How do retail investors use order flow data?

Abstract

We study how retail traders value and learn from order flow data using a randomized online experiment. Sophisticated participants — those with formal financial education — value order flow data in line with a fully optimizing Bayesian investor, while others pay only 59% of the fair value. Overconfidence leads to a 28% higher data valuation. Participants gain limited benefits from order flow data, capturing only 11.8% of a Bayesian trader's value and reducing trading errors by just 6.25% (11.7% for sophisticated traders). Data access mitigates the disposition effect but also fuels excessive trading. Cognitive load rises slightly with data access, particularly for unsophisticated traders.

Keywords: experimental finance, market data, retail trading, order imbalance

JEL Codes: G11, G12, G14

1 Introduction

> "Active retail investors in the U.S. equities market are demanding access to more granular information. Bring your investing decisions to the next level with depth-of-book data."
>
> – Nasdaq TotalView (webpage retrieved on February 12, 2025)

Retail investors increasingly pay for order flow data, yet little is known about how they value and use it. Online brokerage firms, exchanges, and retail-oriented financial platforms monetize market data through subscription-based analytics, premium order flow insights, and tiered access to real-time price feeds. For instance, Robinhood's Gold subscription, priced at $5 per month, provides users with Level II market data from Nasdaq, offering deeper insights into bid and ask prices. Interactive Brokers charges venue-dependent fees for non-professional traders, reaching up to $135 per month for ICE Futures data. At the same time, TD Direct Investing charges up to $32 per month for Level II data on U.S. equity markets, while in Canada, Questrade offers similar access for $89.95 per month.

Retail trading has expanded dramatically over the past decade, reshaping market dynamics. The share of retail volume in U.S. equity markets surged from 10.1% in 2010 to over 23% by January 2023. Between January 2021 and August 2024, Robinhood's assets under custody more than doubled, rising from $63 billion to $140 billion, underscoring the growing influence of retail investors in financial markets. Retail traders have also grown increasingly sophisticated, actively seeking out complex and speculative instruments: Bryzgalova et al. (2023) document that as of 2021, retail investors accounted for up to 60% of trading volume in U.S. options.

Can retail traders effectively learn from market data? Bartlett et al. (2023) and Kwan et al. (2024) leverage machine learning techniques to show that order flow data is highly valuable for algorithmic hedge funds. Professional traders invest heavily in infrastructure to extract short-term signals from order flow and trade on them within milliseconds, which gives them a measurable edge. Retail traders, by contrast, are not natural users of such data. They may lack the speed, automation, and analytical tools required to fully exploit order flow information. Why, then, do they pay for it? One possibility is that retail traders see access to order flow data as a mark of professionalism, aligning themselves with institutional practices even if they cannot replicate the same strategies. Another is that order flow data gives the illusion of control, making traders feel more informed, even if it does not materially improve their returns.

In this paper, we design a randomized controlled trading experiment to assess how retail traders value and use order flow data. We examine their willingness to pay for market data of varying quality and its impact on trading mistakes, the disposition effect, information processing, and cognitive load. The experimental design eliminates selection bias by observing the same participants trading both with and without order flow data. It also allows us to estimate the theoretical value of data by modeling the optimal strategy of a Bayesian investor while varying both data quality and quantity.

We develop an experimental platform in oTree (Chen et al., 2016), building on the classical investment games of Frydman et al. (2014) and Weber and Camerer (1998). The experiment consists of two five-minute trading rounds, preceded by a brief training session. Participants trade a risky asset in real time but cannot short sell or hold more than one unit at a time. They can, however, borrow cash at a zero interest rate. The asset price follows a Markov chain with two highly persistent states and exhibits predictable momentum. This momentum structure allows us to determine the optimal strategy of a Bayesian investor and reinforces the disposition effect in the data — as the optimal strategy involves selling losing stocks more frequently while holding winning stocks longer, creating the reverse of the typical disposition effect.

Participants trade on either a *data-augmented* or *data-free* platform in each round. In data-augmented rounds, they observe order flow, specifically buy and sell marketable orders from algorithmic trading bots. Following Glosten and Milgrom (1985), these bots can be either insiders (informed traders), who know the state of the stock and trade optimally based on this information, or noise traders who submit buy or sell orders at random. Therefore, participants can use the order imbalance between buy and sell orders to update their beliefs on the current state and adjust their trading decisions accordingly. We vary data quality (the share of informed traders) and data quantity (the arrival rate of trading bots) across participants to examine how these factors influence trading behavior.

We recruit 307 participants from Prolific, an online subject pool for experiments, to take part in the trading game on January 29, 2025.[1] Our sample is representative of the U.S. population, stratified by age, gender, and ethnicity using census data. To assess financial sophistication, we record whether participants completed a formal course in finance and administer a 12-question financial literacy quiz based on Fernandes et al. (2014). Further, we elicit participants' self-assessed financial knowledge. Before and after the trading game, we measure willingness to pay (WTP) for order flow data, using certainty calibration, cheap talk, and consequentiality scripts to mitigate hypothetical bias. Finally, after each round, participants complete a short quiz on cognitive effort to assess the mental burden of processing order flow data.

We find that unsophisticated retail investors value order flow data at 41% below the theoretical benchmark of a fully rational Bayesian trader. In contrast, participants with formal financial education price the data more accurately and align their willingness to pay with the Bayesian valuation. A one standard deviation increase in data quality, measured by the share of informed traders, raises willingness to pay by 13 percent. However, the demand for data is inelastic, as participants' willingness to pay increases with quality at a slower rate than the Bayesian benchmark. We also highlight the influence of both cognitive biases and marketing cues as factors shaping investor demand for data. Overconfidence inflates willingness to pay by 28%, while framing data as "premium" increases valuations by 18%.

Participants learn from order flow data, but only to a limited extent. With data access, their decisions better align with Bayesian-optimal strategies: for unsophisticated traders, trading mistakes decline by 6.25%, from 48 percent to 45 percent. Sophisticated traders use data more effectively, reducing mistakes by 11.7%. However, these gains remain modest, as participants capture only 11.8 percent of the theoretical value of the data in terms of additional payoffs. We also find that holding the informational content of data constant, an increase in data volume leads to over-trading (consistent with the cognitive load theory of Achtziger et al., 2020) and, on average, slightly lower payoffs.

[1] See Prolific's website at https://prolific.co/.

We further examine the disposition effect and show that access to order flow data reduces the tendency to hold onto losing positions. In particular, the proportion of realized losses increases from 11% to 17% when traders have access to order flow data, while we find no effect on the propensity of realizing gains. However, the effect is not significantly influenced by data quality, and is significantly weaker when data is framed as a premium product.

Finally, we examine the cognitive load associated with processing order flow data. Participants report a slight increase in cognitive strain when data is available, particularly in frustration, temporal demand (feelings of being rushed), and effort. However, sophisticated participants with formal financial education exert less effort when data is available, suggesting they are better equipped to navigate information-intensive trading environments.

Overall, the findings suggest that while order flow data provides some benefits to retail traders, its value may be overstated since the impact on trading outcomes is limited. Even though unsophisticated traders partly recognize their limitations and underpay relative to the Bayesian benchmark, they still overestimate the value of data given its marginal effect on performance. Meanwhile, overconfident traders and those influenced by premium framing are more likely to pay inflated prices. These results suggest that retail traders' demand for market data is shaped as much by perception, biases, and marketing as by actual trading benefits.

2 Related literature

Our paper contributes to the growing literature on retail investor behavior in modern, technology-driven financial markets. Closest to our work, Chapkovski et al. (2024a,b) find in an experimental setting that gamification of trading platforms increases trading volume, enhances risk-taking, and may reinforce trading mistakes. Using data from 17 major U.S. brokers, Yelagin (2024) shows that gamification increases volatility in retail trader returns but makes retail order flow less toxic, benefiting the broader market.

Other studies highlight how digital engagement tools influence retail trading behavior. Arnold et al. (2022) find that push notifications about large price swings lead to greater risk-taking, as measured by leverage. Similarly, Moss (2022) documents that Robinhood alerts increase retail trading intensity by 25% for 15 minutes after the notification. Kalda et al. (2021) use transaction-level data from two German banks to show that investors take on riskier trades when using smartphones compared to more traditional trading platforms. Our study complements this work by focusing on how retail traders value and use market data. We employ a randomized experimental approach to eliminate selection bias, allowing us to separately measure both willingness to pay for order flow data and the extent to which traders actually learn from it.

Our paper contributes to the resurgent literature on retail trading. Classic studies by Barber and Odean (2000); Barber et al. (2008) document that retail traders exhibit overconfidence, engage in excessive turnover, and favor small, high-beta stocks that capture their attention. Barber et al. (2022) show that Robinhood's "Top Movers" tab directs attention toward high-volatility stocks, contributing to portfolio underperformance.

At the same time, growing evidence suggests that retail order flow may be more sophisticated, implying better use of market data. Kaniel et al. (2008); Kelley and Tetlock (2013) find that retail trades predict future stock returns, with aggressive trades anticipating news while passive orders act as contrarian liquidity providers. Welch (2022) further finds that, in aggregate, retail investors using Robinhood performed well between 2018 and 2020. Dyhrberg et al. (2022) argues that payment for order flow boosts retail trading activity by improving execution prices in small stocks, which are likely to be favored by individual investors. Gao et al. (2020) and Farrell et al. (2022) find that technology-enabled improvements in the information set of retail investors lead to better portfolio performance.

Retail traders now dominate certain market segments, increasingly leaning towards complex, leveraged products. Bryzgalova et al. (2023) and Bogousslavsky and Muravyev (2024) show that they account for 60 percent of U.S. option market volume and are becoming increasingly sophisticated in their use of performance monitoring tools. We contribute to this literature by explicitly examining how investors with varying levels of sophistication, as measured by financial education, value and learn from order flow data.

Finally, we contribute to the literature on the value of financial data. We value data using a revenue approach as in Veldkamp (2023), under the assumption that investors are risk neutral and maximize their expected payoff. Farboodi et al. (2024) develop an indirect framework to value earnings forecast data from I/B/E/S using the statistical distribution of equilibrium asset returns. In our experimental setting, we directly control the quality and flow of information available to investors and find consistent patterns: demand for data is relatively inelastic with respect to quality, and investors exhibit heterogeneous valuations for the same data.

Several studies use machine learning techniques to focus on the value of order flow data. Bartlett et al. (2023) find that sophisticated traders with access to proprietary feeds of odd-lot data have a significant edge over investors without such data. Kwan et al. (2024) document that orders submitted by high-frequency traders contain twice as much information as those from retail-driven brokerages, largely due to their use of order flow information, including the state of the order book and order submission history. In our experiment, we examine whether retail traders can extract meaningful insights from order submission history and incorporate it into their trading decisions.

In our experiment, we rely on trade history data rather than modeling the full order book. Although trade history is not a perfect proxy for Level 2 information, it is among the strongest predictors of short-term returns (Kwan et al., 2024). Full order book models tend to be non-tractable and yield less definitive predictions, while also increasing the cognitive burden on participants. By using the simpler trade history data, we obtain a conservative estimate of overvaluation, as traders should be better able to extract signals compared to more complex market data.

3 Institutional background

The availability and pricing of market data have become central for both institutional and retail investors, shaping how they engage with financial markets. Historically, access to high-quality, real-time market data was the domain of institutional traders and hedge funds. However, with the rise of online brokerage platforms and commission-free trading, retail investors have become more active participants in financial markets, increasing demand for (and supply of) affordable, real-time market data.

3.1 Market data structure, availability, and pricing

Market data is typically categorized into different levels:

- Level 1 data: Includes basic information such as bid-ask prices, last trade prices and volume. This data is generally available to retail investors for free or at minimal cost.
- Level 2 data: Provides more detailed information, including market depth with bid and ask orders beyond the best available price.
- Level 3 data: Used primarily by professional traders and institutions, offers full book transparency and order entry functionality.

Table 1 highlights significant heterogeneity in how market data is priced and made available to retail investors across brokerage platforms and standalone data providers. A key takeaway is that while brokers typically offer Level 1 data for free, the cost and accessibility of Level 2 data vary considerably. Brokers like thinkorswim (by Charles Schwab) and Fidelity offer Level 2 data at no additional cost to qualifying traders, whereas others, such as Interactive Brokers and Webull, charge tiered fees that depend on market selection. Robinhood, in contrast, monetizes Level 2 data directly by bundling it into its paid "Robinhood Gold" subscription, pricing market depth access as part of a broader premium service. This suggests that the availability of detailed market depth information is increasingly being used as a differentiating feature among brokers, with some bundling it into premium services while others tie access to trading activity thresholds.

Standalone market data providers such as Nasdaq TotalView, Databento, and TradingView follow a different pricing model. Rather than bundling data with brokerage services, these providers charge fixed fees for access. Nasdaq TotalView is one of the more affordable options at $15 per month, whereas Databento and Alpaca cater to more advanced traders with pricing that can exceed $3,500 per month for comprehensive datasets. These services offer greater flexibility, allowing traders to integrate data into custom trading systems.

Table 1: **Market data pricing and availability for retail investors**

This table compares market data pricing and availability for non-professional investors across major brokerage platforms and market data providers. The selected U.S. brokers are based on Motley Fool's Best Online Brokers and Trading Platforms list (updated February 2025) and the list of online brokers available in the United States on the BrokerChooser website. Additionally, we include Questrade and TD Direct Investing, ranked as Canada's best online brokerages in 2024 by MoneySense. In addition to brokerage platforms, we include standalone market data providers such as Nasdaq TotalView, Databento, TradingView, Quotestream, and Alpaca, which offer direct access to real-time market data. Market data is categorized by Level 1 and Level 2 data availability, subscription pricing, and free data offerings. The final column describes market data waivers, which refer to the reduction or elimination of fees for traders who meet specific activity thresholds, such as minimum trading volume or commission levels. These waivers do not apply to standalone data providers, as they do not facilitate trading.

