model (1), we regress shift rate on treatment dummies, using a random-effects panel-data OLS regression model with standard errors clustered at the participant level, controlling for round number. In Model (2), we regress shift rate on pooled treatment dummies, using a random-effects panel-data OLS regression model with standard errors clustered at the participant level, controlling for round number and description dummy. Pairwise comparisons of the estimated treatment coefficients were conducted using F-tests with Bonferroni-adjusted  $p$ -values. Results from two-sided tests are reported. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively. (b) The unit of observation is the round which participant received advice. The total number of observations equals the number of rounds who received advice from all participants. (c) Standard errors are reported in parentheses. We report standard errors clustered at the participant level. (d) # indicates Bonferroni-adjusted  $p$ -values.

### **J.3 Change in performance points and net benefit between the first 10 rounds and the last 10 rounds**

In Figure 10 of the manuscript, we compare the change in performance points between the first 10 rounds and the last 10 rounds. When participants did not receive advice, we find no statistically significant difference between the first and last 10 rounds in either the algorithm or expert treatment. However, among participants who received advice, the change in performance points was significantly larger in the first 10 rounds than in the last 10 rounds at the 1% significance level.

This pattern may be partly explained by improvements in participants' initial forecasts over time. Among all participants, average performance points from the initial forecast increased from 12.667 to 14.490 in the algorithm treatments, and from 12.414 to 14.598 in the expert treatments. These increases, approximately 2 points from the first 10 rounds to the last 10 rounds, are statistically significant at the 1% level. Among participants who did not receive advice, the average performance points from the initial forecast increased from 12.584 to 14.514 in the algorithm treatments and from 12.270 to 14.773 in the expert treatments. These increases of approximately 2 points from the first 10 rounds to the last 10 rounds are statistically significant at the 1% significance level. Among participants who received advice, the average performance points from the initial forecast increased from 12.819 to 14.426 in the algorithm treatments and from 12.704 to 14.153 in the expert treatments. These increases of approximately 2 points from the first 10 rounds to the last 10 rounds are statistically significant at the 1% significance level.

Among participants who received advice, the average performance points of the advice in the first 10 rounds were 13.705 in the algorithm treatment and 13.756 in the expert treatment. In the last 10 rounds, the corresponding performance points were 13.460 in the algorithm treatment and 13.073 in the expert treatment. These results indicate that the advice outperformed participants' initial forecasts in the first 10 rounds but underperformed relative to participants' initial forecasts in the last 10 rounds.

After receiving and relying on advice, the average performance points of the final forecast in the first 10 rounds were 13.703 in the algorithm treatment and 14.393 in the expert treatment. In the last 10 rounds, the corresponding values were 14.034 and 14.942, respectively. Therefore, the improvement in performance points was substantially larger in the first 10 rounds than in the last 10 rounds for both treatments.

A similar logic applies to the change in net benefit. Since the price of advice is similar between the first 10 rounds (2.678 for algorithmic advice; 2.726 for expert advice) and the last 10 rounds (2.310 for algorithmic advice; 2.524 for expert advice) in both treatments, the change in net benefit is less negative in the first 10 rounds than in the last 10 rounds. at 5% significance level.

Table J3: Change in performance points and net benefit when comparing with and without receiving advice in Experiment 2

Treatments VARIABLES	(1)	(2) Round 1 to 10		(4)	(5)	(6) Round 11 to 20		(8)
	Algorithm $\Delta$ Perf.	Expert $\Delta$ Perf.	Algorithm $\Delta$ NB.	Expert $\Delta$ NB.	Algorithm $\Delta$ Perf.	Expert $\Delta$ Perf.	Algorithm $\Delta$ NB.	Expert $\Delta$ NB.
Advice received dummy	0.750*** (0.228)	1.634*** (0.212)	-1.887*** (0.257)	-1.086*** (0.249)	-0.384 (0.235)	0.677*** (0.247)	-2.674*** (0.265)	-1.783*** (0.264)
Description dummy	-0.012 (0.188)	0.158 (0.154)	-0.058 (0.200)	0.024 (0.168)	0.096 (0.143)	0.189 (0.193)	0.169 (0.161)	0.021 (0.213)
Constant	0.119 (0.134)	-0.028 (0.111)	0.128 (0.140)	0.037 (0.117)	-0.058 (0.090)	0.013 (0.100)	-0.100 (0.101)	0.079 (0.103)
Observations	1,010	1,000	1,010	1,000	1,010	1,000	1,010	1,000
Cluster at participant level	101	100	101	100	101	100	101	100
Wald $\chi^2$	10.903	63.333	56.375	19.852	4.829	11.243	114.215	46.943
Prob > $\chi^2$	0.004	0.000	0.000	0.000	0.089	0.004	0.000	0.000

*Notes:* (a) We regress the change in performance points, defined as the performance of the final forecast minus the performance of the initial forecast ( $\Delta$ Perf.), and the change in net benefit, defined as the net benefit of the final forecast minus the net benefit of the initial forecast ( $\Delta$ NB.), on the advice-received dummy, controlling for the description dummy in each pooled treatment, using a random-effects panel-data OLS regression model with standard errors clustered at the participant level. The advice-received dummy equals 1 if the participant receives advice for that graph, and 0 otherwise. (b) The unit of observation is the forecast made by each participant in each round. The total number of observations is the total number of forecasts with or without advice received in each pooled treatment. (c) Standard errors clustered at the participant level are reported in parentheses.

Table J4: Change in performance points and net benefit comparing treatments in Experiment 2

Treatments VARIABLES	(1)	(2) Round 1 to 10		(4)	(5)	(6) Round 11 to 20		(8)
	No $\Delta$ Perf.	Yes $\Delta$ Perf.	No $\Delta$ NB.	Yes $\Delta$ NB.	No $\Delta$ Perf.	Yes $\Delta$ Perf.	No $\Delta$ NB.	Yes $\Delta$ NB.
Algorithm dummy	0.050 (0.104)	-0.805*** (0.295)	0.050 (0.104)	-0.762** (0.334)	-0.113 (0.108)	-1.189*** (0.318)	-0.113 (0.108)	-0.972*** (0.371)
Description dummy	0.103 (0.103)	0.019 (0.299)	0.103 (0.103)	-0.242 (0.336)	0.024 (0.107)	0.431 (0.319)	0.024 (0.107)	0.261 (0.369)
Constant	0.000 (0.092)	1.680*** (0.265)	0.000 (0.092)	-0.908*** (0.274)	0.091 (0.080)	0.581* (0.298)	0.091 (0.080)	-1.861*** (0.328)
Observations	1,320	690	1,320	690	1,445	565	1,445	565
Cluster at participant level	199	183	199	183	199	168	199	168
Wald $\chi^2$	1.221	7.459	1.221	6.563	1.227	21.114	1.227	6.878
Prob > $\chi^2$	0.543	0.024	0.543	0.038	0.541	0.000	0.541	0.032

*Notes:* (a) We regress the change in performance points, defined as the performance of the final forecast minus the performance of the initial forecast ( $\Delta$ Perf.), and the change in net benefit, defined as the net benefit of the final forecast minus the net benefit of the initial forecast ( $\Delta$ NB.), on the algorithm dummy, controlling for the description dummy, using a random-effects panel-data OLS regression model with standard errors clustered at the participant level. (b) The unit of observation is the forecast made by each participant in each round. The total number of observations is the total number of forecasts with or without advice received in all treatments. (c) Standard errors clustered at the participant level are reported in parentheses.

Table J5: Change in performance points and net benefit when comparing former ten rounds and latter ten rounds in Experiment 2

Treatments	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Received advice	Algorithm	Expert	Algorithm	Expert	Algorithm	Expert	Algorithm	Expert
VARIABLES	No	No	Yes	Yes	No	No	Yes	Yes
	$\Delta$ Perf.	$\Delta$ Perf.	$\Delta$ Perf.	$\Delta$ Perf.	$\Delta$ NB.	$\Delta$ NB.	$\Delta$ NB.	$\Delta$ NB.
Former ten round dummy	0.113 (0.082)	-0.051 (0.118)	1.273*** (0.285)	0.894*** (0.269)	0.113 (0.082)	-0.051 (0.118)	0.905** (0.357)	0.707** (0.325)
Description dummy	-0.046 (0.102)	0.176 (0.115)	0.232 (0.347)	0.145 (0.348)	-0.046 (0.102)	0.176 (0.115)	0.275 (0.361)	-0.299 (0.365)
Constant	0.013 (0.081)	0.018 (0.083)	-0.508*** (0.192)	0.698** (0.332)	0.013 (0.081)	0.018 (0.083)	-2.840*** (0.359)	-1.558*** (0.336)
Observations	1,380	1,385	640	615	1,380	1,385	640	615
Cluster at participant level	100	100	97	95	100	100	97	95
Wald $\chi^2$	1.996	2.481	24.871	11.101	1.996	2.481	6.632	5.932
Prob > $\chi^2$	0.369	0.289	0.000	0.004	0.369	0.289	0.036	0.052

*Notes:* (a) We regress the change in performance points, defined as the performance of the final forecast minus the performance of the initial forecast ( $\Delta$ Perf.), and the change in net benefit, defined as the net benefit of the final forecast minus the net benefit of the initial forecast ( $\Delta$ NB.), on the former ten round dummy, controlling for the description dummy in each pooled treatment, using a random-effects panel-data OLS regression model with standard errors clustered at the participant level. The former ten round dummy equals 1 if round is from 1 to 10, and 0 otherwise. (b) The unit of observation is the forecast made by each participant in each round. The total number of observations is the total number of forecasts with or without advice received in each pooled treatment. (c) Standard errors clustered at the participant level are reported in parentheses.

## J.4 Factors influence WTP and reliance on advice

Table J6: The effects of beliefs about accuracy and confidence in initial forecasts, past experience, and stock price characteristics on WTP

Variables	(1) Algorithm WTP	(2) Expert WTP	(3) Algorithm WTP	(4) Expert WTP	(5) Algorithm WTP	(6) Expert WTP
Description dummy	-0.260 (0.367)	0.089 (0.414)	-0.599* (0.352)	0.049 (0.420)	-0.596* (0.352)	0.047 (0.419)
Round	-0.041*** (0.011)	-0.031*** (0.010)	-0.054*** (0.016)	-0.038** (0.015)	-0.055*** (0.016)	-0.044*** (0.016)
Evaluation of own accuracy in the initial forecast	-0.366*** (0.089)	-0.423*** (0.086)	-0.352*** (0.097)	-0.399*** (0.081)	-0.350*** (0.097)	-0.398*** (0.082)
<b>Confidence in initial forecasts</b>						
Confidence in the initial forecast	-0.324*** (0.054)	-0.382*** (0.072)	-0.343*** (0.053)	-0.401*** (0.079)	-0.338*** (0.053)	-0.391*** (0.077)
<b>Experiences</b>						
Mean prior net benefit change with purchased advice			0.035 (0.043)	0.026 (0.040)	0.037 (0.044)	0.025 (0.039)
Mean prior net benefit change without purchased advice			-0.252** (0.098)	-0.088 (0.151)	-0.252** (0.098)	-0.079 (0.153)
Number of advice received			0.002 (0.050)	0.005 (0.035)	0.006 (0.050)	0.006 (0.035)
<b>Time Series Characteristics</b>						
Autocorrelation coefficient					0.644 (0.491)	-0.019 (0.426)
Amplitude					-0.004 (0.005)	0.017*** (0.005)
Variance					0.000 (0.001)	-0.002*** (0.001)
MACD reversals					-0.021* (0.012)	0.008 (0.011)
Reversals					0.001 (0.004)	0.001 (0.005)
Constant	5.398*** (0.364)	5.304*** (0.367)	5.908*** (0.404)	5.555*** (0.420)	5.759*** (0.645)	4.886*** (0.643)
Observations	2,020	2,000	1,590	1,475	1,590	1,475
Cluster at participant level	101	100	96	95	96	95
Wald $\chi^2$	91.649	83.976	105.902	73.227	119.218	79.516
Prob > $\chi^2$	0.000	0.000	0.000	0.000	0.000	0.000

Notes: (a) We regress WTP in all rounds on the treatment dummy, evaluation of own accuracy in the initial forecast, confidence in the initial forecast and mean prior net benefit change with purchased advice, mean prior net benefit change without purchased advice, number of received advice from previous rounds, using a random effect panel data OLS regression model with standard errors clustered at participant level, controlling for personal characteristics. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.10, 0.05, and 0.01 levels, respectively. (b) The unit of observation is each subject's WTP in round 1. The total number of observations is the number of subjects in all treatments. (c) Numbers in parentheses are standard errors clustered at participant level.

