Algorithmic Third-Party Advice in Markets with Complex Goods*

Leo Bao† Kenan Kalaycı‡ Ruize Sun§

August 17, 2026

Abstract

Third-party advice is prevalent in markets for complex goods to help consumers compare alternatives, yet its effects remain ambiguous. We develop a simple model in which comparing complex products is cognitively costly and buyers may follow imperfect advice to economize on cognitive effort. We then conduct an experiment that varies product complexity and the availability and accuracy of advice. Consistent with the model, complexity reduces buyer choice accuracy and surplus when no advice is available, while perfect advice restores buyer outcomes to levels close to those in markets with simple goods. Imperfect advice, however, does not improve buyers' outcomes, even though buyers are informed of the imperfection. Seller surplus does not differ significantly across treatments, suggesting that the welfare effects of advice are driven primarily by changes in buyer surplus. These results suggest that the effectiveness of third-party advice hinges on its accuracy and that transparency about its imperfection alone is not sufficient.

Keywords: experiment, bounded rationality, algorithmic advice, product complexity
JEL classification codes: C92, D91, D12, L13

*We thank Klaus Abbink, Johannes Abeler, Zachary Breig, Muruvvet Buyukboyaci, David Byrne, Timothy Cason, Lata Gangadharan, Steffen Huck, Andreas Leibbrandt, Matthew Leister, John List, Birendra Rai, and Klaus Schmidt. We also thank audiences at ESA 2017 (Vienna), ANZWEE 2017 (Melbourne), Middle East Technical University (2018), Sogang University (2022), ARC Behavioural Market Design Workshop, RMIT (2023), Economic Theory Festival, UNSW (2024), and AGORA Workshop on Market Design (2025). Leo is grateful to the faculty and staff at the University of Chicago for their hospitality during his visit when this study was being written. Kenan Kalaycı acknowledges financial support from Australian Research Council Grant DE160101242. Kenan is grateful to the faculty and staff at the University of Oxford for their hospitality during his visit while part of this study was conducted.

†Department of Banking and Finance, Monash University. Email: leo.bao@monash.edu
‡School of Economics, University of Queensland. Email: k.kalayci@uq.edu.au
§Department of Banking and Finance, Monash University. Email: ruize.sun@monash.edu

1 Introduction

Consumers increasingly face alternatives with complex attributes and pricing components that defy easy comparison. A health-insurance consumer, for example, may face dozens or even hundreds of plans that differ not only in premiums, but also in deductibles, copayments, coinsurance, drug formularies, and provider networks. Similar comparison problems arise in mortgages, electricity contracts, and hotel bookings, among others. Evidence shows that complexity leads buyers to make suboptimal choices (Besedeš et al. 2015, Sitzia et al. 2015), and that complexity raises prices and lowers buyer surplus in part by softening firms' incentives to undercut competitors (Gu and Wenzel 2015, Crosetto and Gaudeul 2017, Frank and Lamiraud 2009). These anti-competitive effects leave firms with little motivation to simplify their offerings (Piccione and Spiegler 2012, Ellison and Wolitzky 2012, Chioveanu and Zhou 2013, Ellison and Ellison 2009, Célérier and Vallée 2017).

Various attempts have emerged to address complexity-driven inefficiencies. Direct regulation can improve outcomes when consumers have limited attention (Heidhues et al. 2021), but complexity is harder to quantify and regulate than misleading conduct, and even optimally chosen default-option regulations can backfire by making consumers less attentive (Haan and Linde 2018). Third-party comparison services, such as NerdWallet in the US and iSelect in Australia, offer a market-based alternative, but most such services earn commissions that bias their recommendations toward higher-priced products (Armstrong and Zhou 2011, Inderst and Ottaviani 2012) and rely on ranking algorithms that are not transparent. Despite these limitations, relatively little is known about the effectiveness of third-party advice. This gap is especially important in light of the rapid diffusion of AI technology, which provides consumers with low-cost advice and has the potential to reduce conflicts of interest associated with traditional third-party advice.

We begin with a partial-equilibrium model capturing both product complexity and third-party advice.[1] Given prices, a buyer optimizes choice probabilities to trade off expected surplus against cognitive effort as captured by an entropy cost in a multinomial logit model. Complexity scales this entropy cost: to maintain the same choice accuracy, as measured by the probability of choosing the best, a more complex product environment requires greater cognitive effort to distinguish the surplus-maximizing alternative from inferior ones. When product complexity increases, both choice accuracy and buyer surplus fall. To help buyers handle complexity, we introduce perfect and imperfect advice as revelations of the best and second-best products respectively, whereby we can study the impact of advice quality. Not surprisingly, perfect advice resolves complexity and produces outcomes similar to those under a simple product environment. We then analyze imperfect advice, which is prevalent in real-world advisory settings. Imperfect advice can have an ambiguous impact on buyer surplus. Under moderate complexity, a buyer simply follows the advice, giving up some surplus in exchange for mental comfort. For lower complexity, a buyer excludes the recommended product to find the best product; for higher complexity, expected surplus from manual comparison may or may not outperform that of the second-best choice, despite substantial cognitive effort involved in both cases.

We then conduct a laboratory experiment motivated by the model's economic implications. This approach is well-suited to our purposes given that in real markets buyers' preferences and complete choice sets are rarely observed, the accuracy of algorithmic advice is difficult to measure, and the advice environment is fairly chaotic. The experiment varies the availability and accuracy of algorithmic advice in a duopoly market with differentiated products. Two buyers and two sellers (each offering two goods) trade in continuous time, allowing dynamic buyer-seller interactions. We have four treatment conditions: a simple market condition with no advice (SIMPLE), and three complex market conditions with no (COMPLEX), perfect (PERFECT) or imperfect (IMPERFECT) advice. In the three complex conditions, evaluating a good requires buyers to decipher a five-letter code, sum the numbers corresponding to the five component letters and subtract its price during the 120-second trading stage. We introduce algorithmic third-party advice, framed as a robot, that tells buyers which good currently offers them the highest payoff. In PERFECT the robot always recommends the surplus-maximizing choice; in IMPERFECT it systematically recommends the second-best option; buyers are told which type of robot they face.

Consistent with the theory, we find that complexity worsens buyer outcomes: buyers in the complex environment spend only 40% of their time on the best available choice, compared to 61% in the simple environment, and buyer surplus is 26% lower in complex markets. Perfect algorithmic advice substantially improves buyer choices: buyers spend 66% of their time on the best choice, and buyer surplus rises from A$16.6 to A$20.0 (compared to A$22.3 in simple markets). Imperfect advice, by contrast, fails to improve on no advice. Buyers with imperfect advice spend only 36% of their time on the best choice and forgo 44% of their available surplus. Both outcomes are statistically indistinguishable from the no-advice COMPLEX condition, and substantially worse than under perfect advice. Seller surplus does not differ significantly between treatments, so changes in buyer surplus largely represent changes in total welfare, i.e., the sum of the buyer and seller surplus.

We contribute to several strands of the literature. In behavioral industrial organization, Choi et al. (2010) show in an experiment that investors fail to minimize mutual fund fees, and Brown et al. (2010) show in a field experiment that buyers are insufficiently attentive to shrouded shipping costs on eBay auctions. When some consumers are boundedly rational, firms in a competitive market may have an incentive to take advantage (Gabaix and Laibson 2006, Carlin 2009). Closest to our setting, Kalaycı and Potters (2011) and Kalaycı (2015) demonstrate experimentally that sellers in a duopoly market use product and price complexity to confuse buyers, and that market prices are higher in markets with human buyers than in markets with perfectly rational, computerized buyers. We extend this work in two ways. We show that complexity still harms buyers in a continuous-time market where buyers can change their choices over time. We also show that, across the complex treatments, sellers do not detectably adjust pricing to buyer sophistication, so the welfare gains from accurate advice come primarily through reduced deadweight loss rather than redistribution from sellers.

Two more distant literatures bear on our setup: costly price search and advertisement (Dia-

mond 1971, Stahl 1989), and expert advice under asymmetric information in credence-goods markets (Dulleck and Kerschbamer 2006). We differ from both in that there is no private information: complexity, not hidden attributes, drives buyer mistakes. Compared with most price-search experiments (Morgan et al. 2006, Cason and Friedman 2003), we use human buyers (rather than computerized buyers) and differentiated products. Unlike the credence-goods literature, our advice is automated and either perfect or imperfect by construction, rather than provided by a human subject with conflicts of interest.

On the theoretical side, two related lines of work motivate our experiment: biased intermediation, in which intermediaries divert search to favor higher-margin products, and platform steering, in which recommendation algorithms shape what consumers see and buy. Hagiu and Jullien (2011) show that intermediaries may strategically divert consumer search to increase their own profits or influence seller behavior. de Cornière and Taylor (2019) demonstrate that biased intermediaries can steer consumers toward products that offer lower utility when there is a conflict of interest between the intermediary and the consumer. Their more recent framework shows how data-driven algorithmic recommendations can have pro- or anti-competitive effects depending on how the data are used (de Cornière and Taylor 2025). Heidhues et al. (2023) extend this analysis to intermediaries that steer consumers with systematic mistakes, characterizing the welfare and competition consequences. In contrast, the imperfect robot in our paper is a blunt, fully transparent intermediary supposedly free from such rent-seeking incentives. We show that it does not steer buyers onto the recommendation; the harm operates instead through broadly degraded choice.

Algorithmic advice now shapes many everyday decisions, from navigation to product comparison, and has motivated extensive research on human-machine interaction (Dietvorst et al. 2015, 2018, Logg et al. 2019, Normann and Sternberg 2023). Much of this work studies adoption and trust, including aversion after observed errors (Dietvorst et al. 2015) and a preference for algorithmic over human judgment (Logg et al. 2019). Our setting instead lets us observe consultation directly and link it to welfare. Buyers consult the imperfect robot far less often than the accurate one, consistent with lower willingness to consult an unreliable recommendation. Accurate advice substantially improves buyer choices and welfare; imperfect advice, by contrast, does not improve on no advice, even though its imperfection is disclosed. Given that recommendation imperfection is typically less transparent outside the laboratory, the accuracy of those recommendations deserves regulatory scrutiny.

2 Theoretical Motivation

This section develops a partial-equilibrium model of how product complexity and access to robot advice affect buyer behavior and welfare. The model motivates and disciplines the experimental design.