Platform	Data Types	Price Ranges/month	Free Data Availability	Waivers for Active Traders?
Interactive Brokers	Level 1, 2	1-147 USD (varies by market)[a]	Level 1 (CBOE One & IEX, non-consolidated)	✓
Robinhood	Level 1, 2	5 USD[b]	Level 1	✗
thinkorswim (by Charles Schwab)	Level 1, 2	0[c]	Level 1, 2 (Nasdaq TotalView)	✗
Fidelity	Level 1, 2	0[d]	Level 1, 2	✓
Webull	Level 1, 2	2.99-60.99 USD[e]	Level 1 (NASDAQ Basic), 1-month free Level 2 (Nasdaq TotalView)	✗
Questrade	Level 1, 2	19.95-89.95 CAD[f]	Free snap quotes; Level 1 for TSX, TSXV; CBOE One Summary (NASDAQ, NYSE, AMEX, ARCA, BATS)	✓
TD Direct Investing	Level 1, 2	10-32 USD[g]	Level 1	✓
Nasdaq TotalView	Level 1, 2	15 USD[h]	None	n/a
Databento	Level 1, 2, 3	199-3,500 USD[i]	None	n/a
TradingView	Level 1, 2	12.95-49.95 USD[j]	Free real-time Level 1 (CBOE BZX, limited U.S. equities)	n/a
Quotestream	Level 1, 2	17.94-49.95 USD[k]	None	n/a
Alpaca	Level 1, 2	99 USD[l]	Limited to 30 symbols on IEX	n/a

[a] Source: https://www.interactivebrokers.com/en/pricing/market-data-pricing.php
[b] Robinhood Gold gives access to Level 2 market data from Nasdaq TotalView. Source: https://robinhood.com/us/en/support/robinhood-gold/
[c] Source: https://www.schwab.com/trading/thinkorswim/compare-platforms
[d] Level 2 quotes are only available to individuals who trade more than 72 times a year. Source: https://www.fidelity.com/products/atp/pdf/userguide.pdf
[e] Source: https://www.webull.com/help/faq/126-What-market-data-is-available-on-Webull
[f] Source: https://www.questrade.com/pricing/self-directed-commissions-plans-fees/market-data
[g] Source: https://www.td.com/ca/en/investing/direct-investing/pricing
[h] Source: https://www.nasdaq.com/solutions/level-up-your-retail-trading-with-depth-data
[i] Source: https://databento.com/pricing
[j] Source: https://www.tradingview.com/pricing/
[k] Source: https://quotestream.com/compare
[l] Source: https://alpaca.markets/data

3.2 Market data ecosystem

Retail traders access market data through brokerage platforms and standalone data providers, but the ultimate source of this data is exchanges. Exchanges generate market data and sell it through two main channels:

1. Securities Information Processors (SIPs) – These are consolidated feeds mandated by regulators to ensure public price transparency (e.g., CTA, UTP in the U.S.). They provide Level 1 data but often lack depth-of-book information.

2. Proprietary Data Feeds – Exchanges sell Level 2 and higher-depth data directly to brokers and third-party vendors, typically at a premium. This is where cost disparities arise.

For retail traders, direct exchange data is rarely their primary access point. Instead, brokers act as intermediaries, deciding how to handle these data costs—some absorb them, others charge customers directly, and some offer waivers based on trading activity.

In parallel, third-party data providers such as Databento, TradingView, and Quotestream operate outside the brokerage model, offering market data subscriptions tailored for historical analysis, cross-asset trading, and algorithmic strategies. These platforms serve traders who require more customizable access to raw data, rather than broker-integrated feeds.

The segmentation of market data access across brokers, exchanges, and standalone providers shapes how retail traders incorporate information into their decision-making. Low-cost brokers emphasize accessibility, exchanges monetize depth-of-book data, and third-party providers cater to traders seeking advanced analytics. The result is a fragmented market data landscape, where access varies significantly based on cost, platform, and trading sophistication.

3.3 Regulatory developments and policy discussions

The pricing and distribution of market data have become central to regulatory discussions as policymakers evaluate whether existing market structures create distortions in access, particularly for retail traders. A key concern is that exchanges serve as both trading venues and primary data providers, enabling them to extract rents from market participants through proprietary data feeds. Regulators have debated whether these practices contribute to information asymmetries, as institutional traders are more likely to afford high-cost data products, while retail investors often rely on delayed or basic data.

In response to these concerns, the Securities and Exchange Commission (SEC) has introduced reforms aimed at increasing competition in market data provision. The Market Data Infrastructure Rule (MDI) seeks to modernize the dissemination of consolidated data and reduce dependence on exclusive exchange-operated feeds. A central component of this reform is the introduction of competing consolidators, which would offer alternative distribution channels for market data, potentially mitigating the pricing power of exchanges and increasing accessibility for retail brokers.[2]

Broader industry discussions have also emphasized the economic impact of market data pricing. Reports from Cboe Global Markets and the Cato Institute highlight how rising data costs have contributed to fragmentation in data access, as brokers must balance affordability with providing high-quality, real-time data. Industry advocates have argued that current pricing models disproportionately benefit exchanges, while regulatory oversight has thus far had limited impact on constraining fee growth.[3]

Despite these regulatory initiatives, the effectiveness of policy interventions remains uncertain. Exchanges retain strong financial incentives to maintain control over their proprietary feeds, and past attempts at price regulation have had limited success in reducing costs. As a result, retail investors continue to navigate a fragmented and costly data environment, where access to high-quality market information is largely contingent on a trader's willingness to pay.

3.4 Retail investor trends and market data demand

Retail trading has grown significantly in recent years, driven by commission-free platforms, improved accessibility, and increased retail participation in financial markets. Despite this growth, retail investors do not access market data in the same way as institutional traders.

Survey data suggests that 59.2 percent of retail traders obtain investment ideas from broker apps,

[2]The MDI rules were adopted in December 2020 (`https://www.sec.gov/rules-regulations/2020/12/market-data-infrastructure`), but implementation of the rules has been delayed.

[3]See Cato Institute, Are Market Data Fees Too High?, Spring 2019; Cboe, The Necessity of Real-Time Options Data for Retail Participants, March 2023; Market Structure Partners Report, There's No Market in Market Data, January 2025.

compared to only 16.1 percent who rely on professional advice.[4] The Wall Street Journal reported in January 2024 on the rise of retail quantitative trading, driven by improvements in home computing technology and greater access to market data.[5] As of February 2025, the Reddit community `/r/algotrading` had more than 1.8 million members, reflecting growing interest in data-driven trading among retail investors. By comparison, the more widely known `/r/wallstreetbets` retail trading forum, had 18 million members.

The Ontario Securities Commission (OSC) reports that as of 2021, more than 13% of retail investors requested expanded market data offerings from their brokers. Demand was higher among those with higher self-perceived financial knowledge (18 percent) compared to those with lower perceived knowledge (9 percent), suggesting that interest in enhanced market data is linked to confidence in one's investing ability.[6] In line with this estimate for market data demand from individual investors, Havakhor et al. (2024) document that the shutdown of Yahoo! Finance API in 2017 led to a 10% drop in retail trading volume.

However, the Canadian Securities Administrators and exchange operators such as Cboe note that institutional traders continue to have a major advantage in market data access. Retail investors are often priced out of high-quality order flow data, raising questions about fairness in market structure.[7]

As retail trading activity continues to grow, brokers have responded with promotional trials and fee waivers to incentivize adoption of premium data. For example, Webull and moomoo offer complimentary access to Nasdaq TotalView for a limited period, while platforms such as Schwab and Merrill Edge waive market data fees for high-volume traders or those meeting asset thresholds. These trends suggest a growing recognition of the role market data plays in retail trading.

[4]See Public: The Retail Investor Report 2023.
[5]See WSJ: Meet the Investors Trying Quantitative Trading at Home, January 20, 2024.
[6]Source: Ontario Securities Commission, Self-Directed Investors: Insights and Experiences, April 2021.
[7]See, for example, Investment Executive: Industry calls for market data regulation, March 13, 2023.

4 Experimental design

4.1 Market design

Building on Weber and Camerer (1998) and Frydman et al. (2014), our laboratory market is designed to quantify behavioral trading biases such as the disposition effect — the tendency to sell winners too early and hold onto losers for too long. Our setup is therefore well-suited to determine whether data availability helps traders make better decisions, specifically whether it dampens or reinforces the disposition effect. Following this stream of literature, we emphasize individual decision-making and abstract from market clearing and price discovery considerations.[8]

Market structure and timing. Participants trade a single virtual stock in a laboratory market over four rounds. Each round consists of 40 trials, indexed by t. In turn, each trial represents a stock price update lasting eight seconds. Asset prices and payoffs are denominated in "experimental dollars" (E$), an artificial currency converted to Canadian dollars at the end of the experiment using an exchange rate of E$1 = CA$0.003.

At the start of each round, participants receive E$500 and one unit of the stock, which has an initial price of E$1000 – the total endowment is therefore E$1500. The cash buffer is designed to absorb stock market losses during the round, reducing the likelihood that the limited liability constraint binds.

Participants may hold at most one unit of the stock at any time, and short selling is prohibited. These constraints simplify the strategy space, allowing for sharper identification of the underlying mechanisms. Participants only decide whether to buy the stock if they do not hold it or sell if they do. While trading, they may borrow and carry negative cash balances, but any negative balance at the end of the round is subtracted from the stock portfolio's value to compute the payoff.

Following Frydman et al. (2014), trading is disallowed during the first three trials of each round. This allows participants to observe asset price movements before making trading decisions. From $t = 4$ onward, they can freely buy and sell the stock at any time, subject to position limits.

[8] On January 24, 2025, we filed a pre-analysis plan which is available online at https://osf.io/vbkjh/?view_only=34606b993a0d40b8951cd20d22b8e9e1.

Asset price dynamics. The stock midpoint price is updated each trial through a two-state Markov chain. In the "good" state (g), the midpoint price increases with probability 0.8 and decreases with probability 0.2. In the "bad" state (b), these probabilities are reversed. The magnitude of midpoint price changes is independently drawn with equal probability from the set $\{\text{E\$50}, \text{E\$100}, \text{E\$150}\}$. Conditional on being in state $i \in \{g, b\}$ at trial t, there is an 85% chance of remaining in state i at trial $t+1$ and a 15% chance of switching to state $-i$. Consequently, the stock exhibits momentum: price increases (decreases) are likely followed by further increases (decreases), making the price predictable.

Participants receive information on the process used to generate prices and the transition probabilities, but we do not disclose the state in any given trial. Instead, each participant has to use the history of prices to infer the current state and make predictions about future returns. To facilitate comparison across experimental subjects, we use the same price histories for all participants.

Price grid. As in real-life limit order markets, trades do not execute at the midpoint price. We introduce a fixed half-spread equal to E$0.01 and assume that the market is infinitely deep. That is, participants and trading bots can buy any quantity of the asset at an ask price equal to the midpoint plus E$0.01, and sell any quantity at a bid price equal to E$0.01 below the midpoint. In our setup, the bid-ask spread is fixed (corresponding, for example, to a binding tick size constraint) and very small relative to the midpoint such that it can safely be disregarded by Bayesian investors. Our purpose is not to model trading costs, but to distinguish between buyer-initiated and seller-initiated trades.

Algorithmic traders. At each tick, algorithmic traders arrive to the market with Poisson intensity λ. Upon arrival, they submit either a marketable buy or a marketable sell order for one unit of the asset, which immediately executes against the ask, respectively bid price.

There are two types of algorithmic traders: insiders and noise traders, following Glosten and Milgrom (1985). With probability α, a trading bot is an "insider" who can perfectly infer the current state: If the stock is in a good (bad) state, they always buy (respectively, sell) one unit of the stock. Conversely, with probability $1-\alpha$, the algorithm is a noise trader. Noise traders are uninformed and equally likely to buy or to sell one unit of the asset.

Participant beliefs. To assess participants' perceptions of asset price trends, we follow Weber and Camerer (1998) by directly eliciting beliefs about the current state of the stock. Specifically, for each round, trading is paused for 20 seconds before trial $t = 20$, during which participants are presented with the following questions:

How likely is the stock to go up next? and *How confident are you in this assessment?*

Participants respond using a sliding scale to indicate the perceived probability of an uptick in the next trial, followed by a five-point Likert scale to rate their confidence.

Additionally, as in Chapkovski et al. (2024b), we elicit participant beliefs prior to trial $t = 1$. At the start of each round, with no price history available, the stock is equally likely to be in a good or bad state. Participants are asked, "How likely is the stock to go up in the first period?" and respond using a sliding scale to indicate their perceived probability (denoted by u) and their confidence level. We then derive a measure of prior belief about the stock price using the formula $1 - 2 \times |u - 0.5|$, which assigns a higher value when the participant's belief is closer to the correct answer, $u = 0.5$.

Treatments. Our experimental design combines "within" and "between" treatments. All participants experience rounds with and without order flow data, but each is exposed to only one combination of the parameters α (order flow informativeness, i.e., a measure of data *quality*) and λ (the investor arrival rate, i.e., a measure of data *quantity*).

By having each participant trade on both data-enhanced and data-free platforms, we (i) establish baseline trading behavior at the individual level and (ii) elicit their willingness to pay for data. Keeping the parameters for data informativeness and quantity fixed for each participant ensures they understand the data quality they are paying for, facilitating accurate willingness-to-pay elicitation. The order of data-enhanced and data-free rounds is randomized. Participants are divided into two groups: half start with a round where data is available, and the other half start without data.

Table 2 outlines the four between-subject experimental sessions, each featuring a different combination of order flow informativeness (α) and arrival rate (λ). In addition, to test for potential framing effects, we consider Session V which is identical to Session II with the sole exception that the data is framed as "premium."

Session	Data framing	Share of insiders (α)	Arrival rate (λ)
Session I	Neutral	Low: $\alpha = 0.30$	Low: $\lambda = 0.50$
Session II	Neutral	Low: $\alpha = 0.30$	High: $\lambda = 1.00$
Session III	Neutral	High: $\alpha = 0.70$	Low: $\lambda = 0.50$
Session IV	Neutral	High: $\alpha = 0.70$	High: $\lambda = 1.00$
Session V	Premium	Low: $\alpha = 0.30$	High: $\lambda = 1.00$

Table 2: Experimental sessions

In Sessions I and III, there are few insiders ($\alpha = 0.3$), while Sessions II and IV feature a high proportion of insiders ($\alpha = 0.7$), increasing order flow informativeness. The amount of data also varies between participants: in expectation, one order arrives every two seconds in Sessions I and II ($\lambda = 0.5$), whereas algorithmic orders arrive at a stochastic rate of one trade per second in expectation throughout Sessions III and IV.

This design allows us to examine how different levels of data quality (information content) and data quantity affect trading behavior. As discussed in Section 4.3, the parameters are calibrated such that the value of data is clearly ranked across sessions – as measured by expected payoff and entropy gains. Figure 1 illustrates a screenshot of the trading platform, including the price dynamics, order flow data, and order entry panels.

Round structure and timing. To become familiar with the platform, all participants start with a short training round ("round 0") consisting of 20 trials or price updates. This training round is discarded in the data analysis.

Following the training round, we employ multiple price lists to elicit participants' pre-trade willingness to pay (WTP) for financial data, following Andersen et al. (2006) and Jack et al. (2022). By eliciting WTP for data before any experience, we capture participants' baseline beliefs about the value of data.

