Author/s:
Asparouhova, E;Bossaerts, P;Cai, X;Rotaru, K;Yadav, N;Yang, W

Title:
Humans in charge of trading robots: The first experiment

Date:
2024-07

Citation:
Asparouhova, E., Bossaerts, P., Cai, X., Rotaru, K., Yadav, N. & Yang, W. (2024). Humans in charge of trading robots: The first experiment. Review of Finance, 28 (4), pp.1215-1244. https://doi.org/10.1093/rof/rfae007.

Review of Finance, 2024, **00**, 1–30
https://doi.org/10.1093/rof/rfae007
Advance Access Publication Date: 15 March 2024
Article

Humans in charge of trading robots: the first experiment

Elena Asparouhova[1], Peter Bossaerts[2,*], Xiaoqin Cai[3], Kristian Rotaru[4], Nitin Yadav[5], Wenhao Yang[6]

[1]David Eccles School of Business, University of Utah, Salt Lake City, UT, United States
[2]Faculy of Economics, University of Cambridge, Cambridge, United Kingdom
[3]Gies College of Business, University of Illinois Urbana-Champaign, Champaign, IL, United States
[4]Monash Business School and Monash School of Psychological Sciences, Monash University, Melbourne, Australia
[5]Centre for Brain, Mind and Markets, Faculty of Business and Economics, University of Melbourne, Melbourne, Australia
[6]Belk College of Business, University of North Carolina Charlotte, Charlotte, NC, United States
*Corresponding author: Faculty of Economics, University of Cambridge, Austin Robinson Building, Sidgwick Avenue, Cambridge CB3 9DD, United Kingdom. E-mail: plb32@cam.ac.uk

Abstract

We present results from an experiment where participants have access to automated trading algorithms, which they may deploy at will while still trading manually. Treatments differ in whether robots must not be halted, deployment is compulsory, or robots can be halted and replaced at will. We hypothesize that robot trading would reduce mispricing, and that the effect would be more pronounced as commitment degree increases. Yet, compared to manual trading only, we observe equally large and frequent mispricing and, in early trading, significantly higher bid–ask spreads and more frequent flash crashes/price surges. Participants earn more, provided they combine robot and manual trading. Compared to evidence from archival data, we find significantly higher use of liquidity-taking robots. We attribute this to the inability, in the field, to identify the presence of liquidity takers when they happen not to trade.

Keywords: algorithmic trading; robot-human interaction; robot trading; financial market bubbles; experimental finance.
JEL classifications: G12, G41, C92.

> "I have found that it is seldom a good idea to manually override a model no matter how treacherous the market is looking."
>
> *Dr. E.P. Chan*

1. Introduction

We report results from a financial market experiment in a setting where mispricing of a multi-period security in the form of bubbles is frequently observed (Smith, Suchanek, and Williams 1988). For the first time, in a controlled setting, we provide participants with

Editor: Joel Peress
Received: May 24, 2020. **Accepted:** February 22, 2024

access to simple execution algorithms (trading robots) as a complement to manual trading. We study the impact of the ability to trade in an automated way on the emergence of bubbles.

An execution algorithm[1] is akin to a “strategy” in games: it commits the participants to an action profile. In behavioral game theory, requiring participants to commit beforehand to a strategy is known as the “strategy method” (Alan and Ertac 2015). Commitment forces participants to more carefully deliberate the game at hand: they have to think in advance about potential trading opportunities, without the ability to flexibly adjust strategy as events unfold. The extra care that is needed when committing to a robot therefore imposes discipline on trading. This comes at a cost, however: when the “wrong” robot is chosen, the participant may not be able to timely halt it. The strategy method thus precludes learning “on the job,” that is, as trading unfolds.

Because of the extra care required when choosing a robot, we hypothesize that overall pricing would improve in the presence of robots: the frequency of bubbles would be reduced, and prices would align better with fundamental values. However, when mispricing does emerge, one’s chosen robot may no longer be optimal, and will need to be replaced. If the costs of switching are too high, for instance, because robots cannot be switched off, participants may be deterred from using robots, and price quality may suffer. This leads to a more nuanced hypothesis, namely, that price quality depends on the degree to which robots need to be committed to.

We conduct a controlled experiment to test our hypothesis. We run several treatments distinguished by the degree of commitment to robot deployment, and compare the results with those from a control where trade is exclusively manual. In the first treatment, traders can always stop and change robots at will (No Commitment). In the second treatment, once a robot is deployed, it cannot be stopped (Commitment). The third treatment is like the second one, except that we impose a penalty if one does not deploy a robot within a certain time (Penalty). Because mispricing is generally reduced upon repetition of the same setting with the same cohort (Smith, Suchanek, and Williams 1988), we resort to a between-subjects approach, assigning an individual randomly to only one treatment.

In all treatments, every participant was aware of the potential presence of algorithmic traders. Thus, everyone was induced to carefully consider strategies beforehand, even those participants who opted not to use robots: they may end up trading with a robot. This reinforces our hypothesis, and leads us to expect that the mere potential of encountering algorithmic traders should lead to a reduction in the level of mispricing, regardless of whether robots are actually deployed. This effect has already been verified in an experimental setting before, but the setting was different: mispricing concerned simple pure-arbitrage opportunities, and robots were under the exclusive control of the experimenter (Farjam and Kirchkamp 2018). However, confirming the results of Leal and Hanaki (2018), Angerer, Neugebauer, and Shachat (2023) find the opposite: in their experiment, price quality was unaffected by the announcement of the potential presence of (experimenter-controlled) robots.[2] In our experiment, mispricing does not entail pure arbitrage opportunities but bubbles that, if traded against, expose the traders to risks. In addition, our robots are under control of the participants; the experimenter does not intervene in the marketplace.

[1] Execution algorithms are automated versions of the “program trading” used in the 1980s to arbitrage between options markets and underlying markets. Program trading is thought to be one of the reasons behind the 1987 stock market crash (Furbush 1989). Execution algorithms are to be differentiated from autonomous algorithms, which can adapt strategy online, using techniques such as reinforcement learning (Sutton and Barto 2018).

[2] Angerer, Neugebauer, and Shachat (2023) present a treatment with an announcement in the instructions that there is the possibility of an algorithmic trader participating; however, no algorithmic trader participates. The treatment is introduced as a test for a potential announcement effect as documented by Farjam and Kirchkamp (2018). The treatment and analysis can be found in Supplementary Appendix 1 of the article. Grossklags and Schmidt (2006) introduce an arbitrage algorithm without announcing it to the participants. In that treatment, mispricing is significantly greater than without the participation of the arbitrage trader.

The introduction of algorithmic trading in field markets is thought to have led to improved price quality. The claim is that trade prices have moved closer to fundamental values (Chordia, Roll, and Subrahmanyam 2011; Hendershott, Jones, and Menkveld 2011). However, determining the fundamental value of an asset in the field is challenging, so such evidence remains indirect, and may even be biased.[3] Historical analysis of field data has also established a link between algorithmic trading and liquidity. Bid–ask spreads have been shown to narrow, with corresponding reduction in autocorrelation in transaction price changes; see, for example, Carrion (2013), Brogaard, Hendershott, and Riordan (2014), and Angel, Harris, and Spatt (2011). Hendershott, Jones, and Menkveld (2011) and Boehmer, Fong, and Wu (2021) provide evidence for enhanced liquidity and informational efficiency, but argue that there has been an increase in short-term volatility. Finally, comparing price quality during part of the recent coronavirus disease 2019 (COVID-19) epidemic—when (manual) floor trading was prohibited—with that in normal times, Brogaard, Ringgenberg, and Roesch (2023) have argued that manual trading on the New York Stock Exchange actually contributes significantly to price quality.

Early studies disagree how much money algorithmic trading generates in the field (Kissell and Malamut 2005; Baron, Brogaard, and Kirilenko 2012). There are indications that profits, if any, have eroded gradually, in particular in the domain of High Frequency Trading (HFT; Serbera and Paumard 2016). In addition, HFT imposes negative externalities on other strategies (Goldstein, Kwan, and Philip 2023). Altogether, one is led to conclude that profits depend critically on the type of algorithm and on the presence and nature of other algorithms in the marketplace. This has been confirmed in historical investigation of field data. For instance, Baron, Brogaard, and Kirilenko (2012) find that liquidity-taking algorithms generate above-average profits, while Baron et al. (2019) document that profits depend on algorithm latency. One way to explain the abnormal profits of the liquidity-taking robots is to point to the claim that they are chosen far less frequently. Menkveld (2013) reports that only one out of five trades involving robots are initiated by "taker" robots. The superiority of taker robots reminds us of the outcome of the Santa Fe trading robot competition in the early 90s, where participants designed algorithms to compete against one another in an experimental setting—without manual intervention. There, the winning algorithm was also liquidity-taking. This algorithm has become known as the "Kaplan sniper" (Rust, Miller, and Palmer 1994; LeBaron, Arthur, and Palmer 1999). We hasten to add, however, that the Kaplan sniper did not trade frequently, while the aforementioned field studies tend to focus exclusively on HFT.

In experimental contexts like ours, we can compare inter-trader profits more comprehensively, going beyond HFTs, and eliminating biases caused by inherent limitations in historical data from the field. As to the latter, note that the presence of "taker" robots cannot be detected if, as is the case with data from markets in the field, only quote and trade data are available. Such robots will not submit orders and trade if the conditions that trigger their programmed action do not materialize. They may have been present, and participants in the marketplace may have been aware of their presence and may have adjusted strategies accordingly, but the researcher cannot possibly know because the robots did not act in the sample at hand. In our experiment, instead, we are able to track all robots that are deployed, regardless of whether they end up generating quotes or trades.

The study of robots in experimental markets is not entirely new; for an excellent survey, the reader is referred to Bao et al. (2022). However, except for this study and Aldrich and Vargas (2020), the experimenters were always in control of the robots, not the participants. Aldrich and Vargas (2020) investigate the choices of market-making and liquidity-taking robots in an environment of new information arrivals and competition for speed,

---

[3] To emphasize this point, we employ variance ratio tests, commonly used in these studies, on our experimental data to explore potential biases.

with the goal of comparing the performance of two different market institutions, rather than robot choice *per se*.

Our experimental setting is based on Smith, Suchanek, and Williams (1988) (SSW). There, bubbles are regularly observed and are robust to a number of interventions, including ability to short sell, access to futures markets, inability to re-sell, or having professionals as participants, but not robust to independent repetitions with the same cohort.[4] We conducted thirteen robotic sessions and four manual sessions with students from the University of Melbourne and Monash University. The first series of sessions, conducted in 2017, includes four of the robotic sessions with no commitment and four manual trading sessions.[5] The second series of sessions, conducted in 2022–2023, includes the other nine robotic sessions, with three sessions under each treatment. In Supplementary Appendix 2, we report on four additional pilot sessions conducted at the University of Utah, which were organized to test experiment design and trading software.

The main findings of the study are as follows. We find that, in all of the robotic sessions, bubbles are no smaller and no less frequent than in our manual-trading sessions, rejecting our conjecture that access to algorithmic trading reduces mispricing. Bid–ask spreads are higher with robots in early rounds, and the presence of robots causes more frequent flash crashes/surges in these rounds. Participants use robots intensely—from 2/3 to 4/5 of all transactions involve a robot. These numbers are not unlike those generated in advanced stock markets (De Luca and Cliff 2011). We discover that as much as 40percent of the robots deployed are of the "liquidity-taking" type. Average earnings of participants using robots are not different; however, those deploying robots while trading manually alongside earn more.

The remainder of the article is organized as follows. Section 2 introduces the specific hypotheses on which data analysis is based. Section 3 details the experimental design. Section 4 presents the results. Section 5 concludes the study.

2. Hypotheses

By associating robot choice with the "strategy method" in behavioral game theory, we obtain a framework with which to formalize the deployment of automated trading algorithms (robots) in financial markets. The framework highlights a balance between strategic foresight (advantage) and limited adaptability (disadvantage). While traders are compelled to premeditate their strategy choice (algorithm), the deployed robot lacks the flexibility to adapt swiftly when the chosen strategy is ineffective. In contrast, manual order submission allows traders to adapt in real-time, akin to traditional reinforcement learning in machine learning (Sutton and Barto 2018).