Consider a buyer facing an array of alternative products indexed by $j \in K = \{0, \ldots, n\}$, in which product 0 is the outside option of buying nothing. Denote by P_j and v_j the price and the

fundamental value of product j to the buyer, with $P_0 = v_0 = 0$. Assume for simplicity that the production cost is zero for each product, so the price equals seller surplus (if sold). A rational buyer would choose one product à la the utility maximization problem

$$\max_{j \in K} (v_j - P_j). \tag{1}$$

Under bounded rationality, however, a buyer faces mental costs of comparing products; in the spirit of Anderson et al. (1992) and Matějka and McKay (2015), we capture this via an entropy penalty $\eta H(\rho)$ measuring deviation from uniform randomization, as detailed below, such that the buyer solves a generalized utility maximization problem:

$$\max_{\rho \in \Delta(K)} \left[\left(\sum_{j \in K} \rho_j (v_j - P_j) \right) - \eta H(\rho) \right], \ \eta > 0, \tag{2}$$

where $\Delta(K)$ denotes the set of probability distributions over K such that a generic member $\rho \in \Delta(K)$ specifies the distribution of choices, and

$$H(\rho) = \sum_{j \in K} \left(\rho_j \log \rho_j - |K|^{-1} \log |K|^{-1} \right). \tag{3}$$

$H(\rho) \geq 0$ is the entropy reduction of a buyer's behavior due to an effort to optimize, which reflects mental costs associated with understanding, analyzing, and comparing products. The minimum $H(\rho)$ is 0, obtained when the buyer exerts no effort, that is, when she chooses each $j \in K$ with the same probability $1/|K|$. In contrast, the maximum H occurs when the buyer strives to identify and choose the best for sure, i.e., when $\rho_j = 1$ for some j. Accordingly, η as the coefficient of H measures the overall product complexity, such that the more complex the products are the more difficult the buyer's decision-making problem is. The buyer's utility then equals expected buyer surplus, $\sum_{j \in K} \rho_j u_j$ with $u_j = v_j - P_j$, minus mental costs $\eta H(\rho)$.

The solution of Eq. (2) is given by

$$\rho_j^* = \frac{\exp(\eta^{-1} u_j)}{\sum_{j \in K} \exp(\eta^{-1} u_j)}, \ j \in K. \tag{4}$$

As $\eta \downarrow 0$, ρ^* concentrates on $\arg\max_j u_j$, recovering the deterministic choice under unbounded rationality; as $\eta \to \infty$, ρ^* approaches the uniform distribution. Holding prices fixed, higher complexity therefore makes choice coarser, and it is straightforward to verify that expected buyer surplus and buyer utility are both decreasing in η.

Even with fixed prices, it is indeterminate how seller surplus changes with η. At one extreme, if the prices are such that the buyer captures a fixed fraction of the total surplus of each product, i.e., $u_j = \alpha v_j, \alpha \in (0, 1)$, then seller surplus also decreases in η; at the other extreme, if the products are all unduly expensive in that the best option is buying nothing, then seller surplus is at least

locally increasing in η.

Now suppose the buyer can consult a robot, which returns a recommendation $\hat{j} \in K$ that the buyer can follow at no mental cost. A perfect robot recommends the best option and an imperfect robot recommends the second-best; the perfect case is straightforward since it effectively restores unbounded rationality, so we focus on the imperfect robot.

With the recommendation $\hat{j}$ from the imperfect robot, the buyer can either follow it, or set it aside and then solve a reduced-dimension problem similar to Eq. (2). That is, the outer maximization below chooses between the no-effort option of following $\hat{j}$ and the value of optimizing over the remaining alternatives in $\hat{K}$:[2]

$$\max \left\{ u_{\hat{j}}, \max_{\rho \in \Delta(\hat{K})} \left[\left(\sum_{j \in \hat{K}} \rho_j (v_j - P_j) \right) - \eta \hat{H}(\rho) \right] \right\} \tag{5}$$

where $\hat{K} = K \setminus \{\hat{j}\}$ and

$$\hat{H}(\rho) = \sum_{j \in \hat{K}} \left(\rho_j \log \rho_j - |\hat{K}|^{-1} \log |\hat{K}|^{-1} \right). \tag{6}$$

Figure 1 illustrates these effects with a four-product numerical example (parameters in the caption). Both buyer surplus and buyer utility decrease in product complexity with and without robot access, but for different reasons. Without robot access, the buyer responds to higher product complexity by optimizing more coarsely: her choice distribution flattens toward uniform, placing more weight on inferior products and thereby lowering buyer surplus. With an imperfect robot, the buyer initially uses the robot to exclude the second-best product and then solves the reduced-dimension problem, which is the case when

$$u_{\hat{j}} \leq \max_{\rho \in \Delta(\hat{K})} \left[\left(\sum_{j \in \hat{K}} \rho_j (v_j - P_j) \right) - \eta \hat{H}(\rho) \right]. \tag{7}$$

As product complexity escalates further, however, the buyer prefers to follow the robot recommendation without further comparison, which is the case when the inequality in Eq. (7) is reversed: the region over which the dashed curve remains constant in each panel.

Robot access always raises buyer utility, with the largest gains arising when products are sufficiently complex; its effect on buyer surplus, however, is mixed (Figure 1, panel (a)). At low complexity, an imperfect robot raises buyer surplus by eliminating the second-best option and hence helping the buyer focus on the best one, which echoes the widely documented observation that consumers may exhibit reactance and strategic exploitation of (algorithmic) recommendations (e.g., Friestad and Wright 1994, Fitzsimons and Lehmann 2004, Lee and Lee 2009). At intermediate complexity, a buyer with robot access follows it, forgoing some surplus to avoid mental effort, whereas a buyer without the access exerts substantial effort and ends up with higher surplus but

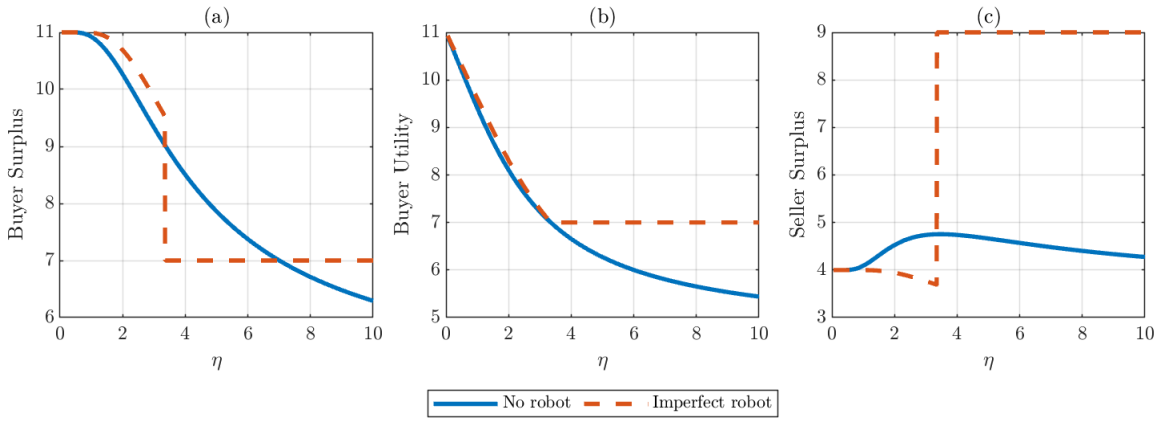


Figure 1: **Partial-Equilibrium Outcomes.** The figure plots buyer surplus, buyer utility, and seller surplus as a function of η, when prices are exogenously given. The parameter values are: $n = 4$, $(v_1, v_2, v_3, v_4) = (15, 16, 7, 3)$, $(P_1, P_2, P_3, P_4) = (4, 9, 3, 2)$; the buyer surplus corresponding to each product is then $(u_1, u_2, u_3, u_4) = (11, 7, 4, 1)$.

lower utility. At high complexity, the robot-using buyer simply continues to follow $\hat{j}$, while the buyer without it settles for a coarse randomization whose expected surplus lies between uniform choice and $u_{\hat{j}}$.

From Figure 1, panel (c), the impact of product complexity and robot access on seller surplus is also mixed. The second-best product is the one whose fundamental value, price, and hence seller surplus are the highest. As product complexity increases from zero, a buyer without robot access occasionally deviates from the best in favor of the second-best, giving rise to higher seller surplus; however, as product complexity further increases, seller surplus declines due to a flatter randomization over all products. The second-best product, though the most profitable to the seller, no longer captures a buyer who misses the best with sufficiently high probability relative to products 3–4, which are poor choices for both buyer and seller. An imperfect robot, at first glance, is likely to increase seller surplus since the second-best product is the most profitable one. Nonetheless, this intuition is true only when the products are so complex that a buyer is willing to follow the robot; for relatively small η, an imperfect robot is exploited to eliminate the second-best product, leaving smaller seller surplus than that under no robot access.

The partial-equilibrium model suggests three testable implications for the experimental treatments. First, expected buyer surplus is decreasing in η without robot access, so Complex reduces buyer surplus relative to the Simple treatment. Second, perfect advice removes the choice problem at the click of a button, so buyers in the Perfect treatment should approach the $\eta \downarrow 0$ benchmark and look like buyers in Simple. Third, under imperfect advice the model predicts a regime-dependent buyer response, with the buyer-surplus response relative to the no-advice baseline depending on the regime (Figure 1, panel (a)): at low η, the buyer uses the recommendation to exclude the second-best and solves a reduced-dimension optimization; at intermediate or higher η, the buyer follows the recommendation and concentrates choice on the second-best option. The

two regimes leave distinct signatures on the time-share allocated to the second-best option, which the within-period analysis in Section 4.1.4 examines.

These predictions are partial-equilibrium: they fix prices and abstract from the strategic seller side. In the full market, buyer behavior interacts with endogenous pricing through inter-seller competition, intra-seller cannibalization, and dynamic feedback. For example, in an equilibrium with multiple buyers, each buyer may exploit an imperfect robot in their own interest, yet this individually rational inattention can produce negative externalities through seller-side exploitation, so that in aggregate each buyer's utility is lower than in an equilibrium without robot access. Because the resulting general equilibrium is difficult to characterize analytically, we turn to the experiment. We therefore use buyer and seller surplus as the welfare metric, since mental costs are unobservable.