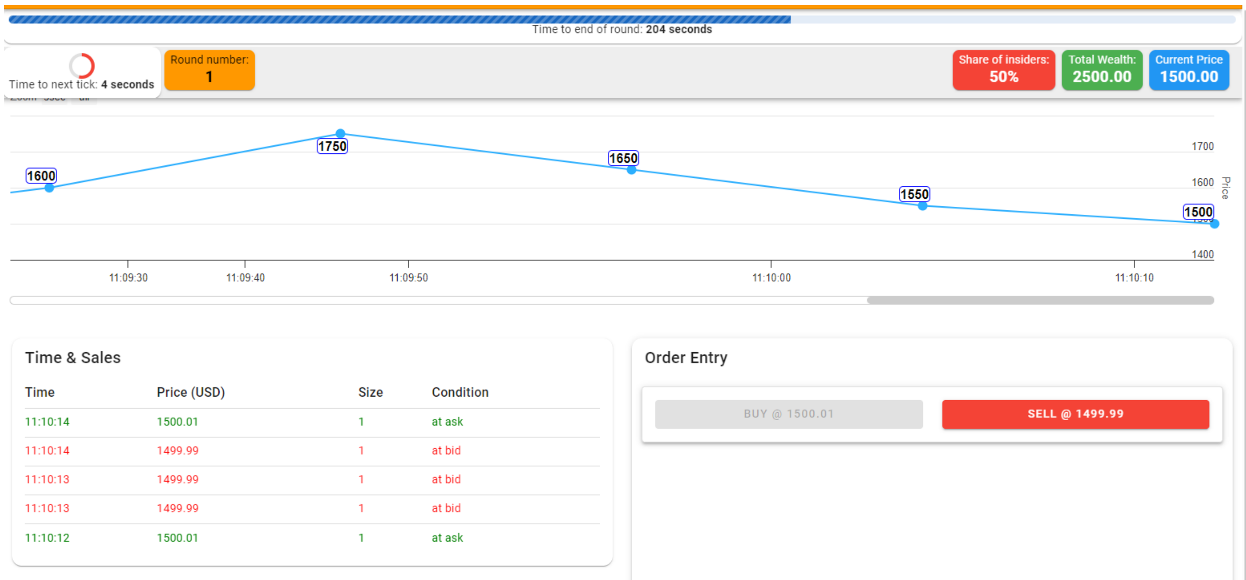


Figure 1: Trading platform design

Next, participants engage in two trading rounds: A data-enhanced round where they can observe the order flow, and a data-free round where they cannot not. The order of these rounds is randomized to control for learning effects. After each trading round, we measure cognitive load using a modified version of the NASA-TLX (Task Load Index) developed by Hart and Staveland (1988). We provide participants with a questionnaire measuring subjective workload across five dimensions. Participants rate each dimension on a 7-point scale, from "very low" (1) to "very high" (7). The scores from each dimension are then averaged to produce a composite cognitive load score.

1. Mental demand: "How mentally demanding was the task in this round ?"
2. Temporal demand: "How hurried or rushed did you feel? "
3. Frustration: "How insecure, discouraged, or irritated did you feel?"
4. Performance: "How successful were you in trading this round?"
5. Effort: "How hard did you have to work to accomplish your performance?"

After the trading rounds, we again employ multiple price lists to elicit participants' post-trade willingness to pay for financial data. One important concern, however, is hypothetical bias – that is, overstating or misrepresent true valuations when participants are presented with non-binding scenarios that do not involve real financial consequences (List and Gallet, 2001). To mitigate hypothetical bias, we simultaneously implement three techniques supported by the literature: cheap talk scripts (Cummings and Taylor, 1999), certainty calibration (Blumenschein et al., 2008), and consequentiality scripts (Bulte et al., 2005).

Finally, participants answer 12 financial literacy questions provided in Appendix A. The questions are adapted from Fernandes et al. (2014) and include the three-question measure developed by Lusardi et al. (2011).

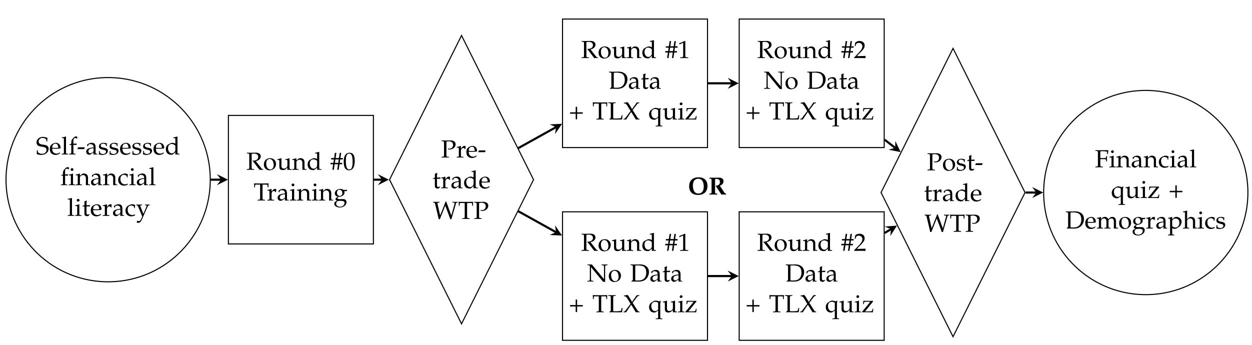
To distinguish between subjective and objective financial literacy, we also ask participants:

> *On a scale from zero to ten, where zero is not at all knowledgeable about personal finance and ten is very knowledgeable about personal finance, what number would you be on the scale?*

This question measures self-assessed (subjective) financial knowledge and is identical to the one used in Cupák et al. (2020). We pose this question before the experiment begins to avoid any influence from the trading game's monetary performance or the quiz's perceived difficulty.

Figure 2 illustrates the experiment's timeline. Appendix B reproduces the consent form. The experimental instructions given to participants are reproduced in Appendix C. Before the trading starts, participants need to correctly answer six comprehension questions, listed in Appendix D. The willingness-to-pay form is reproduced in Appendix E. This allows us to make sure that participants indeed understand the experiment before the trading rounds start. Finally, at the end of the experiment all participants are required to fill in a demographic questionnaire.

Figure 2: **Experiment timing**



Payments. Each participant is expected to receive a fixed compensation of GB£9 (equivalent to approximately CA\$15) per hour. In addition to the fixed amount, participants receive a payment proportional to their performance in the trading game and the financial literacy quiz.

In accordance with standard experimental procedures, the payment round is determined by randomly selecting one of the three trading rounds at the end of the experiment. Participants' earnings are equal to the amount of cash they hold at the end of this randomly chosen payment round plus the end-of-round price of any stock that they own. The exchange rate is set to E\$1 = CA\$0.003.

Besides the payment round profit, participants are also rewarded for correct answers in the post-experimental financial literacy quiz. Specifically, each correct answer is rewarded with 30 experimental dollars, equivalent to CA\$0.09. The monetary quiz payoff is subsequently added to the payment round payoff to determine the total payment.

4.2 The optimal strategy of a Bayesian trader

In this section, we describe the optimal strategy of a Bayesian risk-neutral trader who maximizes the end-of-round expected payoff, as in Frydman and Rangel (2014).

At each tick, the average magnitude of price changes is $\frac{50+100+150}{3} = \text{E\$}100$. Therefore, if the stock is in the good state (g) at time t, it increases by an average of E\$100 with probability 0.85 and decreases by E\$100 with probability 0.15. Therefore, the expected price change conditional on state g is: $(0.85 - 0.15) \times 100 = \text{E\$}70$. Similarly, if the stock is in the bad state (b) at time t, the expected price change is $-\text{E\$}70$.

Let q_t denote the initial holdings of the stock at the start of trial t ($q_t \in \{0, 1\}$) and $\Delta q_t \equiv q_t - q_{t-1} \in \{-1, 0, 1\}$ the direction of the trade (sell, do not trade, or buy, respectively). A Bayesian investor's expected utility can be written as

$$U(q_t, \Delta q_t) = (q_t + \Delta q_t)(2\pi_t - 1)\left(\underbrace{\text{Prob}(g_{t+1} \mid g_t) - \text{Prob}(b_{t+1} \mid g_t)}_{>0}\right) \times \text{E\$}70, \tag{1}$$

where the expression immediately follows from the symmetry of the transition matrix. From equation (1), a Bayesian investors optimally holds the stock if and only if $\pi_t > \frac{1}{2}$.

Therefore, the optimal strategy for a Bayesian investor depends solely on their beliefs about the asset being in a "good" state or not. Investors can learn about the asset state from two sources: that is, past price history (as in Frydman et al., 2014) and from order flow data. We discuss both information channels separately.

Learning from past prices. Let $z_t \in \{1, -1\}$ denote the direction of the price change at trial t. Further, let $s_t \in \{g, b\}$ stand for the state of the Markov process. The estimated probability of being in the good state at trial t using only price history data, that is π_t', evolves as follows:

$$\begin{aligned}
&\pi_t'\left(\pi_{t-1}, z_t\right) \\
&= \frac{\mathbb{P}\left(z_t \mid s_t = g\right)\mathbb{P}\left(s_t = g \mid \pi_{t-1}\right)}{\mathbb{P}\left(z_t \mid s_t = g\right)\mathbb{P}\left(s_t = g \mid \pi_{t-1}\right) + \mathbb{P}\left(z_t \mid s_t = b\right)\mathbb{P}\left(s_t = b \mid \pi_{t-1}\right)} \\
&= \frac{\left(0.5 + 0.3z_t\right)\left(0.85\pi_{t-1} + 0.15\left(1 - \pi_{t-1}\right)\right)}{\left(0.5 + 0.3z_t\right)\left(0.85\pi_{t-1} + 0.15\left(1 - \pi_{t-1}\right)\right) + \left(0.5 - 0.3z_t\right)\left(0.15\pi_{t-1} + 0.85\left(1 - \pi_{t-1}\right)\right)},
\end{aligned} \tag{2}$$

where $q_0 = 0.5$ is the long-run stationary probability of state g.

Learning from order flow. To determine how Bayesian investors update their beliefs upon observing order flow data, let b and s denote the number of buy and sell orders arriving in the market within one tick. Since insiders always trade in the direction of the next price change and noise traders buy or sell with equal probability, the probability of observing a buy order conditional on the next price movement is:

$$\begin{aligned}
\mathbb{P}\left(\text{buy order} \mid \text{good state}\right) = \alpha \times 1 + (1 - \alpha) \times 0.5 = 0.5 + 0.5\alpha \text{ and} \\
\mathbb{P}\left(\text{buy order} \mid \text{bad state}\right) = \alpha \times 0 + (1 - \alpha) \times 0.5 = 0.5 - 0.5\alpha,
\end{aligned} \tag{3}$$

respectively. The conditional probabilities are symmetric for sell market orders.

Therefore, upon observing order flow data with b buys and s sells, a Bayesian trader updates their belief about the stock being in a good state from π_t' to π_t, where

$$
\begin{aligned}
\pi_t \equiv \mathbb{P}\left(\text{good state} \mid \{b,s\}\right) &= \pi_t' \times \frac{\mathbb{P}\left(\{b,s\} \mid \text{good state}\right)}{\pi_t \mathbb{P}\left(\{b,s\} \mid \text{good state}\right) + (1-\pi_t)\,\mathbb{P}\left(\{b,s\} \mid \text{bad state}\right)} \\
&= \frac{(1+\alpha)^b (1-\alpha)^s \pi_t'}{(1+\alpha)^b (1-\alpha)^s \pi_t' + (1-\alpha)^b (1+\alpha)^s (1-\pi_t')} \\
&= \frac{\pi_t'}{\pi_t' + (1-\pi_t')\left(\frac{1-\alpha}{1+\alpha}\right)^{b-s}}. \qquad (4)
\end{aligned}
$$

While investors may not have the capacity to execute the calculations in (2) and (4) within the limited time frame, they can rely on simple heuristics. A sequence of price increases suggests a higher likelihood of being in the good state at the next price update, while a series of price declines implies the opposite. Additionally, if the volume of buy orders significantly exceeds sell orders, investors should have a stronger belief that the asset is in a good state — especially so if the proportion of insiders (α) is larger.

4.3 The informational value of data

We measure the informational value of data in two ways: First, we simulate 10,000 price paths for each parameter combination and compute difference in expected payoff for a Bayesian trader playing the game with and without access to the data. Second, we compute the expected information gain (Kullback and Leibler, 1951), defined as the difference between the entropy of the state process before observing order flow and the entropy conditional on order flow:

$$\text{InformationGain}_t = H_t\left(s_t \mid \{p_t\}\right) - H_t\left(s_t \mid \{p_t, b, s\}\right), \qquad (5)$$

where the entropy of the asset state at tick t is defined as $H_t\left(s_t\right) = -\pi_t \log_2\left(\pi_t\right) - (1-\pi_t)\log_2\left(1-\pi_t\right)$.

Figure 3 shows that a larger data is more valuable when there is a higher share of insiders (larger α) and, conditional on α, when the arrival rate or algorithmic traders is larger (higher λ). The pattern is qualitatively identical regardless on whether we measure data quality by the increase in expected payoff or entropy gain. We calibrate α and λ such that the expected value of data in monetary terms is sensibly equal between Sessions II and III. That is, investors should be roughly indifferent between a highly active market with few insiders and a slower market with a higher share of informed trading.