Table J7: The effects of WTP, beliefs about accuracy and confidence in initial forecasts, change in confidence after receiving the advice, and past experience with and without receiving advices on Shift Rate conditional on personal characteristics

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	Algorithm SHIFT	Expert SHIFT	Algorithm SHIFT	Expert SHIFT	Algorithm SHIFT	Expert SHIFT
WTP	0.031*** (0.012)	0.006 (0.009)	0.072*** (0.024)	0.012 (0.010)	0.068*** (0.024)	0.010 (0.010)
Round number	-0.006 (0.005)	-0.006*** (0.002)	0.021 (0.014)	-0.001 (0.006)	0.019 (0.014)	-0.003 (0.006)
Description dummy	-0.017 (0.073)	-0.019 (0.046)	-0.072 (0.066)	-0.021 (0.057)	-0.060 (0.063)	-0.023 (0.056)
<b>Experiences</b>						
Mean prior net benefit change with purchased advice			0.039* (0.023)	-0.003 (0.012)	0.035 (0.023)	-0.005 (0.012)
Mean prior net benefit change with- out purchased advice			0.196 (0.159)	-0.005 (0.015)	0.194 (0.157)	-0.001 (0.016)
Number of advice received			-0.036** (0.015)	-0.005 (0.012)	-0.032** (0.015)	-0.003 (0.012)
<b>Confidence</b>						
Evaluation of own accuracy in the initial forecast			0.025 (0.044)	-0.004 (0.025)	0.022 (0.043)	-0.003 (0.025)
Confidence in the initial forecast			-0.102*** (0.038)	-0.008 (0.020)	-0.074** (0.036)	0.022 (0.024)
Change in confidence					0.066 (0.043)	0.046** (0.022)
Constant	0.522*** (0.134)	0.810*** (0.068)	0.449** (0.222)	0.779*** (0.104)	0.361 (0.245)	0.683*** (0.106)
Observations	604	569	478	445	478	445
Clustered at participant level	97	94	91	85	91	85
Wald $\chi^2$	16.097	9.083	55.352	5.199	61.329	13.149
Prob > $\chi^2$	0.001	0.028	0.000	0.736	0.000	0.156

*Notes:* (a) We use data for graphs for which participants received advice. We also restrict the sample to observations where the shift rate is between 0 and 1. We regress shift rate on the WTP in each round using a random effect panel data OLS model, with standard errors clustered at the participant level, controlling personal characteristics. WTP indicates the submitted WTP in each round. In Models (3) and (4), we add control variables, including the mean performance change from all previous rounds with purchased advice, the mean performance change from all previous rounds without purchased advice, the number of pieces of advice received in previous rounds, the evaluation of the accuracy of the initial forecast, confidence in the initial forecast. In Models (5) and (6), we add a control variable: the change in confidence from the initial forecast to the final forecast. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively. (b) The unit of observation is a round in which a participant received advice. The total number of observations is the number of SHIFT observations across all such rounds in the algorithm and expert treatments, including both the description and no-description conditions. (c) Numbers in parentheses are standard errors clustered at participant level.

## K Analyses of experimental results conditional on personal characteristics in Experiment 2

Table K1: Summary of personal characteristics

Treatments	Alg-No D	Alg- D	Exp -No D	Exp - D	<i>p</i> -value
Male (%)	58.000 (7.051)	54.000 (7.120)	55.102 (7.179)	73.469 (6.372)	0.134
Undergraduate (%)	70.000 (6.547)	50.000 (7.143)	75.510 (6.207)	69.388 (6.652)	0.082
Science major (%)	56.000 (7.091)	80.000 (5.714)	71.429 (6.521)	77.551 (6.022)	0.055
Stock trading experience (%)	18.000 (5.488)	24.000 (6.101)	24.490 (6.207)	22.449 (6.022)	0.875
Financial literacy score (out of 12)	8.100 (0.350)	8.480 (0.329)	8.061 (0.311)	8.143 (0.327)	0.893
No. of participants	50	51	50	50	

*Notes:* (a) Each personal characteristic was regressed on treatment dummies using an OLS regression model, with robust standard errors. The estimated dummy coefficients were compared using an F-test, with results presented as *p*-values. (b) Numbers in parentheses represent standard errors. (c) In Alg-D, Exp-No D, and Exp-D, one participant in each treatment refused to report their gender. Therefore, Male (%) is calculated based on 50, 49, and 49 participants, respectively.

Table K2: Predicted WTP conditional on personal characteristics

VARIABLES	(1)	(2)
	WTP Round 1	WTP All Rounds
Baseline: Alg-No D dummy		
Alg-D dummy	-0.286 (0.453)	-0.134 (0.386)
Exp-No D dummy	-0.723* (0.432)	-0.247 (0.394)
Exp-D dummy	0.364 (0.511)	0.007 (0.456)
Male dummy	0.068 (0.383)	0.037 (0.335)
Undergraduate student dummy	0.262 (0.353)	0.019 (0.301)
Science major dummy	0.205 (0.380)	0.102 (0.343)
Stock trading experience dummy	-0.498 (0.379)	-0.632* (0.340)
Financial literacy score	-0.100 (0.091)	-0.073 (0.078)
Constant	3.995*** (0.859)	3.922*** (0.757)
Observations	198	3,960
Cluster at participant level	/	198
R-squared	0.060	0.025
Treatments	Predicted WTP Task 1	Predicted WTP Task 2
Alg-No D	3.422 (0.324)	3.292 (0.288)
Alg-D	3.136 (0.325)	3.158 (0.255)
Exp-No D	2.699 (0.285)	3.046 (0.264)
Exp-D	3.786 (0.381)	3.299 (0.335)
Two-sided test	Prob > F	Prob > F
Alg-No D = Alg-D = Exp-No D = Exp-D	0.105	0.906
Alg-No D = Alg-D #	1.000	1.000
Exp-No D = Exp-D #	0.087	1.000
Alg-No D = Exp-No D#	0.381	1.000
Alg-D = Exp-D#	0.815	1.000

*Notes:* (a) We regressed WTP in Round 1 and across all rounds on treatment dummies using OLS regressions, controlling personal characteristics. Model (1) uses robust standard errors, while Model (2) uses robust standard errors clustered at the participant level. Pairwise comparisons of the estimated treatment coefficients were conducted using F-tests with Bonferroni-adjusted  $p$ -values. Results from two-sided tests are reported. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively. (b) The unit of observation is the participant in Model (1) and the participant-round observation in Model (2). The total number of observations equals the number of participants in Round 1 for Model (1) and the number of participant-round observations across all rounds for Model (2). (c) Standard errors are reported in parentheses. Model (1) reports robust standard errors, while Model (2) reports robust standard errors clustered at the participant level. (d) # indicates Bonferroni-adjusted  $p$ -values.

Table K3: The effects of beliefs about accuracy and confidence in initial forecasts, past experience, and stock price characteristics on WTP conditional on personal characteristics

Variables	(1) Algorithm WTP	(2) Expert WTP	(3) Algorithm WTP	(4) Expert WTP	(5) Algorithm WTP	(6) Expert WTP
Description dummy	-0.445 (0.366)	0.198 (0.408)	-0.649* (0.356)	0.080 (0.401)	-0.648* (0.355)	0.080 (0.402)
Round	-0.041*** (0.011)	-0.032*** (0.010)	-0.046*** (0.016)	-0.037** (0.015)	-0.047*** (0.016)	-0.042*** (0.015)
<b>Confidence in initial forecasts</b>						
Evaluation of own accuracy in the initial forecast	-0.371*** (0.088)	-0.424*** (0.086)	-0.335*** (0.094)	-0.390*** (0.081)	-0.333*** (0.094)	-0.390*** (0.081)
Confidence in the initial forecast	-0.325*** (0.055)	-0.375*** (0.074)	-0.341*** (0.052)	-0.397*** (0.081)	-0.335*** (0.053)	-0.387*** (0.079)
<b>Experiences</b>						
Mean prior performance change with purchased advice			0.169*** (0.050)	0.092 (0.057)	0.175*** (0.051)	0.091 (0.055)
Mean prior performance change without purchased advice			-0.219** (0.098)	-0.084 (0.156)	-0.219** (0.098)	-0.076 (0.157)
Number of advice received			-0.011 (0.053)	0.013 (0.034)	-0.008 (0.053)	0.013 (0.034)
<b>Time Series Characteristics</b>						
Autocorrelation coefficient					0.788 (0.490)	0.050 (0.426)
Amplitude					-0.003 (0.005)	0.016*** (0.005)
Variance					0.000 (0.001)	-0.002** (0.001)
MACD reversals					-0.023* (0.012)	0.006 (0.011)
Reversals					-0.002 (0.004)	-0.000 (0.005)
<b>Personal Characteristics</b>						
Male dummy	0.073 (0.407)	0.311 (0.476)	-0.002 (0.374)	0.477 (0.456)	-0.006 (0.373)	0.478 (0.456)
Undergraduate student dummy	0.065 (0.349)	0.182 (0.458)	0.329 (0.344)	0.288 (0.455)	0.329 (0.343)	0.282 (0.454)
Science major dummy	1.232*** (0.385)	-0.748 (0.533)	0.880** (0.391)	-0.780 (0.539)	0.881** (0.390)	-0.784 (0.539)
Stock trading experience dummy	-0.728* (0.394)	-0.541 (0.503)	-0.742** (0.373)	-0.320 (0.492)	-0.741** (0.373)	-0.318 (0.494)
Financial literacy score	-0.054 (0.089)	-0.062 (0.110)	-0.095 (0.082)	-0.067 (0.107)	-0.094 (0.082)	-0.067 (0.107)
Constant	5.193*** (0.794)	6.090*** (1.116)	5.854*** (0.817)	5.960*** (1.133)	5.893*** (0.961)	5.489*** (1.295)
Observations	2,000	1,960	1,572	1,457	1,572	1,457
Cluster at participant level	100	98	95	94	95	94
Wald $\chi^2$	101.241	84.914	126.223	77.510	141.872	90.254
Prob > $\chi^2$	0.000	0.000	0.000	0.000	0.000	0.000

Notes: (a) We regress WTP in all rounds on the treatment dummy, evaluation of own accuracy in the initial forecast, confidence in the initial forecast and mean prior performance change with purchased advice, mean prior performance change without purchased advice, number of received advice from previous rounds, using a random effect panel data OLS regression model with standard errors clustered at participant level, controlling for personal characteristics. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.10, 0.05, and 0.01 levels, respectively. (b) The unit of observation is each subject's WTP in round 1. The total number of observations is the number of subjects in all treatments. (c) Numbers in parentheses are standard errors clustered at participant level.

Table K4: The effects of WTP, beliefs about accuracy and confidence in initial forecasts, change in confidence after receiving the advice, and past experience with and without receiving advices on Shift Rate conditional on personal characteristics

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	Algorithm SHIFT	Expert SHIFT	Algorithm SHIFT	Expert SHIFT	Algorithm SHIFT	Expert SHIFT
WTP	0.024*** (0.006)	0.011* (0.006)	0.024*** (0.008)	0.009 (0.009)	0.024*** (0.008)	0.009 (0.009)
Round number	-0.011*** (0.002)	-0.006*** (0.002)	0.001 (0.005)	-0.002 (0.006)	0.001 (0.005)	-0.003 (0.006)
Description dummy	-0.021 (0.041)	0.043 (0.040)	-0.056 (0.041)	0.062 (0.046)	-0.059 (0.041)	0.061 (0.046)
<b>Experiences</b>						
Mean prior performance change with purchased advice			0.025** (0.012)	0.010 (0.009)	0.026** (0.012)	0.007 (0.009)
Mean prior performance change without purchased advice			-0.004 (0.017)	-0.019* (0.010)	-0.004 (0.017)	-0.016 (0.010)
Number of advice received			-0.013* (0.007)	-0.001 (0.010)	-0.014* (0.008)	0.000 (0.010)
<b>Confidence</b>						
Evaluation of own accuracy in the initial forecast			-0.019 (0.025)	-0.015 (0.025)	-0.018 (0.025)	-0.014 (0.025)
Confidence in the initial forecast			-0.055** (0.022)	-0.021 (0.018)	-0.063** (0.026)	0.002 (0.020)
Change in confidence					-0.015 (0.020)	0.036** (0.017)
<b>Personal characteristics</b>						
Male dummy	0.037 (0.045)	0.013 (0.040)	0.037 (0.049)	0.013 (0.047)	0.040 (0.049)	0.007 (0.047)
Undergraduate student dummy	-0.041 (0.039)	0.055 (0.051)	-0.021 (0.043)	0.050 (0.049)	-0.020 (0.043)	0.043 (0.049)
Science major dummy	0.070 (0.045)	0.033 (0.053)	0.111** (0.047)	0.062 (0.058)	0.114** (0.048)	0.058 (0.057)
Stock trading experience dummy	0.112** (0.050)	-0.056 (0.056)	0.072 (0.063)	-0.048 (0.052)	0.075 (0.064)	-0.033 (0.052)
Financial literacy score	-0.003 (0.010)	0.020 (0.013)	0.003 (0.010)	0.017 (0.014)	0.002 (0.010)	0.017 (0.014)
Constant	0.565*** (0.093)	0.507*** (0.122)	0.565*** (0.131)	0.507*** (0.147)	0.585*** (0.134)	0.449*** (0.149)
Observations	568	519	447	401	447	401
Clustered at participant level	96	92	90	83	90	83
Wald $\chi^2$	63.646	16.277	67.794	18.781	67.780	22.477
Prob > $\chi^2$	0.000	0.039	0.000	0.130	0.000	0.069

*Notes:* (a) We use data for graphs for which participants received advice. We also restrict the sample to observations where the shift rate is between 0 and 1. We regress shift rate on the WTP in each round using a random effect panel data OLS model, with standard errors clustered at the participant level, controlling personal characteristics. WTP indicates the submitted WTP in each round. In Models (3) and (4), we add control variables, including the mean performance change from all previous rounds with purchased advice, the mean performance change from all previous rounds without purchased advice, the number of pieces of advice received in previous rounds, the evaluation of the accuracy of the initial forecast, confidence in the initial forecast. In Models (5) and (6), we add a control variable: the change in confidence from the initial forecast to the final forecast. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively. (b) The unit of observation is a round in which a participant received advice. The total number of observations is the number of SHIFT observations across all such rounds in the algorithm and expert treatments, including both the description and no-description conditions. (c) Numbers in parentheses are standard errors clustered at participant level.