3 Experimental Design

3.1 Market setup

We conduct a continuous-time market experiment that follows the theoretical motivation in Section 2 closely. Four participants form a market: two are randomly selected to be buyers and the other two are sellers. Each buyer's valuation for each good is the sum of five attributes, each drawn from a uniform distribution on $[0, 10000]$, in per-second experimental tokens. We draw these numbers in advance and reuse the same realizations across treatments, holding buyer valuations fixed. Buyers and sellers interact in three stages.

In the first stage, sellers see buyers' valuations over all goods and are given 30 seconds to set their initial prices. At this stage, sellers have no information about the other seller's prices. In the second stage, buyers are given 30 seconds to make their initial choices (they may choose to buy nothing) based on their valuations and the posted prices. In the third stage, buyers and sellers interact continuously for two minutes: buyers can switch their choices and sellers can change the prices of their goods. Throughout stage three, sellers see all prices, valuations, and both buyers' choices, while buyers see only their own valuations or attributes, depending on the treatment, and the four prices.

Payoffs accrue on a per-second basis during the third stage. A buyer's per-second payoff equals the valuation of the chosen good minus its price; a seller's per-second payoff equals the total revenue from goods sold. Cumulative per-second payoffs form a participant's token earnings for the period, which are converted to Australian dollars at 100,000 tokens = A$1. Each participant also receives an A$10 show-up fee.

A price ceiling (unknown to buyers) makes negative buyer earnings unlikely, though still possible if a buyer chooses a good whose valuation falls below its price. We also forbid negative prices, so sellers cannot incur negative earnings.

Each group plays the three-stage game for ten periods so that we can observe how trading behavior evolves over time. We pay participants for one randomly chosen period to avoid wealth

effects. Fixing roles and group composition within markets serves two purposes: to mimic the stability of market participants in real-world settings (external validity), and to obtain repeated observations within stable markets while avoiding re-matching confounds (internal validity). We therefore treat each group as a single independent observation (n=12 per treatment).

Procedures. The experiment was approved by the university's Human Research Ethics Committee, programmed in z-Tree (Fischbacher 2007), and conducted at a university laboratory in Australia. We ran 48 markets across the four treatments (192 participants in total, 12 markets per treatment), recruited via SONA. Each 90-minute session began with consent, instructions read aloud, and comprehension questions; no session proceeded until all participants had answered correctly. Participants earned an average of A$26 in cash (range: A$10 to A$110) and were paid in private.

3.2 Treatments

We vary the treatments along two dimensions. The first varies arithmetic complexity (Simple vs Complex). The second varies the availability of algorithmic advice in the complex environment (no advice, Perfect, or Imperfect). Table 1 summarizes the four treatments.

Table 1: Experimental treatments

	No Advice	Perfect Advice	Imperfect Advice
Simple	$n = 12$	—	—
Complex	$n = 12$	$n = 12$	$n = 12$

This table summarizes the four treatments. Moving down the No Advice column varies complexity; moving across the Complex row varies advice. We collect 12 independent group observations per treatment cell, for a total of 48 groups and 192 participants.

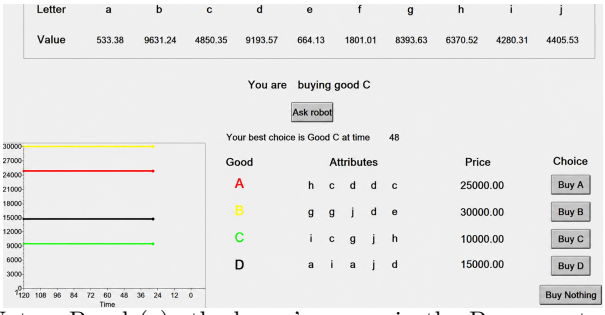
We use the first dimension to test whether complexity harms buyers in a continuous-time market. In the Simple treatment, each buyer is directly informed of their four valuations. In the Complex treatment, the valuation of each good is decomposed into five attributes and encoded as a five-letter string; buyers decipher each string using a lookup table to recover the underlying numbers (Erkal et al. 2011) and sum them to obtain the valuation. Sellers have the same market information across both treatments but know which environment buyers face. Figure 2 contrasts the two environments.

Along the second dimension, we study whether algorithmic advice can mitigate the effects of complexity. Throughout, Complex denotes the complex environment without algorithmic advice and serves as the no-advice baseline against which the two advice treatments are compared. We frame the algorithm as a "robot" and study two variants. In both variants, buyers can click an "Ask Robot" button (illustrated in panel (a) of Figure 2) at any time and at no cost to receive a recommendation. Recommendations are evaluated at the click moment, so a buyer needs to click

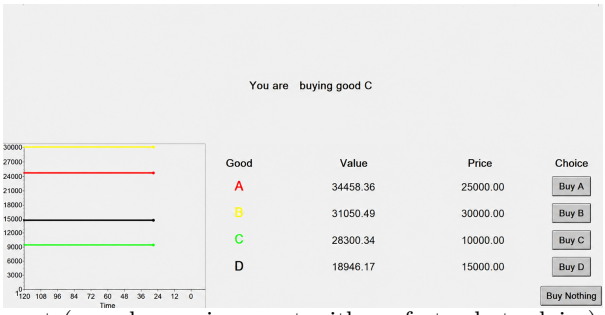
again after a price change to learn the updated recommendation. The two variants differ only in their recommendation rule. In the PERFECT treatment, the robot recommends the good offering the highest payoff (valuation minus price); in the IMPERFECT treatment, it recommends the good offering the *second*-highest payoff. Either variant can also recommend "buy nothing" when that is the best (or second-best) option. Buyers are told explicitly which variant they face, and sellers also know which variant the buyers have. The full experimental instructions are in Appendix A.

Figure 2: Screenshots for buyers

(a) Buyer screen in COMPLEX with perfect robot



(b) Buyer screen in SIMPLE



Notes: Panel (a): the buyer's screen in the PERFECT treatment (complex environment with perfect robot advice). Panel (b): the buyer's screen in the SIMPLE treatment, for the same buyer facing the same goods and prices. The buyer in panel (b) sees her valuation of Good A directly (34458.36). The buyer in panel (a) reads the lookup table at the top (for Good A, letter h represents 6370.52, letter c represents 4850.35, and so on) and sums the five numbers: $6370.52 + 4850.35 + 9193.57 + 9193.57 + 4850.35 = 34458.36$. Full instructions in Appendix A.

3.3 Hypotheses

Hypothesis 1 (Standard). If we assume buyers have zero entropy cost, our model predicts no differences in equilibrium outcomes across treatments, which corresponds to the limiting case $\eta \downarrow 0$. Because sellers observe buyers' valuations, equilibrium prices depend only on these valuations and not on complexity or robot availability; under rational best response by buyers, allocations and surpluses are therefore the same as well. We therefore expect prices, buyer surplus, and seller surplus to be the same across all four treatments.

Hypothesis 2 (Complexity). We expect complexity to reduce buyer surplus and increase seller surplus relative to the simple environment. The buyer-surplus reduction is predicted by the η-comparative statics in Section 2 and consistent with evidence that buyers struggle to compare complex goods (Kalaycı and Serra-Garcia 2016); the seller-surplus prediction is empirically motivated by evidence that duopoly sellers raise prices to exploit this confusion (Kalaycı and Potters 2011, Kalaycı 2015), since the model's seller-surplus response to η is indeterminate.

Hypothesis 3 (Perfect advice). The perfect robot enables buyers to identify the best choice at the click of a button. We expect it to restore buyer surplus in the complex environment toward the level observed in the simple environment.

Hypothesis 4 (Imperfect advice). As shown in Section 2, the impact of the imperfect robot on buyer and seller surplus is ambiguous *a priori*. On one hand, knowing the second-best choice removes one option from contention and could narrow the set of options the buyer must evaluate. On the other hand, the buyer can settle for the second-best to save evaluation effort, which is optimal when products are sufficiently complex. The human factors literature on automation bias likewise suggests that buyers may anchor on a salient automated signal and choose the recommended second-best option even when they know it is imperfect (Parasuraman and Manzey 2010, Goddard et al. 2012). If anchoring dominates, the imperfect robot could reduce buyer surplus below the level in Complex (no advice). Section 2 formalizes these two channels as a regime structure under imperfect advice. At low η, the buyer narrows the set of options by excluding the second-best, leaving the time-share on the second-best no higher than under no advice. At intermediate or higher η, the buyer follows the recommendation, concentrating time on the second-best.

4 Results

Our primary design outcomes are buyer choice accuracy (percentage of time on the surplus-maximizing option) and the proportion of buyer surplus foregone. Buyer surplus, seller surplus, and total welfare are secondary outcomes that are noisier because they combine choice quality with endogenous seller pricing. Unless otherwise specified, all reported p-values come from two-tailed Wilcoxon rank-sum tests, aggregating all dependent observations within a group into one independent observation. With 12 independent groups per treatment, we interpret failures to reject cautiously rather than as evidence of equivalence.[3]

4.1 Buyer Choices

4.1.1 Choice Accuracy

Result 1. Complexity sharply reduces buyer choice accuracy, perfect advice closes the gap, and imperfect advice does not close it.

We measure choice accuracy as the percentage of time (out of the 120-second trading stage) that a buyer spends on the surplus-maximizing choice, given the prevailing prices at each moment. Formally, let $g^*(t) = \arg\max_{g\in\{A,B,C,D,\emptyset\}}\big(v_{ig} - p_g(t)\big)$ denote the best option for buyer i at time t. Here, v_{ig} is buyer i's valuation for good g and $p_g(t)$ is the prevailing price. Choice accuracy for buyer i in a given period is then $\text{Acc}_i = \frac{1}{120}\int_0^{120} \mathbf{1}\big[c_i(t) = g^*(t)\big]\,dt$, with $c_i(t)$ the buyer's selected good at time t. Table 2 reports the average across groups. We compute this measure from the recorded events, re-evaluating the best option at each price change so that it matches the definition above. The same measure is used in all hypothesis tests.

