Competitive Market Behavior: Convergence and Asymmetry in the Experimental Double Auction

Author(s):
Ikica, Barbara; Jantschgi, Simon; Nax, Heinrich; Nuñez Duran, Diego G.; Pradelski, Bary S.R.

Publication date:
2023-08

Permanent link:
https://doi.org/10.3929/ethz-b-000605280

Originally published in:
International Economic Review 64(3), https://doi.org/10.1111/iere.12630

Funding acknowledgement:
324247 - Modeling the Emergence of Social Complexity and Order: How Individual and Societal Complexity Co-Evolve (EC)
654024 - SoBigData Research Infrastructure (SBFI)

INTERNATIONAL ECONOMIC REVIEW
Vol. 64, No. 3, August 2023 DOI: 10.1111/iere.12630

COMPETITIVE MARKET BEHAVIOR: CONVERGENCE AND ASYMMETRY IN THE EXPERIMENTAL DOUBLE AUCTION*

By Barbara Ikica, Simon Jantschgi, Heinrich H. Nax, Diego G. Nuñez Duran, and Bary S. R. Pradelski

University of Zurich & ETH Zurich, Switzerland; CNRS & University Grenoble Alpes, France

We conducted a large number of controlled continuous double auction experiments to reproduce and stress-test the phenomenon of convergence to competitive equilibrium under private information with decentralized trading feedback. Our main finding is that across a total of 104 markets (involving over 1,700 subjects), convergence occurs after a handful of trading periods. Initially, however, there is an inherent asymmetry that favors buyers, typically resulting in prices below equilibrium levels. Analysis of over 80,000 observations of individual bids and asks helps identify empirical ingredients contributing to the observed phenomena including higher levels of aggressiveness initially among buyers than sellers.

1. INTRODUCTION

The double auction (DA) is a ubiquitous market institution which has been at the core of economic investigation, reminiscent of the Walrasian auction, where the notion of competitive equilibrium (CE) originates.[1] A central question in economics has since been whether CE is a good predictor of market outcomes in the DA, and whether deviations from CE ought to be expected temporarily or permanently.

Implemented as continuous DAs for single items of nondurable goods, the experiments by Smith (1962) occupy a special place in the history of economics, providing a first set of experimental counterexamples to the conventional wisdom that ruled at the time according to which CE should not be expected when traders have only purely private—not full—information regarding reservation prices: *"I am still recovering from the shock of the experimental results.*

*Manuscript received May 2020; revised September 2022.

We thank Dan Friedman and Peyton Young for their guidance with experiments and theory. We thank ETH DeSciL's Oliver Braegger, Joris Stemmle, and Stefan Wehrli for lab assistance. We thank Dirk Helbing, the ETH COSS team, and SciOn's demo users for supporting and testing our trading platform. We are grateful for comments by Mikhail Anufriev, David Rojo Arjona, Elena Asparohova, Paul Brewer, John Duffy, Dan Gode, Paul Klemperer, Meg Meyer, Rosemarie Nagel, Tibor Neugebauer, Georg Nöldeke, David Porter, Sven Seuken, Vernon Smith, Bernhard von Stengel, Shyam Sunder, and Roberto Weber, as well as for suggestions from anonymous reviewers and from participants of seminars at UC Irvine, UC Santa Barbara, Ecole Polytechnique, Paris Dauphine, ETH Zurich, the DPG Spring Meetings, the Learning-Evolution-Games Conference, Goethe University Frankfurt, Ca'Foscari University Venice, the Paris Game Theory Seminar at IHP, ISEG Lisbon, University of St. Gallen, the GfeW conferences in Magdeburg and Salzburg, Seoul National University, the First and Second Conferences on Zero Intelligence, the First EUI Workshop on Pricing Technologies, and University of Zurich. Last but not least, we thank the editor, Rakesh Vohra, for his suggestions. The authors gratefully acknowledge support from the European Commission through the ERC Advanced Investigator Grant "Momentum" (No. 324247) and the EU Horizon 2020 Grant "SoBigData" (No. 654024). Heinrich Nax acknowledges support through a Swiss National Science Foundation Eccellenza grant. Bary Pradelski benefited from the support of the ANR grant ALIAS. Please address correspondence to: Barbara Ikica & Heinrich H. Nax, University of Zurich, Andreasstrasse 15, 8050 Zurich, Switzerland. E-mail: *barbar.ikica@gmail.com* & *heinrich.nax@uzh.ch*.

[1] See Walras (1874) and Walras (1883) for the original formulations, and Friedman and Rust (1993) for a discussion of the various strands of research on the DA.

The outcome was unbelievably consistent with competitive price theory," Smith (1991) states 30 years after the experiments.

The continuous single-item DA for nondurables has since become a drosophila of experimental economics, and CE convergence continues to be a common finding that has attained experimental folk result status.[2] Indeed, for private-information settings, comparable to Smith's original experiments, nonconvergence in DA experiments is found only for very skewed or thin markets.[3]

There is indeed something special about DAs as market institutions in terms of their ability to produce convergence to or (very close to) CE quite quickly, not just in experiments involving human subjects. Indeed, even DAs with "zero intelligence" (ZI) trading bots who place bids and asks randomly (with and even without satisfying participation constraints) produce convergence to CE (Gode and Sunder, 1993).[4] By contrast, the very first experiments involving human traders by Chamberlin (1948), which were the experiments that inspired Smith (1962), were run as "open pit" markets, where buyers and sellers bargained directly bilaterally without a central clearinghouse. Those markets did not converge to CE.