Table K5: The effects of beliefs about accuracy and confidence in initial forecasts, past experience (net benefit), and stock price characteristics on WTP conditional on personal characteristics

Variables	(1) Algorithm WTP	(2) Expert WTP	(3) Algorithm WTP	(4) Expert WTP	(5) Algorithm WTP	(6) Expert WTP
Description dummy	-0.445 (0.366)	0.198 (0.408)	-0.641* (0.373)	0.113 (0.408)	-0.640* (0.373)	0.111 (0.409)
Round	-0.041*** (0.011)	-0.032*** (0.010)	-0.048*** (0.016)	-0.037** (0.015)	-0.049*** (0.016)	-0.043*** (0.016)
<b>Confidence in initial forecasts</b>						
Evaluation of own accuracy in the initial forecast	-0.371*** (0.088)	-0.424*** (0.086)	-0.358*** (0.097)	-0.395*** (0.081)	-0.356*** (0.097)	-0.394*** (0.082)
Confidence in the initial forecast	-0.325*** (0.055)	-0.375*** (0.074)	-0.340*** (0.053)	-0.392*** (0.081)	-0.335*** (0.053)	-0.383*** (0.080)
<b>Experiences</b>						
Mean prior net benefit change with purchased advice			0.046 (0.042)	0.024 (0.041)	0.048 (0.043)	0.023 (0.040)
Mean prior net benefit change without purchased advice			-0.238** (0.099)	-0.086 (0.154)	-0.239** (0.099)	-0.079 (0.156)
Number of advice received			-0.013 (0.051)	0.007 (0.035)	-0.011 (0.051)	0.007 (0.035)
<b>Time Series Characteristics</b>						
Autocorrelation coefficient					0.604 (0.494)	-0.013 (0.435)
Amplitude					-0.003 (0.005)	0.016*** (0.005)
Variance					0.000 (0.001)	-0.002** (0.001)
MACD reversals					-0.019 (0.012)	0.008 (0.011)
Reversals					0.000 (0.004)	-0.000 (0.005)
<b>Personal Characteristics</b>						
Male dummy	0.073 (0.407)	0.311 (0.476)	0.069 (0.391)	0.501 (0.466)	0.067 (0.391)	0.503 (0.466)
Undergraduate student dummy	0.065 (0.349)	0.182 (0.458)	0.261 (0.360)	0.290 (0.467)	0.259 (0.360)	0.284 (0.466)
Science major dummy	1.232*** (0.385)	-0.748 (0.533)	0.834** (0.399)	-0.797 (0.549)	0.834** (0.399)	-0.800 (0.549)
Stock trading experience dummy	-0.728* (0.394)	-0.541 (0.503)	-0.734* (0.384)	-0.319 (0.509)	-0.733* (0.384)	-0.316 (0.510)
Financial literacy score	-0.054 (0.089)	-0.062 (0.110)	-0.111 (0.084)	-0.062 (0.108)	-0.111 (0.084)	-0.062 (0.108)
Constant	5.193*** (0.794)	6.090*** (1.116)	6.256*** (0.829)	6.079*** (1.154)	6.123*** (0.986)	5.573*** (1.305)
Observations	2,000	1,960	1,572	1,457	1,572	1,457
Cluster at participant level	100	98	95	94	95	94
Wald $\chi^2$	101.241	84.914	108.360	74.363	122.939	81.001
Prob > $\chi^2$	0.000	0.000	0.000	0.000	0.000	0.000

Notes: (a) We regress WTP in all rounds on the treatment dummy, evaluation of own accuracy in the initial forecast, confidence in the initial forecast and mean prior net benefit change with purchased advice, mean prior net benefit change without purchased advice, number of received advice from previous rounds, using a random effect panel data OLS regression model with standard errors clustered at participant level, controlling for personal characteristics. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.10, 0.05, and 0.01 levels, respectively. (b) The unit of observation is each subject's WTP in round 1. The total number of observations is the number of subjects in all treatments. (c) Numbers in parentheses are standard errors clustered at participant level.

Table K6: The effects of WTP, beliefs about accuracy and confidence in initial forecasts, change in confidence after receiving the advice, and past experience (net benefit) with and without receiving advices on Shift Rate conditional on personal characteristics

Variables	(1) Algorithm SHIFT	(2) Expert SHIFT	(3) Algorithm SHIFT	(4) Expert SHIFT	(5) Algorithm SHIFT	(6) Expert SHIFT
WTP	0.024*** (0.006)	0.011* (0.006)	0.028*** (0.008)	0.011 (0.009)	0.029*** (0.008)	0.010 (0.009)
Round number	-0.011*** (0.002)	-0.006*** (0.002)	-0.000 (0.005)	-0.001 (0.006)	0.000 (0.005)	-0.002 (0.006)
Description dummy	-0.021 (0.041)	0.043 (0.040)	-0.053 (0.042)	0.062 (0.046)	-0.056 (0.042)	0.060 (0.046)
<b>Experiences</b>						
Mean prior net benefit change with purchased advice			0.018** (0.009)	0.000 (0.009)	0.018** (0.009)	-0.002 (0.008)
Mean prior net benefit change without purchased advice			-0.004 (0.017)	-0.019* (0.010)	-0.004 (0.017)	-0.016 (0.010)
Number of advice received			-0.011 (0.008)	-0.002 (0.010)	-0.013 (0.008)	-0.001 (0.010)
<b>Confidence</b>						
Evaluation of own accuracy in the initial forecast			-0.022 (0.025)	-0.017 (0.024)	-0.021 (0.026)	-0.016 (0.025)
Confidence in the initial forecast			-0.052** (0.022)	-0.018 (0.017)	-0.060** (0.026)	0.006 (0.020)
Change in confidence					-0.016 (0.021)	0.039** (0.017)
<b>Personal characteristics</b>						
Male dummy	0.037 (0.045)	0.013 (0.040)	0.041 (0.049)	0.016 (0.047)	0.044 (0.050)	0.009 (0.047)
Undergraduate student dummy	-0.041 (0.039)	0.055 (0.051)	-0.027 (0.042)	0.051 (0.051)	-0.025 (0.043)	0.043 (0.051)
Science major dummy	0.070 (0.045)	0.033 (0.053)	0.111** (0.048)	0.063 (0.059)	0.114** (0.048)	0.059 (0.058)
Stock trading experience dummy	0.112** (0.050)	-0.056 (0.056)	0.072 (0.063)	-0.041 (0.052)	0.074 (0.064)	-0.025 (0.051)
Financial literacy score	-0.003 (0.010)	0.020 (0.013)	0.001 (0.010)	0.018 (0.014)	0.001 (0.010)	0.017 (0.014)
Constant	0.565*** (0.093)	0.507*** (0.122)	0.615*** (0.129)	0.508*** (0.148)	0.638*** (0.133)	0.441*** (0.149)
Observations	568	519	447	401	447	401
Clustered at participant level	96	92	90	83	90	83
Wald $\chi^2$	63.646	16.277	66.493	17.935	66.534	22.370
Prob > $\chi^2$	0.000	0.039	0.000	0.160	0.000	0.071

*Notes:* (a) We use data for graphs for which participants received advice. We also restrict the sample to observations where the shift rate is between 0 and 1. We regress shift rate on the WTP in each round using a random effect panel data OLS model, with standard errors clustered at the participant level, controlling personal characteristics. WTP indicates the submitted WTP in each round. In Models (3) and (4), we add control variables, including the mean net benefit change from all previous rounds with purchased advice, the mean net benefit change from all previous rounds without purchased advice, the number of pieces of advice received in previous rounds, the evaluation of the accuracy of the initial forecast, confidence in the initial forecast. In Models (5) and (6), we add a control variable: the change in confidence from the initial forecast to the final forecast. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively. (b) The unit of observation is a round in which a participant received advice. The total number of observations is the number of SHIFT observations across all such rounds in the algorithm and expert treatments, including both the description and no-description conditions. (c) Numbers in parentheses are standard errors clustered at participant level.

Table K7: Predicted Shift Rate conditional on personal characteristics

VARIABLES	(1) Shift	(2) Shift
Baseline: Alg-No D dummy		
Alg-D dummy	-0.025 (0.041)	
Exp-No D dummy	0.102** (0.041)	
Exp-D dummy	0.146*** (0.041)	
Algorithm dummy		-0.136*** (0.029)
Description dummy		0.009 (0.030)
Round	-0.009*** (0.001)	-0.009*** (0.001)
Male dummy	0.021 (0.032)	0.026 (0.032)
Undergraduate student dummy	-0.001 (0.032)	0.001 (0.033)
Science major dummy	0.063* (0.035)	0.060* (0.035)
Stock trading experience dummy	0.024 (0.039)	0.022 (0.040)
Financial literacy score	0.004 (0.008)	0.004 (0.008)
Constant	0.606*** (0.072)	0.727*** (0.076)
Observations	1,087	1,087
Cluster at participant level	188	188
Wald $\chi^2$	71.925	71.445
Prob > chi2	0.000	0.000
Two-sided test	Prob > F	Prob > F
Alg-No D = Alg-D = Exp-No D = Exp-D	0.000	
Alg-No D = Alg-D #	1.000	
Exp-No D = Exp-D #	1.000	
Alg-No D = Exp-No D#	0.053	
Alg-D = Exp-D#	0.000	

*Notes:* (a) In Model (1), we regress shift rate on treatment dummies, using a random-effects panel-data OLS regression model with standard errors clustered at the participant level, controlling for round number and personal characteristics. In Model (2), we regress shift rate on pooled treatment dummies, using a random-effects panel-data OLS regression model with standard errors clustered at the participant level, controlling for round number, description dummy and personal characteristics. Pairwise comparisons of the estimated treatment coefficients were conducted using F-tests with Bonferroni-adjusted  $p$ -values. Results from two-sided tests are reported. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively. (b) The unit of observation is the round which participant received advice. The total number of observations equals the number of rounds who received advice from all participants. (c) Standard errors are reported in parentheses. We report standard errors clustered at the participant level. (d) # indicates Bonferroni-adjusted  $p$ -values.

Table K8: Change in performance points and net benefit when comparing with and without receiving advice conditional on personal characteristics in Experiment 2

Treatments VARIABLES	(1)		(3)		(5)		(7)		(8)	
	Round 1 to 10				Round 11 to 20					
	Algorithm $\Delta$ Perf.	Expert $\Delta$ Perf.	Algorithm $\Delta$ NB.	Expert $\Delta$ NB.	Algorithm $\Delta$ Perf.	Expert $\Delta$ Perf.	Algorithm $\Delta$ NB.	Expert $\Delta$ NB.		
Advice received dummy	0.808*** (0.228)	1.607*** (0.216)	-1.799*** (0.260)	-1.076*** (0.252)	-0.377* (0.224)	0.672*** (0.257)	-2.638*** (0.247)	-1.759*** (0.273)		
Description dummy	-0.015 (0.184)	0.211 (0.155)	-0.041 (0.186)	0.062 (0.174)	0.078 (0.138)	0.146 (0.190)	0.202 (0.181)	-0.051 (0.219)		
Male dummy	0.176 (0.197)	-0.043 (0.161)	0.142 (0.205)	-0.090 (0.185)	0.203 (0.153)	0.092 (0.206)	0.228 (0.178)	0.092 (0.232)		
Undergraduate student dummy	-0.354* (0.186)	-0.042 (0.155)	-0.399** (0.177)	0.092 (0.200)	-0.154 (0.156)	-0.377 (0.288)	-0.099 (0.184)	-0.328 (0.302)		
Science major dummy	-0.312 (0.207)	-0.122 (0.165)	-0.517** (0.227)	0.072 (0.184)	0.006 (0.129)	0.162 (0.170)	-0.237 (0.176)	0.499** (0.208)		
Stock trading experience dummy	-0.109 (0.235)	0.196 (0.226)	0.032 (0.286)	0.291 (0.236)	-0.061 (0.157)	-0.037 (0.259)	0.105 (0.145)	-0.092 (0.300)		
Financial literacy score	-0.003 (0.037)	0.021 (0.031)	0.020 (0.044)	0.036 (0.040)	-0.001 (0.025)	-0.044 (0.061)	0.016 (0.033)	-0.011 (0.059)		
Constant	0.464 (0.309)	-0.127 (0.336)	0.415 (0.323)	-0.415 (0.425)	-0.052 (0.240)	0.493 (0.610)	-0.187 (0.358)	0.033 (0.609)		
Observations	1,000	980	1,000	980	1,000	980	1,000	980		
Cluster at participant level	100	98	100	98	100	98	100	98		
Wald $\chi^2$	16.697	62.382	76.237	24.250	11.687	12.539	152.209	64.280		
Prob > $\chi^2$	0.020	0.000	0.000	0.001	0.111	0.084	0.000	0.000		

*Notes:* (a) We regress the change in performance points, defined as the performance of the final forecast minus the performance of the initial forecast ( $\Delta$ Perf.), and the change in net benefit, defined as the net benefit of the final forecast minus the net benefit of the initial forecast ( $\Delta$ NB.), on the advice-received dummy, controlling for the description dummy and personal characteristics in each pooled treatment, using a random-effects panel-data OLS regression model with standard errors clustered at the participant level. The advice-received dummy equals 1 if the participant receives advice for that graph, and 0 otherwise. (b) The unit of observation is the forecast made by each participant in each round. The total number of observations is the total number of forecasts with or without advice received in each pooled treatment. (c) Standard errors clustered at the participant level are reported in parentheses.