We find that complexity substantially reduces choice accuracy. In Simple, buyers spend 61% of their time on the best choice, while in Complex the share drops to just 40%, a difference of 21 percentage points ($p < 0.01$). Perfect algorithmic advice raises choice accuracy to 66%, significantly

Table 2: Percentage of time buyers spend on the best available choice (% of 120s)

	No Advice	Perfect Advice	Imperfect Advice
Simple	61 (3.3)	—	—
Complex	40 (2.8)	66 (3.5)	36 (3.4)

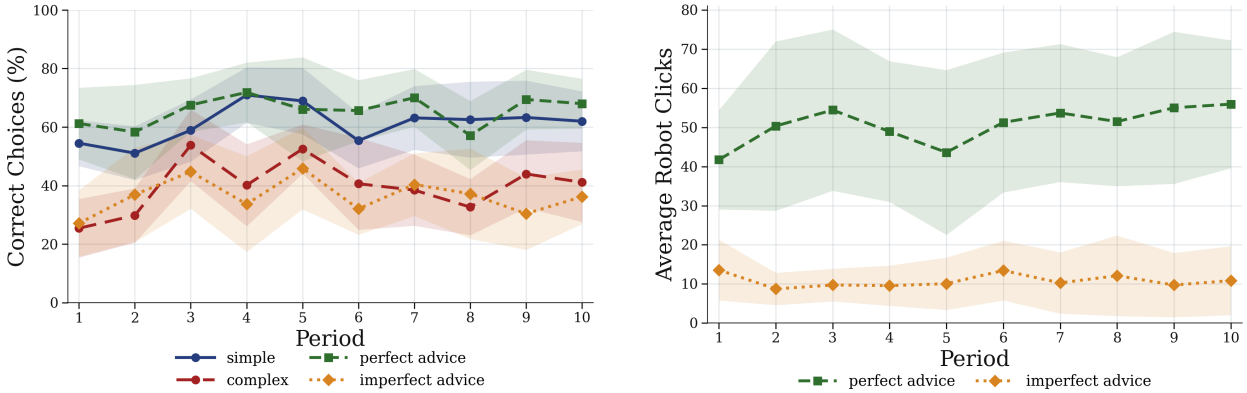
Notes: Each cell reports the average percentage of time during the 120-second trading stage that buyers spend on the option offering the highest payoff, with the best option re-evaluated at each price change. Standard errors of the group-level mean in parentheses; $n = 12$ independent groups per treatment. Dashes indicate treatments not implemented: algorithmic advice was administered only in the complex environment (see Table 1).

more than the 40% in Complex (difference: 26 pp; $p < 0.01$). We cannot reject equality between accuracy under Perfect and the 61% in Simple ($p = 0.42$). Accuracy under Perfect does not reach 100% because prices in a market change about 21 times per trading stage (each of the two sellers adjusts about 10 times; Section 4.3.1), and there is unavoidable latency between a price change and the buyer's next click. As Figure 3(b) shows, buyers in Perfect click the robot more than 50 times per trading stage, far exceeding the frequency of price changes. Imperfect advice does not improve choice accuracy: buyers in Imperfect spend only 36% of their time on the best choice, numerically lower than in Complex without any advice (difference: −3 pp; not statistically significant, $p = 0.39$).

Figure 3(a) plots choice accuracy across the ten periods. Treatment rankings are stable: Perfect is highest, followed by Simple, Complex, and Imperfect, with no evidence that learning closes the gaps; if anything, the Perfect–Imperfect gap grows in later periods.

Figure 3: Choice accuracy and robot use across periods

(a) Percentage of time on the best choice **(b) Average robot use per trading stage**



Notes: Panel (a) plots the average percentage of time that buyers spend on the surplus-maximizing choice in a given period, by treatment. Panel (b) plots the average number of times buyers click the "Ask Robot" button per two-minute trading stage, by treatment and period. Buyers in Simple and Complex (without advice) have no robot button. Shaded bands are 95% confidence intervals based on the 12 group-period means in each treatment-period cell (t-distribution, df = 11).

4.1.2 Robot Use

Buyers in PERFECT click the "Ask Robot" button far more often (about 50 times per trading stage) than in IMPERFECT (about 11), consistent with reduced willingness to consult an unreliable recommendation. Prices in a market change about 20 times per trading stage (roughly ten changes by each of the two sellers; Section 4.3.1). Robot use in PERFECT is more than twice this frequency, so buyers must often click the robot even when prices have not changed since their last click. Robot use in IMPERFECT is roughly half this frequency, so imperfect-advice buyers on average do not re-consult the robot after every price change. Figure 3(b) plots robot-use rates across the ten periods.

4.1.3 Cost of Errors

Result 2. Perfect advice cuts foregone surplus close to the SIMPLE baseline; imperfect advice does not.

We next show that choice inaccuracy translates into foregone surplus. We define the proportion of buyer surplus foregone as the fraction of the available buyer surplus (given sellers' prices) that is not obtained due to suboptimal choices. Formally, let $\pi_i^*(t) = v_{i,g^*(t)} - p_{g^*(t)}(t)$ denote the maximum available payoff for buyer i at time t, and let $\pi_i(t) = v_{i,c_i(t)} - p_{c_i(t)}(t)$ denote the actual payoff from the buyer's chosen good. The proportion of surplus foregone is:

$$\text{Foregone}_i = \frac{\int_0^{120} \left[\pi_i^*(t) - \pi_i(t)\right] dt}{\int_0^{120} \pi_i^*(t)\, dt}.$$

The measure captures how much of the available buyer surplus is lost due to suboptimal choice at each moment, time-weighted across the 120-second trading stage. Table 3 reports the averages.

Table 3: Proportion of available buyer surplus foregone (%)

	No Advice	Perfect Advice	Imperfect Advice
SIMPLE	17 (4.2)	—	—
COMPLEX	28 (4.6)	13 (3.4)	44 (6.8)

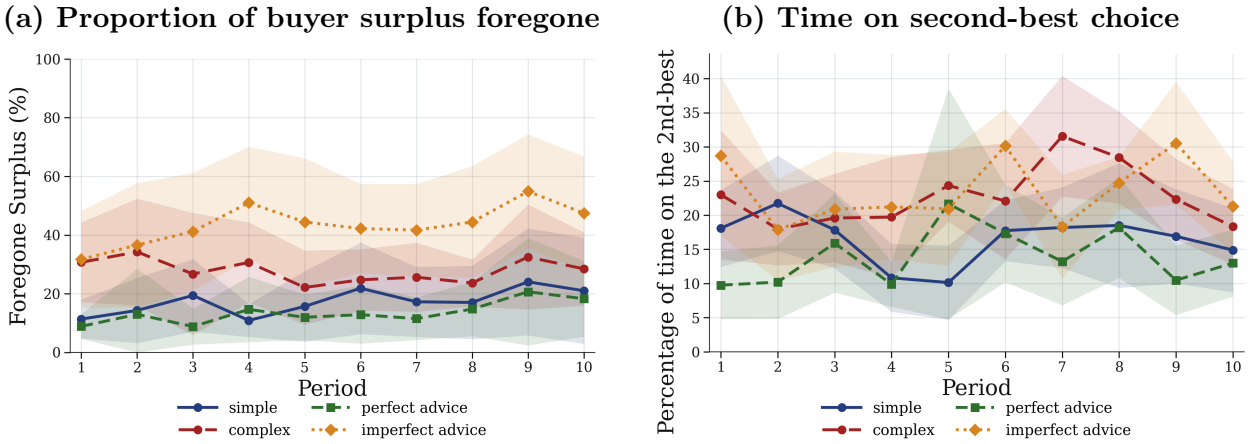
Notes: Each cell reports the average proportion of available buyer surplus that is foregone due to suboptimal buyer choices. Standard errors of the group-level mean in parentheses; $n = 12$ independent groups per treatment. Dashes indicate treatments not implemented (see Table 1).

In SIMPLE, buyers forgo 17% of their available surplus. Complexity raises the share to 28% (difference: 11 pp; $p = 0.07$). Perfect advice reduces foregone surplus to 13%, significantly lower than the 28% in COMPLEX (difference: -15 pp; $p < 0.01$) and we cannot reject equality with the 17% in SIMPLE ($p = 0.53$). Foregone surplus under imperfect advice is 44%, 16 pp above COMPLEX without advice ($p = 0.11$) and more than three times the level under perfect advice ($p < 0.01$).

Figure 4 shows the evolution of foregone surplus across periods. The pattern mirrors the choice

accuracy results in Figure 3(a): PERFECT consistently lowest, IMPERFECT highest, no convergence over time.

Figure 4: Foregone surplus and second-best reliance across periods



Notes: Panel (a) plots the average proportion of available buyer surplus foregone in a given period, by treatment. Panel (b) plots the average percentage of time buyers spend on the second-ranked option in a given period, by treatment. Shaded bands are 95% confidence intervals based on the 12 group-period means in each treatment-period cell (t-distribution, df = 11).

4.1.4 Within-Period Analysis

Result 3. Imperfect advice does not harm buyers by steering them onto the robot's recommendation; the harm comes from broadly degraded choice.

A natural conjecture is that imperfect advice harms buyers primarily by steering them toward the second-best option, the good the robot explicitly recommends. The decomposition does not support this steering conjecture: as under complexity, almost all of the loss comes from time on options even worse than the second-best.

The continuous-time structure of our experiment generates the data behind this decomposition. For each two-minute trading stage we track which good each buyer selects at every moment, compute the rank of the selected good (1st-best through 4th-best, or "buy nothing") given the prevailing prices, and calculate the time spent on each rank. The analysis uses the complete within-period data covering all 48 groups (12 per treatment). Table 4 reports the average time allocation across the five ranks for each treatment.

First, buyers in IMPERFECT spend 23.5% of their time on the second-best option (the good the robot recommends), close to the 22.7% in the no-advice COMPLEX condition; we cannot reject equality of the two ($p = 0.62$). There is no evidence that the robot draws buyers onto its recommendation beyond what complexity alone produces, though with 12 groups per treatment we cannot rule out a small steering effect. Second, complexity spreads time across all ranks: buyers in COMPLEX spend substantial time across the second, third, and fourth-best options (22.7%, 14.3%, 11.7%), with no concentration on any single inferior option. Third, perfect advice concentrates

Table 4: Time allocation by choice rank (% of 120-second trading stage)

	1st (best)	2nd	3rd	4th	Outside option
SIMPLE	61.1	16.5	7.4	6.7	8.4
COMPLEX	39.8	22.7	14.3	11.7	11.4
PERFECT	65.5	14.0	4.6	5.9	10.0
IMPERFECT	36.4	23.5	14.2	17.8	8.1

Notes: Each cell reports the average percentage of time during the 120-second trading stage that buyers spend on the option with the indicated rank, given prevailing prices. Rankings are computed at each event (price change or choice change). The first-rank column is the choice-accuracy measure reported in Table 2 (Section 4.1.1), computed the same way. $n = 12$ independent groups per treatment.

time on the best option: PERFECT buyers spend 65.5% of their time on the best choice (the highest of any treatment) and only 14.0% on the second-best (the lowest).