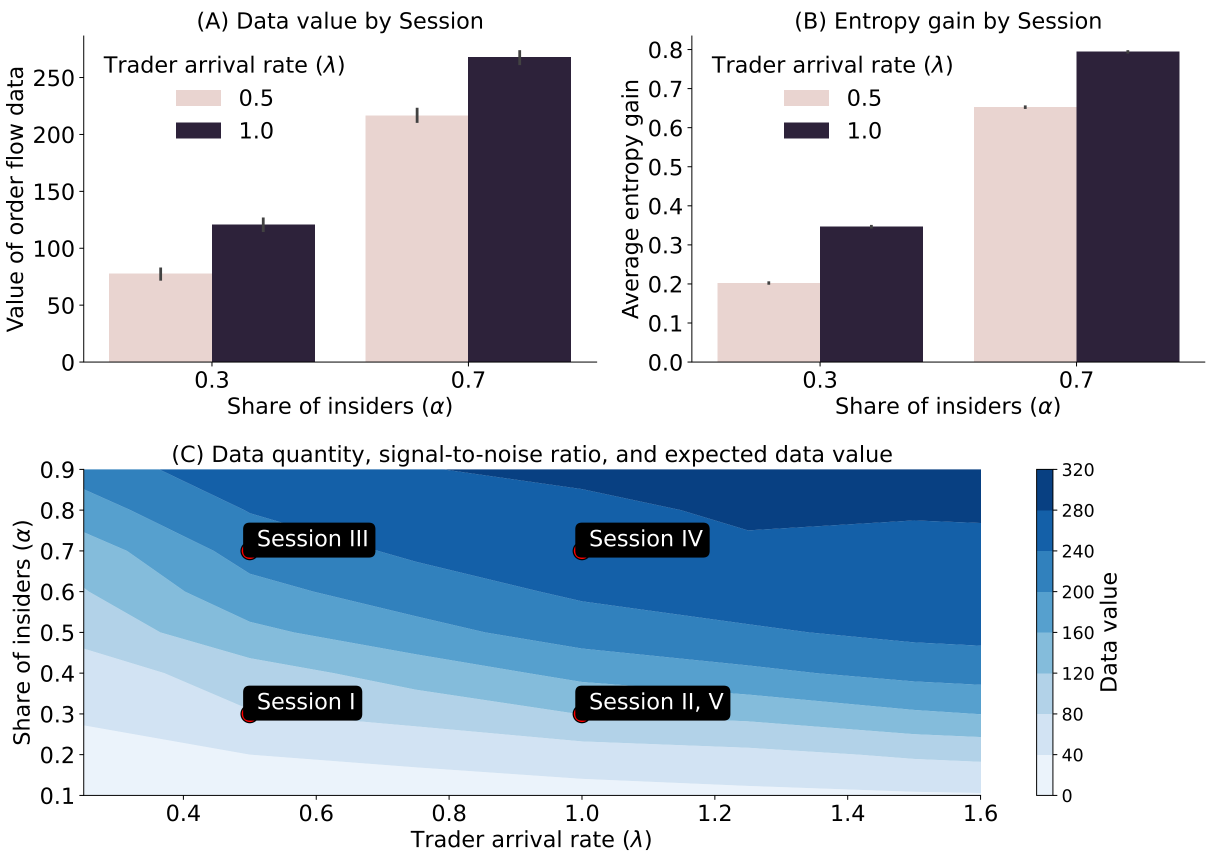


Figure 3: Value of order flow data

Session	Share of insiders (α)	Arrival rate (λ)	Value of data	Entropy gain
Session I	Low: $\alpha = 0.3$	Low: $\lambda = 0.50$	77.44	0.20
Session II, V	Low: $\alpha = 0.3$	High: $\lambda = 1.00$	120.74	0.34
Session III	High: $\alpha = 0.7$	Low: $\lambda = 0.50$	216.62	0.65
Session IV	High: $\alpha = 0.7$	High: $\lambda = 1.00$	267.70	0.79

Table 3: Value of order flow data and entropy gain

5 Results

5.1 Cohort formation

We recruited participants from Prolific, an online subject pool designed for academic research (Palan and Schitter, 2018). To ensure a diverse and representative sample, we targeted United States residents, leveraging Prolific's stratified sampling based on U.S. census data across age, gender, and ethnicity. All sessions took place on January 29, 2025, with participants requiring a median of 30 minutes to complete the experimental tasks.

Table 4 breaks down the sample across experimental sessions. For each session, we aimed for 60 participants. To implement this, individual subjects were assigned sequentially upon arrival to Sessions 1 through 5. However, because participants could exit the experiment at any point after allocation, actual sample sizes deviate slightly from this target. Additionally, two participants were timed out by Prolific but successfully completed all tasks upon data inspection. They were paid in full, and their data is included in the analysis.

Session	Data framing	Share of insiders (α)	Arrival rate (λ)	Participants
Session I	Neutral	Low: $\alpha = 0.30$	Low: $\lambda = 0.50$	58
Session II	Neutral	Low: $\alpha = 0.30$	High: $\lambda = 1.00$	68
Session III	Neutral	High: $\alpha = 0.70$	Low: $\lambda = 0.50$	55
Session IV	Neutral	High: $\alpha = 0.70$	High: $\lambda = 1.00$	66
Session V	Premium	Low: $\alpha = 0.30$	High: $\lambda = 1.00$	60
Total				**307**

Table 4: Sample breakdown

Each participant received a fixed compensation of GB£7.48 per hour (approximately CA$13.30), corresponding to CA$6.65 for the median experiment duration of 30 minutes. In addition to this base payment, participants earned a performance-based bonus tied to their outcomes in a randomly selected trading round and their financial literacy quiz score. The average and median bonus payments are GB£4.44 and GB£4.46 respectively, with bonuses ranging from GB£2.87 to GB£6.15.

Demographics. After completing the experiment, participants filled out a short demographics questionnaire. The sample is gender-balanced, with 153 participants identifying as female (49.83%), 153 as male (49.83%), and one as non-binary. The median age is 45 years, with participants ranging from 18 to 80 years old.

In terms of education, 54.72% (168 participants) have some form of undergraduate education, and 26.71% (82 participants) hold a master's degree, MBA, or doctorate. However, only 24.75% have formal training in finance, economics, or management, and just 8.4% majored in finance.

Trading experience is widespread in the sample, with 57% of participants reporting significant experience and 43.6% familiar with online trading apps. Among experienced traders, 67.8% primarily trade stocks, while 23.6% focus on cryptocurrencies and 5.7% on bonds. Daily traders make up 8.8% of the sample, while 26% trade at least once a month.

At the end of the experiment, participants completed a 12-question financial literacy quiz, detailed in Online Appendix A. The average score is 9.01 (75%), with a standard deviation of 2.41. Scores are positively correlated with formal finance education (9.16 average) and real-life trading experience (9.52 average), suggesting that both education and hands-on market exposure contribute to financial literacy.

5.2 Willingness to pay for order flow data

We begin by examining the determinants of willingness-to-pay (WTP) for order flow data using the regression model:

$$\begin{aligned} \text{WTP}_i = \beta_0 + \beta_1 \,\text{ShareInsiders}_i + \beta_2 \,\text{OrderArrivalRate}_i + \beta_3 \,\text{FinancialEducation}_i + \\ + \beta_4 \,\text{Overconfidence}_i + \text{Controls}_i + \epsilon_i. \end{aligned} \tag{6}$$

Here, WTP_i denotes one of several measures: pre-trade willigness-to-pay (WTP), post-trade WTP, the gap between WTP and a Bayesian trader's valuation, or the change from pre- to post-trade WTP. We proxy data value using both its intrinsic quality (the share of insiders, α) and quantity (order arrival rate, λ). Controls include a dummy variable for having taken a finance course, a standardized overconfidence measure (the difference between self-assessed financial knowledge and financial literacy quiz score), age, gender, real-life trading experience, the difference in realized payoffs between data and no-data rounds, a dummy for whether the participant encountered the data round first, and a dummy for whether the data was framed as a premium product. We present the estimation results in Table 5.

Hypothesis 1. (Data informativeness) *Participants are willing to pay higher amounts for more informative order flow data, as measured by the share of insiders and the order arrival rate.*

We find partial support for this hypothesis. From Model (1), a one standard deviation increase in the share of informed order flow leads to a 12.88% (=12.22/94.84) increase in willingness to pay. We find that participants are only willing to pay more for larger amounts of data if such data is of high quality, as evidenced by the positive and significant interaction between insider share and order arrival rate.

However, participants' willingness to pay for order flow data increases at a slower rate than the Bayesian valuation, suggesting a low demand elasticity with respect to quality. That is, participants do not fully internalize variations in data quality. This result is supported by the negative coefficients on both the insider share and order arrival rate in Models (4) and (5), which examine overpayment for data – that is, the difference between participant WTP and the Bayesian data value.

Notably, Models (6) and (7) indicate that participants' valuation of informed trading remains constant, as their willingness to pay is unchanged pre- and post-trade. In contrast, their willingness to pay for data quantity (i.e., a higher λ) increases sharply after trading, indicating that they update their valuation based on in-game experience.

Hypothesis 2. (Payment relative to Bayesian benchmark) *Participants are willing to pay more for order flow data than a Bayesian trader would.*

To test Hypothesis 2, we compare participants' willingness to pay (WTP_i) with the expected data value for a Bayesian trader ($DataValue_i$) as reported in Table 3. The rationale is that access to data may induce overconfidence, leading participants to overvalue its benefits.

Our findings, however, provide sharp evidence for the contrary. In Model (4), participants on average underpay for financial data by E$66.75—that is, they pay 59 cents on the dollar relative to a fully rational Bayesian trader (computed as $1 - {}^{66.75}/_{66.75+94.84}$). In contrast, participants with formal financial education pay an additional premium of E$76.64 and value the data in line with a Bayesian investor. The result suggests that financial education enables subjects to assess data value more accurately. This evidence is particularly compelling, given that experimental subjects tend to overstate their willingness-to-pay due to hypothetical bias.

Hypothesis 3. (Overconfidence) *Participants with higher overconfidence levels are more likely to overpay for financial data before the trading game starts.*

Participants who self-report a high level of financial literacy are more likely to overvalue their skill in using market data, and therefore might be willing to pay larger amounts. We measure overconfidence as the difference between the financial quiz score and the self-assessed financial literacy at the start of the game.

Table 5: Willingness to pay for financial data

This table presents the estimation results of regression (6):

$$\text{WTP}_i = \beta_0 + \beta_1 \,\text{ShareInsiders}_i + \beta_2 \,\text{OrderArrivalRate}_i + \beta_3 \,,\text{FinancialEducation}_i + \\ + \beta_4 \,\text{Overconfidence}_i + \text{Controls}_i + \epsilon_i.$$

Columns (1)–(3) report coefficients for post-trade WTP (columns 1–2) and pre-trade WTP (column 3), while columns (4)–(5) present estimates for overpayment (defined as WTP minus the Bayesian data value). Columns (6)–(7) report the change in WTP (i.e., post-trade minus pre-trade). Key predictors include the share of insiders, order arrival rate, and their interaction, with controls for financial education, overconfidence, premium framing, financial quiz score, age, gender, order of data exposure, trading experience, and the payoff difference between data and no-data rounds. Standard errors, shown in parentheses, are clustered by participant. Observations are weighted by the certainty calibration score.

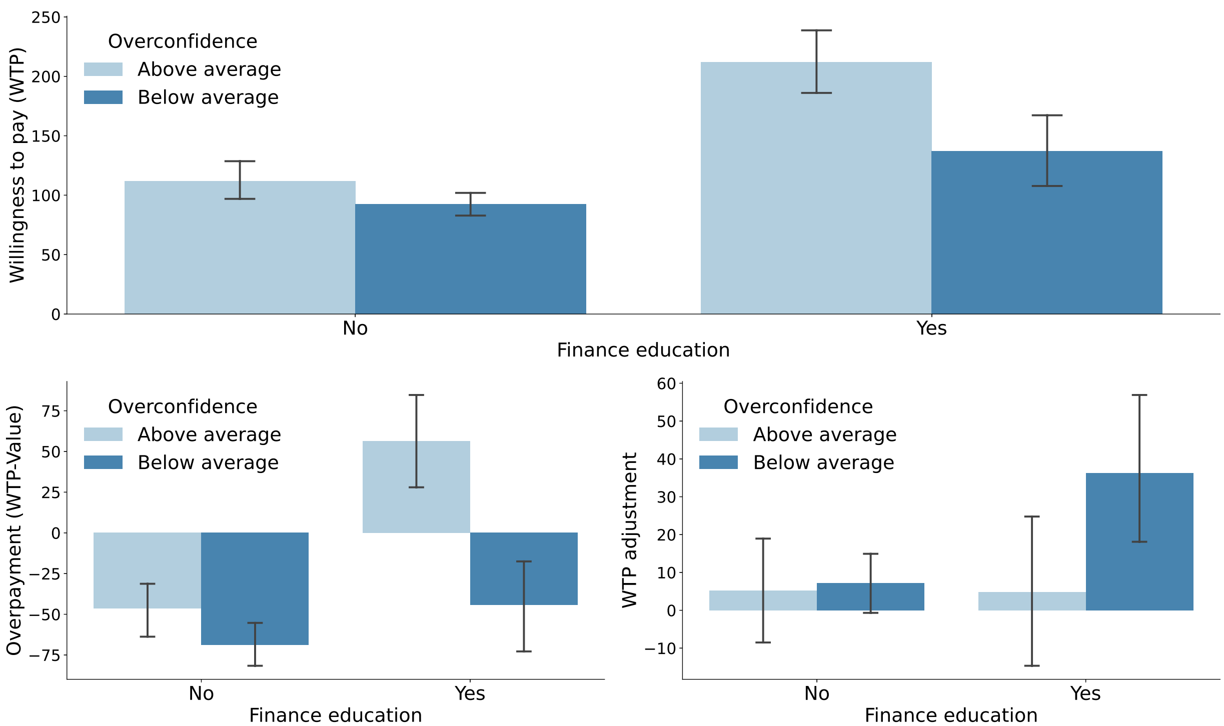
	Willingness-to-pay (WTP)			Overpayment		ΔWTP	
	Post-trade		Pre-trade	WTP−data value		Post vs. pre-trade	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Share insiders	12.22**	8.93	6.41	-58.26***	-61.55***	5.81	7.23
	(2.27)	(1.80)	(1.66)	(-10.81)	(-12.44)	(1.34)	(1.70)
Order arrival rate	-0.02	2.65	-15.35***	-22.41***	-19.74***	15.33**	13.40**
	(-0.00)	(0.48)	(-4.51)	(-3.74)	(-3.54)	(2.77)	(2.70)
Share insiders × order arrival	14.83*	12.74*	14.55***	13.00*	10.91	0.28	1.88
	(2.26)	(1.98)	(4.51)	(1.98)	(1.70)	(0.06)	(0.35)
Finance course taken	76.64***	78.82***	69.68***	76.64***	78.82***	6.96	12.32
	(4.82)	(4.88)	(5.20)	(4.82)	(4.88)	(0.61)	(1.20)
Overconfidence	27.30***	18.89***	30.95***	27.30***	18.89***	-3.65	-7.26
	(3.27)	(5.04)	(3.63)	(3.27)	(5.04)	(-0.44)	(-1.13)
Premium framing	17.34**		25.81*	17.34**		-8.48	
	(2.71)		(2.14)	(2.71)		(-1.17)	
Financial quiz score	14.56		6.69	14.56		7.87	
	(1.64)		(0.77)	(1.64)		(1.13)	
Age	-5.34		2.55	-5.34		-7.89**	
	(-0.77)		(0.53)	(-0.77)		(-2.41)	
Gender (female)	17.13		6.96	17.13		10.18	
	(1.27)		(0.55)	(1.27)		(1.02)	
Data round is first	-2.48		7.83	-2.48		-10.31	
	(-0.29)		(1.57)	(-0.29)		(-1.26)	
Trading experience	-2.44		-5.26	-2.44		2.82	
	(-0.15)		(-0.37)	(-0.15)		(0.37)	
Payoff difference	-9.40		-9.73	-9.40		0.33	
	(-1.06)		(-1.25)	(-1.06)		(0.06)	
Constant	94.84***	103.67***	88.31***	-66.75***	-57.92***	6.53	5.55
	(4.80)	(17.91)	(5.39)	(-3.38)	(-10.01)	(0.62)	(0.98)
Observations	305	305	307	305	305	305	305
R-squared	0.17	0.18	0.14	0.35	0.36	0.04	0.05

We find robust support for this hypothesis. In Model (1), a one standard deviation increase in overconfidence is associated with a 28.78% increase in willingness to pay (i.e., $27.30/94.84$). The effect is slightly larger when WTP is elicited before trading begins, but the difference is not statistically significant.