Table K9: Change in performance points and net benefit comparing treatments conditional on personal characteristics in Experiment 2

Advice received VARIABLES	(1)	(2)		(3)	(4)	(5)	(6)		(7)	(8)
	No $\Delta$ Perf.	Round 1 to 10		No $\Delta$ NB.	Yes $\Delta$ NB.	No $\Delta$ Perf.	Round 11 to 20		No $\Delta$ NB.	Yes $\Delta$ NB.
Algorithm dummy	0.019 (0.098)	-0.810*** (0.283)		0.019 (0.098)	-0.782** (0.342)	-0.120 (0.112)	-1.221*** (0.309)		-0.120 (0.112)	-0.985*** (0.358)
Description dummy	0.122 (0.100)	0.029 (0.286)		0.122 (0.100)	-0.238 (0.326)	-0.007 (0.106)	0.386 (0.317)		-0.007 (0.106)	0.225 (0.382)
Male dummy	0.012 (0.109)	0.230 (0.291)		0.012 (0.109)	0.116 (0.347)	0.154 (0.136)	0.278 (0.311)		0.154 (0.136)	0.268 (0.353)
Undergraduate student dummy	0.024 (0.096)	-0.671** (0.315)		0.024 (0.096)	-0.540 (0.357)	-0.161 (0.152)	-0.472 (0.376)		-0.161 (0.152)	-0.273 (0.432)
Science major dummy	-0.123 (0.137)	-0.349 (0.316)		-0.123 (0.137)	-0.358 (0.395)	0.023 (0.105)	0.237 (0.281)		0.023 (0.105)	0.393 (0.347)
Stock trading experience dummy	-0.218* (0.130)	0.677 (0.430)		-0.218* (0.130)	1.059** (0.504)	-0.115 (0.142)	0.114 (0.475)		-0.115 (0.142)	0.431 (0.492)
Financial literacy score	0.012 (0.023)	0.004 (0.065)		0.012 (0.023)	0.061 (0.084)	-0.030 (0.035)	0.018 (0.071)		-0.030 (0.035)	0.095 (0.073)
Constant	0.008 (0.214)	2.085*** (0.655)		0.008 (0.214)	-1.023 (0.786)	0.380 (0.369)	0.434 (0.679)		0.380 (0.369)	-2.939*** (0.821)
Observations	1,300	680		1,300	680	1,422	558		1,422	558
Cluster at participant level	196	181		196	181	196	166		196	166
Wald $\chi^2$	6.320	19.284		6.320	21.475	2.593	24.903		2.593	10.853
Prob > $\chi^2$	0.503	0.007		0.503	0.003	0.920	0.001		0.920	0.145

*Notes:* (a) We regress the change in performance points, defined as the performance of the final forecast minus the performance of the initial forecast ( $\Delta$ Perf.), and the change in net benefit, defined as the net benefit of the final forecast minus the net benefit of the initial forecast ( $\Delta$ NB.), on the algorithm dummy, controlling for the description dummy and personal characteristics, using a random-effects panel-data OLS regression model with standard errors clustered at the participant level. (b) The unit of observation is the forecast made by each participant in each round. The total number of observations is the total number of forecasts with or without advice received in all treatments. (c) Standard errors clustered at the participant level are reported in parentheses.

Table K10: Change in performance points and net benefit when comparing former ten rounds and latter ten rounds conditional on personal characteristics in Experiment 2

Treatments	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Received advice	Algorithm	Expert	Algorithm	Expert	Algorithm	Expert	Algorithm	Expert
VARIABLES	No	No	Yes	Yes	No	No	Yes	Yes
	$\Delta$ Perf.	$\Delta$ Perf.	$\Delta$ Perf.	$\Delta$ Perf.	$\Delta$ NB.	$\Delta$ NB.	$\Delta$ NB.	$\Delta$ NB.
Former ten round dummy	0.094 (0.080)	-0.060 (0.120)	1.307*** (0.275)	0.889*** (0.270)	0.094 (0.080)	-0.060 (0.120)	0.955*** (0.344)	0.713** (0.331)
Description dummy	-0.035 (0.087)	0.162 (0.115)	0.204 (0.330)	0.185 (0.369)	-0.035 (0.087)	0.162 (0.115)	0.339 (0.344)	-0.243 (0.368)
Male dummy	0.075 (0.092)	0.056 (0.138)	0.453 (0.357)	0.132 (0.351)	0.075 (0.092)	0.056 (0.138)	0.392 (0.364)	0.084 (0.351)
Undergraduate student dummy	-0.001 (0.090)	-0.182 (0.153)	-0.773** (0.366)	-0.331 (0.389)	-0.001 (0.090)	-0.182 (0.153)	-0.815** (0.341)	-0.058 (0.418)
Science major dummy	-0.122 (0.127)	0.066 (0.120)	-0.198 (0.283)	-0.088 (0.386)	-0.122 (0.127)	0.066 (0.120)	-0.899** (0.362)	0.630* (0.360)
Stock trading experience dummy	-0.074 (0.120)	-0.231 (0.182)	-0.035 (0.458)	0.866* (0.498)	-0.074 (0.120)	-0.231 (0.182)	0.492 (0.497)	0.940* (0.493)
Financial literacy score	0.010 (0.021)	-0.032 (0.034)	-0.021 (0.072)	0.022 (0.075)	0.010 (0.021)	-0.032 (0.034)	0.039 (0.088)	0.086 (0.067)
Constant	-0.025 (0.164)	0.391 (0.360)	0.039 (0.550)	0.529 (0.809)	-0.025 (0.164)	0.391 (0.360)	-2.354*** (0.815)	-2.922*** (0.766)
Observations	1,364	1,358	636	602	1,364	1,358	636	602
Cluster at participant level	99	98	96	94	99	98	96	94
Wald $\chi^2$	2.271	7.040	34.969	22.299	2.271	7.040	19.409	18.496
Prob > $\chi^2$	0.943	0.425	0.000	0.002	0.943	0.425	0.007	0.010

*Notes:* (a) We regress the change in performance points, defined as the performance of the final forecast minus the performance of the initial forecast ( $\Delta$ Perf.), and the change in net benefit, defined as the net benefit of the final forecast minus the net benefit of the initial forecast ( $\Delta$ NB.), on the former ten round dummy, controlling for the description dummy and personal characteristics in each pooled treatment, using a random-effects panel-data OLS regression model with standard errors clustered at the participant level. The former ten round dummy equals 1 if round is from 1 to 10, and 0 otherwise. (b) The unit of observation is the forecast made by each participant in each round. The total number of observations is the total number of forecasts with or without advice received in each pooled treatment. (c) Standard errors clustered at the participant level are reported in parentheses.

## L Other Results in Experiment 2

### L.1 WTP trend across rounds

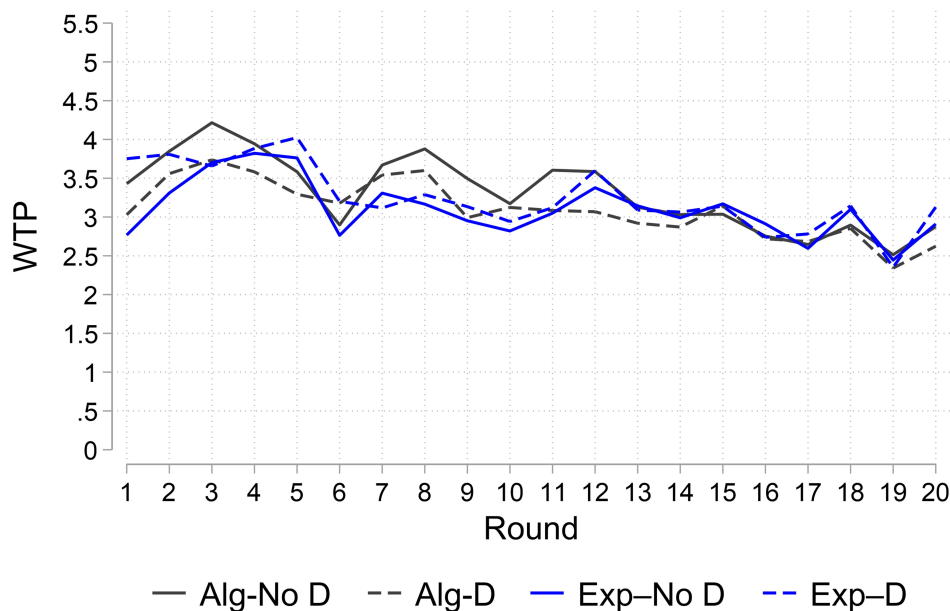
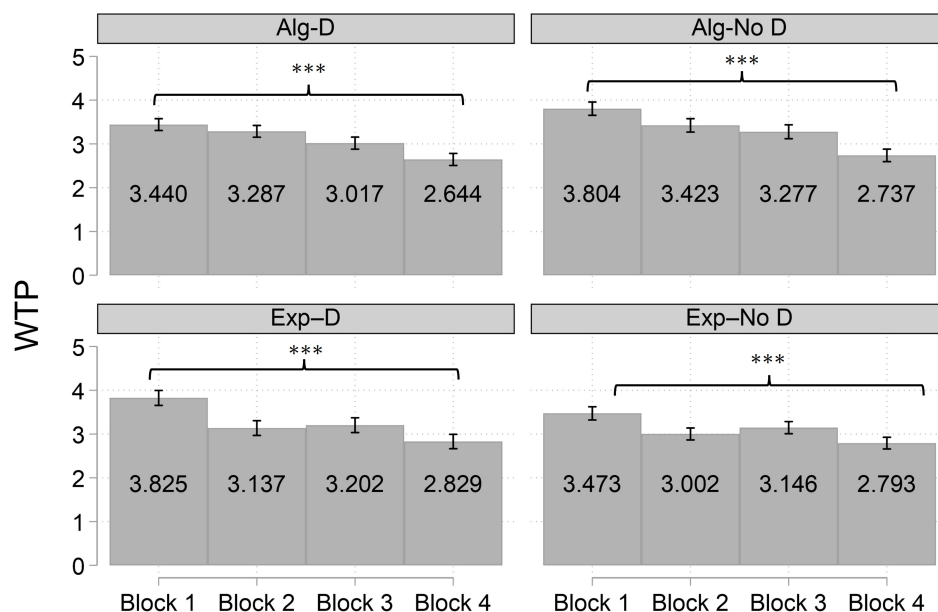


Figure L13: WTP trend in Experiment 2

Figure L13 shows the evolution of WTP across the 20 rounds in the four treatments. WTP follows a broadly similar pattern across treatments. It is relatively high in the early rounds, particularly around Rounds 3 to 5, and gradually declines thereafter. By the final rounds, WTP in all treatments converges to approximately 2.5–3.0.

The algorithm and expert treatments, with and without detailed descriptions, move closely together throughout the experiment. This suggests that providing a detailed description of the information source does not lead to a sustained increase in WTP for either algorithmic or expert advice. Overall, the result indicates that participants slightly reduce their WTP over time, while WTP remains broadly similar across treatments within each round.

## L.2 WTP across Blocks



*Notes:* We regressed the WTP on the block number (i.e., Blocks 1 and 4) by an OLS regression model with robust standard errors clustered at participant level. Pairwise comparisons of estimated dummy coefficients are done by F-test with Bonferroni-adjusted  $p$ -values. Results from two-sided tests are reported. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively.

Figure L14: WTP across blocks in Experiment 2

Table L1: WTP by Block

VARIABLES	(1)	(2)	(3)	(4)
	Block 1 Mean WTP	Block 2 Mean WTP	Block 3 Mean WTP	Block 4 Mean WTP
Baseline: Alg-No D dummy				
Alg-D dummy	-0.364 (0.387)	-0.136 (0.412)	-0.260 (0.431)	-0.093 (0.419)
Exp-No D dummy	-0.332 (0.402)	-0.421 (0.421)	-0.130 (0.433)	0.056 (0.412)
Exp-D dummy	0.021 (0.463)	-0.286 (0.478)	-0.074 (0.483)	0.092 (0.452)
Constant	3.804*** (0.288)	3.423*** (0.318)	3.277*** (0.324)	2.737*** (0.303)
Observations	201	201	201	201
R-squared	0.007	0.005	0.002	0.001

*Notes:* (a) We regressed each participant's mean WTP in each block on treatment dummies using an OLS regression model with robust standard errors. (b) The unit of observation is the participant-level mean WTP in each block. The number of observations equals the total number of participants across all treatments. (c) Robust standard errors are reported in parentheses.

Table L1 reports mean WTP by block. The constant terms indicate that average WTP declines over time in the baseline treatment, from 3.804 in Block 1 to 2.737 in Block 4. Within each block, however, WTP does not differ significantly across treatments. None of the estimated treatment coefficients is statistically significant in any block. These results suggest that WTP decreases over time, while remaining similar across treatments within each block.

### L.3 The Effects of Changes in Performance Points and Net Benefit on WTP

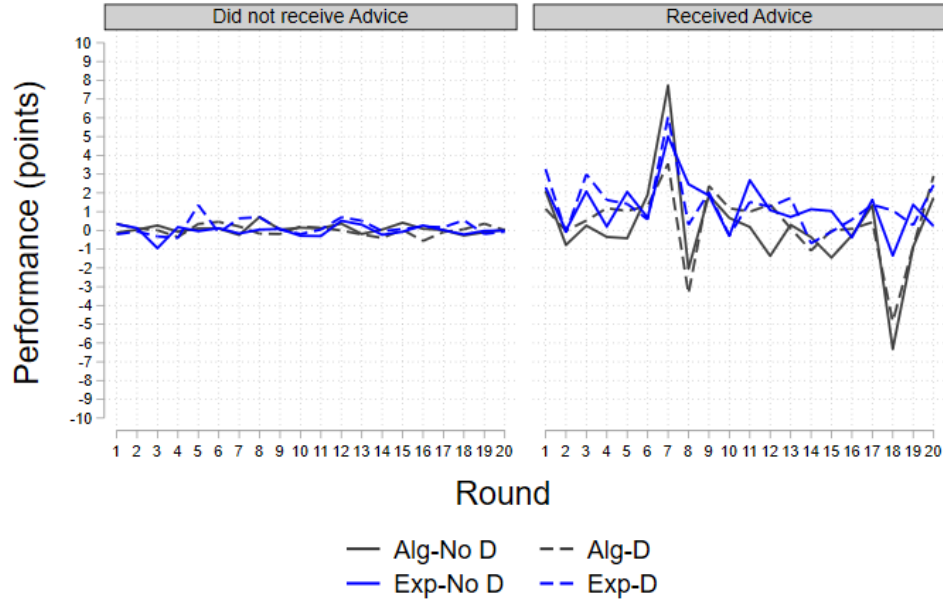


Figure L15: Performance (points) trend in Experiment 2

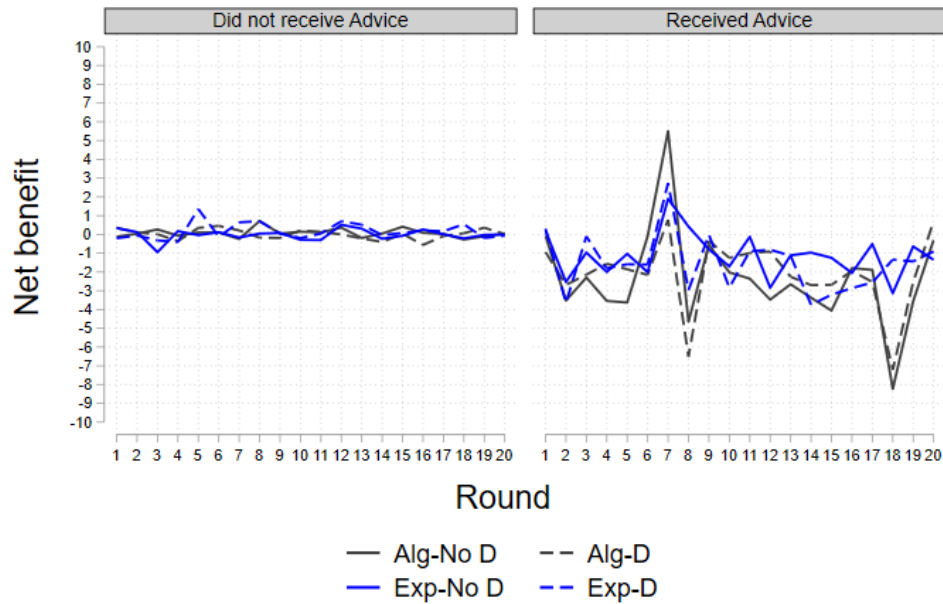


Figure L16: Net benefit trend in Experiment 2

Because detailed descriptions do not significantly affect either WTP or advice utilization, we pool the treatments with and without descriptions in the remaining analyses, while