Table 5: Foregone buyer surplus decomposition (tokens per second, time-weighted)

	Total loss	From 2nd-best	From 3rd+
SIMPLE	5,723	698 (12.2%)	5,025 (87.8%)
COMPLEX	7,360	988 (13.4%)	6,372 (86.6%)
PERFECT	4,949	536 (10.8%)	4,413 (89.2%)
IMPERFECT	9,589	1,178 (12.3%)	8,411 (87.7%)

Notes: Loss is measured in tokens per second, time-weighted across all events within a trading stage. "From 2nd-best" is the loss incurred while buyers hold the second-ranked option. "From 3rd+" is the loss from third, fourth, and fifth-ranked (buy nothing) options. $n = 12$ groups per treatment.

We decompose total foregone buyer surplus into loss from time on the second-best option versus loss from time on third-ranked or worse options (Table 5). Total losses in IMPERFECT are 68% higher than in SIMPLE (9,589 vs. 5,723 tokens/second) and 30% higher than in COMPLEX (7,360). Loss attributable to time on the second-best option is only 12.3% of the total, no larger than its share in the no-advice COMPLEX (13.4%) or SIMPLE (12.2%) conditions, while 87.7% comes from time on third-ranked or worse options. This composition gives no indication that imperfect advice operates by steering buyers onto the robot's recommendation; as under complexity, the loss is spread across far-inferior options. By contrast, PERFECT has both the lowest total loss (4,949) and the lowest share of loss from second-best choices (10.8%).

Section 2 provides a model-consistent interpretation. The imperfect-robot regime structure in Eq. (7) distinguishes two regimes. In the low-η regime, in which evaluation is relatively sharp, the buyer uses the recommendation to exclude the second-best and solves a reduced-dimension optimization. In the intermediate-or-higher-η regime, in which evaluation is noisier, the buyer follows

the recommendation. The within-period evidence does not support the follow regime: time on the second-best in IMPERFECT matches COMPLEX (23.5% vs 22.7%), and the share of loss attributable to second-best choices is no higher (Table 5). We cannot directly distinguish stale following from ignoring or active exclusion, because the data do not record a timestamped recommendation history. The aggregate patterns nevertheless do not support steering as the dominant mechanism, since time on the second-best option in IMPERFECT closely matches the no-advice COMPLEX condition and persists as robot use falls.

Figure 4(b) plots the share of time spent on the second-best choice across the ten periods. The second-best share in IMPERFECT tracks that of the no-advice COMPLEX treatment throughout, with both running well above PERFECT and SIMPLE. The pattern in IMPERFECT does not change as active robot use drops sharply after period 1 (Figure 3(b)), reinforcing the within-period finding that complexity, not the robot's recommendation, drives the elevated second-best share. PERFECT buyers show among the lowest second-best shares in most periods, consistent with perfect advice steering buyers away from suboptimal options.

4.2 Buyer Surplus

Result 4. Complexity reduces buyer surplus, perfect advice partially recovers it, and imperfect advice does not.

Table 6: Average buyer payoffs (A$)

	No Advice	Perfect Advice	Imperfect Advice
SIMPLE	22.3 (2.09)	—	—
COMPLEX	16.6 (2.20)	20.0 (1.85)	14.0 (2.61)

Notes: Each cell reports the average buyer payoff in Australian dollars. Standard errors of the group-level mean in parentheses; $n = 12$ independent groups per treatment. Dashes indicate treatments not implemented (see Table 1).

Complexity reduces buyer surplus by A$5.7 (from A$22.3 to A$16.6, Table 6), a decline of 26% ($p = 0.09$). The drop is economically large: more than a quarter of buyer surplus is lost due to complexity alone.

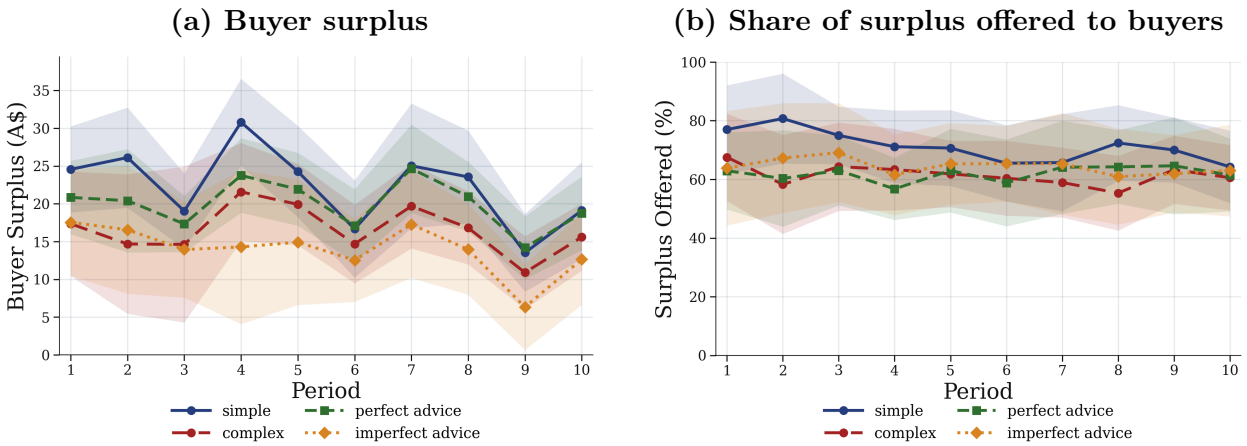
Perfect advice increases buyer surplus from A$16.6 to A$20.0, an increase of A$3.4 (20% above the COMPLEX baseline). The group-level difference is not statistically significant at conventional levels ($p = 0.30$), reflecting the relatively small number of independent group observations ($n = 12$). The individual-level regression in Table 7 is consistent with the perfect-advice gain being concentrated among active robot users: the PERFECT $\times$ advice-use interaction is positive and marginally significant with cluster-robust standard errors ($p = 0.070$).

Buyer surplus under IMPERFECT is A$14.0, A$2.6 below COMPLEX on average, but we cannot reject equality with the no-advice baseline at the group level ($p = 0.33$). The difference between

PERFECT and IMPERFECT is substantial (buyer surplus under perfect advice is roughly 43% higher) and marginally significant ($p = 0.06$).

Figure 5(a) plots buyer surplus across all ten periods. Buyer surplus under PERFECT moves toward the SIMPLE treatment level, consistent with perfect algorithmic advice partially recovering the buyer-surplus loss from complexity. The IMPERFECT treatment yields the lowest buyer surplus in 8 out of 10 periods, with no narrowing of the gap over time.

Figure 5: Buyer surplus and surplus share across periods



Notes: Panel (a) plots the average buyer surplus (in A$) per period, by treatment. Panel (b) plots the average share of the maximum available buyer surplus that sellers' pricing makes available to buyers, by treatment and period; this measure is defined formally in Section 4.3.1. Shaded bands are 95% confidence intervals based on the 12 group-period means in each treatment-period cell (t-distribution, df = 11).

In the buyer-payoffs regression with standard errors clustered at the group level (Table 7), complexity reduces buyer payoffs by A$5.69 ($p = 0.054$). Both advice treatments carry a negative baseline coefficient, but these are evaluated at zero advice use, far below the average usage in both treatments (about 50 clicks per trading stage under perfect advice and 11 under imperfect advice). The PERFECT baseline (-5.19) is close to the COMPLEX coefficient (-5.69): a buyer who does not consult the perfect robot faces the same complexity as a no-advice buyer, so the negative intercept reflects the cost of complexity rather than a cost of robot access. The interaction between PERFECT and advice use is positive and marginally significant (0.057, $p = 0.070$): buyers who consult the perfect robot more often earn more. Because advice use is a buyer choice rather than a randomly assigned treatment, this interaction is associational and cannot separate the causal effect of consulting the robot from selection by more engaged or more able buyers. The corresponding interaction for imperfect advice is not significant: greater use of the imperfect robot shows no detectable association with payoffs.

Table 7: Buyer payoffs regression

	Buyer Payoffs
Period	-0.526***
	(0.18)
Complex	-5.685*
	(2.95)
Perfect Advice	-5.188
	(3.38)
Imperfect Advice	-8.928***
	(3.03)
Perfect Advice × Advice Use	0.057*
	(0.03)
Imperfect Advice × Advice Use	0.061
	(0.05)
Constant	25.159***
	(2.05)
N	960

Notes: Random-effects GLS estimates with standard errors clustered at the group level (48 clusters, the unit of independent sampling). The dependent variable is the buyer's earnings in A$. The omitted category is the SIMPLE treatment. "Advice Use" is the number of times the buyer clicked the robot button in a given period. Standard errors in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

4.3 Seller Surplus

4.3.1 Seller Pricing Strategy

Result 5. Sellers do not adjust pricing in response to buyer sophistication, consistent with the standard benchmark (H1).

To measure the level margin, we examine the share of surplus that sellers' pricing makes available to buyers: the maximum buyer payoff achievable at posted prices, as a fraction of the maximum achievable at the surplus-maximizing prices. Because this measure does not depend on which good the buyer actually selects, it isolates the supply-side pricing channel from buyers' choice quality.

Figure 5(b) shows that sellers in SIMPLE offer a numerically larger share of surplus to buyers (around 70%) than in the complex treatments (around 60–65%), but this difference is not statistically significant (SIMPLE vs COMPLEX, $p = 0.17$). No pairwise difference in the share offered reaches significance ($p > 0.10$ for all comparisons), including between COMPLEX and PERFECT: even though buyers in PERFECT are well-informed, sellers do not extract a larger share of the surplus. Differences in buyer surplus across the complex treatments therefore reflect buyers failing to select the best option, not sellers exploiting buyer confusion through differences in the share of surplus made available.