An open question is whether the convergence patterns that have been observed in DAs refute nonconvergence hypotheses in specific settings, or whether—as is folk wisdom—they indeed represent positive evidence for convergence in a more general sense. In particular, an open question is whether it is the DA's central clearing mechanism or the associated centralized dissemination of market information that drives convergence. To test this, we conducted different DA experiments under private information in which we vary the amount of centralized order-book access and price information that is available to traders. As convergence occurs across all treatments, our experiments thus highlight that it is centralized market clearing, not centralized market information, that drives convergence.

Our experimental design required running 104 individual market experiments, each with 10 or more trading rounds (involving 1.7k+ subjects in total). The resulting data are well-documented and freely accessible at the Open Science Framework (OSF) (under https://osf.io/gu62n/) and on GitHub (under https://github.com/ikicab/Trading-in-a-Black-Box). All markets are run with private information, of which over half with less than full access to order-book information and to price histories. These low information treatments include "Black Box" treatments, where subjects do not receive any feedback from the order book about others' bids and asks or realized prices. Their only source of feedback is whether their own bids/asks result in deals and, if so, at what price, representing even less market feedback than in Chamberlin's open pit markets, where they get bilateral information. Unlike in Chamberlin's bilateral setting, however, bids and asks can be matched and result in trade with anyone on the other market side whose bid or ask is compatible through the DA's central clearinghouse, but nobody gets any information about them other than through a deal price when realizing a trade. As we find significant evidence for convergence in all treatments including Black Box, our results provide evidence that it is not the centralized price information that drives convergence in DAs but rather centralized clearing.[5]

The nature of the typical convergence pattern is of interest in itself: initial trading periods are inefficient with prices below CE levels, and subsequent equilibration dynamics occur through rising prices. The same pattern was already present in the experiments by Smith (1962, 1964), and what has since become an experimental folk result associated with CE convergence is that equilibration dynamics tend to favor buyers initially.[6] Convergence from

[2] A large number of studies is reviewed in Davis and Holt (1994).

[3] Nonconvergent dynamics—such as "bubbles"—are associated with retrade speculation (Smith et al., 1988).

[4] See, for example, Farmer et al. (2005) for further work.

[5] If anything, in the prior literature, it appears that *more*-than-private information (instead of *less*-than-full information) produces nonconvergence (Kimbrough and Smyth, 2018).

[6] *"A persistent regularity is that overall buyers are better than sellers in bargaining over the division of surplus through market trading,"* Smith noted and asked in an e-mail to Nax in February 2017, *"are you finding this, and what explains?,"* which actually motivated several elements of the research design.

below has been observed when the market structure is symmetric or not too skewed against buyers. The phenomenon disappears when markets are skewed substantially toward sellers (Smith and Williams, 1982), but there is still evidence of an initial bias favoring buyers as market skews favoring sellers have a smaller effect than market skews of the same magnitude favoring buyers.[7] Holt et al. (1986) provide related evidence, as they obtain different directions of convergence depending on which market side has more power.[8]

Two complementary forces contribute to the phenomenon: first, buyers are more aggressive in their bidding than sellers are in their asking, and, second, sellers concede more, and more quickly, than buyers when making adjustments. Smith (1964) and Walker and Williams (1988) examined the latter channel and found evidence that, by comparing bid auction, offer auction, and DA, sellers concede more, and more quickly than buyers, thus leading to more favorable prices for buyers on average. Evidence that buyers are more aggressive with their initial bids than sellers with their asks is one of our contributions. Indeed, we find that it is the initial aggressiveness that drives the effect more so than the subsequent yielding dynamics. A disclaimer might be opportune here that we, in line with Smith (1964) and Walker and Williams (1988), treat evidence of such asymmetries as "explanations" for the purpose of this article. What explains these asymmetries at a deeper psychological level is not the purpose of our investigation.

The paper most closely related to ours is Lin et al. (2020). They also "go big" in terms of the number of markets that they analyze, also motivated by the goal of elevating the convergence result from experimental folk knowledge to more formal evidence. Their data are from an online tool for running standardized classroom economics experiments, and they also present evidence in favor of CE convergence. Their data source and findings differ in three ways from ours. One, their classroom experiments are not controlled by the experimenter, typically not financially incentivized, and market structures are run (mostly in default settings) without controlled variations. In particular, we as analysts do not know whether the experimenter informed subjects of the market structure and equilibrium prices before trading began and what kind of communication took place during and between trading periods. Two, the types of markets that are investigated are quite different: their experiments permit "losses" (without financial incentives); the majority of the markets are small (less than 10 subjects); most markets last one, two, or three trading rounds (only one market lasts for 10 rounds); most CE ranges are rather large relative to the range of possible market outcomes; most markets are multi-item DAs.[9] Finally, their results are different from ours. In contrast to our equilibration pattern, which occurs significantly from below and takes a handful of periods, their initial deals favor sellers but convergence basically takes no time at all.[10] Our focus is on larger markets, where the equilibrium price range is relatively small compared with the range of feasible market outcomes, and we investigate how convergence occurs in terms of speed and direction.