controlling for the presence of descriptions with a treatment dummy.<sup>9</sup>

Table L2: The Effects of Changes in Performance Points on WTP

Variables	(1) Algorithm WTP	(2) Expert WTP	(3) Algorithm WTP	(4) Expert WTP
Mean prior performance change with purchased advice	0.207*** (0.054)	0.100* (0.056)	0.199*** (0.055)	0.092 (0.060)
Mean prior performance change without purchased advice			-0.252** (0.122)	-0.076 (0.141)
Number of received advice from previous rounds	0.070 (0.047)	0.055 (0.035)	0.023 (0.050)	0.038 (0.035)
Round number	-0.079*** (0.016)	-0.069*** (0.016)	-0.072*** (0.016)	-0.057*** (0.015)
Description dummy	-0.390 (0.359)	0.128 (0.407)	-0.537 (0.357)	0.095 (0.404)
Constant	3.995*** (0.302)	3.713*** (0.289)	4.158*** (0.321)	3.630*** (0.291)
Observations	1,663	1,541	1,590	1,475
Cluster at participant level	97	95	96	95
Wald $\chi^2$	62.414	24.693	60.064	22.585
Prob > chi2	0.000	0.000	0.000	0.000

*Notes:* (a) If a participant had not received advice in any previous round, the mean prior performance change with purchased advice is undefined. Similarly, if a participant had received advice in all previous rounds, the mean prior performance change without purchased advice is undefined. In Models 1 and 2, we regress WTP on the mean change in performance points from all previously purchased advice, using a random-effects panel-data OLS regression model with standard errors clustered at the participant level. The change in performance points is calculated as the points awarded for the final forecast minus the points awarded for the initial forecast in each round, excluding the endowment and the price of advice. In Models 3 and 4, we add the independent variable measuring the mean change in performance points from all previous rounds without purchased advice. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.10, 0.05, and 0.01 levels, respectively. (b) The unit of observation is a participant-round in which the participant received at least one piece of advice. The total number of observations is the number of such participant-round observations across all rounds in the algorithm and expert treatments, including both the description and no-description conditions. (c) Numbers in parentheses are standard errors clustered at participant level.<sup>10</sup>

<sup>9</sup>This pooled-treatment approach was not pre-registered.

<sup>10</sup>Models (1) and (2) were partly pre-registered, except that we pooled the treatments and added the number of received advice from previous rounds as a control variable. Models (3) and (4) were not pre-registered.

<sup>11</sup>Models (1) and (2) were partly pre-registered, except that we pooled the treatments and added the number of received advice from previous rounds as a control variable. Models (3) and (4) were not pre-registered.

Table L3: The Effects of Net Benefit on WTP

Variables	(1) Algorithm WTP	(2) Expert WTP	(3) Algorithm WTP	(4) Expert WTP
Mean prior net benefit change with purchased advice	0.042	0.000	0.061	0.027
	(0.041)	(0.044)	(0.044)	(0.044)
Mean prior net benefit change with- out purchased advice			-0.274**	-0.078
			(0.125)	(0.140)
Number of received advice from previous rounds	0.071	0.042	0.021	0.032
	(0.047)	(0.036)	(0.048)	(0.035)
Round number	-0.083***	-0.068***	-0.075***	-0.057***
	(0.017)	(0.017)	(0.016)	(0.016)
Description dummy	-0.372	0.169	-0.530	0.127
	(0.375)	(0.417)	(0.371)	(0.413)
Constant	4.207***	3.876***	4.398***	3.802***
	(0.295)	(0.293)	(0.330)	(0.306)
Observations	1,663	1,541	1,590	1,475
Cluster at participant level	97	95	96	95
Wald $\chi^2$	44.629	22.848	44.687	22.574
Prob > chi2	0.000	0.000	0.000	0.000

*Notes:* (a) If a participant had not received advice in any previous round, the mean prior net benefit change with purchased advice is undefined. Similarly, if a participant had received advice in all previous rounds, the mean prior net benefit change without purchased advice is undefined. In Models 1 and 2, we regress WTP on the mean change in net benefit from all previously purchased advice, using a random-effects panel-data OLS regression model with standard errors clustered at the participant level. The change in net benefit is calculated as the net benefit from the final forecast minus the net benefit from the initial forecast in each round. The net benefit from the initial forecast equals the reward from each graph plus the points earned in the slider task (i.e., 10 points). The net benefit from the final forecast equals the reward from each graph plus the points earned in the slider task (i.e., 10 points), minus the price of advice. In Models 3 and 4, we add the independent variable measuring the mean change in net benefit from all previous rounds without purchased advice. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.10, 0.05, and 0.01 levels, respectively. (b) The unit of observation is a participant-round in which the participant received at least one piece of advice. The total number of observations is the number of such participant-round observations across all rounds in the algorithm and expert treatments, including both the description and no-description conditions. (c) Numbers in parentheses are standard errors clustered at participant level.<sup>11</sup>

Table L4: The Effects of Changes in Performance Points in one round and two rounds ahead on WTP

Variables	(1) Algorithm WTP	(2) Expert WTP	(3) Algorithm WTP	(4) Expert WTP	(5) Algorithm WTP	(6) Expert WTP
Change in performance points at $t - 1$	0.005 (0.011)	-0.015 (0.016)	-0.042** (0.018)	-0.047** (0.022)	-0.035* (0.019)	-0.040** (0.018)
Received advice dummy at $t - 1$	-0.026 (0.105)	-0.006 (0.113)	-0.032 (0.105)	-0.044 (0.108)	-0.008 (0.110)	-0.063 (0.109)
Change in performance points $\times$ received advice dummy at $t - 1$			0.068*** (0.026)	0.059** (0.026)	0.071*** (0.027)	0.049** (0.023)
Change in performance points at $t - 2$					-0.029 (0.019)	0.011 (0.024)
Received advice dummy at $t - 2$					-0.021 (0.074)	0.093 (0.087)
Change in performance points $\times$ received advice dummy at $t - 2$					0.066** (0.028)	-0.017 (0.024)
Round number	-0.067*** (0.011)	-0.050*** (0.010)	-0.066*** (0.011)	-0.049*** (0.010)	-0.066*** (0.012)	-0.051*** (0.010)
Description dummy	-0.203 (0.390)	0.104 (0.431)	-0.206 (0.389)	0.107 (0.431)	-0.207 (0.393)	0.082 (0.434)
Constant	4.045*** (0.303)	3.681*** (0.299)	4.041*** (0.304)	3.672*** (0.299)	4.043*** (0.322)	3.682*** (0.317)
Observations	1,919	1,900	1,919	1,900	1,818	1,800
Cluster at participant level	101	100	101	100	101	100
Wald $\chi^2$	44.07	28.08	46.23	30.83	49.95	47.29
Prob > chi2	0.000	0.000	0.000	0.000	0.000	0.000

*Notes:* (a) We regress WTP on the change in performance points in one round ahead and two rounds ahead, using a random-effects panel-data OLS regression model with standard errors clustered at the participant level. The change in performance points at  $t - 1$  is calculated as the points awarded for the final forecast minus the points awarded for the initial forecast in one round ahead, excluding the endowment and the price of advice. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.10, 0.05, and 0.01 levels, respectively. (b) The unit of observation is each participant's WTP in a given round, conditional on the participant having received at least one piece of advice. The total number of observations is the number of WTP observations satisfying this condition across all rounds in the algorithm and expert treatments, including both the description and no-description conditions. (c) Numbers in parentheses are standard errors clustered at participant level.<sup>12</sup>

<sup>12</sup>Models (1) and (2) were partly pre-registered, except that we pooled the treatments. Models (3) to (6) were not pre-registered.

Table L5: The Effects of Changes in Net Benefit in one round ahead on WTP

Variables	(1) Algorithm WTP	(2) Expert WTP	(3) Algorithm WTP	(4) Expert WTP	(5) Algorithm WTP	(6) Expert WTP
Change in net benefit at $t-1$	-0.026* (0.014)	-0.031** (0.015)	-0.043** (0.019)	-0.047** (0.022)	-0.036* (0.019)	-0.038** (0.018)
Received advice dummy at $t-1$	-0.072 (0.095)	-0.067 (0.102)	-0.060 (0.097)	-0.056 (0.103)	-0.017 (0.101)	-0.076 (0.102)
Change in net benefit $\times$ received advice dummy at $t-1$			0.022 (0.021)	0.027 (0.026)	0.025 (0.020)	0.026 (0.024)
Change in net benefit at $t-2$					-0.031 (0.020)	0.010 (0.024)
Received advice dummy at $t-2$					-0.009 (0.072)	0.056 (0.082)
Change in net benefit $\times$ received advice dummy at $t-2$					0.029 (0.021)	-0.032 (0.023)
Round number	-0.068*** (0.010)	-0.050*** (0.010)	-0.068*** (0.010)	-0.050*** (0.009)	-0.070*** (0.011)	-0.051*** (0.010)
Description dummy	-0.202 (0.388)	0.102 (0.430)	-0.203 (0.388)	0.105 (0.431)	-0.201 (0.393)	0.078 (0.432)
Constant	4.055*** (0.302)	3.683*** (0.298)	4.055*** (0.302)	3.681*** (0.298)	4.074*** (0.320)	3.694*** (0.317)
Obs.	1,919	1,900	1,919	1,900	1,818	1,800
Cluster at participant level	101	100	101	100	101	100
Wald $\chi^2$	44.24	29.25	44.40	29.76	44.54	43.42
Prob > chi2	0.000	0.000	0.000	0.000	0.000	0.000

*Notes:* (a) We regress WTP on the change in net benefit in one round ahead and two rounds ahead, using a random-effects panel-data OLS regression model with standard errors clustered at the participant level. The change in net benefit at  $t-1$  is calculated as the net benefit from the final forecast minus the net benefit from the initial forecast in one round ahead. The net benefit from the initial forecast equals the reward from each graph plus the points earned in the slider task (i.e., 10 points). The net benefit from the final forecast equals the reward from each graph plus the points earned in the slider task (i.e., 10 points), minus the price of advice. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.10, 0.05, and 0.01 levels, respectively. (b) The unit of observation is each participant's WTP in a given round, conditional on the participant having received at least one piece of advice. The total number of observations is the number of WTP observations satisfying this condition across all rounds in the algorithm and expert treatments, including both the description and no-description conditions. (c) Numbers in parentheses are standard errors clustered at participant level.<sup>13</sup>

<sup>13</sup>Models (1) and (2) were partly pre-registered, except that we pooled the treatments. Models (3) to (6) were not pre-registered.

## L.4 The Characteristics of Stock Price

Table L6: The Effects of Characteristics of the stock price series on WTP

Variables	(1)	(2)	(3)	(4)
	Panel OLS Algorithm WTP	Panel OLS Expert WTP	OLS Algorithm WTP	OLS Expert WTP
Autocorrelation coefficient	-0.290 (0.520)	-0.023 (0.448)	-0.290 (0.520)	-0.023 (0.448)
Amplitude	0.004 (0.005)	0.018*** (0.005)	0.004 (0.005)	0.018*** (0.005)
Variance	0.000 (0.001)	-0.001*** (0.001)	0.000 (0.001)	-0.001*** (0.001)
MACD reversals	-0.019 (0.012)	-0.010 (0.010)	-0.019 (0.012)	-0.010 (0.010)
Reversals	0.004 (0.004)	0.000 (0.004)	0.004 (0.004)	0.000 (0.004)
Round number	-0.057*** (0.010)	-0.049*** (0.010)	-0.057*** (0.010)	-0.049*** (0.010)
Description dummy	-0.213 (0.382)	0.145 (0.425)	-0.213 (0.382)	0.145 (0.425)
Constant	3.699*** (0.594)	3.227*** (0.579)	3.699*** (0.594)	3.227*** (0.579)
Observations	2,020	2,000	2,020	2,000
Cluster at participant level	101	100	101	100
Wald $\chi^2$	49.350	45.840		
Prob > chi2	0.000	0.000		
R-squared			0.025	0.016

*Notes:* (a) We regress WTP on the characteristics of stock price time series, using a random-effects panel-data OLS regression model with standard errors clustered at the participant level in Model (1) and (2). We regress WTP on the characteristics of stock price time series, using a OLS regression model with robust standard errors clustered at the participant level in Model (3) and (4). The symbols \*, \*\*, and \*\*\* indicate significance at the 0.10, 0.05, and 0.01 levels, respectively. (b) The unit of observation is each subject's WTP in a given round. The total number of observations is the number of WTP observations across all rounds in the algorithm and expert treatments, including both the description and no-description conditions. (c) Numbers in parentheses are standard errors clustered at participant level.<sup>14</sup>

## L.5 Shift rate without restriction

In the manuscript, our analysis focuses on observations with shift rates between 0 and 1. Shift rates below 0 or above 1 capture movements away from the advice, or beyond it, and therefore reflect different behavioral responses.