Figure 4 illustrates the interaction between overconfidence and financial education as it impacts willingness-to-pay (WTP) for market data. In the top panel, we show that overconfident participants exhibit a higher WTP on average, with the effect being strongest for those who have taken a finance course.

Figure 4: **Overconfidence, Financial Education, and Willingness-to-Pay for Market Data**

This figure illustrates the relationship between overconfidence, financial education, and willingness-to-pay (WTP) for market data. Financial education is measured as a binary indicator equal to one if the participant has taken at least one formal finance course and zero otherwise. Overconfidence is defined as the difference between a participant's normalized self-assessment of financial literacy and their actual score on the financial quiz. The top panel shows WTP across groups, while the bottom panels depict overpayment (WTP minus Bayesian data value) and WTP adjustment (post-trade minus pre-trade). Error bars indicate 95% confidence intervals.



The bottom-left panel shows that overconfident participants with financial education are the only subgroup that, on average, overpays for financial data. Meanwhile, the bottom-right panel indicates that financial education moderates WTP adjustments post-trade, with financially educated participants revising their valuations more sharply upwards when they are *not* overconfident. These findings suggest that while financial education enhances participants' ability to assess data value, it may also amplify overconfidence-driven demand.

Hypothesis 4. (Framing Bias) *Participants are willing to pay more for identical data when it is framed as "premium" rather than presented neutrally.*

We hypothesize that labeling or describing the data as "premium" influences participants' perceived value and willingness to pay, even when data quality remains unchanged. The results confirm this effect: participants exposed to the "premium" framing are willing to pay 18.28% more for order flow data (i.e., $17.34/94.84$ based on Model 1 estimates).

This finding highlights that marketing and presentation strategies can significantly impact the perceived value of information. A salient implication is that online brokers and data providers may leverage such biases to enhance revenue, even without improving the underlying quality of financial data.

5.3 Order flow data and trading behavior

We next examine whether access to order flow data influences trading behavior and outcomes for retail traders. To do so, we estimate a series of panel regressions across participants i and rounds t:

$$
\begin{aligned}
y_{i,t} = {} & \beta_0 + \beta_1 \mathrm{d}_{\text{data},it} + \beta_2 \mathrm{d}_{\text{data},it}\text{ShareInsiders}_i + \beta_3 \mathrm{d}_{\text{data},it}\text{OrderArrivalRate}_i + \\
& + \beta_4 \text{ FinancialEducation}_i + \beta_5 \text{ } \mathrm{d}_{\text{data},it}\text{FinancialEducation}_i + \beta_6 \text{ Overconfidence}_i + \\
& + \text{Controls}_{it} + \epsilon_i, \qquad (7)
\end{aligned}
$$

Interaction terms capture whether the impact of data access depends on the share of insiders, order arrival rate, or financial education. We include control variables such as subject age and gender and round fixed effects to account for participant and round-level heterogeneity.

Hypothesis 5. (Tick-by-tick trading decisions) *Access to order flow data aligns participants' trading decisions with those of a Bayesian investor. The effect is stronger if data is more informative.*

This hypothesis reflects the economic value of data: when the data is both informative and timely, participants can align their actions with rational, profit-maximizing behavior. At the end of each tick, we compare the participants' position in the stock (either hold or do not hold) with the position of a rational Bayesian trader. That is, we define the variable *BayesianPosition* as follows:

$$\text{BayesianPosition}_{i,t} = \begin{cases} 1, & \text{if participant } i \text{ takes the Bayesian-optimal position at tick } t, \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

The *BayesianRate* for participant i in round r is then calculated as the proportion of ticks in which the participant's action is aligned with the Bayesian-optimal strategy:

$$\text{BayesianRate}_{i,r} = \frac{\sum_{t=1}^{T} \text{BayesianPosition}_{i,t}}{T}, \tag{9}$$

where T is the total number of ticks in round r.

The estimation results in Models (1) and (2) of Table 6 provide strong empirical support for Hypothesis 5. In the absence of data, participants' positions align with the Bayesian benchmark 52% of the time, as indicated by the intercept in Model (2). Access to order flow data increases alignment by 3 percentage points (5.7%), with the effect being 2 percentage points (3.8%) stronger when data quality is higher – as proxied by the share of informed traders. Participants with formal financial education exhibit an additional 3 percentage point increase in their alignment with Bayesian strategy when data is available. Conversely, overconfident traders are 2 percentage points less aligned with Bayesian optimal trades, consistent with the idea that overconfidence distorts optimal decision-making.

The effects are statistically significant but economically modest. Participants learn from order flow data and move closer to Bayesian-optimal trades, but the effect remains marginal. Retail traders do not fully internalize the value of order flow data, as evidenced by the lack of a significant payoff increase in Models (3) through (5) when data is available. The average data value for a Bayesian trader, considering all realized price and order flow paths, is E$115. However, experimental participants realize only 11.8% of this value in additional payoffs (computed as $13.56/115$ from

Model 3). This suggests that while access to data improves trade alignment, it does not necessarily translate into higher profits.

Table 6: Order flow data and trading behavior

This table presents the estimation results of regression (7):

$$
\begin{aligned}
y_{i,t} = \beta_0 + \beta_1 \mathrm{d}_{\mathrm{data},it} + \beta_2 \mathrm{d}_{\mathrm{data},it}\mathrm{ShareInsiders}_i + \beta_3 \mathrm{d}_{\mathrm{data},it}\mathrm{OrderArrivalRate}_i + \\
+ \beta_4\ \mathrm{FinancialEducation}_i + \beta_5\ \mathrm{d}_{\mathrm{data},it}\mathrm{FinancialEducation}_i + \beta_6\ \mathrm{Overconfidence}_i + \\
+ \mathrm{Controls}_{it} + \epsilon_i,
\end{aligned}
$$

where $y_{i,t}$ represents, in turn, the alignment to Bayesian position (Columns 1–2), trade payoff (Columns 3–4), and trade count (Columns 5–6). The key independent variable, $\mathrm{d}_{\mathrm{data},it}$, is a dummy equal to one if participant i has access to order flow data in round t and zero otherwise. We include controls for financial education, overconfidence, premium framing, financial quiz score, age, gender, order of data exposure, trading experience, and round fixed effects. Standard errors, shown in parentheses, are clustered by participant, round, and experimental session.

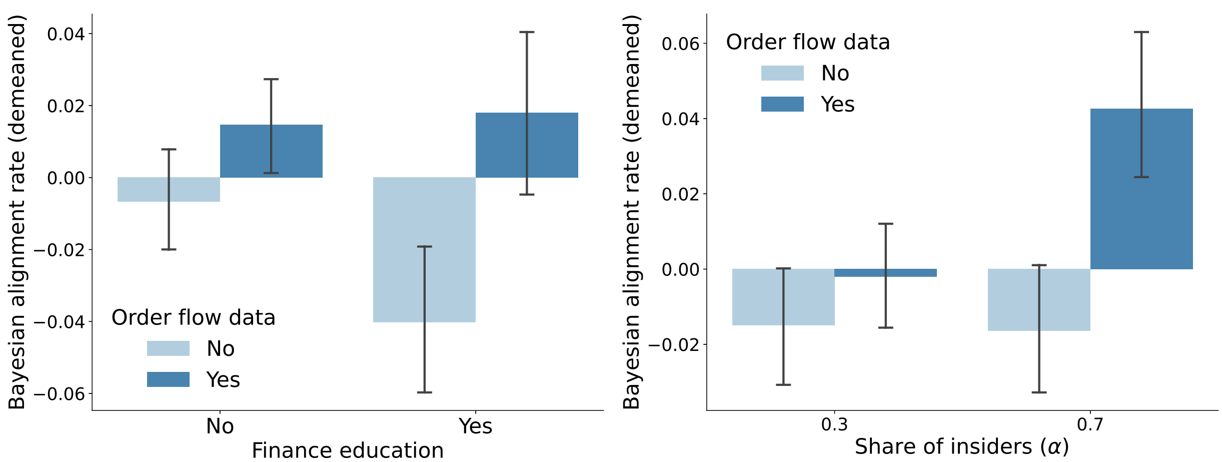
	Bayesian rate		Trade payoff			Trade count	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Data available	0.03***	0.03***	13.56	17.22	3.23	0.39***	0.48*
	(5.29)	(3.35)	(0.48)	(0.82)	(0.10)	(4.39)	(1.94)
Data available × order rate	0.01	0.01		-38.06***	-36.47**	0.23	0.39***
	(1.03)	(1.14)		(-3.34)	(-2.34)	(1.19)	(2.95)
Data available × share insiders	0.02***	0.02**		37.52**	34.28	-0.06	-0.26**
	(3.04)	(2.41)		(2.43)	(1.67)	(-0.32)	(-2.39)
Data available × order rate × share insiders	0.01	0.01		24.61*	27.27	0.07	-0.14
	(1.61)	(1.43)		(2.00)	(1.73)	(0.42)	(-1.21)
Finance course taken		-0.03*			-76.15		1.12*
		(-1.74)			(-1.52)		(2.03)
Data available × financial course		0.03**			82.90		-0.16
		(2.30)			(1.34)		(-0.20)
Overconfidence		-0.02***			-48.24*		0.49
		(-2.87)			(-2.04)		(1.37)
Premium framing		0.01			30.75		-1.13***
		(1.07)			(0.79)		(-3.17)
Premium × data available		-0.02			-38.07		-0.43
		(-1.10)			(-0.63)		(-1.52)
Financial quiz score		-0.01			-8.98		-0.33
		(-1.14)			(-0.45)		(-0.95)
Age		-0.01			7.28		0.50*
		(-1.32)			(0.41)		(1.88)
Gender (female)		-0.01			-42.54*		0.68
		(-0.65)			(-1.83)		(1.35)
Data round is first		-0.01			-18.25		0.57
		(-1.04)			(-0.92)		(1.49)
Trading experience		0.01			-0.17		0.03
		(0.76)			(-0.01)		(0.04)
Constant	0.51***	0.52***	2,421.03***	2,420.97***	2,465.18***	6.15***	5.45***
	(170.51)	(44.95)	(157.98)	(158.05)	(73.83)	(25.90)	(9.44)
Observations	614	614	614	614	614	614	
R-squared	0.08	0.11	0.02	0.04	0.01	0.07	

Figure 5 illustrates how access to order flow data influences participants' alignment with

Bayesian-optimal trading strategies, conditional on financial education (left panel) and data quality (right panel). The left panel shows that financial education improves Bayesian alignment, but only when participants have access to order flow data. Without data, financially educated participants exhibit greater variability in alignment, suggesting that they rely on other heuristics or strategies. The right panel reveals that data availability has a stronger impact when the share of insiders (α) is high, reinforcing the idea that participants internalize the value of data more effectively in high-quality information environments. However, even with access to data, alignment remains imperfect, indicating that participants do not fully adopt Bayesian strategies.

Figure 5: **Order flow data and alignment with Bayesian trader**

This figure illustrates the impact of order flow data on participants' alignment with Bayesian-optimal trading decisions. The $BayesianRate$ for participant i in round r is then calculated as the proportion of ticks in which the participant's action is aligned with the Bayesian-optimal strategy. Financial education is measured as a binary indicator equal to one if the participant has taken at least one formal finance course and zero otherwise. The left panel examines the effect of order flow data on Bayesian alignment, splitting the sample by financial education. The right panel assesses the impact of data availability across different levels of data quality, proxied by the share of insiders (α). Error bars represent 95% confidence intervals.



Hypothesis 6. (Cognitive Load and Trading Frequency) *Participants exposed to a higher volume of data will trade more frequently.*

Achtziger et al. (2020) find that individuals under higher cognitive load exhibit faster response times, shifting to an intuitive "fast thinking" framework. In our setting, we expect participants to engage more actively with trading platforms featuring large volumes of data, conditional on data quality. The estimation results in Models (5) and (6) of Table 6 support this conjecture. Participants execute 0.48 more trades (an 8.8% increase) in rounds where data is available, with the effect nearly doubling in treatments with a high order flow arrival rate (0.87 more trades, or a 16% increase over the unconditional rate).

However, this increased trading activity comes at a cost. In rounds with high data volume, participants earn 1.47% lower payoffs relative to no-data rounds ($36.47/2465.18$, Model 4). Meanwhile, higher-quality data improves trading payoffs, though the effect is not always statistically significant. These findings suggest that, holding quality constant, a higher volume of data can lead to over-trading and worse outcomes for retail traders.

Figure 6: **Order flow data, trade count, and payoffs**

This figure illustrates the impact of order flow data on participants' trading activity and outcomes. The number of trades (left panel) is measured as the total executed trades per round, while trade payoff (right panel) represents participants' end-of-round portfolio value. Order arrival rate indicates the frequency of new orders entering the market (i.e., $\lambda = 0.5$ or $\lambda = 1$). The sample is split by order flow data availability, with dark blue bars representing participants with access to order flow data and light blue bars representing those without. Error bars represent 95% confidence intervals.

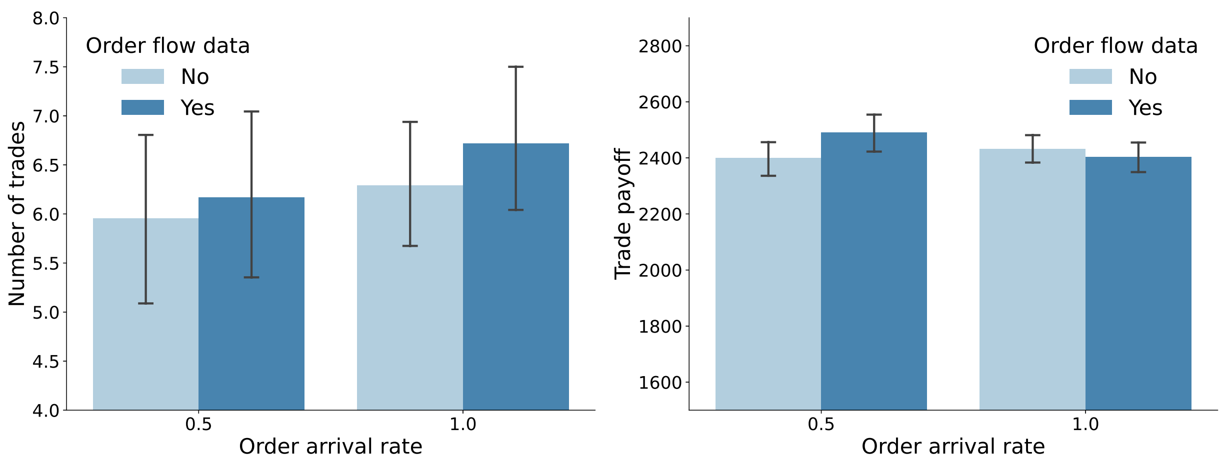


Figure 6 illustrates the impact of order arrival rate on trading behavior and outcomes. The left panel shows that participants trade more frequently when data is available, with the effect being stronger at higher order arrival rates. This suggests that access to data encourages greater market engagement, particularly in fast-moving environments. The right panel examines trade payoffs, revealing that while data access slightly improves payoffs when order arrival rates are low, this advantage disappears at higher arrival rates. This pattern is consistent with the idea that while order flow data stimulates trading activity, its benefits to performance are limited and may even lead to over-trading in high-frequency settings.