Sellers also do not differentiate on the adjustment margin: average price changes per seller per

trading stage are similar across treatments (SIMPLE 9.1, COMPLEX 9.2, PERFECT 10.4, IMPERFECT 9.8; Figure 6(b)), with no pairwise Wilcoxon contrast significant ($p > 0.45$). The second-margin null reinforces the interpretation based on the standard benchmark.

4.3.2 Seller Payoffs

Result 6. We find no detectable effect of algorithmic advice on seller surplus.

Table 8: Average seller payoffs (A$)

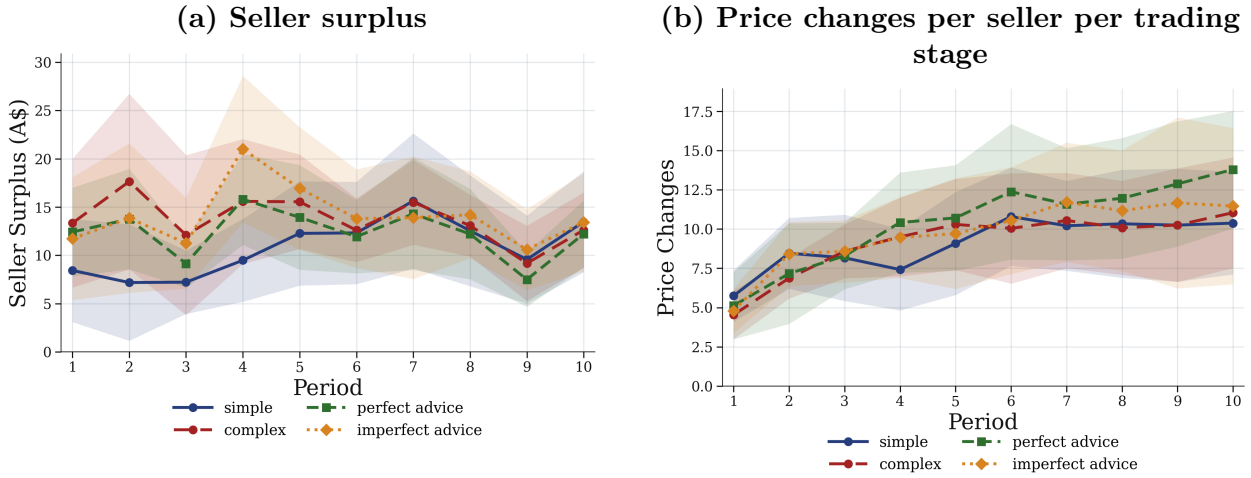
	No Advice	Perfect Advice	Imperfect Advice
SIMPLE	10.8 (1.84)	—	—
COMPLEX	13.7 (1.80)	12.3 (1.57)	14.1 (2.00)

Notes: Each cell reports the average seller payoff in Australian dollars. Standard errors of the group-level mean in parentheses; $n = 12$ independent groups per treatment. Dashes indicate treatments not implemented (see Table 1).

Average seller payoffs (Table 8) do not differ significantly between COMPLEX (A$13.7) and SIMPLE (A$10.8; $p = 0.39$). Across the three complex treatments, seller surplus is not significantly affected by the type of algorithmic advice available to buyers: payoffs are A$12.3 under perfect advice and A$14.1 under imperfect advice, neither significantly different from the A$13.7 in COMPLEX without advice.

Figure 6(a) plots seller surplus across periods. Seller surplus is initially lower in SIMPLE but converges toward the complex treatments by period six, after which all four treatments exhibit roughly similar seller surplus levels.

Figure 6: Seller surplus and price adjustment frequency across periods



Notes: Panel (a) plots the average seller surplus (in A$) per period, by treatment. Panel (b) plots the average number of price changes per seller per two-minute trading stage, by treatment and period. Shaded bands are 95% confidence intervals based on the 12 group-period means in each treatment-period cell (t-distribution, df = 11).

4.4 Total Welfare

Result 7. The welfare ranking mirrors buyer choice accuracy: perfect advice nearly eliminates the loss from complexity, while imperfect advice does not.

Total welfare is the sum of all four participants' payoffs in a group. Because each price is a transfer from a buyer to a seller, prices cancel in this sum, so realized welfare equals the total realized valuation of the goods buyers hold, independent of how surplus is divided between buyers and sellers. Welfare therefore depends on the valuation of the good a buyer holds, not on the buyer's payoff net of price, and is highest when buyers hold the goods they value most. Welfare differences across treatments thus reflect allocative efficiency, not the split of surplus.

Table 9: Average total welfare (A$)

	No Advice	Perfect Advice	Imperfect Advice
Simple	66.2 (1.20)	—	—
Complex	60.6 (1.49)	64.6 (2.39)	56.1 (2.90)

Notes: Each cell reports the average total welfare (sum of both buyers' and both sellers' payoffs) in Australian dollars. Standard errors of the group-level mean in parentheses; $n = 12$ independent groups per treatment. Dashes indicate treatments not implemented (see Table 1).

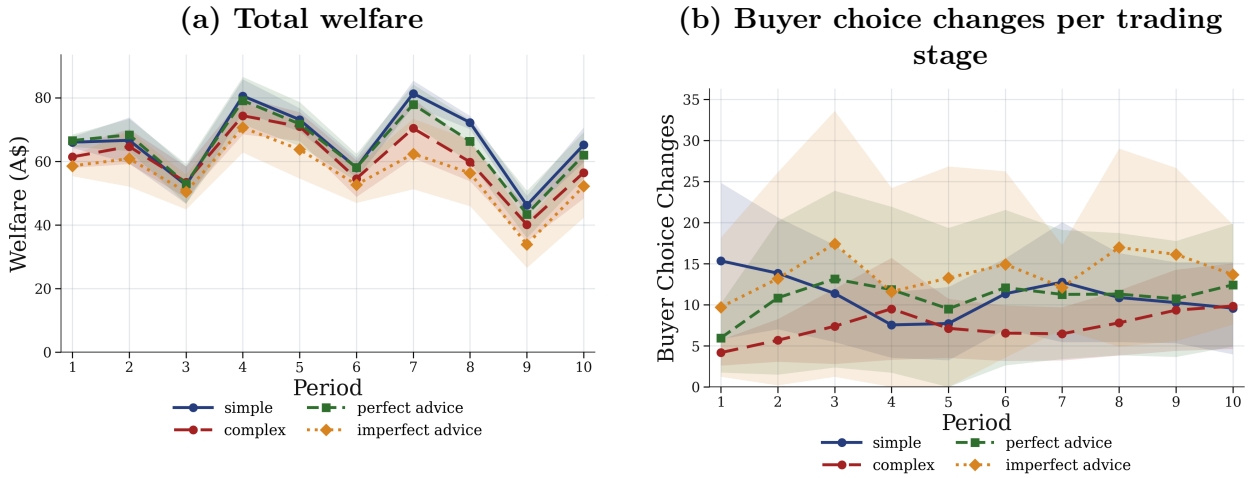
Complexity creates a significant deadweight loss of A$5.6 per market (from A$66.2 to A$60.6, Table 9; $p = 0.02$). Perfect advice nearly eliminates this loss, raising welfare to A$64.6 ($p = 0.01$ versus Complex). Welfare under Imperfect is A$56.1, A$4.5 below Complex on average, but we cannot reject equality with the no-advice baseline at the group level ($p = 0.25$); the contrast against Perfect is significant ($p < 0.01$).

The Simple–Complex welfare comparison is more precisely estimated than the corresponding buyer-surplus comparison ($p = 0.02$ versus $p = 0.09$) because total welfare aggregates all four participants' payoffs per group, averaging out idiosyncratic variation. The welfare ranking across treatments (Simple > Perfect > Complex > Imperfect) mirrors the buyer choice accuracy ranking, consistent with buyer decision quality being an important channel through which complexity and algorithmic advice affect welfare.

Figure 7(a) plots total welfare across periods, with the ranking stable across nearly all periods.

Figure 7(b) shows the frequency of buyer choice changes per trading stage. Buyers in Imperfect switch choices most frequently, particularly in later periods, consistent with buyers attempting to identify the right option when the robot's advice proves unreliable. Buyers in Simple start with the most switching but reduce it over time, consistent with learning.

Figure 7: Total welfare and buyer choice switching across periods



Notes: Panel (a) plots the average total welfare (in A$) per period, by treatment. Panel (b) plots the average number of buyer choice changes per two-minute trading stage, by treatment and period. Shaded bands are 95% confidence intervals based on the 12 group-period means in each treatment-period cell (t-distribution, df = 11).

5 Discussion

We conduct a laboratory experiment disciplined by a simple model to study how the availability and accuracy of algorithmic advice affect buyer behavior in the presence of product complexity. We find that algorithmic advice can substantially mitigate the welfare cost of product complexity, but only when the advice is accurate. Perfect advice improves buyer choices and total welfare relative to no advice, recovering choice accuracy to near simple-market levels. One might expect even an imperfect recommendation to help, since a buyer told which option is second-best could use that information to narrow the choice set. We find no such benefit. Imperfect advice leaves buyer choices, surplus, and welfare no better than no advice, even though buyers are told exactly how the robot is imperfect. Accurate advice, by contrast, improves buyer outcomes and total surplus.

Our results have implications for both policy and managerial practice. First, accuracy standards matter alongside transparency requirements. Our design is a best case for disclosure-based regulation, since buyers know exactly what the robot does, yet disclosure alone does not protect them (Cain et al. 2005); recommendation accuracy therefore also merits regulatory attention. Second, the welfare gains from better advice need not be zero-sum, since perfect advice raises total welfare, suggesting that the increase in buyer surplus overcompensates for the potential decrease in seller surplus. Third, for firms that provide recommendation tools, two less obvious points follow. Disclosing a tool's imperfection does not substitute for making it accurate, since buyers here knew exactly how the robot was imperfect and were still no better off than with no tool at all. And an unreliable tool tends to be used less rather than worked around, since buyers consulted the imperfect robot far less than the accurate one.

In an era of inexpensive computing power and easily accessible artificial intelligence, algorithmic advice will mediate a growing share of consumer choice in markets in which complexity is itself

a strategic instrument. Our results suggest that the welfare effects of these systems depend on recommendation accuracy, and that transparency about imperfection is not a sufficient regulatory safeguard.