Our hypotheses span from market-level outcomes (mispricing, liquidity) to phenomena at the individual level (earnings disparities; robot selection). Thus, they are grouped into two categories: (1) Market-level Outcomes and (2) Individual Decisions on Robot Choice.

2.1 Market-level outcomes

Hypothesis A.1: Mispricing will be reduced when robots are available.

We hypothesize that the presence of robots will diminish mispricing as they encourage market participants to adopt the "strategy method." This method requires the participants to take a forward-looking perspective whereby they attempt to anticipate the evolution of future prices. This is difficult as many future pricing scenarios could be considered. Still,

---

[4] See King et al. (1993) for a study using corporate executives and professional traders, and Palan (2013) for a general survey.

[5] The first series also included a session intended to be with robots, so participants went through the corresponding instructions, but when there was a technical glitch, they ended up trading manually only. We do not include the session in our main analysis.

pricing at the fundamental value of the asset, which equals the sum of remaining expected dividends,[6] is both focal—it is given to the participants during the instruction period—and it minimizes losses in case the anticipation is incorrect. Consequently, it is expected that participants would deploy robots with algorithms that are based on the expectation that prices should remain close to fundamental values. Buy robots are utilized to exploit offers to sell below the fundamental values, while sell robots will take advantage of offers to buy above the fundamental values. If everyone thinks alike, prices will be forced to stay close to fundamental values, in a self-fulfilling equilibrium. Our prediction that the presence of robots will reduce mispricing is rendered even more plausible since, in simple games with strategic uncertainty, participants resort to risk minimization to deal with strategic uncertainty (Embrey, Fréchette, and Yuksel 2017). In our game, trading strategies that use the fundamental value as reference point are also risk-minimizing since even if someone deviates from it, payoffs from trading against the deviation always have positive expected value.

Hypothesis A.2: Mispricing will be lowest in the Commitment Treatment.

Deployment of strategies (algorithms) comes with advantages and challenges. As discussed in the Introduction, commitment to a strategy exposes one to repercussions of strategic missteps, as any errors become more financially consequential once a strategy is set in motion and can no longer be halted. Given the elevated risks associated with rigid commitment, it is reasonable to assume that participants will exercise greater caution and be more thoughtful when determining their algorithmic parameters. Hence, we surmise that the *Commitment Treatment will generate the lowest levels of mispricing*.

Our prediction appears less convincing in the *Penalty Treatment*. The urgency instigated by penalties for robot deployment postponement may undermine strategic decision quality. Hastily made choices, spurred by the aim to sidestep penalties, could inadvertently exacerbate mispricing. Time pressures combined with the urgency to swiftly deploy a robot might erode the quality of robot selection, a phenomenon echoed in Madan, Spetch, and Ludvig (2015).

Hypothesis A.3: With robots, a higher number of price crashes/surges are expected.

Theory predicts that strategy choices based on price expectations that deviate from the fundamental value can lead to excessive volatility (Biais, Foucault, and Moinas 2011). Flash crashes are one way excessive volatility is expressed, the most notorious being the flash crash of 2010 (Kirilenko et al. 2017). Flash crashes are thought to contribute to lower or fleeting liquidity (Golub, Keane, and Poon 2012). From historical analysis of field data, one may conclude that algorithmic trading has increased the incidence of flash crashes, though this conclusion is controversial; see, for example, Chaboud et al. (2014), Brogaard, Hendershott, and Riordan (2014), or Jarrow and Protter (2012).

Following Biais, Foucault, and Moinas (2011), our hypothesis is that *strategic miscoordination will cause excessive price volatility*. We would expect such miscoordination to mostly occur in the early rounds of our experiments, while the participants are still coordinating strategies. Accordingly, we conjecture that more flash crashes and price surges emerge in early rounds. It is worth noting that this effect might be somewhat tempered in the *Commitment Treatment*. The reason is that traders might exhibit a heightened sense of caution, preferring robots that capitalize on clear mispricing (buying below and selling above the fundamental values) rather than those that aggressively ride price volatility.

---

[6] Formally, the fundamental value should be the sum of the discounted future payoffs; in a short experimental session, the discount factor is 1.

Note that the joint hypothesis of less mispricing and more flash crashes/surges with robot use does not lead to a contradiction. The flash crashes/surges are expected to happen only in the early rounds. After a period of attempts to coordinate, we expect prices to be close to fundamental values.

Hypothesis A.4: Manual orders affect bubbles and crashes.

Drawing on the earlier discussion about strategic uncertainty and coordination challenges resulting from robot deployment, one recognizes that manual trading may play a distinct role: while commitment to robots necessitates a forward-looking perspective, and while ineffective robots cannot swiftly be disarmed and replaced, manual trades offer participants with real-time flexibility to adapt to changes. But adaptability may come at a cost. If driven by reactive and potentially myopic decisions, this excessive volatility could become a catalyst for bubbles and crashes. Otherwise, one can expect manual trading to improve pricing, that is, lead the market away from bubbles and crashes.

Thus, the direction in which manual orders could affect prices is an empirical question. Therefore, our hypothesis is that there will be an effect, but we refrain from specifying a direction: *bubbles and crashes will be affected by manual orders*.

Hypothesis A.5: With robots, bid–ask spreads will be higher in early periods.

As discussed in the Introduction, the evidence from historical data from the field favors the claim that algorithmic trading increases liquidity, and hence, reduces the bid–ask spread, although the effect varies with market conditions. However, there could be many confounding factors in those studies. One is that the emergence of algorithmic trading in the field has not been exogenous.[7] In addition, the reduction of bid–ask spreads was already a trend before the advent of algorithmic trading.[8]

There is a fundamental reason why robot trading may *not* lead to a lower bid–ask spread. There are two countervailing effects. For the first effect, the argument mirrors that for Hypothesis A.3. The choice of the right strategy requires participants to think hard about what strategies others will play: one with correct pricing, or with underpricing, or maybe overpricing? While learning the strategies of others, a process that likely happens in the first periods, more conservative strategies, those using robots with larger individual bid–ask spreads in the robot configuration, are likely to be used. Such strategies soften the competition among robots (and especially so when there is a penalty as even more conservative strategies might be chosen simply as a way to avoid the imposition of a penalty). The second, countervailing effect is that robot participation, through its commitment mechanism (of varying degrees across the treatments), exacerbates the rigidity of traders with different reference prices. This could result in lower market best bid–ask spread, as follows. Consider the benchmark where all traders use the same robots, and hence, the same bids and offers. Now consider the case of heterogeneous robots. Under heterogeneity, the minimum ask is likely to become lower than under homogeneity, and the highest bid is likely to be higher than its counterpart. It is an empirical question which of the effects dominates. However, in early periods of trading, we conjecture that the first effect would be stronger than in later periods. Hence, we hypothesize that bid–ask spreads will be higher in early periods.[9] The

---

[7] Therefore, the causal effect of algorithmic trading on liquidity has been analyzed using instrumental variable (IV) estimation. In Hendershott, Jones, and Menkveld (2011), for instance, autoquote data of an eight-month period is used as the IV for autotrading (AT) with the identification assumption that autoquote affects liquidity only via its effect on AT.

[8] The trend has been attributed to causes such as regulatory changes and increased competition in the brokerage industry, availability of real-time online trading, etc. (Chordia, Roll, and Subrahmanyam 2001), and not the advent of algorithmic trading.

[9] We thank an anonymous referee for the above observations.

increase is expected to be exacerbated when robot adjustment is impossible because commitment is required, in the second (Commitment) and the third (Penalty) treatments.

Hypothesis A.6: Commitment causes a decrease in trading volume.

Robot deployment in trading inherently implies a certain degree of commitment, as it limits the traders' real-time adaptability. When this limitation spans the entire trading period, as in the Commitment Treatment, we anticipate a more cautious approach to deploying robots. Such extended commitment naturally restricts traders from trying out different strategies, thus limiting the scope for experiential learning. *We hypothesize that this will lead to a reduction in trading volume*. The situation is ambiguous in the Penalty Treatment: excessive trading because of ill-chosen robots may cause losses, but these offset the penalty avoided because of robot deployment. Ill-chosen robots may lead to increased trading volume.

Hypothesis A.7: An equilibrium mix of robots emerges.

A priori, no distinct advantage is evident between taker and maker robots. Rather than any inherent superiority of one over the other, we therefore expect some equilibrium level of deployment of each. Specifically, an abundance of taker robots (liquidity-takers) creates an environment conducive for the deployment of maker robots (liquidity-suppliers), and the inverse holds true as well. *We hypothesize that robots will never be all of one kind; an increase of one type in a period would likely cause an increase of the other type in the subsequent period until an equilibrium is reached*.

Hypothesis A.8: Robot use causes higher wealth inequality.

By design, average earnings must be the same in all sessions except for randomness in realized dividends.[10] But there could be increased earnings inequality because of robot use. Here is why.

The availability of robots expands the range of strategies that participants can entertain. A foreseeable outcome is therefore that *wealth inequality increases when robots are available*, especially when robot utilization is compulsory, as in the *Penalty Treatment*. With penalties for non-deployment within a strict time limit, traders accustomed to on-the-fly adaptability are asked to strategize ahead of trading, a departure from their familiar approach, thereby putting them at a disadvantage, further potentially exacerbating inequality.

2.2 Individual decisions on robot choice

Hypothesis B.1: Unless robot use is mandatory, high/low use of robots in one period leads to either a reduction or an increase in use in the subsequent period.

Robots provide a distinct advantage: they can execute trades swiftly, allowing traders to center their attention on strategy formulation rather than execution. Yet, this advantage is not without limitations. While they excel in speed and efficiency, robots may limit the flexibility traders need to adapt to market changes. The more robots are deployed, the more advantageous it becomes to remain flexible, and hence, not to deploy robots. Vice versa if too few robots are deployed in a period, it will lead to an increase in the next.

Our hypothesis suggests that *as the prevalence of robot usage increases/decreases within a trading period, traders will reduce/increase their reliance on them in the subsequent*

---

[10] Note that the total earnings in sessions with robots (in experimental dollars) are exactly the same as with manual trading; only the conversion rates to legal tender (AUD) were different.

period. This pattern is expected to change only in situations where the use of robots is a mandated requirement, as in the Penalty Treatment.

Hypothesis B.2: Participants specialize in one type of robot.

Given the distinct nature of strategies associated with liquidity provision and liquidity taking, it is plausible to anticipate that participants might gravitate toward consistency in their robot selection. We thus expect specialization: *Instead of oscillating between robot types, participants are more likely to reveal a preference for one specific type throughout the trading sessions*.

Hypothesis B.3: Robot use declines over time.

Robot deployment entails commitment to a strategy, temporarily (in the No Commitment Treatment) or for one period (Commitment Treatment and Penalty Treatment). If a robot's strategy proves to be suboptimal, a new strategy—a new robot—has to be chosen. In contrast, manual trading offers more adaptability, allowing participants to tweak their strategies in real-time as they become more attuned to the unfolding market conditions. With this reasoning, we anticipate that *robot use will decline over time*.

Hypothesis B.4: Robot use declines when mispricing increases.

Our robots execute simple buy and/or sell strategies with respect to a predetermined reference price. Adjusting to mispricing events, such as navigating through a bubble, would necessitate the dynamic modification of this reference price based on updated expectations regarding the duration and magnitude of the mispricing. This could entail frequent switching between different robots. Against time-consuming switching of robots, manual trading offers a quicker approach to adapt one's strategy. Therefore, we hypothesize that *as mispricing becomes more pronounced* and flexible adaptation is called for, the strategic rigidity of robots becomes a liability, and hence *reliance on robots diminishes*.

Hypothesis B.5: Earnings will be higher for robot users.

Our experiments provide a unique opportunity to measure any differences in earnings, not only between manual and robot trading, but also with respect to the type of robot used.