Hypothesis 7. (Disposition Effect) *The disposition effect is reduced on data-enhanced platforms. That is, the proportion of losses (gains) realized is higher (lower) in rounds featuring order flow data, defined as in Odean et al. (1998). The effect is stronger for higher values of α and λ.*

Next, we measure the disposition effect, following Odean et al. (1998). At each price update, we classify investors' positions based on realized and unrealized gains and losses. A realized gain (loss) occurs when a stock is sold at a higher (lower) price than its purchase price, while an unrealized (paper) gain (loss) is recorded when the participant holds the stock in their portfolio at the end of a trial and the market price is higher (lower) than the purchase price.

To quantify the disposition effect, we compute the following ratios, aggregating across all stocks and trials in round t:

$$\text{PGR}_{i,r} = \frac{\text{Realized Gains}_{i,r}}{\text{Realized Gains}_{i,r} + \text{Paper Gains}_{i,r}}, \quad \text{PLR}_{i,r} = \frac{\text{Realized Losses}_{i,r}}{\text{Realized Losses}_{i,r} + \text{Paper Losses}_{i,r}}. \tag{10}$$

We conjecture that access to order flow data increases participants' tendency to realize gains. Moreover, we expect that higher-quality data – proxied by the share of insiders and order arrival rate – plays a greater role in mitigating behavioral biases.

We report the results in Table 7. In line with our hypothesis, data availability reduces the disposition effect. From models (1) and (2), the difference between PLR and PGR decreases significantly when order flow data is available, indicating that participants are less prone to holding onto losses while realizing gains. The effect is entirely driven by a strong increase in realized losses from 11% to 17% (Model 3), while the impact on PGR (profit realization rate) is smaller and statistically weaker (Columns 5–6).

Table 7: Order flow data and the disposition effect

This table presents the estimation results of regression (7):

$$y_{i,t} = \beta_0 + \beta_1 \mathrm{d}_{\mathrm{data},it} + \beta_2 \mathrm{d}_{\mathrm{data},it}\mathrm{ShareInsiders}_i + \beta_3 \mathrm{d}_{\mathrm{data},it}\mathrm{OrderArrivalRate}_i + \\ + \beta_4\ \mathrm{FinancialEducation}_i + \beta_5\ \mathrm{d}_{\mathrm{data},it}\mathrm{FinancialEducation}_i + \beta_6\ \mathrm{Overconfidence}_i + \\ + \mathrm{Controls}_{it} + \epsilon_i,$$

where $y_{i,t}$ represents, in turn, the difference between profit and loss realization rates (PLR-PGR, Columns 1–2), the profit realization rate (PGR, Columns 3–4), and the loss realization rate (PLR, Columns 5–6). The key independent variable, $\mathrm{d}_{\mathrm{data},it}$, is a dummy equal to one if participant i has access to order flow data in round t and zero otherwise. We include controls for financial education, overconfidence, premium framing, financial quiz score, age, gender, order of data exposure, trading experience, and round fixed effects. Standard errors, shown in parentheses, are clustered by participant, round, and experimental session.

	PLR-PGR		PLR		PGR	
	(1)	(2)	(3)	(4)	(5)	(6)
Data available	0.04**	0.05***	0.06***	0.06***	0.01**	0.01
	(2.63)	(3.48)	(3.22)	(4.17)	(2.39)	(1.26)
Data available × share insiders	0.00	-0.01	-0.00	-0.01	-0.00	-0.00*
	(0.02)	(-1.07)	(-0.28)	(-1.55)	(-0.70)	(-1.87)
Data available × order rate	-0.01	-0.00	-0.00	0.00	0.00	0.01***
	(-0.57)	(-0.03)	(-0.29)	(0.61)	(1.70)	(3.22)
Data available × order rate × share insiders	0.03**	0.02***	0.03*	0.02**	-0.00	-0.01**
	(2.77)	(3.99)	(1.91)	(2.11)	(-1.22)	(-2.44)
Finance course taken		0.01		0.02		0.01
		(0.39)		(0.99)		(1.25)
Data available × financial course		-0.00		-0.00		-0.00
		(-0.07)		(-0.01)		(-0.11)
Overconfidence		-0.01		-0.00		0.01
		(-0.85)		(-0.15)		(1.42)
Premium framing		-0.01		-0.03		-0.02**
		(-0.49)		(-1.60)		(-2.33)
Premium × data available		-0.05*		-0.05***		-0.00
		(-1.86)		(-3.32)		(-0.24)
Financial quiz score		-0.01		-0.01		0.00
		(-0.60)		(-0.47)		(0.48)
Age		0.01		0.02		0.01*
		(0.88)		(1.48)		(1.98)
Gender (female)		0.03		0.03		0.00
		(1.42)		(1.42)		(0.16)
Data round is first		0.06***		0.06***		0.00
		(4.83)		(4.21)		(0.56)
Trading experience		0.02		0.01		-0.00
		(0.85)		(0.54)		(-0.18)
Constant	-0.01	-0.06***	0.11***	0.05**	0.11***	0.10***
	(-0.78)	(-3.04)	(8.17)	(2.67)	(28.05)	(10.04)
Observations	484	484	484	484	614	614
R-squared	0.04	0.08	0.04	0.08	0.01	0.04

We do not find evidence that the mitigating effect of order flow data on the disposition effect depends on data quality, as the interaction terms with the share of insiders and order arrival rate are small and mostly insignificant. However, the three-way interaction (Column 2) suggests that when both data volume and quality are high, the reduction in the disposition effect is slightly dampened, possibly due to increased trading activity in large order flow environments leading to participants realizing more losses.

Interestingly, framing data as a "premium" product strengthens the disposition effect, suggesting that participants become more reluctant to realize losses when the data is perceived as exclusive or high-value.

Figure 7: **Order flow data and the disposition effect**

This figure illustrates the proportion of losses realized (PLR) across different participant characteristics, comparing those with and without access to order flow data. The left panel splits the sample by financial education, measured as a binary indicator equal to one if the participant has taken at least one formal finance course. The middle panel examines the effect of premium framing, where data is labeled as a "premium" product. The right panel splits participants by overconfidence, defined as the difference between self-assessed financial knowledge and actual quiz performance. Dark blue bars represent participants with access to order flow data, while light blue bars represent those without. Error bars indicate 95% confidence intervals.

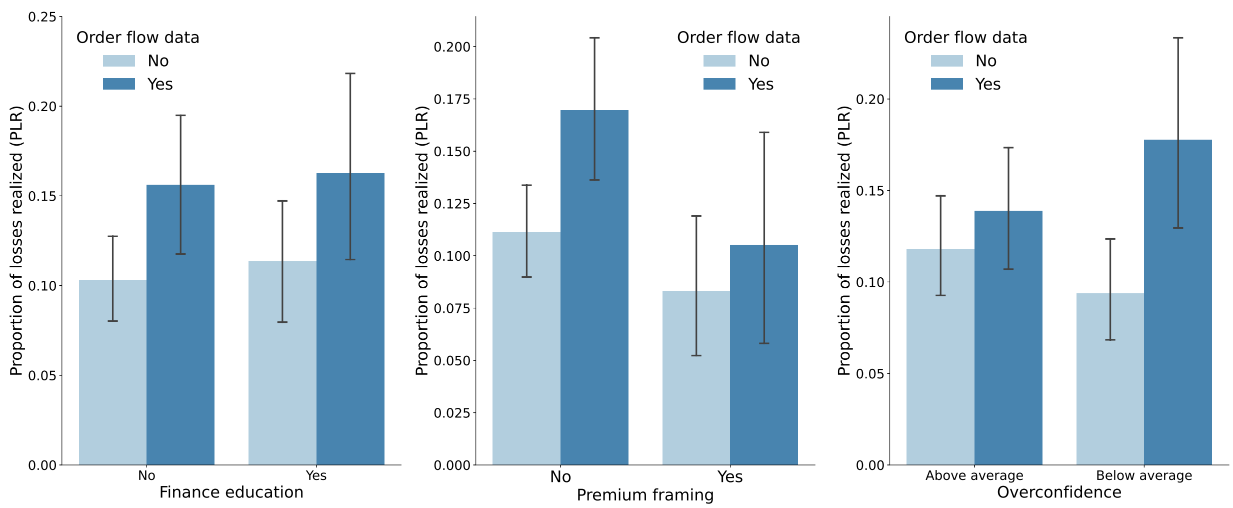


Figure 7 illustrates how access to order flow data influences participants' willingness to realize losses, depending on financial education, premium framing, and overconfidence. The left panel shows that data availability increases the proportion of losses realized (PLR) for both financially educated and non-educated participants, with a similar effect size. This suggests that financial education does not significantly alter how traders respond to data when deciding to cut losses. The middle panel highlights that premium framing dampens the effect of data on PLR, as participants exposed to premium-labeled data remain more reluctant to realize losses. This finding suggests that perceiving data as exclusive or high-value may make traders less likely to act on negative signals. Finally, the right panel reveals that overconfident participants react less strongly to data availability, realizing a lower share of losses compared to less overconfident traders. This suggests that overconfidence reduces responsiveness to order flow data, possibly by making traders less willing to update their beliefs in the face of losses.

Hypothesis 8. (Beliefs) *Participants hold more accurate beliefs in rounds featuring order flow information than in rounds without order flow information. The effect is stronger for participants with higher levels of financial literacy.*

Belief accuracy is computed by mapping participants' probability estimates to a Bayesian benchmark derived from optimal trading strategies. To ensure comparability, Bayesian probabilities π_t are transformed into a five-point Likert scale using quantile thresholds, distinguishing between cases with and without access to order flow data. Since Bayesian inference becomes more precise with order flow data, the distribution of benchmark probabilities is wider, allowing beliefs to converge more strongly toward extreme values.

Accuracy is then measured as one minus the normalized absolute difference between a participant's response and the corresponding Bayesian probability, $\mathcal{L}(\pi_t)$, as defined in Equation (11):

$$\text{Prediction accuracy} = 1 - \frac{1}{4}\left|\mathcal{L}\left(\pi_t\right) - \text{investor belief}\right|. \tag{11}$$

This formulation penalizes deviations symmetrically while ensuring a scale from zero (completely inaccurate) to one (fully Bayesian). For example, if a participant assigns a low probability to a good state (e.g., a Likert score of 1), but the Bayesian estimate suggests $\pi_t = 0.85$, mapped to a score of 4, the prediction accuracy is $1 - \frac{|1-4|}{4} = 0.25$.

Table 8: Order flow data and accuracy of beliefs

This table presents the estimation results of regression (7):

$$
\begin{aligned}
\text{PredictionAccuracy}_{i,t} = {} & \beta_0 + \beta_1 \text{d}_{\text{data},it} + \beta_2 \text{d}_{\text{data},it}\text{ShareInsiders}_i + \beta_3 \text{d}_{\text{data},it}\text{OrderArrivalRate}_i + \\
& + \beta_4 \text{ FinancialEducation}_i + \beta_5 \text{ d}_{\text{data},it}\text{FinancialEducation}_i + \beta_6 \text{ Overconfidence}_i + \\
& + \text{Controls}_{it} + \epsilon_i,
\end{aligned}
$$

Prediction accuracy measures the alignment between participants' stated beliefs and Bayesian-optimal probabilities, with confidence-weighted versions incorporating participants' self-reported confidence levels. The key independent variable, $\text{d}_{\text{data},it}$, is a dummy equal to one if participant i has access to order flow data in round t and zero otherwise. Interaction terms capture the role of financial education, overconfidence, and premium framing in shaping belief accuracy. We include additional controls for financial quiz scores, age, gender, order of data exposure, trading experience, and round fixed effects. Columns (1) through (3) weigh the prediction accuracy by reported confidence in the prediction, while Columns (4) through (6) use equally weighted observations. Standard errors, shown in parentheses, are clustered by participant, round, and experimental session.

	Prediction accuracy					
	Confidence-weighted			Non-weighted		
	(1)	(2)	(3)	(4)	(5)	(6)
Data available	0.18***	0.16***	0.17***	0.17***	0.15***	0.16***
	(10.78)	(6.97)	(6.90)	(10.75)	(7.50)	(7.20)
Finance course taken		-0.07	-0.08*		-0.07	-0.07
		(-1.55)	(-1.79)		(-1.60)	(-1.71)
Data available × financial course		0.06	0.06		0.07	0.06
		(1.31)	(1.40)		(1.35)	(1.42)
Overconfidence			-0.02			-0.02
			(-0.99)			(-1.39)
Premium framing			-0.00			0.01
			(-0.04)			(0.28)
Premium × data available			-0.02			-0.03
			(-0.85)			(-1.52)
Financial quiz score			0.03*			0.02
			(2.05)			(1.71)
Data value (E$)			0.00			0.00
			(0.20)			(0.48)
Age			-0.02*			-0.02*
			(-2.08)			(-1.83)
Gender (female)			-0.02			-0.02
			(-0.90)			(-0.95)
Data round is first			-0.04**			-0.05***
			(-2.40)			(-3.06)
Trading experience			0.02			0.01
			(0.80)			(0.52)
Constant	0.60***	0.62***	0.65***	0.62***	0.64***	0.67***
	(43.15)	(32.28)	(20.31)	(50.03)	(37.61)	(23.41)
Observations	591	591	591	610	610	610
R-squared	0.15	0.16	0.21	0.16	0.16	0.20

The estimation results in Table 8 show that access to order flow data significantly improves prediction accuracy, both in confidence-weighted (Models 1–3) and non-weighted (Models 4–6) specifications. Based on point estimates from Model (3), accuracy increases by 16 percentage points, from 62% to 78%, representing a 25% improvement. This suggests that participants incorporate order flow data effectively into their belief formation, leading to a closer alignment with Bayesian-optimal probabilities. The interaction between financial education and data availability is positive but not significant, implying that the benefits of data access apply broadly, regardless of financial expertise.