Our design has limitations that point to natural extensions. Our complexity manipulation is computational: the relevant information is available but costly to process. This maps well to markets in which comparison services are especially relevant, such as energy tariffs, telecommunications plans, and financial products (Wilson and Waddams Price 2010, Kalaycı 2015, Choi et al. 2010), but may not capture complexity arising from qualitative uncertainty or from product attributes consumers genuinely value, such as variety, discounts, or partitioned prices (Simonson 1990, Lichtenstein et al. 1990, Morwitz et al. 1998). Relatedly, because buyers choose under a fixed clock, our time-based accuracy measure combines two channels: avoiding mistakes and saving computation time. Future work could separately identify these mechanisms and measure which product features or recommendation cues become salient to buyers, especially since salient cues may mitigate the effects of complexity (Dertwinkel-Kalt and Köster 2024). Our imperfect robot gives a deterministic second-best recommendation, whereas real comparison services may exhibit more elusive imperfections such as omitting options or adding noise; whether such less transparent forms of imperfection amplify the demand-side harm we document is a natural next question.

Notes

[1]A general equilibrium involving multiple strategic buyers and sellers lacks tractability and clean comparative statics, so we do not go in this direction.

[2]Observe that

$$\max\left\{u_{\hat{j}}, \max_{\rho\in\Delta(\hat{K})}\left[\left(\sum_{j\in\hat{K}}\rho_j(v_j-P_j)\right)-\eta\hat{H}(\rho)\right]\right\} \geq \max_{\rho\in\Delta(K)}\left[\left(\sum_{j\in K}\rho_j(v_j-P_j)\right)-\eta H(\rho)\right],$$

that is, the better of following the robot and solving the reduced-dimension problem is weakly preferred to not consulting the robot at all.

[3]This experiment was not pre-registered. Power calculations were based on choice accuracy, for which treatment differences are large and precisely estimated.

Data Availability Statement. The experimental data, z-Tree programs, Stata analysis scripts, and full replication package for this study are available at `https://github.com/kenankalayci/robot-advice`.

References

Simon P. Anderson, André De Palma, and Jacques-François Thisse. *Discrete choice theory of product differentiation*. MIT Press, 1992.

Mark Armstrong and Jidong Zhou. Paying for prominence. *Economic Journal*, 121(556):F368–F395, 2011.

Tibor Besedeš, Cary Deck, Sudipta Sarangi, and Mikhael Shor. Reducing choice overload without reducing choices. *Review of Economics and Statistics*, 97(4):793–802, 2015.

Jennifer Brown, Tanjim Hossain, and John Morgan. Shrouded attributes and information suppression: Evidence from the field. *Quarterly Journal of Economics*, 125(2):859–876, 2010.

Daylian M. Cain, George Loewenstein, and Don A. Moore. The dirt on coming clean: Perverse effects of disclosing conflicts of interest. *Journal of Legal Studies*, 34(1):1–25, 2005.

Bruce Ian Carlin. Strategic price complexity in retail financial markets. *Journal of Financial Economics*, 91(3):278–287, 2009.

Timothy N. Cason and Daniel Friedman. Buyer search and price dispersion: a laboratory study. *Journal of Economic Theory*, 112(2):232–260, 2003.

Claire Célérier and Boris Vallée. Catering to investors through security design: Headline rate and complexity. *Quarterly Journal of Economics*, 132(3):1469–1508, 2017.

Ioana Chioveanu and Jidong Zhou. Price competition with consumer confusion. *Management Science*, 59(11):2450–2469, 2013.

James J. Choi, David Laibson, and Brigitte C. Madrian. Why does the law of one price fail? An experiment on index mutual funds. *Review of Financial Studies*, 23(4):1405–1432, 2010.

Paolo Crosetto and Alexia Gaudeul. Choosing not to compete: Can firms maintain high prices by confusing consumers? *Journal of Economics & Management Strategy*, 26(4):897–922, 2017.

Alexandre de Cornière and Greg Taylor. A model of biased intermediation. *The RAND Journal of Economics*, 50(4):854–882, 2019.

Alexandre de Cornière and Greg Taylor. Data and competition: A simple framework. *The RAND Journal of Economics*, 56(4):494–510, 2025.

Markus Dertwinkel-Kalt and Mats Köster. Salient cues and complexity. *Management Science*, 71 (1):428–469, 2024.

Peter A. Diamond. A model of price adjustment. *Journal of Economic Theory*, 3(2):156–168, 1971.

Berkeley J. Dietvorst, Joseph P. Simmons, and Cade Massey. Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1):114–126, 2015.

Berkeley J. Dietvorst, Joseph P. Simmons, and Cade Massey. Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, 64(3):1155–1170, 2018.

Uwe Dulleck and Rudolf Kerschbamer. On doctors, mechanics, and computer specialists: The economics of credence goods. *Journal of Economic Literature*, 44(1):5–42, 2006.

Glenn Ellison and Sara Fisher Ellison. Search, obfuscation, and price elasticities on the Internet. *Econometrica*, 77(2):427–452, 2009.

Glenn Ellison and Alexander Wolitzky. A search cost model of obfuscation. *The RAND Journal of Economics*, 43(3):417–441, 2012.

Nisvan Erkal, Lata Gangadharan, and Nikos Nikiforakis. Relative earnings and giving in a real-effort experiment. *American Economic Review*, 101(7):3330–3348, 2011.

Urs Fischbacher. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178, 2007.

Gavan J. Fitzsimons and Donald R. Lehmann. Reactance to recommendations: When unsolicited advice yields contrary responses. *Marketing Science*, 23(1):82–94, 2004.

Richard G. Frank and Karine Lamiraud. Choice, price competition and complexity in markets for health insurance. *Journal of Economic Behavior & Organization*, 71(2):550–562, 2009.

Marian Friestad and Peter Wright. The persuasion knowledge model: How people cope with persuasion attempts. *Journal of Consumer Research*, 21(1):1–31, 1994.

Xavier Gabaix and David Laibson. Shrouded attributes, consumer myopia, and information suppression in competitive markets. *Quarterly Journal of Economics*, 121(2):505–540, 2006.

Kate Goddard, Abdul Roudsari, and Jeremy C. Wyatt. Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association*, 19(1):121–127, 2012.

Yiquan Gu and Tobias Wenzel. Strategic obfuscation and consumer protection policy. *Journal of Industrial Economics*, 63(3):397–428, 2015.

Thomas Haan and Jona Linde. 'good nudge lullaby': Choice architecture and default bias reinforcement. *Economic Journal*, 128(610):1180–1206, 2018.

Andrei Hagiu and Bruno Jullien. Why do intermediaries divert search? *The RAND Journal of Economics*, 42(2):337–362, 2011.

Paul Heidhues, Johannes Johnen, and Botond Kőszegi. Browsing versus studying: A pro-market case for regulation. *Review of Economic Studies*, 88(2):708–729, 2021.

Paul Heidhues, Mats Köster, and Botond Kőszegi. Steering fallible consumers. *The Economic Journal*, 133(652):1430–1465, 2023.

Roman Inderst and Marco Ottaviani. How (not) to pay for advice: A framework for consumer financial protection. *Journal of Financial Economics*, 105(2):393–411, 2012.

Kenan Kalaycı. Price complexity and buyer confusion in markets. *Journal of Economic Behavior & Organization*, 111:154–168, 2015.

Kenan Kalaycı and Jan Potters. Buyer confusion and market prices. *International Journal of Industrial Organization*, 29(1):14–22, 2011.

Kenan Kalaycı and Marta Serra-Garcia. Complexity and biases. *Experimental Economics*, 19(1): 31–50, 2016.

Gyudong Lee and Won Jun Lee. Psychological reactance to online recommendation services. *Information & Management*, 46(8):448–452, 2009.

Donald R. Lichtenstein, Richard G. Netemeyer, and Scot Burton. Distinguishing coupon proneness from value consciousness: An acquisition-transaction utility theory perspective. *Journal of Marketing*, 54(3):54–67, 1990.

Jennifer M. Logg, Julia A. Minson, and Don A. Moore. Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151: 90–103, 2019.

Filip Matějka and Alisdair McKay. Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–298, 2015.

John Morgan, Henrik Orzen, and Martin Sefton. An experimental study of price dispersion. *Games and Economic Behavior*, 54(1):134–158, 2006.

Vicki G. Morwitz, Eric A. Greenleaf, and Eric J. Johnson. Divide and prosper: Consumers' reactions to partitioned prices. *Journal of Marketing Research*, 35(4):453–463, 1998.

Hans-Theo Normann and Martin Sternberg. Human-algorithm interaction: Algorithmic pricing in hybrid laboratory markets. *European Economic Review*, 152:104347, 2023.

Raja Parasuraman and Dietrich H. Manzey. Complacency and bias in human use of automation: An attentional integration. *Human Factors*, 52(3):381–410, 2010.

Michele Piccione and Ran Spiegler. Price competition under limited comparability. *Quarterly Journal of Economics*, 127(1):97–135, 2012.

Itamar Simonson. The effect of purchase quantity and timing on variety-seeking behavior. *Journal of Marketing Research*, 27(2):150–162, 1990.

Stefania Sitzia, Jiwei Zheng, and Daniel John Zizzo. Inattentive consumers in markets for services. *Theory and Decision*, 79(2):307–332, 2015.

Dale O. Stahl. Oligopolistic pricing with sequential consumer search. *American Economic Review*, 79(4):700–712, 1989.

Chris M. Wilson and Catherine Waddams Price. Do consumers switch to the best supplier? *Oxford Economic Papers*, 62(4):647–668, 2010.

A Instructions

We provide the instructions for PERFECT here. The instructions for the other treatments only differ in the description of valuation (complexity) and types of robot and they are available upon request.

Instructions

Welcome! Please put your phone on silent and put all your belongings on the floor. The experiment will start soon. You will be paid for participating in the experiment in private and in cash at the end of the experiment. The amount of money you earn depends on the decisions that you and the other participants make and on chance. If you have any questions, please raise your hand. Please do not communicate with other participants during the experiment. During the experiment, you are not allowed to use a calculator other than the calculator provided by the experimenter.