The decision to deploy a robot demands meticulous analysis and strategy formulation. At the same time, manual trading affords the flexibility needed to adapt to unforeseen events such as offers outside the expected range of prices, and, in the case of treatments with commitment, offset the actions of one's robots.[11] While deployment of robots may lead the agent to more carefully consider the trading game, manual trading provides cognitive flexibility. When used in conjunction, robot and manual trading optimize "cognitive adaptability" (Glöckner, Hilbig, and Jekel 2014). We thus expect that *those who combine robot and manual trading will earn most*.

3. Experimental design

Out of a total of 137 participants, 33 were recruited from the University of Melbourne, through online website and on-site advertisements. The remaining 104 participants were recruited at Monash University. All sessions were in person, conducted between August 2017 and April 2023. The long hiatus between the sessions was caused by COVID-19

---

[11] Our trading software allows participants to cancel orders submitted by their own robot.

Table 1. Experimental sessions.

Listed are: Session ID (column 1), Session name (column 2), Location (column 3), Date (column 4), and Robot Navigation Characteristics (column 5).

(1) ID	(2) Session	(3) Location	(4) Date	(5) Can Stop/Penalty	(6) # Participants
1	Baseline 1	Monash	September 2017		8
2	Baseline 2	Monash	September 2017		8
3	Baseline 3	Monash	September 2017		8
4	Baseline 4	Monash	September 2017		8
5	No Commitment 1	Melbourne	August 2017	Y/N	8
6	No Commitment 2	Melbourne	September 2017	Y/N	8
7	No Commitment 3	Melbourne	September 2017	Y/N	8
8	No Commitment 4	Melbourne	October 2017	Y/N	9
9	No Commitment 5	Monash	October 2022	Y/N	8
10	No Commitment 6	Monash	October 2022	Y/N	8
11	No Commitment 7	Monash	April 2023	Y/N	8
12	Commitment 1	Monash	October 2022	N/N	8
13	Commitment 2	Monash	October 2022	N/N	8
14	Commitment 3	Monash	April 2023	N/N	8
15	Penalty 1	Monash	April 2023	N/Y	8
16	Penalty 2	Monash	April 2023	N/Y	8
17	Penalty 3	Monash	April 2023	N/Y	8

lockdowns and resulting restrictions on in-person experiments. The experiment was approved by the Human Research Ethics Committee at the University of Melbourne and Monash University. Rudimentary demographic information for each participant was collected, together with a written informed consent form, in accordance with ethics rules. In all sessions except one, eight subjects participated; because of a mis-communication, nine subjects were allowed into Session 8. Table 1 summarizes the sessions and the treatments. Below, we first describe the main setup and then the treatments and control.

3.1 Trading game

The design of the control experiment closely follows that of the original experiment by Smith, Suchanek, and Williams (1988). Eight participants are endowed with cash and a certain number of units of a security called Stock. The trading game lasts fifteen periods. At the end of each period, each unit of the Stock pays a random dividend, equal to 0, 0.25, 0.5, or 1.25 experimental dollars, equally likely.[12] As such, the expected dividend is \$0.50 per period, and the expected payoff of the Stock, or its fundamental value, is \$7.50 in period 1 (=15*\$0.5), and declines by \$0.50 in every period, respectively. At the end of the game, having paid all its dividends, the Stock expires worthless. In the 2017 sessions, the length of each trading period is between 2 and 5 min, longer in initial periods, and gradually declining to 2 min (minimum) in later periods. In the later series of sessions, all trading periods were set to 3 min.

Half of the participants have initial endowment of 20 units of the Stock and 100 experimental dollars, while the other half receive 12 units of the Stock and 160 experimental dollars. With the Stock evaluated at fundamental value, both types start with initial portfolios worth 250 experimental dollars. At the end of the trading game, after the Stock has paid its final dividends, participants hold only cash. The compensation from the experiment equals to the amount of this cash, converted to Australian dollars (AUD) at a pre-announced

---

[12] Immediately after the end of each trading period, an online random number generator (https://www.calculator.net/random-number-generator.html) is used to determine the dividend.

exchange rate.[13] Participants are also given an extra ten AUD as show-up fee, independent of performance. The actual remuneration varied depending on performance, but was bounded between 40 and 80 AUD for the treatments and between 15 and 55 AUD for the controls.

3.2 Robot Sessions (5–17)

In sessions with access to robots, the first half is dedicated to training, followed by a short break, after which the actual experimental sessions start, where participants are paid for performance. The training portion of each session consists of two parts. In the first part, participants familiarize themselves with manual trading on the online trading platform, Flex-E-Markets.[14] The second part involves learning how to deploy, (possibly) stop, and change the parameters of an algorithmic trader. Figure 1 displays the timeline of a typical session with access to robots.

In the first four sessions (Sessions 5–8), robots were provided via a management platform called Algohost (described at the end of this section). In the remaining nine sessions (Sessions 9–17), robots were integrated into the trading platform, Flex-E-Markets. (We elaborate on the trading platform later.) The assets traded during the training sessions are one-period, random-payoff assets, meant to teach the mechanics of trading, and not how to value a multi-period security.[15] The instructions presented Sessions 9–17 differ in one important respect: they are provided as a video rather than a handout. The content is essentially the same, but the video format ensures that delivery is harmonized across sessions.[16] The interested reader can find the video online at https://tinyurl.com/sswrobots. The video was projected on a big screen via the above link (with the appropriate modifications for each treatment). When participants had questions, those were answered by the experimenters present in the laboratory.

3.3 Control (manual) Sessions (1–4)

The trading environment in these four sessions is identical to the one in the robotic sessions, except that participants do not have access to robots. Both the training and the actual trading in those sessions take less time due to the simpler environment, and the exchange rate used to calculate take-home earnings was adjusted upward, as discussed before.[17] Instructions for the manual sessions were delivered in paper format. These instructions are an abbreviated version of those for the Robot Sessions 5–9 (see Supplementary Appendix 5), with the sections pertaining to robot use deleted.

---

[13] In some of the sessions, the conversion was 5:1, while for the others it was 10:1 (manual sessions and later robot sessions where the deployment of robots was faster). The reason for different conversion rates is the different expected lengths of the respective sessions. The exchange rates are designed in a way to provide similar hourly average payments across treatments.

[14] Flex-E-Markets is a SaaS platform where users can flexibly design marketplaces with multiple simultaneous public and private markets using the continuous open-book mechanism, and invite participants to trade directly through an online manual user interface or indirectly through robots that exploit the API. Developed by Quantahm, Salt Lake City, UT (USA). See quantahm.com.

[15] An additional session was initially intended to be conducted with robot trading, but due to unforeseen technical issues, the robots were not functioning properly after the completion of the training period. As a result, the participants completed the trading game with manual trading only. The results of this session are excluded from the analysis, but we report them in Supplementary Appendix 1, where the session is referred to as "Session M." In Supplementary Appendix 2, we also report on four pilot sessions conducted in 2014 at the University of Utah. The results are qualitatively the same, although the early software did not have the ability to record whether an order was submitted manually or through a robot. The latter could only be deduced from the frequency of the submitted orders. The development of the robot management platform Algohost allowed for the differentiation of robotic and manual orders. In the sessions run after 2017, this functionality was built into Flex-E-Markets.

[16] Instructions in paper format are usually read aloud and commented on by the experimenter, causing slight deviations in delivery from session to session.

[17] We only conducted four control sessions because the results from them are well established and ours are no different. More examples of the control sessions can be found in Bossaerts et al. (2023).

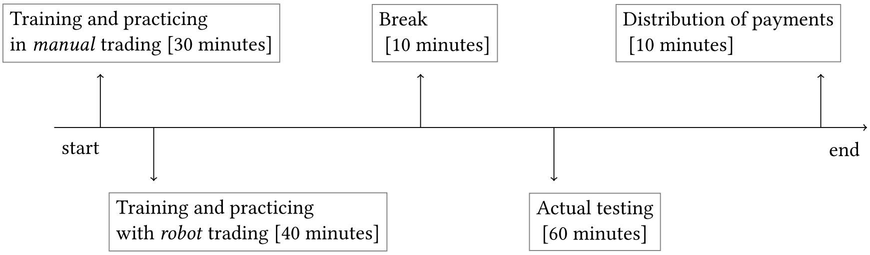


Figure 1. Timeline of a typical robot session.

3.4 Manual trading and robot deploying mechanism

The online marketplace is organized as a continuous, double-sided open book system, like most of the real-world electronic financial markets.[18] Only limit orders can be used, and strict time/price priority is adhered to. Order submission and trade reporting are anonymous. Among others, this means that participants cannot know whether their counterparty is a human trader or a robot. An illustration of the user interface used in Sessions 9–17 is provided in Figure 2.

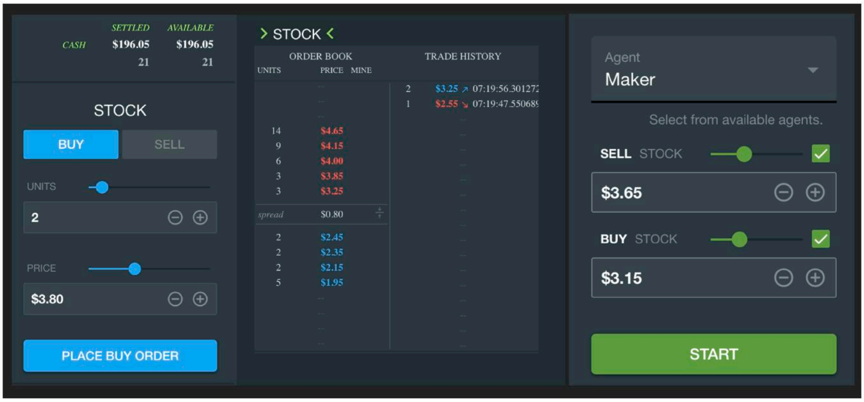
Robot choice and deployment are integrated into the user interface (see left panel in Figure 2). Participants choose the parameters of a robot (function: Maker/Taker; role: buyer/seller/both buyer and seller). They then launch the robot by clicking on "Start." They can follow their robot's actions in the book in the middle of the user interface. Robots are halted by clicking on "STOP" in the robot interface panel (which replaces the "START" button once a robot is launched). The following robots could be launched:

- *Maker robots*: Once the option "Maker" is selected, each participant can choose if they want to take a "buy" or "sell" side, or both. Consequently, a bid and/or an ask price must be selected with the condition that the ask exceeds the bid. Once started, the maker robot monitors its own orders and at all time maintains one buy/sell order in the book, depending on the selection being made regarding buy and/or sell. The trading platform stops accepting robot orders once the robot owner runs out of cash (for buy orders) or securities (for sell offers). The same rule applies to manually submitted orders.
- *Taker robots*: Like for the Maker robot, once Taker is selected, the participant has to select side and price. Once started, the Taker robot monitors the book, and when an order arrives that meets the robot's price condition, the robot immediately sends a limit order against it. If somehow the limit order does not trade (e.g., because someone else scoops up the profitable order), the order is added to the book. As for the Maker robots, the trading platform stops accepting Taker robot orders once the robot owner runs out of cash or securities.

In Sessions 5–8, a separate interface for robot choice and deployment was used. See Supplementary Appendix 3 for details. Here, we provide a brief description. Robots were uploaded to the "Algohost" server from the participant's computer desktop through a web browser window (see https://algohost.bmmlab.org/). On their computer desktop, participants were given a folder with python scripts. They had to choose between these scripts. Each script represented a robot with a particular function (Maker/Taker), role

---

[18] There is one substantive difference with field markets, namely: the system accepts orders only if they satisfy the submitter's budget constraint. For example, if a participant held $10 in cash, and submits one bid for $6 that ends in the book as a standing bid, no more bids above $4 could be made until the first one either executes or is canceled.



	SETTLED	AVAILABLE
CASH	$196.05	$196.05
	21	21

UNITS: 2

PRICE: $3.80

UNITS	PRICE	MINE
14	$4.65	
9	$4.15	
6	$4.00	
3	$3.85	
3	$3.25	
spread	$0.80	
2	$2.45	
2	$2.35	
2	$2.15	
5	$1.95	

2	$3.25 ↗	07:19:56.30127
1	$2.55 ↘	07:19:47.55068

Agent: Maker

Select from available agents.