In this section, we present results without restricting the shift rate to lie between 0 and 1. Specifically, we include all observations, including those with shift rates below 0 or above 1. The shift rate ranges from -19.216 to 22.223. Unlike in Experiment 1, there are no extreme shift-rate values.

The number of observations with shift rates greater than 1 is 51 in the Algorithm treatment (25 in Alg - No D and 26 in Alg - D), 80 in the Expert treatment (37 in Exp -

<sup>14</sup>Models (1) to (2) were not pre-registered. Models (3) and (4) were partly pre-registered, except that we pooled the treatments.

No D and 43 in Exp - D).

The number of observations with shift rates below 0 is 17 in the Algorithm treatment (7 in Alg - No D and 10 in Alg - D), 4 in the Expert treatment (1 in Exp - No D and 3 in Exp - D).

The number of observations with shift rates between 0 and 1 is 572 in the Algorithm treatment (284 in Alg - No D and 288 in Alg - D), 531 in the Expert treatment (264 in Exp - No D and 267 in Exp - D).

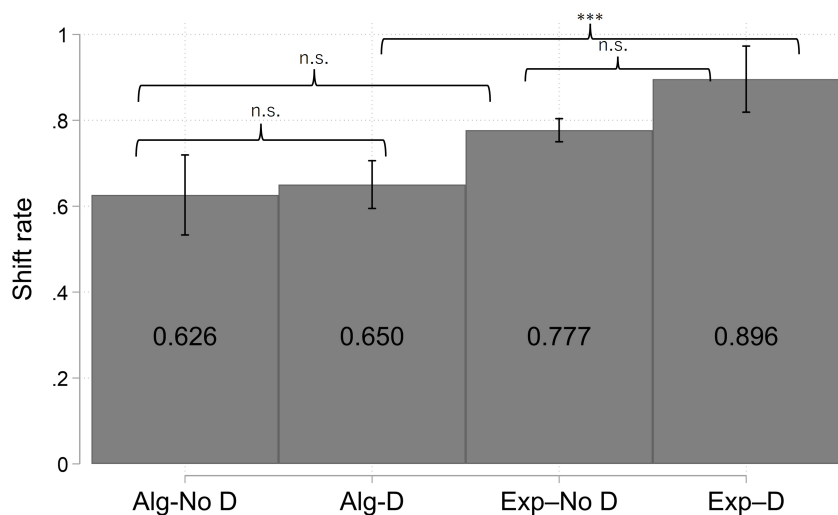


Figure L17: Shift rate in Experiment 2

*Notes:* We consider all shift rate observations without restriction. We regress shift rate on treatment dummies, using a random-effects panel-data OLS regression model with standard errors clustered at the participant level, controlling for round number. Pairwise comparisons of estimated dummy coefficients are done by F-test with Bonferroni-adjusted  $p$ -values. Results from two-sided tests are reported. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively.

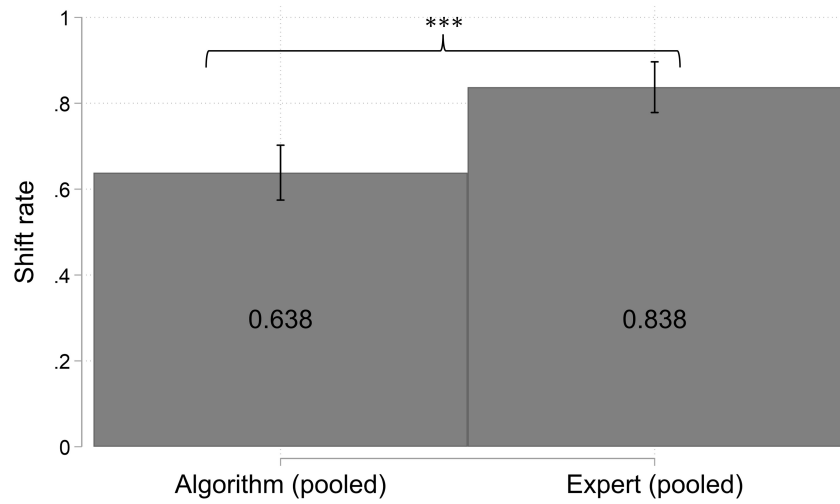


Figure L18: Shift rate between algorithm and expert in Experiment 2

*Notes:* We consider all shift rate observations without restriction. We regress shift rate on pooled treatment dummies, using a random-effects panel-data OLS regression model with standard errors clustered at the participant level, controlling for round number and description dummy. Pairwise comparisons of estimated dummy coefficients are done by F-test. Results from two-sided tests are reported. The symbols \*, \*\*, and \*\*\* indicate significance at the 0.1, 0.05, and 0.01 levels, respectively.

## M Experimental Instruction in Experiment 2 (English Translation)

# Experimental Instruction

1

## Welcome to the experiment

- We will now begin the experiment. In addition to the participation fee of 500 yen for taking part in this experiment, you will also receive reward based on the choices you make during the experiment.
- You are not allowed to communicate with other participants during the experiment. Additionally, please turn off all electronic devices such as smartphones and iPads during the experiment.
- If you have any questions during the experiment, please raise your hand. We will answer them individually.

2

## **Experiment details**

The experiment consists of three parts:

Part 1: Questionnaire

Part 2: Experiment Task 1

Part 3: Experiment Task 2

The total duration of the experiment is approximately one hour.

3

## **Experiment details: Questionnaire**

- In the questionnaire, you will be asked mainly about your experience with stock investments and your knowledge of financial literacy to assess your knowledge and experience related to financial investment.
- Your responses will not affect your reward, but please answer honestly according to your personal situation.
- The estimated time for completing the questionnaire is 10 minutes.
- If the time limit is exceeded, you will be automatically moved to the next step.
- Once you have completed the questionnaire, click the 'Next' button to begin Task 1.

4

# Experiment details: Task 1

- Use your mouse to drag the slider handle so that the slider points to the value **50**.
- There are **20 sliders** in total.
- When you have finished, click the **“Complete”** button.
- You cannot click the **“Complete”** button unless you have completed **100%** of the task, so you will not be able to proceed to the next page.
- You will receive **0.5 points** for each slider you complete.
- If you complete **100%** of the task, you will receive **10 points**.
- The **10 points** you earn will be used to pay for additional information in **Task 2**.
- How you can use the **10 points** you earn will be explained in detail later.

5

# Experiment details: Task 1

## Slider Task

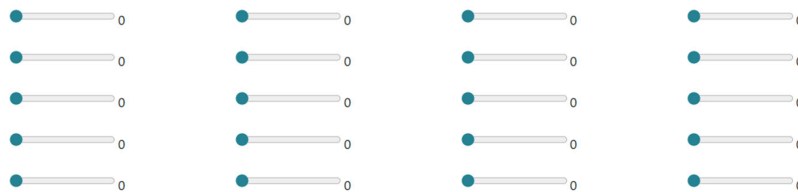
Use your mouse to drag each slider handle so that the slider points to the value **50**.

There are **20 sliders** in total.

Number of sliders currently set to the correct position: **0**

Current completion rate: **0.00%**

When you have finished, please click the **“Complete”** button.



Complete

6

## Experiment details: Task 2

- You will be asked to predict future stock prices based on past stock price information. Your reward will depend on how accurate your predictions are, as will be explained later.
- 20 questions will be presented. The stock price graph shown in each question will be different.
- In each question, you will be shown a graph displaying the closing prices of a particular stock over a 12-month period. You will then be asked to predict the stock price 30 days after the last day shown on the graph.
- The period for each graph is randomly selected from the time between January 1, 2008, and June 30, 2018, and will cover a 12-month period starting from that randomly selected date.
- The stock will be randomly chosen from the S&P 500.
- The stock name and the exact period will not be displayed.

7

## Experiment details: Task 2

- The flow of each question is as follows.

Step	In each question	What you will do
1	First half (Initial Forecast)	Read the graph of a stock and make a prediction
2	Confidence 1	Indicate how confident you are in your initial forecast
3	Evaluation of Additional Information	Evaluate the additional information
4	WTP	Indicate how much you are willing to pay (WTP) for the additional information
5	Second half (Final Forecast)	Read the graph and the additional information, then make a prediction
6	Confidence 2	Indicate how confident you are in your final forecast

- More details will be provided later.

8

## Experiment details : The first half of each question

- In the first half, as mentioned earlier, you will read a graph displaying the closing prices of a particular stock over a 12-month period and predict the stock price 30 days after the last day shown on the graph.
- All graphs are normalized so that the stock price on the first day of the graph is set to 100.
- Please submit your predicted value with up to three decimal places.
- In the next slide, a sample question will be displayed.

9

### Sample Question

Please read the below graph displaying the closing prices of a particular stock over a 12-month period and predict the stock price 30 days after the last day shown on the graph. The closing price on the last day is 110.693.



Your prediction:

Note: Please submit your prediction up to three decimal places.

Next

10

### Experiment details : The second half of each question

- In the second half of each question, the same graph as in the first half will be shown, and you will make a prediction again.
- However, in the second half, you may have the opportunity to obtain additional information to assist with your predictions.
- Whether you can obtain the additional information depends on the amount you submit between the first and second halves as your willingness to pay (WTP, the maximum amount you are willing to pay to purchase the additional information) and the randomly determined price ( $p$ ) of the additional information.

11

### Experiment details : The second half of each question

- **Participants who obtained additional information:** they can refer to both the graph and the additional information to predict the stock price again.
- **Participants who did not obtain additional information:** they will predict the stock price again using only the graph, just as in the first half.
- As in the first half, please submit your predicted value with up to three decimal places.

12

## Obtaining additional information

- Before the second half begins, an explanation will be provided on how the availability of additional information is determined, based on the submitted willingness to pay (WTP) and the price ( $p$ ) of the additional information, which is randomly set for each question.
- In this experiment, the WTP that can be submitted is limited to numbers **from 0.0 to 10.0, with one decimal place.**
- The price of the additional information ( $p$ ) ranges **from 0.1 to 10.0 and is randomly determined** for each of the 10 questions in the second half. All prices are chosen with the same probability.

13

## Purchase of additional information

- In a given graph, if the WTP is greater than or equal to the price of the additional information ( $p$ ) ( $WTP \geq p$ ):
  - You will pay  $p$  and obtain the additional information.
- Conversely, if the WTP is less than  $p$  ( $WTP < p$ ):
  - You will not obtain the additional information, and no payment will occur.
- The WTP is submitted before the second half of each question begins.
- The price of the additional information ( $p$ ) will be randomly determined separately for each question.

14

## Purchase of additional information

- Example 1

- For instance, suppose you submit a WTP of 10.0.
- In this case, since the price of the additional information determined randomly for each graph is at most 10.0, you will obtain additional information for all graphs and pay the price of the additional information for each graph.

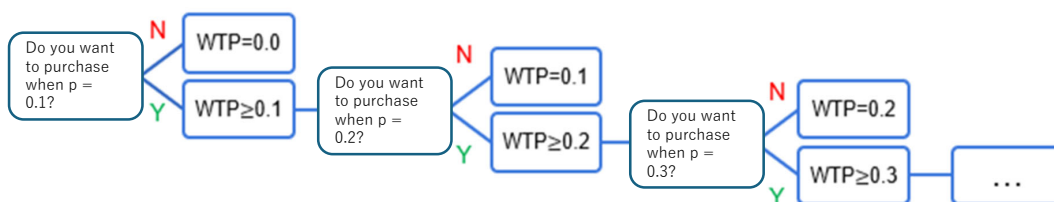
- Example 2

- For instance, suppose you submit a WTP of 0.0.
- In this case, since the minimum price of the additional information determined randomly for each graph is 0.1, you will not obtain additional information for any graph, and no payment for the additional information will occur either.

15

## How to determine the willingness to pay (WTP)

When determining the WTP, first consider the following questions:



Continue this until your answer changes from 'yes' to 'no'.

The price at which your answer changes is your WTP.

16

## Endowment for purchasing additional information

- The 10 points earned in Task 1 will be used to pay for additional information in Task 2.
- **Example 1:** If you submit **0.0** as your WTP and do not purchase any additional information, the **10 points** will remain unchanged.
- **Example 2:** If you submit **10.0** as your WTP, obtain the additional information, and pay its price of **9.0 points**, you will have **1 point** remaining.

17

## Example of Purchasing Additional Information and Point Calculation

Question	WTP	Price of additional information(p)	Can the additional information be purchased?	Remaining points
Q1	10.0	3.5	Yes	$10 - 3.5 = 6.5$
Q2	0.0	6.2	No	10
Q3	5.1	0.3	Yes	$10 - 0.3 = 9.7$
Q4	7.3	5.1	Yes	$10 - 5.1 = 4.9$
...	...	...	...	...

18

## Calculation of rewards

- One question will be randomly selected from the 20 questions. Then, either the prediction from the first half or the prediction from the second half will be randomly selected.
- Your additional reward will be determined based on the accuracy of the selected prediction.
- Please refer to the table below for how the additional reward is calculated for predictions in the first half and the second half.

Degree of deviation from the true value	Points
±0%	20
±0% ~±20%	20- Degree of deviation×100
±20%~	0

19

## Example of point calculation

For example, suppose the true value for a particular question 118.890.

- If your prediction is 130.123 :

The degree of deviation from the true value is

$$\left| \frac{130.123 - 118.890}{118.890} \right| \approx 0.09448 \text{ (9.448\%)}$$

You get  $20 - |0.09448 \times 100| \approx 10.552$  points.