5.4 Order flow data and cognitive load

We now examine whether access to order flow data increases cognitive load and which specific components of mental effort are affected.

Hypothesis 9. (Order flow data and cognitive load) *Participants will report significantly higher NASA-TLX scores in rounds with order flow data than in rounds without data.*

Following Hypothesis 9, we expect participants to report significantly higher NASA-TLX scores in rounds with order flow data than in rounds without. Cognitive load is measured using six NASA-TLX subcomponents: frustration, mental demand, temporal demand, performance, and effort, with higher scores indicating greater cognitive strain.

The estimation results in Table 9 support the hypothesis that order flow data increases cognitive load, though the effect is small and varies across dimensions. In Column (1), overall NASA-TLX scores increase by 0.28 points (1.33%) in rounds with order flow data, suggesting that participants experience a mild increase in cognitive strain when processing market information. This effect is primarily driven by higher frustration (0.15, Column 2, a 3.92% increase), temporal demand (0.13, Column 4, a 3.43% increase), and effort (0.07, Column 6, a 1.48% increase). These results indicate that participants find trading more demanding and rushed when order flow data is available, but the overall cognitive burden remains moderate.

Table 9: Market data and cognitive load

This table presents the estimation results of regression (7):

$$
\begin{aligned}
\text{CognitiveLoad}_{i,t} = \beta_0 &+ \beta_1 \mathrm{d}_{\text{data},it} + \beta_2 \mathrm{d}_{\text{data},it}\text{ShareInsiders}_i + \beta_3 \mathrm{d}_{\text{data},it}\text{OrderArrivalRate}_i + \\
&+ \beta_4 \text{ FinancialEducation}_i + \beta_5 \text{ d}_{\text{data},it}\text{FinancialEducation}_i + \beta_6 \text{ Overconfidence}_i + \\
&+ \text{Controls}_{it} + \epsilon_i,
\end{aligned}
$$

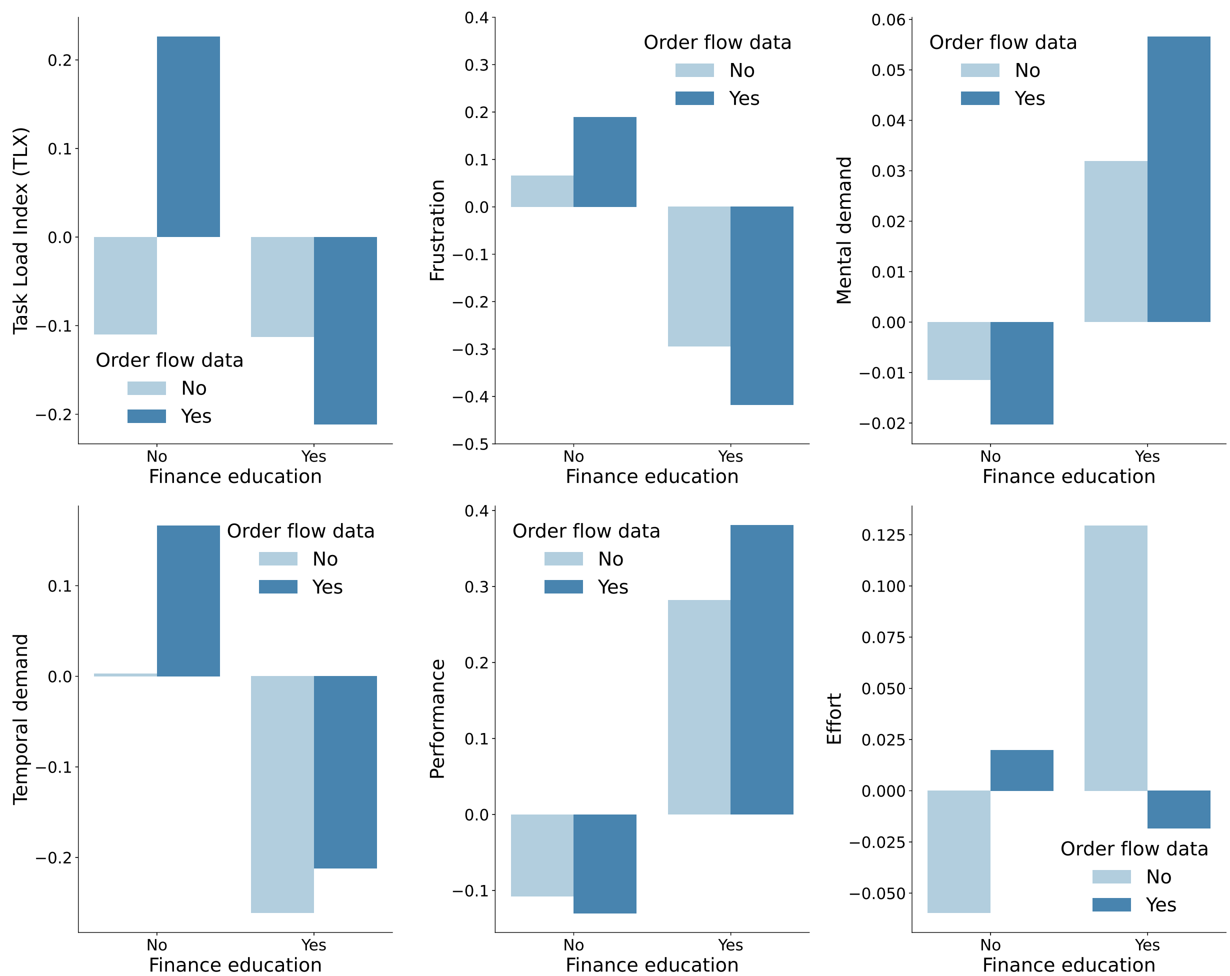
where the dependent variable is, in turn, total NASA-TLX cognitive load score (Column 1), frustration (Column 2), mental demand (Column 3), temporal demand (Column 4), performance self-assessment (Column 5), and effort (Column 6). The key independent variable, $\mathrm{d}_{\text{data},it}$, is a dummy equal to one if participant i has access to order flow data in round t and zero otherwise. Interaction terms capture the role of financial education, overconfidence, and premium framing in shaping belief accuracy. We include additional controls for financial quiz scores, age, gender, order of data exposure, trading experience, and round fixed effects. Standard errors, shown in parentheses, are clustered by participant, round, and experimental session.

	TLX score	Frustration	Mental demand	Temporal demand	Performance	Effort
	(1)	(2)	(3)	(4)	(5)	(6)
Data available	0.28**	0.15**	-0.01	0.13***	-0.06	0.07**
	(2.64)	(2.30)	(-0.27)	(2.94)	(-0.52)	(2.31)
Data available × share insiders	-0.08	0.01	-0.07***	-0.15**	0.14*	-0.02
	(-0.61)	(0.54)	(-2.92)	(-2.28)	(1.81)	(-0.55)
Data available × order rate	-0.21	-0.10***	0.03	-0.04	-0.11	-0.00
	(-1.56)	(-5.41)	(1.45)	(-0.57)	(-1.61)	(-0.08)
Data available × order rate × share insiders	0.26**	-0.05***	0.16***	-0.02	0.08	0.09**
	(2.49)	(-3.22)	(5.20)	(-0.41)	(1.23)	(2.41)
Finance course taken	0.13	-0.11	0.19	-0.19	0.03	0.22
	(0.20)	(-0.38)	(1.19)	(-0.78)	(0.14)	(1.64)
Data available × financial course	-0.39	-0.24	0.03	-0.07	0.12	-0.22***
	(-1.64)	(-0.85)	(0.44)	(-0.89)	(0.55)	(-3.37)
Overconfidence	0.41	-0.27*	0.06	0.10	0.35***	0.16*
	(1.23)	(-2.03)	(0.66)	(0.82)	(4.43)	(1.90)
Premium framing	0.29	0.27**	-0.04	-0.13	0.18	0.01
	(0.88)	(2.14)	(-0.34)	(-0.95)	(0.89)	(0.06)
Premium × data available	0.44*	-0.18**	0.13*	0.13	0.24	0.12*
	(1.85)	(-2.32)	(1.85)	(1.03)	(1.23)	(1.86)
Financial quiz score	-0.19	-0.25**	-0.11	0.03	0.16	-0.02
	(-0.56)	(-2.41)	(-1.30)	(0.21)	(1.65)	(-0.12)
Age	0.45	-0.05	0.24***	0.18	-0.07	0.14**
	(1.64)	(-0.49)	(3.21)	(1.73)	(-1.03)	(2.13)
Gender (female)	0.42	0.27	0.33**	0.22	-0.53***	0.12
	(0.84)	(1.70)	(2.53)	(1.58)	(-3.49)	(0.85)
Data round is first	0.57	-0.08	0.08	0.28*	0.03	0.26**
	(1.54)	(-0.60)	(0.74)	(1.98)	(0.25)	(2.15)
Trading experience	-0.22	-0.21	-0.03	-0.12	0.28**	-0.13
	(-0.33)	(-0.86)	(-0.16)	(-0.44)	(2.27)	(-0.83)
Constant	21.05***	3.82***	4.75***	3.78***	4.00***	4.70***
	(39.95)	(21.68)	(30.36)	(19.31)	(26.24)	(32.37)
	(38.92)	(21.88)	(28.18)	(18.83)	(27.67)	(33.08)
Observations	614	614	614	614	614	614
R-squared	0.03	0.07	0.06	0.04	0.14	0.04

Interestingly, data quality moderates cognitive load in different ways. The three-way interaction between data availability, order arrival rate, and insider share is positive and significant in overall TLX scores (Column 1), mental demand (Column 3), and effort (Column 6), indicating that high-frequency, high-quality data environments impose the greatest cognitive burden. Finally, participants with formal financial education experience lower effort when data is available, as seen in the negative interaction in Column (6), suggesting that they may be better equipped to handle information-intensive trading environments. Figure 8 graphically illustrates the results.

Figure 8: **Order flow data and cognitive load**

This figure illustrates the impact of order flow data on NASA-TLX cognitive load components, split by financial education. The top-left panel shows the overall Task Load Index (TLX), while the remaining panels display individual components: frustration (top-center), mental demand (top-right), temporal demand (bottom-left), performance (bottom-center), and effort (bottom-right). Dark blue bars represent rounds where participants had access to order flow data, while light blue bars represent rounds without data.



6 Conclusion

Our study provides new insights into how retail investors value and use order flow data. Sophisticated traders are willing to pay larger amounts for market data than participants without financial education, although the demand for data is relatively inelastic with respect to quality. At the same time, behavioral factors such as investor overconfidence and framing significantly inflate valuations. Despite some learning effects, improvements in trading outcomes remain modest – participants capture only a fraction of the theoretical value of data. Access to order flow data partially mitigates the disposition effect, but also leads to excessive trading activity.

From a regulatory and policy standpoint, these results raise important questions about market structure and retail investor protection. As financial markets become more complex and data-intensive, understanding how retail investors interact with market data is crucial for designing appropriate oversight frameworks. The findings could inform discussions on data pricing, market access, and the importance of financial education. Ensuring that investors understand the value of data feeds provided by brokerages and exchanges is a key concern for regulators.

For the investment industry, our results offer valuable insights into retail investor behavior and demand for market data. Brokerages and financial technology firms can use these insights to design more effective trading platforms and develop data products that better serve retail investors. Understanding whether investors derive value from premium data services – and how their behavior changes in response to different data offerings – could lead to more efficient pricing models, improved product offerings, and better educational tools to help retail traders make informed decisions.

Ultimately, our findings suggest that while market data can improve decision-making, its benefits to retail traders are often overstated. Behavioral biases and marketing cues shape demand as much as actual utility, raising important questions about whether retail investors are paying for an edge they cannot fully use.

References

Achtziger, A., Alós-Ferrer, C., Ritschel, A., 2020. Cognitive Load in Economic Decisions. SSRN Electronic Journal .

Andersen, S., Harrison, G. W., Lau, M. I., Rutström, E. E., 2006. Elicitation using multiple price list formats. Experimental Economics 9, 383–405.

Arnold, M., Pelster, M., Subrahmanyam, M. G., 2022. Attention triggers and investors' risk-taking. Journal of Financial Economics 143, 846–875.

Barber, B. M., Huang, X., Odean, T., Schwarz, C., 2022. Attention-Induced Trading and Returns: Evidence from Robinhood Users. Journal of Finance 77, 3141–3190.

Barber, B. M., Odean, T., 2000. Trading is hazardous to your wealth: The common stock investment performance of individual investors. Journal of Finance 55, 773–806.

Barber, B. M., Odean, T., Berk, J., Blake, D., French, K., Gervais, S., Griffin, J., Karolyi, A., Mullainathan, S., Rubinstein, M., Trueman, B., 2008. All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. The Review of Financial Studies 21, 785–818.

Bartlett, R. P., McCrary, J., O'Hara, M., 2023. The Market Inside the Market: Odd-Lot Quotes. The Review of Financial Studies .

Blumenschein, K., Blomquist, G. C., Johannesson, M., Horn, N., Freeman, P., Cull, M., Stephens, D., McIntosh, A., Salyer, K., Dawson, A., Hamm, G., Claywell, L., Wallace, J., Hawkins, M., 2008. Eliciting Willingness to Pay Without Bias: Evidence from a Field Experiment. The Economic Journal 118, 114–137.

Bogousslavsky, V., Muravyev, D., 2024. An Anatomy of Retail Option Trading. SSRN Electronic Journal .

Bryzgalova, S., Pavlova, A., Sikorskaya, T., 2023. Retail Trading in Options and the Rise of the Big Three Wholesalers. The Journal of Finance 78, 3465–3514.

Bulte, E., Gerking, S., List, J. A., De Zeeuw, A., 2005. The effect of varying the causes of environmental problems on stated WTP values: evidence from a field study. Journal of Environmental Economics and Management 49, 330–342.

Chapkovski, P., Khapko, M., Zoican, M., 2024a. Gamified Risk-Taking .

Chapkovski, P., Khapko, M., Zoican, M., 2024b. Trading Gamification and Investor Behavior. Management Science Forthcoming.

Chen, D. L., Schonger, M., Wickens, C., 2016. oTree-An open-source platform for laboratory, online, and field experiments. Journal of Behavioral and Experimental Finance 9, 88–97.

Cummings, R. G., Taylor, L. O., 1999. Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method. American Economic Review 89, 649–665.

Cupák, A., Fessler, P., Hsu, J. W., Paradowski, P. R., 2020. Confidence, financial literacy and investment in risky assets: Evidence from the Survey of Consumer Finances. Finance and Economics Discussion Series 2020.