SELL STOCK ☑ $3.65

BUY STOCK ☑ $3.15

Figure 2. Flex-E-markets user interface. The book of orders (red: asks; blue: bids) and trades (red: seller initiated; blue: buyer initiated) is in the middle. To the left is the order form for manual trading, and a list of holdings (cash, stock; "available" equals current ("settled") holdings minus commitments implicit in standing bids (cash) or standing asks (stock)). To the right is the algorithmic trader control panel, allowing the user to configure a robot, launch and halt, and replace it. Manual trading can continue even if the user has launched a robot.

(Buyer, Seller, Two-Sided), and reference value. Once uploaded to Algohost, the participant could launch, pause or stop, and replace the robot script. The Algohost server maintained the interface between the chosen python script and Flex-E-Markets without the participant having to know Python or understand Application Programming Interface (API).

Supplementary Appendix 3 provides more details on the design, interface, and instructions for Sessions 5–8. Instructions for the Manual (Control) Sessions were an abbreviated form of those for Sessions 5–8, as mentioned before.

4. Results

4.1 Market-level outcomes: overall mispricing and market quality

Hypothesis A.1: Mispricing will be reduced when robots are available.

Figure 3 provides an illustration of the average mispricing across the fifteen periods for all treatments (including control). Each of the dots represents the average mispricing for all replications (sessions) in a period/treatment. Given $N(p,s)$ trades in a given period $p = 1,\dots,15$, of a session $s = 1,\dots,S$, average mispricing is computed as follows:

$$Avg.Mispricing_{p,s} = \frac{\sum_{n=1}^{N(p,s)} (P_n - FV_p)}{N(p,s)}. \qquad (1)$$

The per-treatment measure of a given period's average mispricing is then the simple average across the $S(T)$ session averages for the treatment T:

$$Avg.Mispricing_{p,T} = \sum_{s=1}^{S(T)} \frac{Avg.Mispricing_{p,s}}{S(T)}.$$

It is evident in Figure 3 that quintessential SSW bubble behavior is observed in all treatments: Prices generally start below fundamental values, reach the fundamental values

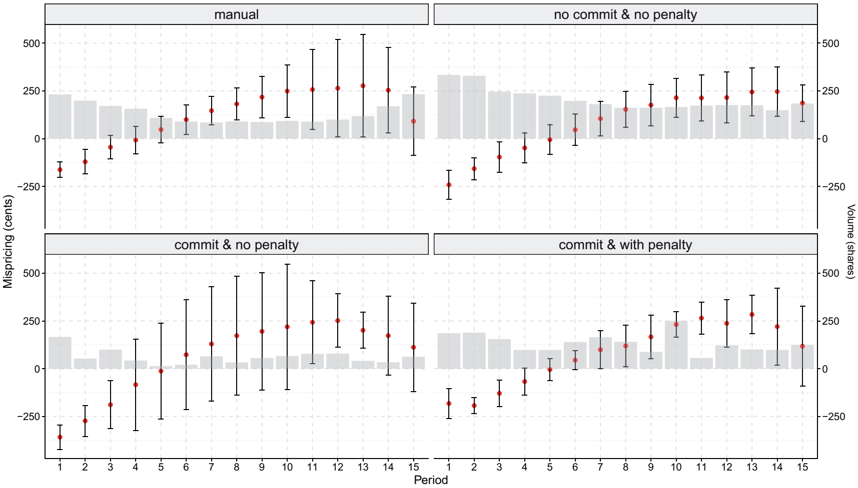


Figure 3. Mispricing. The red dots indicate the level of mispricing measured as the deviations from fundamental values (in cents). The gray bars are trading volume at the corresponding price level (in shares). The error bars are based on two times standard errors range computed across sessions for a given period.

around period 5, followed by a period of severe overpricing for most of the remaining periods, peaking around period 10; correction ensues by the last period. The error bars identify the locus of average mispricing for sessions when this average is within two standard deviations from the mean, while the bars indicate the volume of trade.

While there is substantial heterogeneity across sessions, the primary difference between robotic and manual sessions is that more severe under-pricing occurs in the early periods under the robotic treatment. No further salient differences emerge in later periods.

Table 2 presents the treatment-level variables for average mispricing, absolute mispricing, as well as another widely used measure of mispricing, namely the Absolute Mispricing Ratio; see Palan (2013). The formulas corresponding to Equation (1) for absolute mispricing and the mispricing ratio are as follows:

$$Abs.\,Mispricing_{p,s} = \frac{\sum_{n=1}^{N(p,s)} |P_n - FV_p|}{N(p,s)}$$

$$Abs.\,Mispricing\,Ratio_{p,s} = \frac{Abs.\,Mispricing_{p,s}}{FV_p}$$

Between-treatment comparison of the measure in Table 2 fails to reveal statistical significance (all P levels are above the 10 percent cutoff and thus not reported in the table). Table 3 presents the regression results of the three different mispricing measures on the treatment indicators using the Manual treatment as the baseline, allowing for period fixed effects. Confirming the results from Table 2, Table 3 demonstrates that only average mispricing is marginally impacted, and only in two of the three treatments. However, the average mispricing measure simply reflects that there is higher under-pricing in early periods in robotic sessions than in manual ones. Once absolute mispricing is considered, no mispricing differences can be discerned between manual and any of the robotic treatments. The conclusion is that Hypothesis A.1 fails to hold: mispricing is not reduced in the robotic treatments.

Table 2. Descriptive statistics on mispricing.

Each variable listed in the first column is computed as the average across session-periods for the control (Manual) and the three robotic treatments (No Commitment, Commitment, and Penalty). None of the pairwise differences across treatments is significant at $P = 0.10$.

Variable	Manual	No commitment	Commitment	Penalty
Average mispricing	116.7	83.5	56.9	80.6
Absolute mispricing	179.0	172.1	208.6	160.7
Absolute mispricing ratio	0.9	0.9	0.8	0.8

Table 3. Mispricing level across treatments.

Shown are regressions of per-period average mispricing measures on treatment dummies, allowing for period fixed effects. Robust standard errors are reported in parentheses.

	Dependent variable:		
	Avg mispricing (1)	Abs mispricing (2)	Abs mispricing ratio (3)
No commitment	−33.175	−6.891	−0.005
	(21.54)	(18.24)	(0.16)
Commitment and no penalty	−59.737*	29.593	−0.053
	(31.39)	(27.20)	(0.20)
Commitment with penalty	−36.036*	−18.316	−0.109
	(21.61)	(18.89)	(0.19)
Period fixed effects	✓	✓	✓
Observations	255	255	255
Adjusted R^2	0.573	0.188	0.435

Note: $^*P < 0.1$.

In studies of archival data from field markets, mispricing is often studied using variance ratio tests, absent a clearly defined fundamental value of the asset at hand (Litzenberger, Castura, and Gorelick 2012). Since in our experiment, the fundamental value is a design feature, and known a priori, we can verify whether the variance ratio test leads to correct inference about mispricing. The analysis can be found in Supplementary Appendix 4. There, we report that, when applied to our data, variance ratio tests lead to incorrect inference.

Hypothesis A.2: Mispricing will be lowest in the Commitment Treatment.

The results from Table 2 do not account for period fixed effects, while the large variability in average pricing in Figure 3 suggests that such effects may be present. Table 3 shows that there is a marginally significant reduction in average mispricing ($0.05 \leq P < 0.10$) when robots are deployed with commitment, regardless of the presence of a penalty for non-deployment. However, the results are statistically insignificant when “absolute mispricing” and “absolute mispricing ratio” are considered. Overall, our evidence is not supportive (or mixed at best) of Hypothesis A.2.

Hypothesis A.3: With robots, a higher number of price crashes/surges are expected.

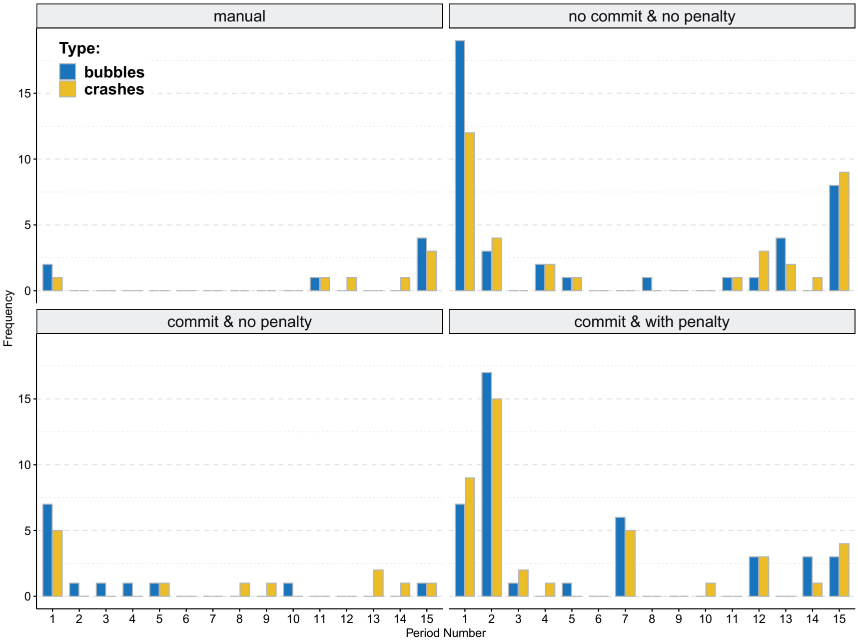


Figure 4. Frequency (Total Number) of flash surges and crashes, per period and treatment. Flash surges (respectively, crashes) are defined as within-period transaction price increases (respectively, decreases) of more than two standard deviations where the latter is computed from transaction price changes within a period.

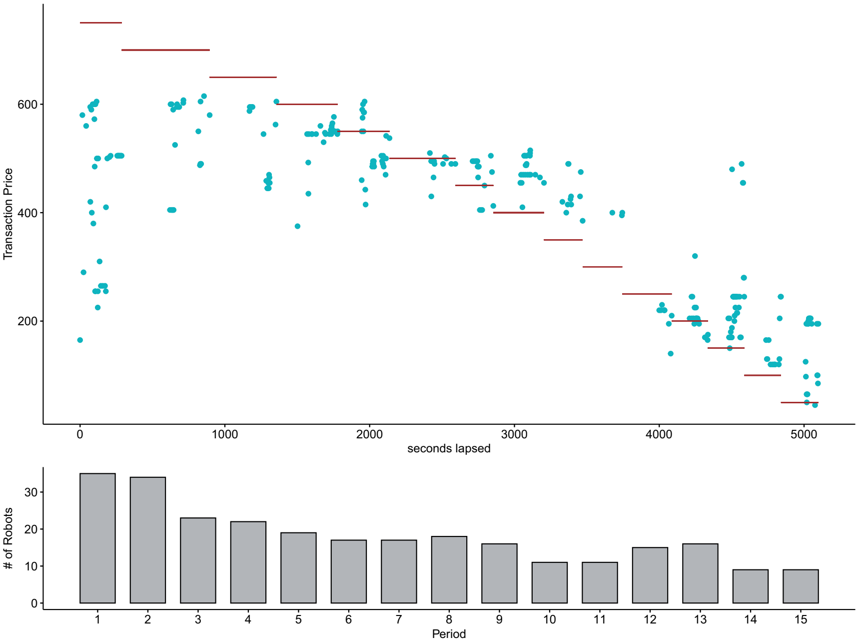


Figure 5. Transaction prices (dots), fundamental values (red line segments) and number of robots launches (bars; per period) in the first robot session of the no commitment treatment.

Flash crashes/surges are defined as outliers in log price changes (returns) that were more than 2 Sds away from the sample mean for the entire session. The cutoff of two standard deviations is inspired by analyses of bubbles in the literature. Siegel (2003) proposes an operational definition of a bubble "as any time the realized asset return over given future period is more than two standard deviations from the expected return." In the context of the SSW experimental design, Noussair, Robin, and Ruffieux (2001) used an analogous definition: bubbles emerge when "the average price is at least two standard deviations greater than the fundamental value for five consecutive periods." Here, we study sudden anomalous transaction price movements within a single period, distinguishing between price increases (flash surges) and price decreases (flash crashes).