- If your prediction is 90.002 :

The degree of deviation from the true value is

$$\left| \frac{90.002 - 118.890}{118.890} \right| \approx 0.243 \text{ (24.3\%)}$$

You get no points.

20

## N Screenshots in Task 1 in Experiment 2 (English Translation)

[Screen 1]

### Task 1: Introduction

Use your mouse to move the blue slider handle so that the slider points to the value 50.

The total number of sliders is 20.

When you have finished, click the “Complete” button.

You cannot click the “Complete” button unless you have completed 100% of the task, so you will not be able to proceed to the next page.

You will receive 0.5 points for each slider you complete.

If you complete 100% of the task, you will receive 10 points.

The 10 points you earn will be used to pay for additional information.

Next

[Screen 2]

### Slider Task

Use your mouse to drag each slider handle so that the slider points to the value 50.

There are 20 sliders in total.

Number of sliders currently set to the correct position: 0

Current completion rate: 0.00%

When you have finished, please click the “Complete” button.



Complete

[Screen 3] Results

Number of sliders correctly positioned: 20.0

Completion rate: 100.0

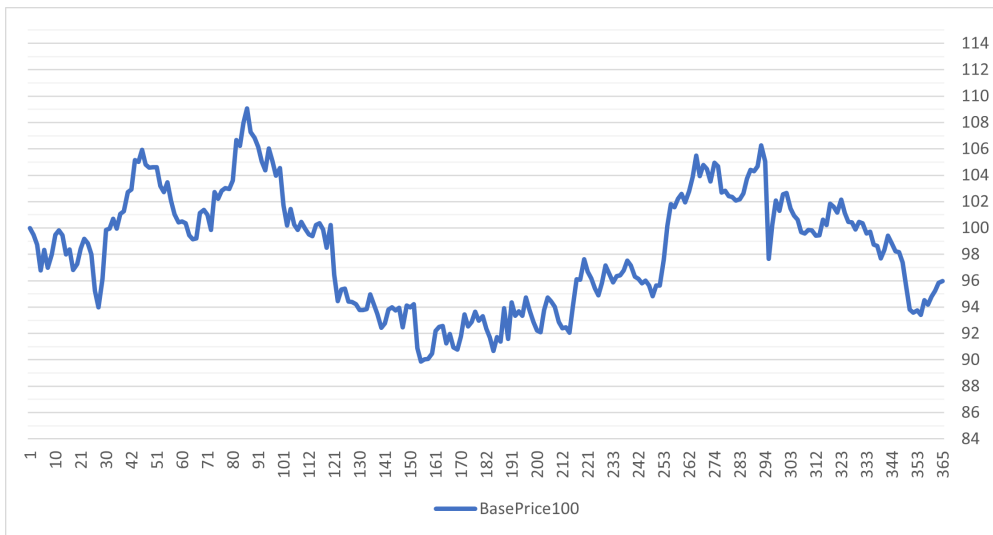
Slider task reward =  $20.0 \times 0.5$  points = 10 points

Next

## O Screenshots in Task 2 in Experiment 2 (English Translation)

[Screen 1] Question 1

Please read the below graph displaying the closing prices of a particular stock over a 12-month period and predict the stock price 30 days after the last day shown on the graph. The closing price on the last day is 95.962.



Your prediction:

Note: Please submit your prediction up to three decimal places.

Next

[Screen 2]

Question 1: Confidence

Please evaluate how confident you are in your initial forecast for the stock price graph in this question.

- Not confident at all
- Not very confident
- Neither confident nor unconfident
- Somewhat confident
- Very confident

Next

[Screen 3] (shown in algorithm without description treatment)

### **Question 1: Explanation and Evaluation of the Information Source**

The additional information provided here is the stock price 30 days after the final day shown in the graph, submitted by [an algorithm](#) crafted to predict future stock prices, based on the same stock price graph used in this experiment.

In this question, how do you think the accuracy of your initial prediction compares with the accuracy of the prediction based on the additional information?

- The accuracy of my initial prediction is much lower than that of the additional information.
- The accuracy of my initial prediction is somewhat lower than that of the additional information.
- The accuracy of my initial prediction is about the same as that of the additional information.
- The accuracy of my initial prediction is somewhat higher than that of the additional information.
- The accuracy of my initial prediction is much higher than that of the additional information.

Next

[Screen 3] (shown in algorithm with description treatment)

### **Question 1: Explanation and Evaluation of the Information Source**

The additional information provided here is the stock price 30 days after the final day shown in the graph, submitted by [an algorithm](#) crafted to predict future stock prices,

based on the same stock price graph used in this experiment.

This algorithm makes a future stock price forecast by learning the historical stock price information, from January 1st, 2000, or from the initial public offering (IPO) day to January 1st, 2020, of 83 target companies ranked at the top in their capital market sectors (i.e., Basic Materials, Consumer Goods, Healthcare, Services, Utilities, Conglomerates, Financial, Industrial Goods, Technology).

In this question, how do you think the accuracy of your initial prediction compares with the accuracy of the prediction based on the additional information?

- The accuracy of my initial prediction is much lower than that of the additional information.
- The accuracy of my initial prediction is somewhat lower than that of the additional information.
- The accuracy of my initial prediction is about the same as that of the additional information.
- The accuracy of my initial prediction is somewhat higher than that of the additional information.
- The accuracy of my initial prediction is much higher than that of the additional information.

Next

[Screen 3] (shown in expert without description treatment)

### **Question 1: Explanation and Evaluation of the Information Source**

The additional information provided here is the stock price 30 days later, submitted by [human experts](#) based on the stock price graph used in this experiment.

In this question, how do you think the accuracy of your initial prediction compares with the accuracy of the prediction based on the additional information?

- The accuracy of my initial prediction is much lower than that of the additional information.
- The accuracy of my initial prediction is somewhat lower than that of the additional information.
- The accuracy of my initial prediction is about the same as that of the additional information.

- The accuracy of my initial prediction is somewhat higher than that of the additional information.
- The accuracy of my initial prediction is much higher than that of the additional information.

Next

[Screen 3] (shown in expert with description treatment)

### **Question 1: Explanation and Evaluation of the Information Source**

The additional information provided here is the average of stock price predictions submitted by 198 [experts \(CMAs\)](#) who participated in a similar experiment in the past, based on the stock price chart used in this experiment, for the stock prices 30 days ahead.

CMA stands for Certified Member Analyst of the Securities Analysts Association of Japan. The CMA is a qualification granted to those who have taken the prescribed training courses and passed the examinations based on these courses and fulfilled certain requirements; it is a sign of an expertise in the fields of finance and investments. The CMAs' investment valuations are based on the calculation of the corporate value of an investment and the forecasting of its future value. The particularly crucial knowledge required can be broadly categorized into three main areas: evaluation of investment data, decision making on investment policies, and construction and management of portfolios. In this question, how do you think the accuracy of your initial prediction compares with the accuracy of the prediction based on the additional information?

- The accuracy of my initial prediction is much lower than that of the additional information.
- The accuracy of my initial prediction is somewhat lower than that of the additional information.
- The accuracy of my initial prediction is about the same as that of the additional information.
- The accuracy of my initial prediction is somewhat higher than that of the additional information.
- The accuracy of my initial prediction is much higher than that of the additional information.

Next

[Screen 4] (shown in algorithm without description treatment)

**Question 1: Explanation of Information Source and WTP**

The additional information provided here is the stock price 30 days after the final day shown in the graph, submitted by [an algorithm](#) crafted to predict future stock prices, based on the same stock price graph used in this experiment.

The 10 points earned in the slider task will be used to pay for the additional information.

How much are you willing to pay, up to a maximum of 10.0, for information from the algorithm?

Note: Submissions can range from 0 to 10, up to one decimal place.

Next

[Screen 4] (shown in algorithm with description treatment)

**Question 1: Explanation of Information Source and WTP**

The additional information provided here is the stock price 30 days after the final day shown in the graph, submitted by [an algorithm](#) crafted to predict future stock prices, based on the same stock price graph used in this experiment.

This algorithm makes a future stock price forecast by learning the historical stock price information, from January 1st, 2000, or from the initial public offering (IPO) day to January 1st, 2020, of 83 target companies ranked at the top in their capital market sectors (i.e., Basic Materials, Consumer Goods, Healthcare, Services, Utilities, Conglomerates, Financial, Industrial Goods, Technology).

The 10 points earned in the slider task will be used to pay for the additional information.

How much are you willing to pay, up to a maximum of 10.0, for information from the algorithm?

Note: Submissions can range from 0 to 10, up to one decimal place.

Next

[Screen 4] (shown in expert without description treatment)

**Question 1: Explanation of Information Source and WTP**

The additional information provided here is the stock price 30 days later, submitted by [human experts](#) based on the stock price graph used in this experiment.

The 10 points earned in the slider task will be used to pay for the additional information.

How much are you willing to pay, up to a maximum of 10.0, for information from the algorithm?

Note: Submissions can range from 0 to 10, up to one decimal place.

Next

[Screen 4] (shown in expert with description treatment)

**Question 1: Explanation of Information Source and WTP**

The additional information provided here is the average of stock price predictions submitted by 198 [experts \(CMAs\)](#) who participated in a similar experiment in the past, based on the stock price chart used in this experiment, for the stock prices 30 days ahead.

CMA stands for Certified Member Analyst of the Securities Analysts Association of Japan. The CMA is a qualification granted to those who have taken the prescribed training courses and passed the examinations based on these courses and fulfilled certain requirements; it is a sign of a expertise in the fields of finance and investments. The CMAs' investment valuations are based on the calculation of the corporate value of an investment and the forecasting of its future value. The particularly crucial knowledge required can be broadly categorized into three main areas: evaluation of investment data, decision making on investment policies, and construction and management of portfolios.

The 10 points earned in the slider task will be used to pay for the additional information.

How much are you willing to pay, up to a maximum of 10.0, for information from the algorithm?

Note: Submissions can range from 0 to 10, up to one decimal place.

Next

[Screen 5] (shown in the case  $WTP \geq p$ )

**Question 1: The price of additional information**

In Question 1, the Price of additional information is 1.1.  
Your willingness to pay is 4.0.

Because  $WTP \geq p$ , you can obtain additional information.

Next

[Screen 5] (shown in the case  $WTP < p$ )

**Question 1: The price of additional information**

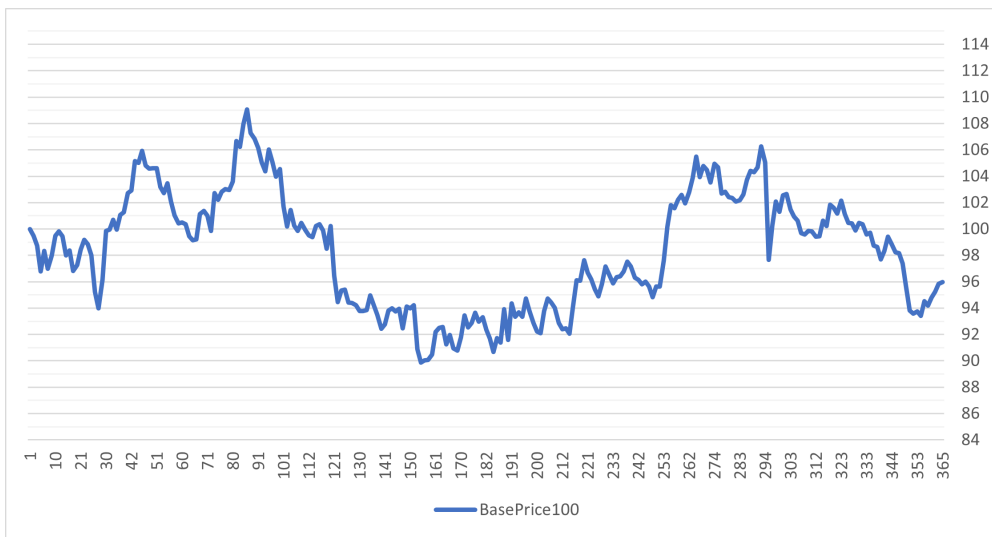
In Question 1, the Price of additional information is 8.0.  
Your willingness to pay is 4.0.

Because  $WTP < p$ , you cannot obtain additional information.

Next

[Screen 6] **Question 1: Final Forecast**

Please read the below graph displaying the closing prices of a particular stock over a 12-month period and predict the stock price 30 days after the last day shown on the graph. The closing price on the last day is 95.962.



Additional information: The prediction value from the algorithm is 97.026. (shown in algorithm treatment and in the case  $WTP \geq p$ )

Additional information: The prediction value from the experts is 96.434. (shown in expert treatment and in the case  $WTP \geq p$ )

Your prediction in first half is 100.0.

Your prediction:

Note: Please submit your prediction up to three decimal places.

Next

[Screen 7]

### Question 1: Confidence

Please evaluate how confident you are in your final forecast for the stock price graph in this question.

- Not confident at all
- Not very confident
- Neither confident nor unconfident
- Somewhat confident
- Very confident

Next

[Screen 8] The following is an example of the feedback given at the end of each question.

### Question 1: Results

Your willingness to pay (WTP): 10.0

Price of the additional information: 4.2

Because  $WTP \geq p$ , you can obtain the additional information.

Your initial prediction: 100.0

Your final prediction: 111.0

Prediction from the additional information: 96.434

The realized stock price: 96.288

The degree of deviation from the true value =  $\frac{|\text{Your forecast} - \text{Realized stock price}|}{\text{Realized stock price}}$

Your initial forecast's degree of deviation: 0.039

Your final forecast's degree of deviation: 0.153

The additional information's degree of deviation: 0.002

Reward points =  $20 - (\text{Degree of deviation} \times 100)$

※ If the reward points are less than 0, they are set to 0.