We believe that our evidence, together with Lin et al. (2020)'s, both of which were collected in parallel over the past four years, helps understand experimental DAs significantly better and that these findings are of general interest to interpret equilibration dynamics. We, compared with Lin et al. (2020), draw some of the same conclusions in terms of final convergence but also opposite conclusion as our early realized prices are generally below equilibrium and subsequently rise throughout a handful of trading periods, whereas their results suggest that prices would quickly fall within the very first trading period and then stabilize. These are

[7] We find a similar phenomenon.

[8] See also a recent paper by Alós-Ferrer et al. (2022) that confirms the asymmetry à la Holt et al. (1986) and focuses on traders' preferences for market institutions.

[9] Out of a total of 5,809 markets, 3,807 last a single round, most others two or three rounds, and only 33 markets have more than five rounds, whereby no market has more than 10 rounds. Ninety-six of the markets are—like ours—repeated single-item DAs, in which participants stay on the same market side with the same valuation across all the trading rounds; only nine of these have five rounds or more.

[10] Our markets start more than 5% below CE in round one and move within 2% of CE after round five, whereas theirs start over 15% above CE in terms of opening prices but move within 2% during round one already.

economically relevant predictions that merit further investigation, as these differences might be the result of several differences in experimental design related to incentivization and information but also in terms of other factors such as framing.

The remainder of the article is structured as follows: Next, we describe the experimental framework. Section 3 contains our market-level results. Sections 4 and 5 contain our analyses of individual-level behaviors. Finally, Section 6 concludes.

2. METHODS

2.1. *Underlying Single-Item DAs.* Buyers $b_i \in B = \{b_1, b_2, \ldots, b_n\}$ and sellers $s_j \in S = \{s_1, s_2, \ldots, s_m\}$ participate in a two-sided single-item economy. Each buyer has a maximal *reservation price* (also called *valuation*) $\overline{\beta}_i$, and similarly each seller has a minimal reservation price $\underline{\sigma}_j$, below/above which they are willing to trade a single item of a homogeneous good. Hence, trade between a buyer b_i and a seller s_j is individually rational only if $\overline{\beta}_i \geq \underline{\sigma}_j$. For any such buyer–seller pair, $\overline{\beta}_i - \underline{\sigma}_j$ is the *gain of trade* between the two, which is split between them depending on what price $P \in [\underline{\sigma}_j, \overline{\beta}_i]$ is implemented. If a buyer b_i buys for price P from seller s_j, then buyer b_i's resulting payout $\overline{\beta}_i - P$ results in a payoff/utility of $u_i(\overline{\beta}_i - P)$ and, conversely, in a utility for seller s_j of $u_j(P - \underline{\sigma}_j)$.[11],[12]

2.2. *CE and Convergence Metrics.* A *market outcome* is defined by a price P along with a set $B \times S$ of buyer–seller pairs involved in trade at that price.

Definition 1. (Walrasian) CE is a market outcome with a single price P^* such that (1) trade is *individually feasible and optimal*, that is, every trader who actively trades weakly prefers trading at that price to all other feasible alternative trades as well as to not trading at that price at all, and every trader who does not actively trade weakly prefers not trading over trading at the given price level, and (2) the *market clears*, that is, there are equal numbers of buyers and sellers actively involved in trade.[13]

The LHS of Figure 1 illustrates a single-item economy as in Smith (1964). CE prices P^* always exist, but need not be unique, in which case we shall denote the minimum CE price by $\underline{P}^*$ and the maximum CE price by $\overline{P}^*$. Any market clearing allocation $B^* \times S^*$, determining who trades in equilibrium and who does not, supported by CE prices maximizes the total gains of trade (GOT), that is,

$$\text{(1)} \qquad \mathrm{GOT}^* = \sum_{i:b_i \in B^*} \overline{\beta}_i - \sum_{j:s_j \in S^*} \underline{\sigma}_j \geq \mathrm{GOT}' = \sum_{k:b_k \in B'} \overline{\beta}_k - \sum_{l:s_l \in S'} \underline{\sigma}_l$$

holds for every feasible market allocation $B' \times S'$. In the LHS of Figure 1, for instance, the maximal GOT that can be achieved (highlighted in gray) amount to 245, attained at any CE price in the range of 113–118.