Across treatments (including control), Figure 4 charts the number of flash surges and crashes, stratified per period. The average number of flash crashes and surges in robot sessions is a multiple of that in manual sessions, consistent with Hypothesis A.3.

Robot sessions thus stand out in that they generated many flash crashes/surges in the first period. Figure 5 illustrates that these can be attributed to miscoordination, as we argued when formulating Hypothesis A.3. The figure shows, for the first robot session in the No Commitment Treatment: (1) the evolution of transaction prices (dots) and fundamental values (red line segments) (Top); (2) Total number of robots launched each period (Bottom). At thirty-five, the number of robots started was highest in the first trading period. The number dropped subsequently, to settle between ten and fifteen after period five. Hence, in early periods, many robots were started, but stopped very quickly, and replaced. In fact, only five robots were active at any point in time. The robot switches caused high price volatility, as shown in the Top panel (transaction price evolution).

Because large overpricing is common in the SSW setting, one can expect flash crashes in later periods, coinciding with the bursting of the bubbles—which tends to be sudden. Indeed, Figure 4 shows that, in manual and robot sessions alike, many flash crashes/surges occur after the tenth period. Interestingly, bubbles do not simply disappear after a single flash crash: prices at times surge again, leading to subsequent flash crashes. That is, when a bubble bursts, crashes and surges often go together. The evidence suggests that flash crashes and surges that accompany the bursting of bubbles are not unique to robot sessions.

Figure 6 refines this evidence in the form of the following regression:

$$\#\,\text{of Bubbles or Crashes} = \sum_i \beta_i \text{Treatment}_i + \sum_{k \neq 7} \gamma_k \text{Period}_k + \varepsilon_i, \tag{2}$$

where Treatment_i is a dummy variable for the treatment i (the Manual control is thus used as baseline), and Period_k is a period dummy variable referring to the k-th Period (using Period 7 as the baseline). We separately analyze price surges (bubbles) and drops (crashes). Vertical line segments cover 95% confidence intervals.

Treatment "Penalty" displays a statistically significantly higher number of flash bubbles and crashes than the manual treatment. Regardless of treatment, we observe significantly higher numbers of bubbles and crashes in the first two periods and in the last one period. The formal evidence from Figure 6 forces us to qualify our conclusion: the availability of algorithmic trading *per se* does not lead to significantly higher numbers of flash events; instead, the compulsory use of robots does.

Hypothesis A.4: Manual orders affect bubbles and crashes.

Table 4 examines the role of manual orders in the formation of bubbles and crashes. Since the data tag each order as submitted manually or by a robot, we use a logistic regression to examine the likelihood of a flash bubble/crash as a function of the presence of a manual order in the trade. The regression includes treatments as confounding factors, as

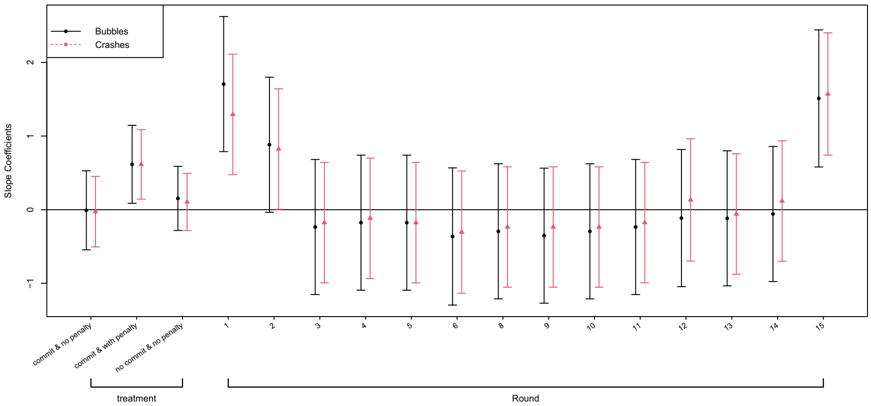


Figure 6. Effects of treatment and period on number of flash bubbles/crashes. This figure plots the slope coefficients in regression 2 (dots). Vertical line segments cover 95% confidence intervals. Baseline for Treatment: Manual Control; Baseline for Round: Period 7.

well as period fixed effects. Panel *A* of Table 4 indicates that the propensity of crashes is decreased significantly ($P< 0.001$) when manual orders are involved, but those orders have no effect on bubbles. Panel *B* of Table 4 zooms in on the types of manual orders, distinguishing between "large" and "small" and between "sitting" and "coming." In the first case, we categorize a manual order as large if it exceeds the size of the transaction. For example, if a single-unit transaction involved a manual order for more than a unit, we would categorize it as large. All other orders are deemed "small." The second categorization is based on whether the order in the transaction was standing ("sitting") or came in second and hence caused the trade (marketable orders or "coming orders"). The results in Panel *B* show that manual orders have an attenuating effect on crashes, mostly through large coming and through sitting orders. Small coming manual orders, instead, contribute to bubble formation.

Consequently, manual trading does affect the propensity of bubbles and crashes, but differentially so for large and small orders and for sitting and coming orders.

Hypothesis A.5: With robots, bid–ask spreads will be higher in early periods.

We here report the quoted bid–ask spread, defined as follows.[19] Every second, we identify the (market) best bid and ask and compute the spread as the scaled difference of the best ask (the lowest outstanding limit sell order price in the book) and best bid (the highest outstanding limit buy order price in the book):

$$Spread = \frac{Best\ Ask - Best\ Bid}{0.5(Best\ Ask + Best\ Bid)}. \quad (3)$$

We then average *Spread* over the last five seconds of a period (round) to obtain an estimate of the liquidity at the end of a period. Figure 7 plots the average end-of-period spread period-by-period across different session types pooling all treatments, for the (manual) control, and for the treatments separately. The left panel separates the spreads between manual sessions and robotic sessions. When pooling the treatments (robot sessions), the spread is

---

[19] A previous version of the article used the effective bid–ask spread as measured from the autocovariance of consecutive transaction price changes (Roll 1984). This analysis produces qualitatively indistinguishable results; it can be replicated with the programs and data posted in the Github repository bmmlab/RobotsSSW.

Table 4. Effect of manual orders on the likelihood of flash bubbles/crashes.

This table estimates the likelihood of a flash bubble or crash, as a function of the presence of manual orders in the transactions, using logistic regression. In Panel *A* the variable "Manual Order in Trade" is a dummy variable that equals 1 if the flash event involves a manual order. The remaining variables are treatment dummies. Period fixed effects are present in all estimations. In Panel *B* manual orders are further divided into "small" and "large," and "sitting" and "coming." Sitting order is a limit order, coming order is a marketable limit order, large order is one that is larger than the size of the transaction, the rest of the orders are small.

Panel A: Effect of manual orders

	Bubble (1)	Crash (2)	Bubble or Crash (3)
Manual order (in a trade)	0.0325	−1.060***	−0.5296*
	(0.4777)	(0.3461)	(0.3066)
No commitment	−0.3770	−0.2698	−0.3426
	(0.2554)	(0.3785)	(0.2401)
Penalty	0.0240	0.1076	0.0594
	(0.4205)	(0.2165)	(0.2975)
Observations	2,968	3,137	3,137
Pseudo R^2	0.08721	0.08772	0.10159

Panel B: Effect of large and small manual orders

	Bubble (1)	Crash (2)	Bubble or Crash (3)
Manual × Large Coming	−0.3272	−1.028*	−0.6519*
	(0.4222)	(0.5258)	(0.3793)
Manual × Small Coming	0.7248***	0.1011	0.4538*
	(0.2344)	(0.3241)	(0.2499)
Manual × Sitting	0.0203	−0.5095*	−0.2383
	(0.3184)	(0.2625)	(0.1571)
No Commitment	−0.2536	−0.2069	−0.2478
	(0.2574)	(0.3834)	(0.2452)
Penalty	0.1860	0.1848	0.1912
	(0.4636)	(0.1943)	(0.3152)
Round fixed effects	✓	✓	✓
Observations	2,968	3,137	3,137
Pseudo R^2	0.10029	0.08253	0.10844

distinctly larger up to period 10; this difference is significant at $P = 0.01$ for all periods (two-sample t test). This confirms Hypothesis A.5. When comparing the end-of-period average spread across treatments, however, no distinct pattern emerges.

Hypothesis A.6: Commitment causes a decrease in trading volume.

Total order submissions exhibit considerable similarity between sessions involving robots and manual-only sessions, except for the session with commitment, consistent with the hypothesis; see Table 5, row 1. But the drop in total order submission is insignificant ($P > 0.10$). Notice also that the number of robot activations, robot-initiated orders, and trades is also lowest in the Commitment Treatment (rows 2, 3, and 11).

Hypothesis A.7: An equilibrium mix of robots emerges.

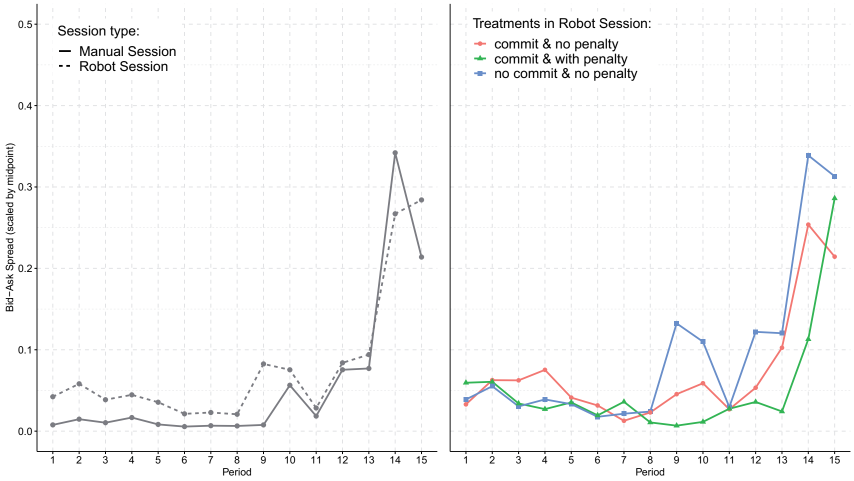


Figure 7. Bid–ask spreads computed based on the average quoted bid–ask spread over the last 5 s of each period, scaled by the spread midpoint. Left panel: Averages of manual sessions (solid line) and averages when pooling all robot sessions (dashed line). Right panel: Averages for each (robot) treatment separately.

Table 5. Descriptive statistics of robot deployment.

Shown are averages across session-periods for the control (Manual) and the three robot treatments (No Commitment, Commitment, Penalty). Robot types (Maker, Taker; Buy, Sell, Two-Sided) do not always add to Total Robot Activations because of rounding error.

		Treatment			
		Manual	**No commitment**	**Commitment**	**Penalty**
Orders:	Total	162.2	126.9	92.9	171.0
	Robots	–	28.8	22.1	66.3
Activations:	Total	–	11.2	4.7	8.0
	Makers	–	7.4	2.6	4.0
	Takers	–	3.8	2.2	4.0
	Buyers	–	3.2	0.9	1.1
	Sellers	–	4.1	1.2	1.5
	Two-Sided	–	3.8	2.6	5.3
Halts:	Total	–	9.4	–	–
	Switches	–	1.0	–	–
Trades		33.7	29.5	20.3	44.7

As already discussed in the discussion of Table 5, there is substantial heterogeneity in participants' selection of robots. On the whole, market-making robots and robots that function on both buy and sell side are most frequently employed. Notably, the usage of liquidity-taking robots is relatively high, at 40 percent of all robot activations. We emphasize, again, that the impact of taker robots cannot be assessed with a dataset that retains only orders and trades; such datasets are typical in the analysis of historical data from the field: they lack data on deployment, and presence is inferred only indirectly, when the taker submits a market order. This leads to biased inference and misdirected out-of-sample forecasts. Here, we track deployment, regardless of order submission, and discover that the use of Taker robots can be rather high: about 1/3 in the No Commitment Treatment, and 1/2 in the Commitment and Penalty Treatments.

Table 6. Predicting robot choices.