Your initial forecast's reward points: 16.1 points

Your final forecast's reward points: 4.7 points

The additional information's reward points: 19.8 points

Your initial forecast's final reward (the earned points in the slider task + reward points): 26.1 points

Your final forecast's final reward (the earned points in the slider task + reward points – the price [if you received the additional information]): 10.5 points

If you rely fully on the additional information (additional information's reward points – initial forecast's reward points): 3.7 points

The change in reward points (final forecast's reward points – initial forecast's reward points – the price [if you received the additional information]): -15.6 points

Next

[Screen 9] The following is an example of the final reward feedback given at the end of the experiment.

### The end of experiment and reward announcement

Thank you for participating in the experiment today.

We will inform you about the rewards.

Final reward based on the initial forecast = 10 points earned in Task 1 + points earned based on prediction accuracy

Final reward based on the final prediction = 10 points earned in Task 1 + points earned based on prediction accuracy – price of the additional information, if obtained

Round	Initial forecast's reward points	Final forecast's reward points	WTP	Price	Purchased?	Initial forecast's final reward	Final forecast's final reward
1	19.8	19	0.5	8.9	No	29.8	29
2	0.4	3.7	3.5	9	No	10.4	13.7
3	18.4	15.3	5	2.3	Yes	28.4	23
4	7.5	12.4	2	0.7	Yes	17.5	21.7
5	10.6	10.6	1.5	7.1	No	20.6	20.6
6	7.9	7.9	3	8.8	No	17.9	17.9
7	9	1.7	3	5.7	No	19	11.7
8	19.5	16.1	6	6.3	No	29.5	26.1
9	17.1	15.7	3.5	7.6	No	27.1	25.7
10	14	16.6	8	1.5	Yes	24	25.1
11	14.4	14	4.2	9.8	No	24.4	24
12	12.3	14.2	3.5	6.9	No	22.3	24.2
13	15.8	13.3	5.5	7.2	No	25.8	23.3
14	12.5	11.6	5.6	7.6	No	22.5	21.6
15	6.1	13.8	8.6	5.6	Yes	16.1	18.2
16	14	13.1	7.8	2	Yes	24	21.1
17	14.4	16.6	7.8	1.2	Yes	24.4	25.4
18	16.7	7.3	7.6	6	Yes	26.7	11.3
19	15.1	14	5.3	3.4	Yes	25.1	20.6
20	15.7	18	7.3	2.3	Yes	25.7	25.7

The prediction selected for payment was the initial forecast from Round 7.

Your payment is calculated as: 19 points × 60 JPY per point + 1,000 yen

Therefore, your total payment is 2140 JPY.

Thank you for your participation. Please click the “Next” button to proceed to the questionnaire.

Next

## P Post-experiment survey in Experiment 2

### Questionnaire Question 1 (shown in treatments with description)

#### **Explanation of the algorithm:**

The additional information provided here is the stock price 30 days after the final day shown in the graph, submitted by **an algorithm** crafted to predict future stock prices, based on the same stock price graph used in this experiment.

This algorithm makes a future stock price forecast by learning the historical stock price information, from January 1st, 2000, or from the initial public offering (IPO) day to January 1st, 2020, of 83 target companies ranked at the top in their capital market sectors (i.e., Basic Materials, Consumer Goods, Healthcare, Services, Utilities, Conglomerates, Financial, Industrial Goods, Technology).

#### **Explanation of human experts:**

The additional information provided here is the average of stock price predictions submitted by 198 **experts (CMAs)** who participated in a similar experiment in the past, based on the stock price chart used in this experiment, for the stock prices 30 days ahead.

CMA stands for Certified Member Analyst of the Securities Analysts Association of Japan. The CMA is a qualification granted to those who have taken the prescribed training courses and passed the examinations based on these courses and fulfilled certain requirements; it is a sign of a expertise in the fields of finance and investments. The CMAs' investment valuations are based on the calculation of the corporate value of an investment and the forecasting of its future value. The particularly crucial knowledge required can be broadly categorized into three main areas: evaluation of investment data, decision making on investment policies, and construction and management of portfolios.

In the stock forecast task you just completed, if you had to choose either the additional information from the algorithm or the additional information from the human experts, which would you prefer?

- Strongly prefer the additional information from the human experts
- Somewhat prefer the additional information from the human experts
- Neither
- Somewhat prefer the additional information from the algorithm
- Strongly prefer the additional information from the algorithm

### Questionnaire Question 1 (shown in treatments without description)

In the stock forecast task you just completed, if you had to choose either the additional information from the algorithm or the additional information from the human experts, which would you prefer?

- Strongly prefer the additional information from the human experts
- Somewhat prefer the additional information from the human experts
- Neither
- Somewhat prefer the additional information from the algorithm
- Strongly prefer the additional information from the algorithm

### Questionnaire Question 2 (shown in algorithm treatments with and without description)

**Explanation of the algorithm, that is, information about how the additional information was generated:**

The additional information provided here is the stock price 30 days after the final day shown in the graph, submitted by **an algorithm** crafted to predict future stock prices, based on the same stock price graph used in this experiment.

This algorithm makes a future stock price forecast by learning the historical stock price information, from January 1st, 2000, or from the initial public offering (IPO) day to January 1st, 2020, of 83 target companies ranked at the top in their capital market sectors (i.e., Basic Materials, Consumer Goods, Healthcare, Services, Utilities, Conglomerates, Financial, Industrial Goods, Technology).

**No explanation of the algorithm:**

The additional information provided here is the stock price 30 days after the final day shown in the graph, submitted by **an algorithm** crafted to predict future stock prices, based on the same stock price graph used in this experiment.

After being shown the above explanation of the algorithm, that is, information about how the additional information was generated, how did your willingness to pay for the algorithm change?

- My willingness to pay increased substantially.
- My willingness to pay increased somewhat.
- It did not change.

- My willingness to pay decreased somewhat.
- My willingness to pay decreased substantially.

### **Questionnaire Question 2 (shown in expert treatments with and without description)**

#### **Explanation of the human experts, that is, information about how the additional information was generated:**

The additional information provided here is the average of stock price predictions submitted by 198 **experts (CMAs)** who participated in a similar experiment in the past, based on the stock price chart used in this experiment, for the stock prices 30 days ahead.

CMA stands for Certified Member Analyst of the Securities Analysts Association of Japan. The CMA is a qualification granted to those who have taken the prescribed training courses and passed the examinations based on these courses and fulfilled certain requirements; it is a sign of an expertise in the fields of finance and investments. The CMAs' investment valuations are based on the calculation of the corporate value of an investment and the forecasting of its future value. The particularly crucial knowledge required can be broadly categorized into three main areas: evaluation of investment data, decision making on investment policies, and construction and management of portfolios.

#### **No explanation of the human experts:**

The additional information provided here is the stock price 30 days later, submitted by **human experts** based on the stock price graph used in this experiment.

After being shown the above explanation of the human experts, that is, information about how the additional information was generated, how did your willingness to pay for the human experts change?

- My willingness to pay increased substantially.
- My willingness to pay increased somewhat.
- It did not change.
- My willingness to pay decreased somewhat.
- My willingness to pay decreased substantially.

### **Questionnaire Question 3**

Over the past 20 rounds, what mainly determined your willingness to pay (WTP) for the additional information?

For each item below, please rank its importance from 1 = most important to 4 = least important. Please use each number only once.

I adjusted my WTP based on the feedback.

I increased my WTP when I judged that the prediction was difficult.

I increased my WTP when my confidence in my initial prediction was low.

I increased my WTP when I judged that the additional information would be more useful than my own prediction.

#### Questionnaire Question 4

Over the past 20 rounds, which types of feedback influenced your willingness to pay for the additional information?

You may select multiple answers.

- Actual stock price
- Your initial forecast's reward points
- Your final forecast's reward points
- Additional information's reward points
- Price of the additional information
- Your initial forecast's final reward = points earned in the slider task + reward points of your initial forecast
- Your final forecast's final reward = points earned in the slider task + reward points of your final forecast price of the additional information, if obtained
- Change in points if you rely fully on the additional information = Additional information's reward points Your initial forecast's reward points
- The change in reward points = Your final forecast's reward points Your initial forecast's reward points price of the additional information, if obtained
- Other:

## Q Quiz in Experiment 2

1. What can you do in this experiment?
  - 1 View the stock price graph of a company you specify
  - 2 Submit your predictions
  - 3 Decide the part used for calculating additional rewards
  - 4 Choose the source of additional information you want to obtain

[Correct answer: 2]

2. In which of the following cases do you obtain additional information?

Case	Willingness to pay (WTP)	Price of additional information (p)
1	10	5
2	4	6
3	8	8
4	3	9

- 1 Case 1 and Case 3
- 2 Case 2 and Case 4
- 3 Case 1 and Case 4
- 4 Case 2 and Case 3

[Correct answer: 1]

3. Which of the following is true about the willingness to pay (WTP)?
  - 1 It is possible to submit more than once for each question.
  - 2 It is possible to submit 12.
  - 3 It is not possible to submit 5.5.
  - 4 It is the maximum number of points you are willing to pay to acquire additional information.

[Correct answer: 4]

4. Which of the following is true about the price of additional information (p)?
  - 1  $p = 12$  is possible.
  - 2  $p = 0$  is possible.
  - 3 For each question,  $p$  is randomly determined.
  - 4 You can set it yourself.

[Correct answer: 3]

5. Please select the correct option regarding the purchase of additional information.

- 1 If  $WTP \geq p$ , you decide whether to buy it yourself.
- 2 If  $WTP < p$ , you cannot purchase, but you will still pay the points.
- 3 If  $WTP = p$ , you cannot purchase.
- 4 If  $WTP \geq p$ , you automatically pay  $p$  points.

[Correct answer: 4]

## R Clustering the stock price into blocks

To examine whether characteristics of the stock-price series affected WTP, we organized the 20 time series into blocks based on statistical and technical features following Bao et al. (2023). Using k-means clustering, we grouped the time series into four clusters and constructed 4 blocks of 5 rounds containing a balanced mix of stock-price patterns across blocks.

Table R1: Stock Data

Graph	Stock Ticker	Period Date (Start from June)	Closing Price at Last Date (Normalized)	Closing Price One Month Later (Normalized)
1	STX	2010-2011	116.183	107.676
2	LKQ	2011-2012	135.314	125.498
3	UPS	2009-2010	116.381	109.475
4	DLTR	2009-2010	136.241	138.273
5	SBUX	2015-2016	105.113	108.656
6	FOXA	2017-2018	142.146	183.223
7	TSN	2012-2013	130.098	132.576
8	COP	2009-2010	105.073	102.479
9	F	2013-2014	104.847	110.204
10	FISV	2016-2017	118.940	115.456
11	EL	2017-2018	158.742	151.572
12	ADM	2014-2015	117.637	108.027
13	NKE	2016-2017	95.962	96.288
14	TRIP	2015-2016	88.537	83.505
15	WM	2013-2014	106.869	106.940
16	MLM	2016-2017	118.546	117.107
17	HES	2013-2014	91.097	97.096
18	SIVB	2016-2017	130.597	136.070
19	SPGI	2008-2009	78.589	73.714
20	FMC	2014-2015	73.448	67.383

Table R2: Stock Price Graph Characteristics

Round	Block	Stock price graph	K-mean cluster	Autocorrelation	Amplitude	Variance	MACD reversals	Reversals
1	1	13	Group 1	0.6879	19.20	18.98	14	128
2	1	6	Group 2	0.8939	51.81	292.95	19	131
3	1	16	Group 3	0.7714	38.52	107.32	11	134
4	1	8	Group 4	0.7995	42.30	90.53	17	127
5	1	2	Group 4	0.7855	58.25	171.66	14	136
6	2	17	Group 1	0.5468	14.86	11.44	16	116
7	2	19	Group 2	0.8356	69.75	441.27	13	121
8	2	12	Group 3	0.4948	23.92	33.52	13	134
9	2	4	Group 4	0.8219	50.46	167.18	20	129
10	2	10	Group 4	0.8085	27.09	42.08	18	137
11	3	15	Group 1	0.6222	16.84	13.31	16	122
12	3	11	Group 2	0.8861	62.94	401.36	20	119
13	3	5	Group 3	0.6128	25.22	29.53	14	138
14	3	14	Group 4	0.7270	51.34	153.52	17	130
15	3	3	Group 4	0.8059	45.40	98.11	17	124
16	4	9	Group 1	0.6238	20.47	25.32	19	123
17	4	7	Group 2	0.9045	58.44	328.72	22	132
18	4	1	Group 3	0.7130	55.88	163.12	11	135
19	4	20	Group 4	0.7344	34.95	58.82	15	128
20	4	18	Group 4	0.8561	41.79	108.79	16	136

## References

- BAILEY, P. E., T. LEON, N. C. EBNER, A. A. MOUSTAFA, AND G. WEIDEMANN (2023): “A meta-analysis of the weight of advice in decision-making,” *Current Psychology*, 42, 24516–24541, <https://doi.org/10.1007/s12144-022-03573-2>.
- BAO, T., B. CORGNET, N. HANAKI, Y. E. RIYANTO, AND J. ZHU (2023): “Predicting the unpredictable: New experimental evidence on forecasting random walks,” *Journal of Economic Dynamics and Control*, 146, 104571, <https://doi.org/10.1016/j.jedc.2022.104571>.
- BONACCIO, S., AND R. S. DALAL (2006): “Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences,” *Organizational Behavior and Human Decision Processes*, 101, 127–151, <https://doi.org/10.1016/j.obhdp.2006.07.001>.
- HARVEY, N., AND I. FISCHER (1997): “Taking advice: Accepting help, improving judgment, and sharing responsibility,” *Organizational behavior and human decision processes*, 70, 117–133, <https://doi.org/10.1006/obhd.1997.2697>.
- SNIEZEK, J. A., G. E. SCHRAH, AND R. S. DALAL (2004): “Improving judgement with prepaid expert advice,” *Journal of Behavioral Decision Making*, 17, 173–190, <https://doi.org/10.1002/bdm.468>.