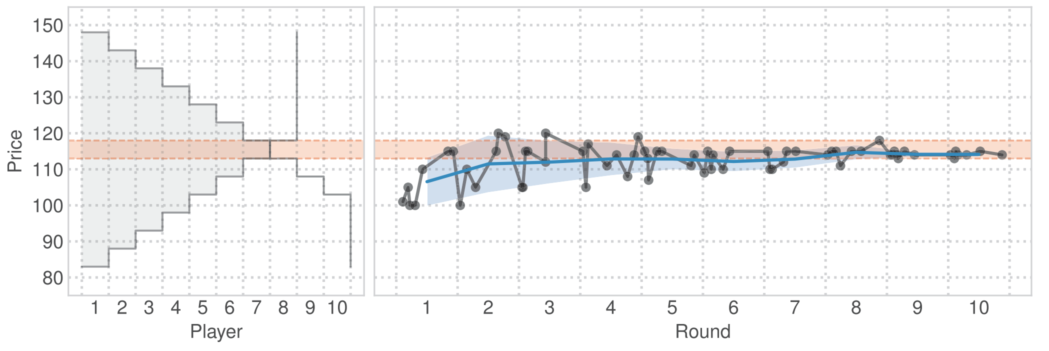
We shall evaluate the degree of convergence of a given market mainly through three metrics. The first one measures the distance to CE in terms of prices[14]:

[11] We make standard assumptions regarding utilities s.t. $u' > 0$ and $u'' < 0$.

[12] For simplicity, we assume for now that the surplus is fully split between the two trading parties.

[13] These conditions are often summarized as *equating demand and supply*, which leaves open some issues related to tie-breaking, etc. (see Jantschgi et al., 2022, for details).

[14] Note that we chose an absolute measure of distance here. Our finding of equilibration from below would only be accentuated by other reasonable measures that would scale down large distances (measures in log scale, relative distances, etc.). Hence, setting the measure up against the result, we chose to report results based on the absolute measure of distance. Given that our experimental platform permits sellers to propose prices several orders of magnitude higher than the corresponding CE prices, this makes our findings even more surprising and robust (sellers were per-



NOTES: The LHS summarizes the demand–supply schedule through ordered reservation prices for buyers and sellers. The gray area highlights the gains of trade realized in CE, and the red area the CE price range. The RHS shows realized prices (black dots) during one of the experimental sessions with a one-standard-deviation interval around the average per round (blue shade). Realized prices enter the CE range from below within a few rounds, their volatility goes down, and the maximum number of deals is reached after the initial couple of rounds.

FIGURE 1

A TYPICAL MARKET AND A TYPICAL SESSION

DEFINITION 2. *Price distance to CE* (DCE) of a market outcome with price P is the signed distance between P and the closest CE price, that is, $\text{DCE} = 0$ if $P \in [\underline{P}^*, \overline{P}^*]$, $P - \underline{P}^*$ if $P < \underline{P}^*$, and $P - \overline{P}^*$ if $P > \overline{P}^*$.

The second measure, introduced by Plott and Smith (1978) (see also Friedman and Rust, 1993), evaluates the performance of a market with respect to allocative efficiency:

DEFINITION 3. *Realized allocative efficiency* (AEff) is the realized GOT over the maximum that is achievable, $\text{AEff} = \text{GOT}/\text{GOT}^*$.

Finally, another useful measure of convergence, as used, for example, in Dolgopolov et al. (2019), is the predictive success index (PSI) by Selten and Krischker (1983) (see also Selten, 1991), which takes into account predictive accuracy adjusted for precision:

DEFINITION 4. The PSI of a model is $\text{PSI} = h - a \in [-1, 1]$, where $h \in [0, 1]$ denotes the *hit rate*, that is, the relative frequency of correct predictions, and $a \in [0, 1]$ corresponds to the *area*, the cardinality of the predicted set relative to the cardinality of the set of all possible outcomes.

In our single-item economy, the hit rate corresponds to the number of prices realized within the CE range over the total number of realized prices and the area to the number of prices comprising the CE range divided by the number of prices that can feasibly be realized.[15]

mitted to propose prices as high as 99,999, whereas the CE prices in our experiments were narrowly centered around 100).

[15] In our application, hit rates are straightforward to compute, but areas are not. An individually rational trade can take place between *any* buyer and seller with compatible reservation prices *who have not traded yet* and at *any* price between the two reservations. Hence, to rigorously compute area, one would have to take into account all feasible allocations and all feasible prices for a given allocation and compute the spread between reservation prices for a given trading pair, averaged across transactions weighted by their likelihood. Since every transaction depends on the trades so far, this is not trivial to compute. Dolgopolov et al. (2019) resolves this issue by estimating PSI through sophisticated simulations. Because our experiments involve a large number of different settings, the same kind of simulation exercise would become overwhelming, which is why we opted for a crude approximation, whereby we assume that all buyer–seller pairs and all prices between their reservations are equally likely. Accordingly, we count the number of feasible prices (i.e., compute the spread between the reservation prices) for all compatible buyer–seller pairs, and

Positive values indicate predictive success (hit rates above pure chance) and negative values conversely predictive failure.