Shown are results from regressions where the dependent variable is related to individual robot choices, and where the independent variables are constructed from market measures of the prior period. The first two and the fifth models are estimated using data from all robotic sessions (subject to data availability, for example, we can only obtain the "Robot midpoint" prices for two-sided robots). The third and fourth models are estimated using the "No Commitment" sessions only, in which subjects can stop and switch robots. Regressor definitions can be found in the discussion of Table 5. We include session fixed effects in all models.

	Takers (Number) (1)	Robot Bid–Ask Midpoint (2)	Activations (Number) (3)	Stops (Number) (4)	Switches (Number) (5)
Absolute Mispricing	0.3647***	0.0544	0.1457	0.6551*	0.1940
Ratio	(0.1228)	(0.1433)	(0.1484)	(0.3317)	(0.1247)
Trading Volume	0.0112	0.0141	0.0042	−0.0131	0.0113
	(0.0072)	(0.0105)	(0.0137)	(0.0478)	(0.0192)
Price Volatility	0.0029	−0.0129***	0.0012	0.0068	0.0089
	(0.0041)	(0.0043)	(0.0048)	(0.0134)	(0.0055)
Number of Robots	0.1814**	0.0300	0.5923***	0.4182***	0.1052***
Activated	(0.0725)	(0.0280)	(0.0668)	(0.0902)	(0.0313)
Number of Orders	−0.0034	−0.0060	−0.0017	0.0143	−0.0008
Submitted	(0.0026)	(0.0037)	(0.0046)	(0.0157)	(0.0066)
Trading Period	−0.1059***	−0.1557***	−0.1046**	−0.3062**	−0.0703*
	(0.0292)	(0.0421)	(0.0488)	(0.1271)	(0.0404)
Observations	236	166	236	98	98
Within R^2	0.27646	0.32228	0.60735	0.65702	0.48815

Note: $^{*}P<.1$. $^{**}P<.05$. $^{***}P<.01$.

Another way to test this is by regressing the number of deployments of one type of robot (say, Taker), on the number of robot activations in the prior period. This is done in Table 6. The left-most column of the table shows that the number of Taker robots in a period increases significantly when the total number of robots activated in the prior period increases. One can interpret this as follows: if the overall number of activated robots increases, all categories need to follow suit; among others, Takers. This is indeed the case: the subsequent period, there are more Takers.

From the same regression, however, we can see that, if absolute mispricing increases in a period, the number of Takers are also increased in the subsequent period. Participants react to mispricing by choosing robots that do not reveal their intentions through submission of standing orders (Maker robots); they prefer Takers, whose presence will not be revealed until a profitable counter-offer enters the book.

We have already commented on how the most significant predictor of robot trading halts is the number of robots activated in the prior round (third column): there is mean reversion in the number of robots activated. Putting the results of the left-most and third regression together, we notice that robot choice converges to some kind of equilibrium in terms of number of robots rather than the kind, contrary to the hypothesis: the period after an increase in robots, the number of robots decreases, but the number of Taker robots increases.

Hypothesis A.8: Robot use causes higher wealth inequality.

A metric of statistical dispersion, the Gini coefficient (Gini 1936), is used here to measure the degree of inequality in the final distribution of wealth among participants. The average Gini coefficient (standard deviation in parentheses) for robot sessions is 0.163 (0.065),

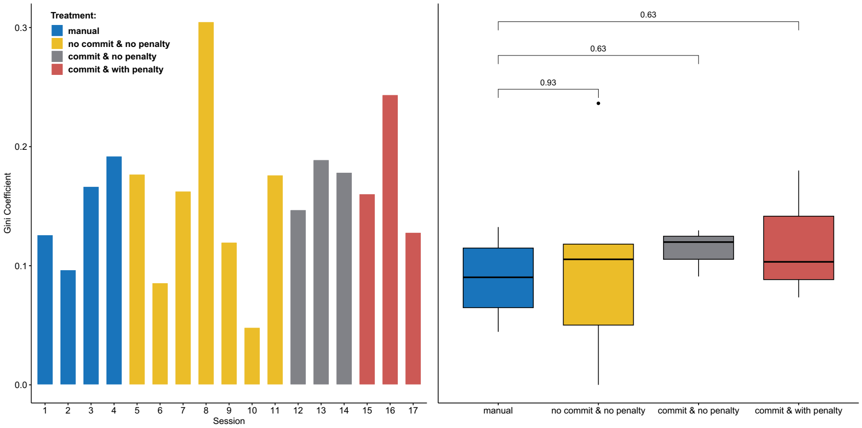


Figure 8. Left panel: Gini coefficients per session. Right panel: boxplots of Gini coefficients per treatment (including manual control); numbers on top are *P*-values based on Wilcoxon rank-sum test of differences in average Gini coefficients across pairs of treatments.

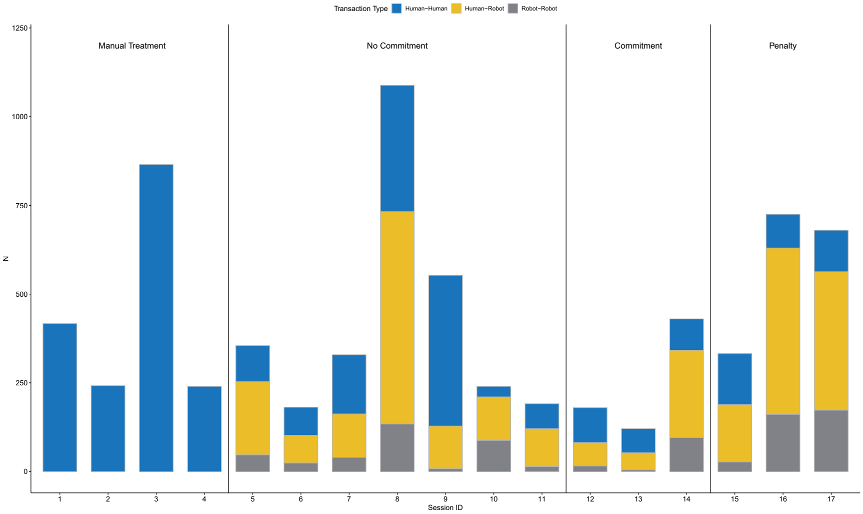


Figure 9. Robot involvement in trades. The identity of the two parties to transactions is charted here for the thirteen robot sessions (Sessions 5–17). (a) Robot deployment by role. (b) Robot deployment by function.

slightly, but insignificantly ($P > 0.10$) higher than its value of 0.145 (0.042) for the manual (control) sessions. Disaggregated data per session and per treatment (including control) are displayed in Figure 8. Overall, wealth inequality appears not to change with usage of robots whether purely voluntarily, or constrained by commitment or penalty.

4.2 Individual decisions on robot choice

We observe ample interaction between humans and robots in the thirteen robot sessions. Figure 9 shows that the percentage of trades involving a robot on one side and a human on the other side ranges from 22 to 65 percent, with an average of 48 percent across all sessions. Interestingly, these numbers resemble the situation in major European and US equity

Table 7. Robot activation.

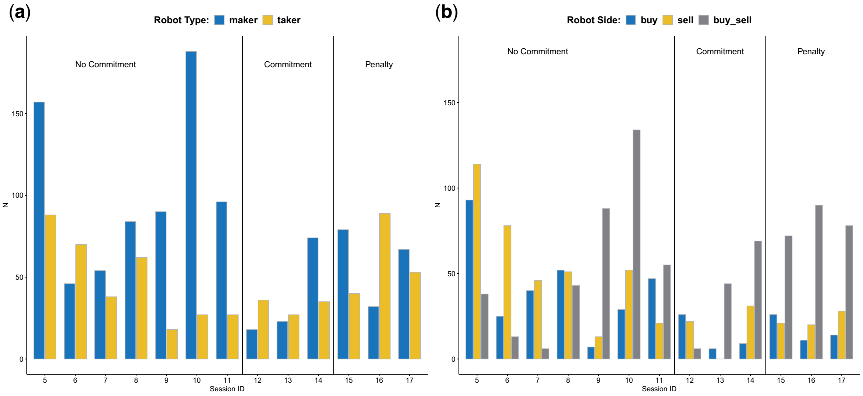
This table presents a regression with a dependent variable that is a dummy of whether a trader deployed a robot in a given trading period t. The independent variables are constructed from the previous period $t-1$, and include whether the trader activated a robot in that round, whether they made profit, as well as indicators of treatments. To compute profits in period t, we use the trader's net trades for that period and take the difference between expected dividends and price paid as profit (for the buyers; we reverse the sign for the sellers). The *ex ante* profit and *ex post* profit differ solely on whether we use the expected or the distributed (realized) dividend for period t (note that the computation can only include expected dividends for periods $t+1 \ldots 15$, thus the only difference between the two measures is for the dividend in period t component) when computing the profits. More precisely, *profit = (net change in cash) + (net change in stock) × Dividend per share (expected or realized).* The standard errors are clustered at the session level.

Regressor	***Dependent variable:***		
	Used Robot		
	(1)	(2)	(3)
Used Robot (lag)	0.6904*** (0.0724)	0.6904*** (0.0722)	0.6907*** (0.0722)
Made Profit (*ex ante*)		0.0011 (0.0188)	
Made Profit (*ex post*)			−0.0122 (0.0182)
Commit and No Penalty	0.0215 (0.0441)	0.0215 (0.0441)	0.0216 (0.0442)
Commit and Penalty	0.1709*** (0.0390)	0.1709*** (0.0401)	0.1721*** (0.0399)
Trading period	−0.0032** (0.0013)	−0.0032** (0.0013)	−0.0032** (0.0013)
Constant	0.1646*** (0.0383)	0.1642*** (0.0432)	0.1685*** (0.0422)
Observations	945	945	945
Adjusted R^2	0.58764	0.58721	0.58737

Note: $^{*}P < .1$; $^{**}P < .05$; $^{***}P < .01$.

exchanges, where the proportion of robot-executed trades is known to be within the range of 30–70 percent (De Luca and Cliff 2011; Goldstein, Kumar, and Graves 2014).

Table 5 reports period-average number of orders and robot activations/halts as well as trading volume, across all sessions in a treatment and in the (manual) control. With robots, we record a decrease in order submission compared to manual trading, except in the Penalty Treatment (first row). Robots submit a relatively small number of orders (row 2) compared to their share in trades (see Figure 9); they are very "productive." With voluntary deployment but commitment (Commitment Treatment), robot activations are the lowest, unsurprisingly (row 3). Rows 4–8 suggest that a rich mix of robots is used. Without commitment, about one-third (9.4/28.8) of robot deployments are halted by the participant (row 9); in a small number of cases (1.0/9.4) does the participant choose to re-deploy with another robot ("Switches"; row 10). Except under the Penalty Treatment, the ratio of robot orders to trades is close to 1 (row 11); since the majority of order submission is manual (e.g., under No Commitment: 28.8/126.99 = 23%; see row 1), this means that robots in principle should be involved in a small fraction of trades; yet, as already pointed out, robots are far more productive: Figure 9 shows that they are involved in more than 50 percent of the trades, except in Session 9.



Robot deployment by role.

Robot deployment by function.

Figure 10. Number of times a robot of a particular role (Maker, Taker; LEFT panel) and function (Buyer, Seller, Two-Sided; RIGHT panel) is launched, average across robot sessions (Sessions 5–17).

The regression analysis reported in Table 7 delineates the individual decisions underpinning the activation of robots during our experimental sessions. A distinct serial correlation between rounds is observable, with the historical usage of robots ("Used Robot (lag)") being a substantial and statistically significant predictor of subsequent activation decisions. This suggests that participants' past experience with robot activation is a strong predictor of their future decisions to activate robots. The treatment condition "Commit & Penalty" promotes a higher incidence of robot activation relative to the baseline "No Commit & No Penalty." Notably, the profitability of previous rounds, whether anticipated (*ex ante*) or realized (*ex post*), does not exert a significant influence on the decision to activate robots. The insignificance of profit outcomes from preceding rounds in influencing robot activation decisions suggests that such decisions may be predicated on a broader array of strategic considerations than immediate prior financial returns.