Dyhrberg, A. H., Shkilko, A., Werner, I. M., 2022. The Retail Execution Quality Landscape. SSRN Electronic Journal .

Farboodi, M., Singal, D., Veldkamp, L., Venkateswaran, V., 2024. Valuing Financial Data. The Review of Financial Studies .

Farrell, M., Green, T. C., Jame, R., Markov, S., 2022. The democratization of investment research and the informativeness of retail investor trading. Journal of Financial Economics 145, 616–641.

Fernandes, D., Lynch, J. G., Netemeyer, R. G., 2014. Financial literacy, financial education, and downstream financial behaviors. Management Science 60, 1861–1883.

Frydman, C., Barberis, N., Camerer, C., Bossaerts, P., Rangel, A., 2014. Using Neural Data to Test a Theory of Investor Behavior: An Application to Realization Utility. The Journal of Finance 69, 907–946.

Frydman, C., Rangel, A., 2014. Debiasing the disposition effect by reducing the saliency of information about a stock's purchase price. Journal of Economic Behavior & Organization 107, 541–552.

Gao, M., Huang, J., Almeida, H., Akbas, F., Brockman, P., Faccio, M., Giannetti, M., Hanley, K., He, J., Hirshleifer, D., Hu, J., Irvine, P., Johnson, T., Kelley, E., Koch, P., Kronlund, M., Li, L., Li, L., Li, W., Nanda, V., Pearson, N., Pollet, J., Pontiff, J., Spiegel, M., Subrahmanyam, A., Tchistyi, A., Tetlock, P., Nieuwerburgh, S. V., Wagner, A., Wang, C., Wang, R., Weisbenner, S., Yang, B., Yang, L., Zhang, H., Zhao, H., Zhao, M., Zhu, H., Zhu, W., 2020. Informing the Market: The Effect of Modern Information Technologies on Information Production. The Review of Financial Studies 33, 1367–1411.

Glosten, L., Milgrom, P., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. Journal of Financial Economics 14, 71–100.

Hart, S. G., Staveland, L. E., 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. Advances in Psychology 52, 139–183.

Havakhor, T., Rahman, M. S., Zhang, T., Zhu, C., 2024. Tech-Enabled Financial Data Access, Retail Investors, and Gambling-Like Behavior in the Stock Market. https://doi.org/10.1287/mnsc.2021.01379 71, 1646–1670.

Jack, B. K., McDermott, K., Sautmann, A., 2022. Multiple price lists for willingness to pay elicitation. Journal of Development Economics 159, 102977.

Kalda, A., Loos, B., Previtero, A., Hackethal, A., 2021. Smart(Phone) Investing? A within Investor-time Analysis of New Technologies and Trading Behavior. .

Kaniel, R., Saar, G., Titman, S., 2008. Individual investor trading and stock returns. Journal of Finance 63, 273–310.

Kelley, E. K., Tetlock, P. C., 2013. How Wise Are Crowds? Insights from Retail Orders and Stock Returns. The Journal of Finance 68, 1229–1265.

Kullback, S., Leibler, R. A., 1951. On Information and Sufficiency. https://doi.org/10.1214/aoms/1177729694 22, 79–86.

Kwan, A., Philip, R., Shkilko, A., 2024. The Conduits of Price Discovery: A Machine Learning Approach. SSRN Electronic Journal .

List, J. A., Gallet, C. A., 2001. What experimental protocol influence disparities between actual and hypothetical stated values? Environmental and Resource Economics 20, 241–254.

Lusardi, A., Mitchell, O. S., Washington, G., 2011. Financial Literacy around the World: An Overview .

Moss, A., 2022. How do brokerages' digital engagement practices affect retail investor information processing and trading? Working paper .

Odean, T., Barber, B., Klein, P., Leland, H., Lyons, R., Modest, D., Nofsinger, J., Poterba, J., Rubinstein, M., Ruud, P., Sansing, R., Thaler, R., Trueman, B., 1998. Are Investors Reluctant to Realize Their Losses? The Journal of Finance 53, 1775–1798.

Palan, S., Schitter, C., 2018. Prolific.ac—A subject pool for online experiments. Journal of Behavioral and Experimental Finance 17, 22–27.

Veldkamp, L., 2023. Valuing Data as an Asset. Review of Finance 27, 1545–1562.

Weber, M., Camerer, C. F., 1998. The disposition effect in securities trading: an experimental analysis. Journal of Economic Behavior & Organization 33, 167–184.

Welch, I., 2022. The Wisdom of the Robinhood Crowd. The Journal of Finance 77, 1489–1527.

Yelagin, E., 2024. Gamification of Stock Trading: Losers and Winners .

Online Appendix

March 1, 2025

A Financial literacy quiz

1. Suppose you had $100 in a savings account and the interest rate was 3% per year. After 4 years, how much do you think you would have in the account if you left the money to grow?
   (a) More than $112
   (b) Exactly $112
   (c) Less than $112
   (d) Don't know / Not sure
   (e) Prefer not to say
2. Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, with the money in this account, would you be able to buy
   (a) More than today
   (b) Exactly the same as today
   (c) Less than today
   (d) Don't know / Not sure
   (e) Prefer not to say
3. Do you think that the following statement is true or false? "Bonds are normally riskier than stocks."
   (a) True
   (b) False
   (c) Don't know / Not sure
   (d) Prefer not to say
4. 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.
   (a) True
   (b) False
   (c) Don't know / Not sure
   (d) Prefer not to say
5. Normally, which asset described below displays the highest fluctuations over time?
   (a) Savings accounts
   (b) Stocks
   (c) Bonds
   (d) Don't know / Not sure
   (e) Prefer not to say

6. When an investor spreads his money among different assets, does the risk of losing a lot of money:
   (a) Increase
   (b) Decrease
   (c) Stay the same
   (d) Don't know / Not sure
   (e) Prefer not to say
7. Considering a long time period (for example, 10 or 20 years), which asset described below normally gives the highest return?
   (a) Savings accounts
   (b) Stocks
   (c) Bonds
   (d) Don't know / Not sure
   (e) Prefer not to say
8. Do you think that the following statement is true or false? "If you were to invest $1,000 in a stock mutual fund, it would be possible to have less than $1,000 when you withdraw your money."
   (a) True
   (b) False
   (c) Don't know / Not sure
   (d) Prefer not to say
9. 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."
   (a) True
   (b) False
   (c) Don't know / Not sure
   (d) Prefer not to say
10. Which of the following statements is correct?
   (a) Once one invests in a mutual fund, one cannot withdraw the money in the first year
   (b) Mutual funds can invest in several assets, for example invest in both stocks and bonds
   (c) Mutual funds pay a guaranteed rate of return which depends on their past performance
   (d) None of the above
   (e) Don't know / Not sure
   (f) Prefer not to say
11. Which of the following statements is correct? If somebody buys a bond of firm B:

(a) She owns a part of firm B
(b) She has lent money to firm B
(c) She is liable for the company's debts
(d) None of the above
(e) Don't know / Not sure
(f) Prefer not to say

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?

    (a) Less than 5 years
    (b) Between 5 and 10 years
    (c) Between 10 and 15 years
    (d) Never
    (e) None of the above
    (f) Don't know / Not sure
    (g) Prefer not to say

B Consent form

Dear Participant,

You are invited to take part in an economic experiment, which is part of a project studying retail trading platforms funded by the Social Sciences and Humanities Research Council of Canada. You will be asked to trade assets on a virtual financial market.

Your participation in this study is entirely voluntary. Participation presents no known risks or burdens. You will receive a compensation for participating in the study, announced on the Prolific platform, as well as a monetary compensation proportional to the profits you earn both in the process of trading and for the completion of the non-trading task. You can withdraw from the study at any time without penalty. However, if you withdraw before the end of the experiment the compensation will be forfeit, as per Prolific's rules.

Please be assured that all data generated during this study will remain confidential. You will be asked to provide some basic demographic information about yourself, but none of this information will in any way be used to personally identify you and no personal information will appear in any published study.

Your decision to participate in this study will remain completely confidential and only the researchers will have access to any of the individual data. We do not collect any data that can personally identify you. Any demographics data will be stored electronically on a web-based server until it is downloaded for the purpose of analysis. This server will only be accessible by the research team using a secure password. Data will be retained in de-identifiable format for 7 years and discarded afterwards. At that time, electronic files will be deleted and written files will be shredded. This is a standard practice for the American Psychological Association. As required by the guidelines of American Economic Association, during the publication process the data might be shared with other investigators for re-analysis. However, to ensure security and confidentiality of the data files only details of the computations sufficient to permit replication might be shared.

This study has been reviewed by the Research Ethics Boards at the University of Calgary. If you are interested, you can receive a summary of the research study by emailing the primary researchers:

- Prof. Marius Zoican at marius.zoican@ucalgary.ca
- Prof. Mariana Khapko at mariana.khapko@utoronto.ca
- Dr. Philipp Chapkovski at filipp.chapkovskii@uni-due.de

The research study you are participating in may be reviewed for quality assurance to make sure that the required laws and guidelines are followed. If chosen, (a) representative(s) of the Human Research Ethics Program (HREP) may access study-related data and/or consent materials as part of the review. All information accessed by the HREP will be upheld to the same level of confidentiality that has been stated by the research team.

If you wish to obtain further information regarding your rights as a participant you may contact the Conjoint Faculties Research Ethics Board (CFREB) at University of Calgary at cfreb@ucalgary.ca or (403) 220-8640. You may also contact Prof. Marius Zoican (marius.zoican@ucalgary.ca), Mariana Khapko (mariana.khapko@utoronto.ca) and Dr. Philipp Chapkovski (filipp.chapkovskii@uni-due.de) for further information or to obtain answers to any questions about this research.

I agree to participate in research being conducted by Prof. Marius Zoican of the Haskayne School of Business at University of Calgary, Prof. Mariana Khapko of the Department of Management at University of Toronto Scarborough and Dr. Philipp Chapkovski of the University Duisburg-Essen. I understand that I can withdraw my consent to participate at any time and by giving consent I am not giving up any of my legal rights.

I have read and understood the information above.

C Instructions

Welcome to this trading experiment! Please read these instructions carefully. Your payment will be based on your performance in the trading game and your answers to a financial literacy quiz.

Overview You will trade a single stock over two rounds, with a training round at the start to help you get comfortable. Each round includes:

- 40 price updates ("trials"), each lasting 8 seconds
- Starting balance of E$500 in cash
- One unit of stock worth E$1000
- A short questionnaire after each round about your trading experience

Trading Rules

- You can own either 0 or 1 unit of stock at any time.
- You can only buy when you don't own the stock.
- You can only sell when you own the stock.
- You can borrow money (have negative cash), but this will reduce your final payoff.
- No trading is allowed during the first three price updates of each round.

How Stock Prices Move The stock price updates every 8 seconds and follows two possible states:

- **Good State (like economic expansion):**
  - 80% chance price goes up
  - 20% chance price goes down
- **Bad State (like recession)::**
  - 80% chance price goes down
  - 20% chance price goes up

When the price changes, it will move by either E$50, E$100, or E$150 (randomly chosen).

The state can change after each price update:

- 85% chance the state stays the same
- 15% chance it switches to the other state

The initial state is randomly chosen.

Market Data You will trade on two different platforms (in random order):

1. **Basic Platform:**
   - Shows only price information
2. **Data-Enhanced Platform:**
   - Shows price information
   - Shows other traders' buying and selling activity
   - These traders include:
     - Informed traders who know the current state (they buy in good states and sell in bad states)
     - Random traders who trade without any information
   - You'll know what percentage of traders are informed, but not which trades come from which type
   - The arrival rate of traders and the share of insiders is the same for each round.
   - The information about buying and selling activity clears after every price update.

Before and after trading, you'll be asked how much you would pay to access the market data.

Payment

- One round will be randomly selected for payment.
- Your earnings at the end of the experiment will be equal to the amount of cash you hold at the end of the randomly chosen payment round plus the end-of-round price of the stock if you own it.
- Your earnings will be converted from experimental dollars (E$) to real money at a rate of E$1 = CA$0.003.
- You will also earn CA$0.09 for each correct answer in the financial literacy quiz.
- You will receive a fixed payment of CA$15 per hour in addition to your performance-based earnings.

Important Notes

- Watch the market carefully to try to identify whether the stock is in a good or bad state.
- Your trading screen will show the current stock price, your cash balance, and your stock holdings.
- In rounds with market data, you'll also see recent buying and selling activity.
- Trading is paused briefly at certain points during a round to ask about your beliefs regarding price movements.

Training Before the main rounds, you'll complete a 10-trial practice round to familiarize yourself with the trading interface.

D Comprehension quiz

1. When you receive market data about other traders, which of the following is true?
    - You can always tell which trades come from informed traders.
    - You know the percentage of informed traders but not which trades are theirs.
    - All traders in the market are informed traders.
    - You can see future price movements from the trading data.
2. In the good state, what happens to stock prices?
    - Prices always go up.
    - Prices have a 50% chance of going up or down.
    - Prices have an 80% chance of going up and 20% chance of going down.
    - Prices have a 15% chance of going up.
3. The stock is more likely to be in the good state if:
    - There are more sell (at bid) than buy (at ask) orders.

- There are more buy (at ask) than sell (at bid) orders.
- There are more orders overall.
- There is no relationship between order flow and the state of the stock.

4. How do informed traders behave in the market?

    - They buy in good states and sell in bad states.
    - They buy in bad states and sell in good states.
    - They trade randomly like noise traders.
    - They never trade at all.

5. How does borrowing money affect your final payoff?

    - It has no effect.
    - You pay interest on borrowed amounts.
    - Any negative balance is subtracted from your end-of-round portfolio value.
    - Borrowing is not allowed.

6. What happens when noise traders place orders?

    - They always buy.
    - They always sell.
    - They buy or sell randomly with equal probability.
    - They follow informed traders' actions.

E Willingness-to-Pay for market data

Important Information. Research has shown that people often state they would pay more in surveys than they actually would in real situations. This is because it's easy to be generous when you're not actually spending your money. Before you answer the following questions, please imagine that you are actually paying real money from your own pocket.

Your answers to these questions will help inform the pricing of market data services. While this particular choice won't directly affect your payment today, similar decisions about market data access and pricing are regularly made based on research like this.

For each price below, indicate whether you would purchase access to market data:

Price	Choice	
E$25	Yes	No
E$50	Yes	No
E$100	Yes	No
E$150	Yes	No
E$200	Yes	No
E$250	Yes	No
E$300	Yes	No
E$400	Yes	No

How certain are you about your choices? Please choose a number from 0 (Very uncertain) to 10 (Very certain).

Please review your choices carefully. Remember that while these choices are hypothetical, they should reflect what you would actually be willing to pay for market data access.