2.3. *Real-Time Trading Experiments on "SciOn".* We built our own real-time trading platform "SciOn" to run the experiments.[16] In total, 104 individual market experiments were run with 1,751 subjects who submitted a total of 86,386 bids and asks. Subjects were recruited online via Amazon's Mechanical Turk. Play was incentivized with a fixed $1 show-up component that was paid to every player and by an incentive bonus from one randomly selected trading round per player per every 10 rounds. Bonuses ranged from zero to a maximum of $7.50 per 10 rounds of play depending on trade success.[17] Experiments lasted between 7 and 25 minutes. Average total earnings were $2.52, in a range from $1.00 to $19.74.[18] We do not include classroom experiments in our analysis, because we wanted to control and incentivize trading as much as possible. We ran experiments online instead of in the laboratory as our focus was to recruit a large number of subjects, and our budget did not permit to incentivize comparable numbers of subjects in the lab. Although some differences in behavior between lab and online experiments have been noted, in particular concerning more randomness in online than in lab data, the overall relative quality and consistency that has been confirmed in recent studies convinced us that data quality would be sufficient for our purposes, especially as more randomness would set things up against our hypothesis, which is to test convergence and asymmetries.[19]

As in Smith (1962)'s original experiments, our experiments follow the standard "induced value" approach (Smith, 1976) with the following economic interpretation. A buyer represents a shopper who is given a deal-dependent budget by us, the experimenter: the shopper can buy the product for at most the budget and gets to keep whatever he/she did not spend. Otherwise, when not making a deal, the shopper returns the budget and makes zero profit. Conversely, a seller acts as a reseller who is given a deal-dependent opportunity to buy from us at a fixed reservation price: if the reseller manages to sell the product above that price on the market to a shopper, then he/she gets to keep that amount. Again, no deal means no money (and incurs no loss). Importantly, shoppers and resellers cannot make losses in this setup, which is what has to be ensured given the experimental guidelines governing our experiments. Consequently, we forbid bids/asks beyond/below reservation prices.[20]

In the experiments, participants are randomly allocated positions as buyers or sellers with fixed reservation prices. The market participants then engage in trading over a finite number of subsequent "rounds," indexed by $T \in \mathbb{N}$, each lasting a certain amount of time, typically ca. 2 minutes. During the course of time t_T in round T (where $t_T \in \mathbb{R}^+$ denotes a measure of continuous "time" during round T) market participants place bids $\beta_i^T(t_T)$ and asks $\sigma_j^T(t_T)$, whereby buyers can place any bid between one and their valuation, that is, $\beta_i^T(t_T) \in [1, \overline{\beta}_i]$, and sellers can place any ask between their valuation and a maximum input of 99,999, that is,

take "the number of prices that can feasibly be realized" to be the arithmetic mean of these counts. More details can be found in the Appendix.

[16] We have been developing https://scienceexperiment.online/scienceexperiment.online ("SciOn") since 2016, supporting lab, online, and classroom experiments without any commercial interests. All materials relevant to the project, such as instructions, screenshots, data and analysis, lab policies (governing the ETH Decision Science Laboratory—https://www.descil.ethz.ch/https://www.descil.ethz.ch/), ethics approvals (via the German Association for Experimental Economic Research), registration, preregistration, and platform details are available at our project Web site on the OSF (under https://osf.io/gu62n/) and in the Appendix.

[17] We followed recent pay-one recommendations (Charness et al., 2016; Azrieli et al., 2018).

[18] Evaluated at equilibrium prices, our monetary incentives were designed to correspond to payments above (roughly thrice) average MTurk wages (and any minimum wage criteria).

[19] See recent comparisons of online and lab data (Hauser and Schwarz, 2016; Arechar et al., 2018; Snowberg and Yariv, 2021).

[20] Admittedly this is an important restriction, but it was unavoidable for us to ensure that we could commit to incentivizing all behaviors, on the one hand, and that nobody could lose money by trading, on the other.

$\sigma_j^T(t_T) \in [\underline{\sigma}_j, 99, 999]$.[21] Furthermore, we use $\beta_{i,k}^T$ and $\sigma_{j,k}^T$ to denote the kth bid and ask submitted by buyer b_i and seller s_j in trading round T, respectively, and $\beta_{i,k}$ and $\sigma_{j,k}$ to denote their kth bid or ask overall. Bids and asks automatically expire after 10 seconds and can be overwritten anytime with a new one until a subject trades. We added this expiry feature to increase the chance that all traders will stay actively involved in the market and to stress the continuous time feature of the trading environment mirroring online the "bazaar feel" of the oral classroom experiments by Smith (1962).[22] Whenever a bid and an ask cross (i.e., the bid is greater than or equal to the ask), a trade occurs, and the two involved traders leave the market until the next round starts. If multiple trades are profitable, the one with the highest bid–ask spread takes place, thus favoring larger GOT. If there is still a tie, the one placed earlier has priority First In, First Out (FIFO). The round ends when the time runs out, or when all players have been matched, whichever occurs first, and then—unless it is the final round—the next round begins.

2.4. *Treatments.* Our baseline treatment is a symmetric market with "equal sides" involving 10 buyers and 10 sellers whose reservation prices are allocated randomly in 5-unit steps between 103 and 148 for buyers and between 73 and 118 for sellers, resulting in CE involving eight sellers and eight buyers at prices between 108 and 113. Subjects have *private information*, that is, they know only their own reservation prices but not those (nor the distribution) of the others. Trading takes place in 10 market rounds, and during each trading round subjects have full access to the live order book in the form of sorted lists of all currently active bids (from high to low) and asks (from low to high). A player's own bid/ask is always clearly marked, and in addition to the order book, there is an instant 3-second broadcast to all market participants every time a deal is made, which also informs them about the corresponding price. Prices are determined by the *first price* rule so that the bid/ask that was recorded first determines the price when a bid and an ask cross and are cleared.