> **Hypothesis B.1:** Unless robot use is mandatory, high/low use of robots in one period leads to a reduction/increase of their use in the subsequent period.

Figure 10 demonstrates that average robot use fluctuates across periods. Shown are number of robots deployed by role (Maker, Taker) or function (Buyer, Seller, or Two-Sided).

Formal evidence about the usage of robots as a function of past usage is presented in column (3) of Table 6. As this table also sheds light on the remaining hypotheses regarding robot choice, we first describe its overall aim. Table 6 uses a number of prior-period statistics to predict robot use in the subsequent period: number of Takers deployed, midpoint of bid and ask chosen for two-sided robots, overall robot activation, number of robot halts, and number of robot switches. The fourth and fifth regressions pertain to the No Commitment Treatment only. All regressions include session fixed effects.

The regression with number of robot activations as the dependent variable (column (3) of Table 6) shows that there is a strong and significant effect of prior robot activation on subsequent activation, suggesting that robot usage between consecutive periods is positively serially correlated, a result that is contrary to the hypothesis. As the number of robot usage increases after high prior usage, so do the halts of robots as documented in column (4) of the table. Additionally, we record a significant increase in the use of Taker robots (column (1)); we elaborate later.

Figure 10 confirms that the proportion of each robot type (Maker or Taker; Buy, Sell, or Two-Sided) varies quite substantially across periods.

Hypothesis B.2: Participants specialize in one type of robot.

Column (5) in Table 6 shows that participants indeed rarely switch robots, regardless of conditions in the prior round. Switches are more prevalent in early rounds. As such, traders stick to their chosen bots or de-activate them (third column).

Hypothesis B.3: Robot use declines over time.

We use several regression specifications to study the effect of the variable "Trading Period" on robot use. The dependent variable of the regression presented in column (3) of Table 6 is the number of robots activated at the session-period level, while it is a dummy variable for robot activation for the regressions presented in Table 7. In all these specifications, the "Trading Period" variable carries a negative and significant coefficient. As such, robot use declines over time, consistent with our hypothesis. This result does not seem to depend on whether the usage of robots was profitable or not, as shown in Table 7.

Hypothesis B.4: Robot use declines when mispricing increases.

Consistent with the hypothesis, the regression with "Number of Stops" as regressand (column (4) of Table 6) shows that increase in absolute mispricing in one period causes an increase in robot halts in the subsequent period, though the coefficient is only marginally significant ($0.05 \leq P < 0.10$).

Curiously, volatility in a period appears to be interpreted as a sign of overpricing: increased volatility in a period decreases the midpoint of the bid and ask offers of two-way robots in the subsequent period; see second column: regression with Robot Bid–Ask midpoint as regressand. The same regression shows that the midpoint decreases over time, which is to be expected, as the fundamental value decreases.

Hypothesis B.5: Earnings will be higher for robot users.

The final performances of participants in the manual control sessions (Sessions 1–4) and the robot (treatment) sessions (Sessions 5–17) are plotted in Figure 11. The *ex ante* expected payoff (value of initial holdings plus cash) is represented by the dashed line. Performance ranges from a maximum of 413 and a minimum of 51 experimental dollars, with mean equal to 248 and standard deviation 70.[20] Both the lowest and the highest individual performance scores are recorded in the robot sessions. Visual inspection of performance ranges reveals no distinguishable pattern. Formal comparison between the standard deviations around average earnings confirms that there are no statistically significant differences ($P \geq 0.10$).

We now investigate participant earnings. We look at three measures of robot use: (1) Number of robots used, calculated as number of times a participant starts a robot (including re-launching a robot previously deployed); (2) percentage of trades completed by robots, calculated as the per-participant frequency that a trade originated with the participant's robot rather than manually, and (3) percentage of deployed robots that are liquidity-taking.

No significant correlations were found between any metric and performance. However, in a regression analysis, where participants' total number of orders submitted is controlled

---

[20] As pointed out in Section 3, the expected average pay is calibrated at 250 but the empirical average depends on the realized dividends.

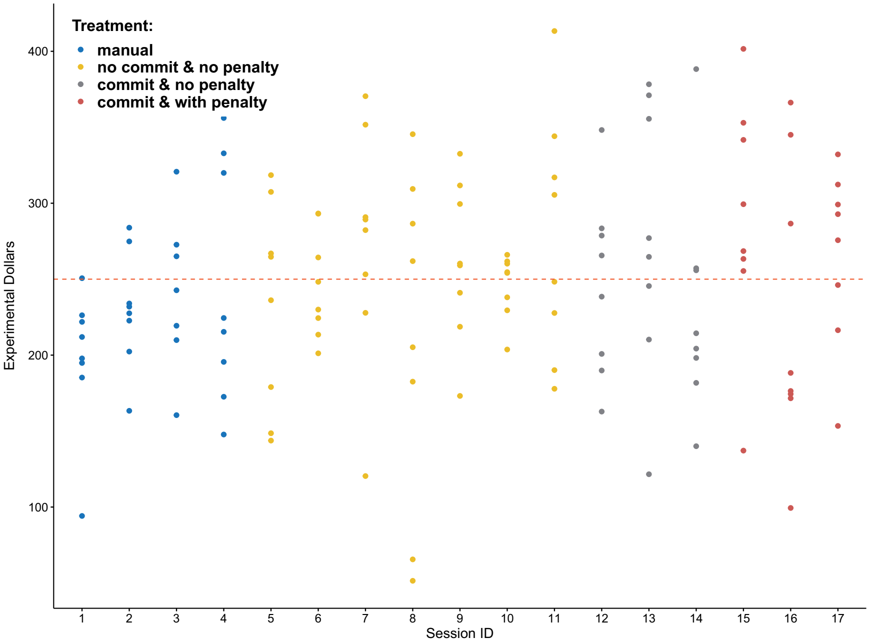


Figure 11. Individual earnings. Total experimental dollars earned per participant in the Control sessions (blue dots), the robotic sessions with No Commitment (yellow dots), three robot Commitment sessions (gray dots), and the sessions with Penalty for not deploying robots within a time limit (red dots). Dashed line indicates expected payoff, which is constant across Treatments/Sessions; actual average earnings vary across sessions because dividend payouts are random.

Table 8. Earnings regression.

Regression of earnings on Number of Robots Launched, Percentage of (Total) Orders (of Participant) from Robots and Percentage of (Total) Taker Robot Deployments (by Participant). We include session fixed effects and report the robust standard errors (in parentheses).

Regressor	*Dependent variable* Earnings
Number Robots Launched	1.78**
	(0.696)
Percentage Orders from Robots	−1.44**
	(0.594)
Percentage Taker Robots	0.23
	(0.270)
Observations	90
Adjusted R^2	0.092

Note: $^{*}P<.1$; $^{**}P<.05$; $^{***}P<.01$.

for, both measures contributed significantly to explaining earnings. The regressions results are displayed in Table 8. We find the following.

- Number of robots used increased earnings: starting an extra robot led to a 1.78 cents increase in final performance ($P < 0.05$);

- Percentage of trades completed by robots reduced earnings: a 50 percent points increase in the fraction of individual trades completed by robots, rather than manually, decreased earnings by 72 cents ($P < 0.05$).

Altogether, this implies that robot use increases earnings, consistent with our hypothesis, but only if robots are used alongside manual trading. Evidently, participants who know *when* to deploy robots and *when* to trade manually earn above-average. Interestingly, the use of Taker Robots, comparable to the "Kaplan Sniper" (see discussion in Introduction) does not increase earnings.

We thus find that robot use *per se* does not increase earnings; only judicious use of robots does. One way to interpret our results is that skill that allows one to deploy the right robots at the right time also makes one perform better when trading manually. Alternatively, the results could be interpreted as suggesting that low-frequency robots (robots that do not trade significantly more frequently than manual traders) are superior.

4.3 Summary of results

We studied robot choice and pricing in a canonical multi-period markets experiment which reliably generates mispricing when participants can only trade manually. We hypothesized that the availability of a set of trading robots would force participants to explicitly consider future price scenarios. Participants who select to deploy robots would choose algorithms that are consistent with fundamental values, and participants who opt to trade manually would expect robots in the marketplace to trade with reference to the fundamental values. Hence, we expected less mispricing when participants have access to algorithmic trading. This access comes in three forms, and thus three treatments, with increasing level of commitment, from availability of robots and ability to start and stop them and change parameters at any time, to an environment where once robots are started they cannot be stopped within a round, to the last treatment that forces participants to choose robots (by imposing a penalty otherwise) and robots only changed between periods. We hypothesized that the reduction of mispricing would be more pronounced the stronger the commitment to a strategy.

Contrary to our expectations, no reduction in mispricing was observed relative to the control (manual trading only), and the three treatments had similar levels of mispricing. The only marked difference between the treatments was in terms of trading volume, which was reduced when robot deployment was voluntary but robots could not be halted once deployed.

We found that, while robots were used extensively, robot usage was a substitute, not a complement to manual trading. While average earnings did not increase for participants who used robots, they were significantly higher for participants who deployed robots in ways that kept manual trading a significant fraction of their total trading.

We also discovered that the bid–ask spread was significantly higher in early rounds of trading in robot sessions. This is consistent with our conjecture that robot trading leads to mis-coordination in the early periods, especially in sessions with strict commitment. Additionally, in the presence of robots, flash crashes/surges were frequent in the first trading period. We attribute this to failed attempts to coordinate robot algorithms to similar reference values in the early stages of the trading game. The most surprising finding is that mispricing was not reduced with robots. The finding is significant because participants had to choose robots (or had to anticipate others' choosing robots) that submitted orders relative to a predetermined reference point. The most obvious (focal) reference point was the fundamental value.

The literature contains only a handful of interventions that have been shown to eliminate bubbles in our setting (SSW): repeating the session with exactly the same participants, exposing participants to very detailed instructions, starving the market of cash; a summary can be found in Powell and Shestakova (2016). Our evidence shows that the presence of robot trading does not eliminate bubbles.

5. Conclusion

Our experiment demonstrates that it is possible to study robot use in financial markets in a controlled setting. This adds a dimension to the existing body of work, as prior studies have focused on archival data from the field. Such studies are limited because there is no control of the setting, and robot deployment cannot be directly observed but has to be inferred from quote and trade data.

Lack of control implies that fundamental values cannot be determined, and hence, quality of pricing cannot be measured directly, but instead evaluated based on tests such as variance ratios. We reported that, when applied to our data, variance ratio tests lead to incorrect inference.

Because robot deployment cannot be measured in field studies unless the robot trades, it may be harder to detect liquidity-taking robots. We confirm here that this biases the findings: we found that 40 percent of the deployed robots were liquidity-taking, far superior to numbers reported in field studies.

Access to unbiased data on robot deployment will be crucial to predict market behavior in crisis situations, when robots that have lurked in the background may all at once become active and generate outcomes that could not have been predicted based on prior activity only. To solely rely on historical analysis of field data is tantamount to pretending that all possible crises have already happened. Therefore, we consider experiments, like the one presented here, to be an essential tool in the study of algorithmic trading.

In the broader context of the behavioral sciences, our experiment sheds light on the use of ‘tools’ among humans (and nonhuman primates such as capuchins); see Moura and Lee (2004). Trading algorithms can indeed be considered “tools” that humans can deploy in order to extend their manual capabilities. Humans readily resort to physical tool use when this makes their work more efficient. Here, we study whether this extends to *software tools*.[21] The participants in our experiment were only minimally trained in the use of algorithms during trading. The intensity with which they are willing to use and engage with robots provides a lower bound on how eager professionals would be to resort to robot trading.

As such, we view robots as a way to enhance human abilities. In the context of the SSW environment, cognitive ability in particular has been a much-studied topic. Hanaki et al. (2017) demonstrate that mispricing in the SSW environment is larger if participants are aware of the heterogeneity of cognitive ability in their cohort. In a study meant to uncover the effects of gender composition on mispricing in the SSW setup, Cueva and Rustichini (2015) report that those with higher cognitive ability have higher earnings and also trade at prices closer to fundamentals. Following up on these studies, Bosch-Rosa, Meissner, and Doménech (2018) report on a two-part study where participants are separated into two groups with high and low cognitive ability, after which they participate in separate market experiments. The study documents classic bubble and crash patterns in markets populated by subjects with low levels of cognitive sophistication, while no bubbles emerge with sophisticated subjects.