Procedures

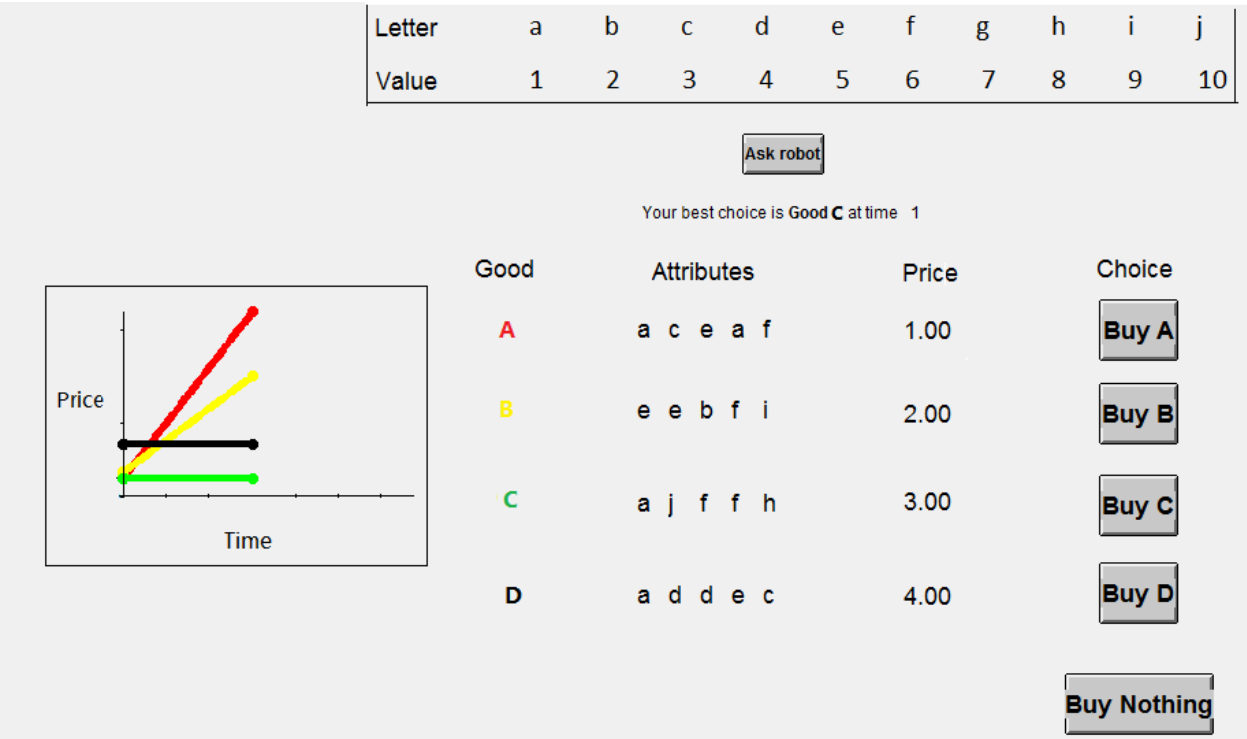
The experiment consists of a short quiz, ten market rounds, and a questionnaire. Each of the ten market rounds follow the same instructions below.

Market structure

You will be randomly matched with three other players to form a market of four people. Two of you will be sellers and the other two will be buyers. The roles of buyers and sellers are randomly assigned by the computer, and they are fixed during the experiment. You will be informed about whether you are a seller or a buyer before the beginning of the first market round. Each seller sells two goods, while each buyer buys at most one good at one time.

If you are a buyer:

You will buy at most one of the four goods in the market at a time. Below is an example of the screen you will see during the market rounds.



Your valuation for each good is the *summation* of its five attributes, which are randomly drawn by the computer. Your payoff (points) for each second is your valuation *minus* the price of the good that you are buying. Please note that your payoff can be negative if the price is higher than your valuation. If you buy nothing, then you will earn 0 points per second.

Example 1 (the case in the figure): Letter a is worth 1, letter b is worth 2, letter c is worth 3 points and so on (please see the figure). Good A has attributes *aceaf*, so your valuation of Good A is the *summation of its five attributes*$= 1+3+5+1+6 = 16$ points. The price of Good A is 1 and if you buy it, then your payoff at this second is *your valuation minus the price*$= 16-1 = 15$ points.

Example 2 (the case in the figure): Suppose the values of attributes are the same as in example 1. If Good D has attributes *addec*, then your valuation of Good D is the *summation of its five attributes*$= 1+4+4+5+3 = 17$ points. The price of Good D is 4 and you buy it, then your payoff at this second is *your valuation minus the price*$= 17-4 = 13$ points.

Ask robot button

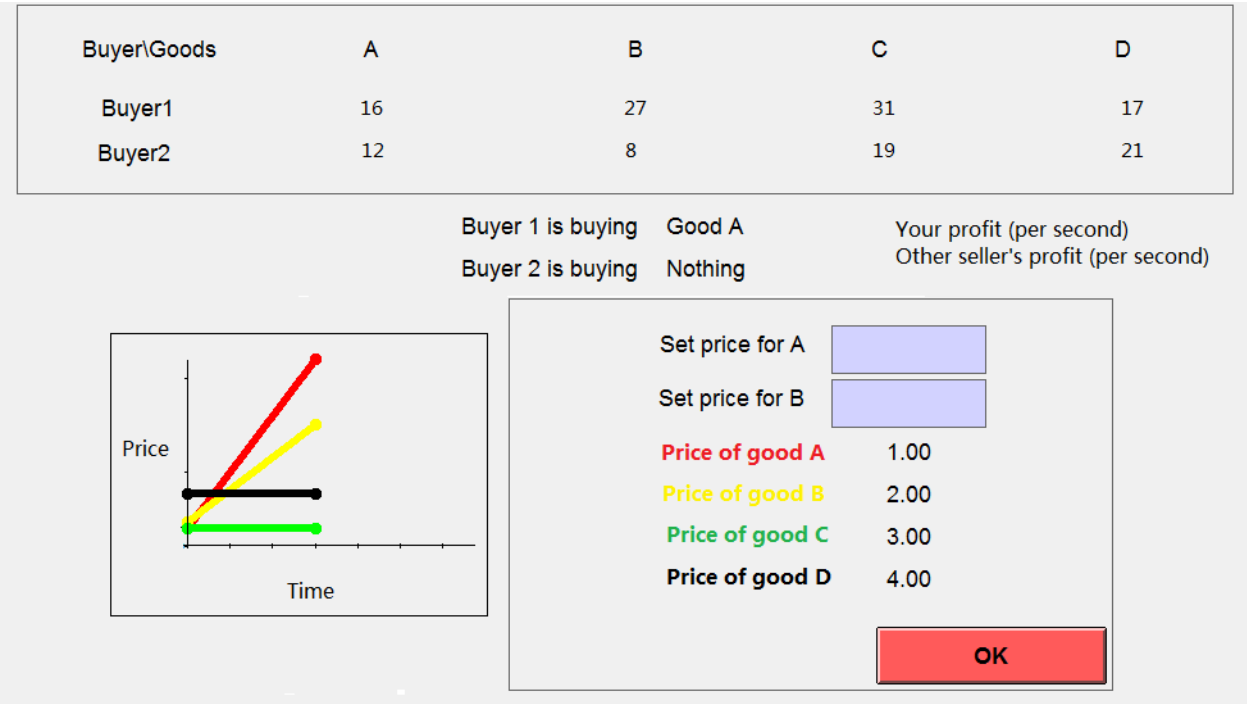
If you are a buyer, you have the option to ask a robot for recommendation. The robot is honest, and it will always tell you the best choice (i.e. the good that offers you the highest payoff) at the time you click the button. The **robot** is free of charge. Be aware that if the prices of any of the goods change your best choice may also change, and you will need to click the ask robot button again if you want to learn your new best choice.

At the left side of the interface, there is a price chart to document price history. The color of the line identifies the good. In particular, the red line represents the price of Good A, the yellow line represents the price of Good B, the green line represents the price of Good C, and the black

line represents the price of Good D.

If you are a seller:

You will be selling two goods (either Good A and B or Good C and D). You have to choose a price for each of the two goods you are selling. You will see the following window during the experimental rounds.



The upper table in the graph indicates the valuation of each good by each buyer. For instance, the valuation of Good A by Buyer 1 (16 points) that appears here is the same as the sum of the attributes of Good A of Buyer 1 (see example 1 in the buyer's instruction). Your payoff (points) for each second is the prices of the goods you are selling times quantity purchased by the buyers.

Example 1 (the case in the figure). Suppose you are selling Goods A and B, and your price for Good A is 1 and your price for Good B is 2. If Buyer 1 buys Good A and Buyer 2 buys nothing, then your payoff at this second is $price\ of\ A \times quantity\ of\ A + price\ of\ B \times quantity\ of\ B = 1 \times 1 + 2 \times 0 = 1$ points.

Example 2 (the case in the figure). Suppose you are selling Goods C and D, and your price for Good C is 3 and your price for Good D is 4. If Buyer 1 buys Good A and Buyer 2 buys nothing, then your payoff at this second is $price\ of\ C \times quantity\ of\ C + price\ of\ D \times quantity\ of\ D = 3 \times 0 + 4 \times 0 = 0$ points.

At the left side of the interface, there is a price chart to document price history. Like in the buyer's screen, the color of the line identifies the good.

Stage one

There are three stages in each round. In stage one, sellers are given 30 seconds to set the initial prices for their own goods. Both sellers know the two buyers' valuations over all four goods, but a seller cannot see the other seller's prices at this stage.

Stage two

After sellers have set the initial prices, buyers are given 30 seconds to decide which good to buy. If a buyer does not make a choice within 30 seconds it will be assumed that she chose the Buy Nothing option. After 30 seconds, stage three starts.

Stage three

In this stage, buyers and sellers interact continuously (i.e. sellers can change their prices and buyers can switch their choices) for two minutes. Sellers know the two buyers' valuations over all four goods (as in stage 1) in addition to the prices of the other seller and the two buyers' choices (i.e. sellers have full market information). Buyers however only know the attributes for four goods (i.e. they can use this to calculate their valuations of the four goods) and prices for the four goods (as in stage 2). Your payoff for the round accumulates only in this stage.

Payments

You will be paid for only ***one*** randomly chosen market round among the **ten** rounds. The total amount of money you will earn is the A$10 show-up fee *plus* accumulated points over the two minutes at **stage three** of the chosen round *divided by* the exchange rate of **100,000**.