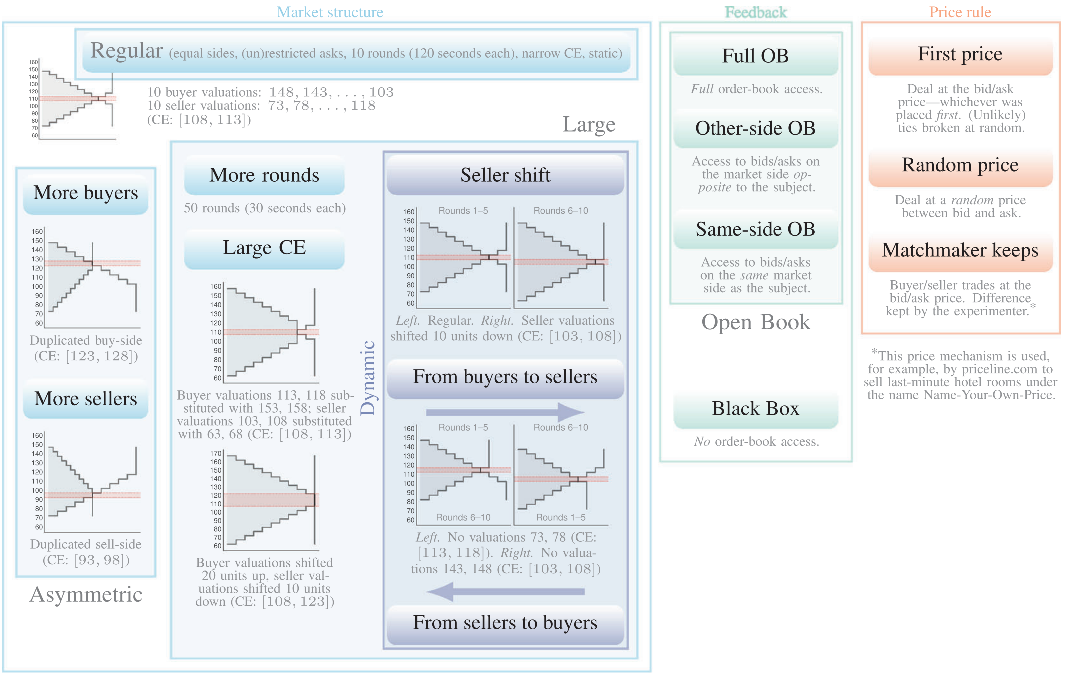
We vary this baseline market in various ways concerning all constituent building blocks of the underlying market, each resulting in a treatment with a ceteris paribus variation. Other than that, the markets function as before.

One set of variations concerns the underlying market structure. Some treatments preserve the market symmetry but vary the size of the CE price range or the size of the maximal GOT and number of subjects clearing the market (referred to as *Large CE*), other treatments involve more rounds or result in asymmetric markets (favoring either buyers or sellers) shifting the CE price range up or down compared with the baseline markets. Note that all markets continue to have more than a single buyer and/or seller and are set up to place CE as relatively narrow ranges somewhere within the interquartile range of the reservation prices.

Other variations involve withholding some or all of the order book and price feedback and varying the price rule. Note also that we did not run all possible combinations of settings regarding the market structure, pricing rule, and feedback information, but we varied them

[21] Note that we occasionally write t for continuous time and omit T if we speak of a single trading round, or when it is clear which trading round we are referring to. Moreover, if the timing of the bid/ask placement is not relevant, we drop the time argument t_T altogether.

[22] Contrary to the levels of "ghosting" subjects found in online experiments elsewhere (e.g., Arechar et al. 2018), we had relatively few inactive subjects, whereby we define a subject to be inactive in a trading round if they neither throughout the given round nor throughout any subsequent round submitted any bids/asks or accepted any deals. By the end of second round, less than 1% of subjects, with buyers and sellers roughly equally represented, were inactive. This percentage increased but was, by the end of round five, still below 2% in Open Book and less than 4% in Black Box. In both cases, just under 1% of buyers and just under 1% of sellers who were not "out of money" and would actually be able to trade had the total GOT been maximized, were inactive. Ghosting in our experiments does not exceed 6% even by the final rounds, by which time many of the inactive subjects are out-of-the-money subjects who realized that they would never deal anyway. While subject absence required appropriate adjustments to the computations, note that the low levels of inactivity, which are also symmetric around CE prices, do not substantially bias or alter the results qualitatively. All our implementations and further details are publicly available at our OSF registry under https://osf.io/gu62n/.



NOTES: A market is determined by its structure, form of order-book feedback, and price rule.

FIGURE 2

EXPERIMENTAL SETUP

so that every treatment was comparable to the baseline. See Figure 2 for a conceptual figure placing the different treatments and Table 1 providing details for each variation.