Our experimental framework should prompt not only further work on why and how humans interact with algorithms, and how this enhances cognitive capabilities. The framework can also be seen as a testing ground for new types of algorithms (such as fully

---

[21] The literature has recently been referring to the use of software as a tool in terms of “conjoined agency,” distinguishing between software that merely executes pre-defined algorithms, and software that autonomously adjusts algorithms in order to reach a certain goal. Here, we study the former. Asparouhova et al. (2022) have been studying the latter. For an introduction to conjoined agency, see Murray, Rhymer, and Sirmon (2021).

autonomous, intentional agents, as opposed to our execution-only robots), or a training ground for people interested in algorithmic trading.[22]

Acknowledgments

Asparouhova and Bossaerts acknowledge funding from the US National Science Foundation (SES-1061824). Bossaerts acknowledges funding from the Australian Research Council (DP180102284), his R@MAP Chair at the University of Melbourne, and an International Professorship grant to the University of Cambridge from the Leverhulme Foundation. Rotaru thanks the Monash Business School and the Australian Research Council (ARC LIEF grant LE130100112) for funds associated with the Monash Business Behavioural Laboratory equipment used in this study. The article benefited from discussions during the CAFIN-GRU Conference on Market Design and Regulation in the Presence of High-Frequency Trading (2017, Hong Kong), the 2018 Algorithmic Trading Conference at the University of Luxembourg, the 2019 Behavioral Finance and Capital Markets Conference (Melbourne), and comments and suggestions from the Editor and two anonymous reviewers. Some of the results of the first eight experimental sessions were included as part of the 2017 Honours Thesis of Tingxuan Wang at the University of Melbourne.

Supplementary material

Supplementary material is available at *Review of Finance* online.

Funding

Funding support for this article was provided by the US National Science Foundation (SES-1061824), Australian Research Council (DP180102284, LE130100112), University of Melbourne (R@MAP Chair), Leverhulme Trust (Leverhulme International Pofessorship).

Conflicts of interest: The authors have no conflicts of interest to declare.

Data availability

All data, programs, and materials needed to replicate this study are made available on Github, in the public repository archive pbossaerts/RobotsSSW.

References

Alan, S., and S. Ertac. 2015. “Patience, Self-control and the Demand for Commitment: Evidence from a Large-scale Field Experiment.” *Journal of Economic Behavior & Organization* 115: 111–122.

Aldrich, E. M., and K. L. Vargas. 2020. “Experiments in High-frequency Trading: Comparing Two Market Institutions.” *Experimental Economics* 23: 322–352.

Angel, J. J., L. E. Harris, and C. S. Spatt. 2011. “Equity Trading in the 21st Century.” *The Quarterly Journal of Finance* 1: 1–53.

Angerer, M., T. Neugebauer, and J. Shachat. 2023. “Arbitrage Bots in Experimental Asset Markets.” *Journal of Economic Behavior & Organization* 206: 262–278.

Asparouhova, E. N., X., Cai, D., Meloso, J., Nielsen, C.A., Parlour, W., Yang 2022. “For Better or For Worse: Algorithmic Choice in Experimental Markets.” Working Paper, SSRN. https://ssrn.com/abstract=4308398.

---

[22] See 2020 Innovation in Teaching Award Winner, Financial Management Association: https://www.fma.org/iita2023.

Bao, T., E., Nekrasva, T., Neugebauer, Y. E., Riyanto. 2022. "Algorithmic Trading in Experimental Markets with Human Traders: A Literature Survey." In *Handbook of Experimental Finance*, edited by Sascha Füllbrunn and Ernan Haruvy, 302–322. Cheltenham: Edward Elgar Publishing.

Baron, M., J. Brogaard, B. Hagströmer, and A. Kirilenko. 2019. "Risk and Return in High-Frequency Trading." *Journal of Financial and Quantitative Analysis* 54: 993–1024.

Baron, M., J. Brogaard, and A. Kirilenko. 2012. "The Trading Profits of High Frequency Traders (Risk and Return in High Frequency Trading)." *8th Annual Central Bank Workshop on the Microstructure of Financial Markets*, Vol. 25. Ottawa (Ontaria), Canada: Bank of Canada.

Biais, B., T. Foucault, and S. Moinas. 2011. "Equilibrium High Frequency Trading." In *International Conference of the French Finance Association (AFFI)*. Paris, France: French Finance Association.

Boehmer, E., K. Fong, and J. J. Wu. 2021. "Algorithmic Trading and Market Quality: International Evidence." *Journal of Financial and Quantitative Analysis* 56: 2659–2688.

Bosch-Rosa, C., T. Meissner, and A. Bosch-Domènech. 2018. "Cognitive Bubbles." *Experimental Economics* 21: 132–153.

Bossaerts, P., F. Fattinger, K. Rotaru, and K. Xu. 2023. "Emotional Engagement and Trading Performance." *Management Science*. https://doi.org/10.1287/mnsc.2023.4883

Brogaard, J., T. Hendershott, and R. Riordan. 2014. "High-frequency Trading and Price Discovery." *The Review of Financial Studies* 27: 2267–2306.

Brogaard, J., M. C. Ringgenberg, and D. Roesch. 2023. "Does Floor Trading Matter." *Journal of Finance* http://dx.doi.org/10.2139/ssrn.360907.

Carrion, A. 2013. "Very Fast Money: High-frequency Trading on the NASDAQ." *Journal of Financial Markets* 16: 680–711.

Chaboud, A. P., B. Chiquoine, E. R. I. K. Hjalmarsson, and C. Vega. 2014. "Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market." *The Journal of Finance* 69: 2045–2084.

Chordia, T., R. Roll, and A. Subrahmanyam. 2001. "Market Liquidity and Trading Activity." *The Journal of Finance* 56: 501–530.

Chordia, T., R. Roll, and A. Subrahmanyam. 2011. "Recent Trends in Trading Activity and Market Quality." *Journal of Financial Economics* 101: 243–263.

Cueva, C., and A. Rustichini. 2015. "Is Financial Instability Male-driven? Gender and Cognitive Skills in Experimental Asset Markets." *Journal of Economic Behavior & Organization* 119: 330–344.

De Luca, M., and D., Cliff. 2011. "Human-agent Auction Interactions: Adaptive-aggressive Agents Dominate." In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, Vol. 22, 178. Menlo Park, CA, USA: IJCAI.

de A. Moura, A. C., and P. C. Lee. 2004. "Capuchin Stone Tool Use in Caatinga Dry Forest." *Science* 306: 1909–1909.

Embrey, M., G. R. Fréchette, and S. Yuksel. 2017. "Cooperation in the Finitely Repeated Prisoner's Dilemma." *The Quarterly Journal of Economics* 133: 509–551.

Farjam, M., and O. Kirchkamp. 2018. "Bubbles in Hybrid Markets: How Expectations about Algorith- mic Trading Affect Human Trading." *Journal of Economic Behavior & Organization* 146: 248–269.

Furbush, D. 1989. "Program Trading and Price Movement: Evidence from the October 1987 Market Crash." *Financial Management* 18: 68–83.

Gini, C. 1936. "On the Measure of Concentration with Special Reference to Income and Statistics." *Colorado College Publication, General Series* 208: 73–79.

Glöckner, A., B. E. Hilbig, and M. Jekel. 2014. "What is Adaptive about Adaptive Decision Making? A Parallel Constraint Satisfaction Account." *Cognition* 133: 641–666.

Goldstein, M., A. Kwan, and R. Philip. 2023. "High-frequency Trading Strategies." *Management Science* 69: 4413–4434.

Goldstein, M. A., P. Kumar, and F. C. Graves. 2014. "Computerized and High-Frequency Trading." *The Financial Review* 49: 177–202.

Golub, A., J. Keane, and S. H. Poon. 2012. "High Frequency Trading and Mini Flash Crashes." Working Paper, SSRN. https://ssrn.com/abstract=2182097.

Grossklags, J., and C. Schmidt. 2006. "Software Agents and Market (in) Efficiency: A Human Trader Experiment." *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 36: 56–67.

Hanaki, N., E., Akiyama, Y., Funaki, R., Ishikawa. 2017. "Diversity in Cognitive Ability and Mispricing in Experimental Asset Markets". HAL SHS working paper halshs-01202088.

Hendershott, T., C. M. Jones, and A. J. Menkveld. 2011. “Does Algorithmic Trading Improve Liquidity.” *The Journal of Finance* 66: 1–33.

Jarrow, R. A. and P. Protter. 2012. “A Dysfunctional Role of High Frequency Trading in Electronic Markets.” *International Journal of Theoretical and Applied Finance* 15: 1250022.

King, R. R., V. Smith, W.A. Wlliams, M. Van Boening. 1993. “The Robustness of Bubbles and Crashes in Experimental Stock Markets.” In *Nonlinear Dynamics and Evolutionary Economics*, Chapter 13, edited by Richard Hollis Day and Ping Chen, 183–200. Oxford: Oxford University Press.

Kirilenko, A., A. S. Kyle, M. Samadi, and T. Tuzun. 2017. “The Flash Crash: High-frequency Trading in an Electronic Market.” *The Journal of Finance* 72: 967–998.

Kissell, R., and R. Malamut. 2005. “Understanding the Profit and Loss Distribution of Trading Algorithms”. New York: JP Morgan.

Leal, J., and S. N. Hanaki. 2018. “Algorithmic Trading, What if it is Just an Illusion? Evidence from Experimental Financial Markets.” In *Technical Report, Groupe de Recherche en Droit, Economie, Gestion (GREDEG CNRS)*. France: Université Côte d’Azur.

LeBaron, B., W. B. Arthur, and R. Palmer. 1999. “Time Series Properties of an Artificial Stock Market.” *Journal of Economic Dynamics and Control* 23: 1487–1516.

Litzenberger, R., J. Castura, and R. Gorelick. 2012. “The Impacts of Automation and High Frequency Trading on Market Quality.” *Annual Review of Financial Economics* 4: 59–98.

Madan, C. R., M. L. Spetch, and E. A. Ludvig. 2015. “Rapid Makes Risky: Time Pressure Increases Risk Seeking in Decisions from Experience.” *Journal of Cognitive Psychology* 27: 921–928.

Menkveld, A. J. 2013. “High Frequency Trading and the New Market Makers.” *Journal of Financial Markets* 16: 712–740.

Murray, A., J. Rhymer, and D. G. Sirmon. 2021. “Humans and Technology: Forms of Conjoined Agency in Organizations.” *Academy of Management Review* 46: 552–571.

Noussair, C., S. Robin, and B. Ruffieux. 2001. “Price Bubbles in Laboratory Asset Markets with Constant Fundamental Values.” *Experimental Economics* 4: 87–105.

Palan, S. 2013. “A Review of Bubbles and Crashes in Experimental Asset Markets.” *Journal of Economic Surveys* 27: 570–588.

Powell, O., and N. Shestakova. 2016. “Experimental Asset Markets: A Survey of Recent Developments.” *Journal of Behavioral and Experimental Finance* 12: 14–22.

Roll, R. 1984. “A Simple Implicit Measure of the Effective Bid-ask Spread in an Efficient Market.” *The Journal of Finance* 39: 1127–1139.

Rust, J., J. H. Miller, and R. Palmer. 1994. “Characterizing Effective Trading Strategies: Insights from a Computerized Double Auction Tournament.” *Journal of Economic Dynamics and Control* 18: 61–96.

Serbera, J. P., and P. Paumard. 2016. “The Fall of High-frequency Trading: A Survey of Competition and Profits.” *Research in International Business and Finance* 36: 271–287.

Siegel, J. J. 2003. “What is an Asset Price Bubble? An Operational Definition.” *European Financial Management* 9: 11–24.

Smith, V. L., G. L. Suchanek, and A. W. Williams. 1988. “Bubbles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets.” *Econometrica* 56: 1119–1151.

Sutton, R. S., and A. G., Barto. 2018. *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.