Note that our *Black Box* (BB) feedback is the one where the least information is provided to the subjects. In Black Box, subjects do not have any access to feedback from the order book. Subjects know only their role and reservation price as a buyer or seller, and, as trading proceeds, each subject knows his/her own history of past bids/asks only, as well as which of these resulted in a deal and at what price. Subjects are also not told how many others they would be interacting with.[23] To differentiate between Black Box and the rest of the treatments with partial/full order-book feedback (i.e., treatments where there is some access to order information apart from one's own), we shall simply use *Open Book* (OB) as an umbrella term for the latter. When the terms *Black Box* (BB) and *Open Book* (OB) are used in isolation, they refer to pooling all available Black Box and Open Book data across all the corresponding treatments, respectively.

To ensure robust results, we ran each resulting market experiment five times (in two cases only four times[24]) using the same combination of settings. See Table 1 for a summary of all the markets we ran, their exact setup, number of participants, and number of submitted bids/asks.[25] In addition, to avoid end-game effects, the number of trading rounds and their duration were also unknown to the subjects, but they did know how long the experiment would roughly take based on the recruitment (HIT) description that was used on MTurk.

[23] In contrast to experiments conducted in a classroom, where they can count the number of participants, they could also not deduce the market size given the Black Box information and the online setting of the experiment.

[24] Two markets failed to initiate due to technical problems.

[25] Note that the final number of subjects active in an experimental market may differ from the number of subjects initially recruited for it either due to inactivity of the subjects or due to disconnections before trading begins, thus giving us some random variation in the underlying markets. See also the Appendix and our OSF registry (https://osf.io/gu62n/) for more details.

TABLE 1
EXPERIMENTAL SESSIONS: OVERVIEW OF ALL THE MARKETS THAT WE RAN

	Treatment		Experimental Data	
Market Structure	Feedback	Price Rule	Subjects (Markets)	Observations
Regular				
	Full OB	First price	73 (4)	4810
	Other-side OB	First price	70 (5)	4238
	Same-side OB	First price	71 (5)	4069
		Matchmaker keeps	64 (4)	4605
	Black Box	First price	84 (5)	4495
		Random price	78 (5)	4249
		Matchmaker keeps	88 (5)	5334
Restricted asks	Full OB	First price	76 (5)	2802
	Black Box		71 (5)	2446
Asymmetric				
More buyers	Full OB	First price	130 (5)	5683
	Black Box		113 (5)	4792
More sellers	Full OB	First price	140 (5)	8012
	Black Box		124 (5)	5236
Large				
More rounds	Black Box	First price	77 (5)	7392
Large CE	Full OB	First price	183 (13)	6999
	Black Box		74 (8)	3259
Seller shift	Full OB	First price	80 (5)	2467
From buyers to sellers			74 (5)	2380
From sellers to buyers			81 (5)	3118
		Total	**1,751 (104)**	**86,386**

NOTE: The final number of subjects active in an experimental market may differ from the number of subjects initially recruited for it either due to inactivity of the subjects or due to technical disconnections before trading began. See also our OSF registry under https://osf.io/gu62n/.

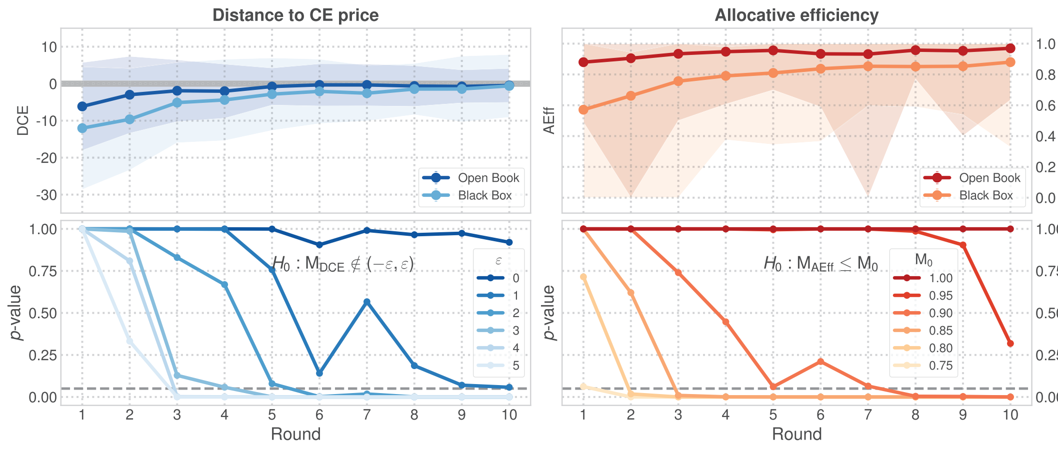
The RHS of Figure 1 illustrates our prototypical equilibration dynamics as observed in one of the four sessions of the regular market with private information and first price (see also the Appendix and our OSF registry for all sessions individually), with the LHS corresponding to the ordered reservation prices.

3. MARKET-LEVEL RESULTS

3.1. *CE Convergence and Allocative Efficiency.* We now turn to empirical results. First, we shall investigate the convergence properties of the markets in terms of DCE and realized AEff. Figure 3 illustrates that convergence generally occurs within a handful of trading rounds at most. This occurs for Black Box and Open Book markets, but it is visible that convergence is slower and more volatile in Black Box.[26] What is more, convergence in terms of prices occurs from below in all cases. Note that the volatility of DCE decreases over time (particularly in Open Book), whereas the volatility of AEff remains constant. In fact, these patterns are found in all treatments, see Figures A.6 and A.7 in the Appendix, and we shall confirm them by formal hypothesis testing.

To formally test whether market prices converge (from below) to CE prices and whether allocation becomes more efficient throughout the rounds, we conduct, for each round separately, three one-sample hypothesis tests. For the purpose of each of these hypothesis tests, we compute, for each experimental session separately, the market statistic under test relative to the given round, and treat the thus obtained market statistics as individual observations,

[26] We shall report on the statistically significant differences between Black Box and Open Book markets in our analysis in Section 4 (refer to the statistical test results).



NOTES: *Top*: *Left*. DCE with a one-standard-deviation interval around the average per round (blue shade) evolving over time. *Right*. AEff with an interval indicating the least and the most efficient markets around it (red shade) over time. *Bottom*: *p*-Values corresponding to the hypothesis tests pooling across all (Open Book and Black Box) treatments. The horizontal dashed lines indicate the 0.05 significance level. *Left*. Convergence to CE price range. *Right*. Convergence to allocative efficiency.

FIGURE 3

OVERALL CONVERGENCE PATTERNS OVER THE FIRST 10 ROUNDS (TOP ROW) AND HYPOTHESIS TESTS (BOTTOM ROW)

whereby we pool across all (Open Book as well as Black Box) treatments.[27] Accordingly, the observations are independent from each other for each of the conducted statistical tests.

Hypothesis 1. Per-session mean prices converge to the corresponding CE prices after the first five trading rounds.

First, we test the null hypothesis that the median value of the per-session mean price distance to CE, hereafter denoted by $\mathrm{M}_{\mathrm{DCE}}$, is either greater than or equal to $\varepsilon \geq 0$ or smaller than or equal to $-\varepsilon$:

$$H_0 : \mathrm{M}_{\mathrm{DCE}} \notin (-\varepsilon, \varepsilon) \quad \text{versus} \quad H_a : \mathrm{M}_{\mathrm{DCE}} \in (-\varepsilon, \varepsilon).$$

We report results from 10 equivalence tests, one for each round. Specifically, we use a two one-sided test (TOST) approach where the two individual tests are Wilcoxon signed-rank tests, with Pratt (1959) corrections applied to account for the cases where the mean price distances to CE are zero. Each test is applied to the means of price distances to CE attained in the corresponding round, which are computed relative to the individual experimental sessions.

One signed-rank test concerns the null hypothesis $\mathrm{M}_{\mathrm{DCE}} \leq -\varepsilon$, whereas the other one considers $\mathrm{M}_{\mathrm{DCE}} \geq \varepsilon$. The relevant *p*-value for the equivalence test, the one reported in the bottom left panel of Figure 3, is the larger of the *p*-values from these two individual tests. For all 10 equivalence tests, it so happens that the reported *p*-value comes from the individual test of $\mathrm{M}_{\mathrm{DCE}} \leq -\varepsilon$.

In the tests, we vary the equivalence margin ε from 0 (no margin) to 5 (substantial margin) using unit increments in order to express deviations from CE prices. CE prices range from 63 to 148 with a mean value of 108, so units are roughly interpretable as percentage deviations.

[27] Here, we use the full power of all market experiments combined instead of splitting by market structure, information type, or otherwise. Note that, depending on the round considered, the sample size (i.e., the number of experimental sessions taken into account) may range from 101 to 104 due to some trading rounds ending without a deal within particular experimental sessions.

The results indicate that the mean deal prices tend to CE with high precision (i.e., not further away than 2%) from the sixth round onward at a significance level of 0.05. Hence, convergence (of the mean price plus minus 2%) occurs after round 5. ■

Hypothesis 2. Per-session mean prices are below the corresponding CE prices in the first four trading rounds.

The same series of one-sided Wilcoxon signed-rank tests with the Pratt correction applied to the same samples can be used to assess whether convergence occurs from below, that is, whether the median values of $\mathrm{M}_{\mathrm{DCE}}$ are bounded above by a small constant of $-\varepsilon$, with ε again assuming values between 0 and 5:

$$H_0 : \mathrm{M}_{\mathrm{DCE}} \geq -\varepsilon \quad \text{versus} \quad H_a : \mathrm{M}_{\mathrm{DCE}} < -\varepsilon.$$

Recall that testing Hypothesis 1 involved testing the null hypothesis $\mathrm{M}_{\mathrm{DCE}} \leq -\varepsilon$. We now test the null hypothesis that $\mathrm{M}_{\mathrm{DCE}} \geq -\varepsilon$. As such, the relevant p-value is simply 1 minus the p-value reported in Figure 3 (remember that all p-values reported in Figure 3 come from the individual tests of $\mathrm{M}_{\mathrm{DCE}} \leq -\varepsilon$).

Initial prices are 5% below CE, and still 2% below by period 4. This confirms a significant downward asymmetry by at least 2% (initially by at least five) up until the fourth round. Hence, per-session mean prices are initially substantially below CE. ■

Hypothesis 3. Allocative efficiency is reached over time.

An analogous series of one-sided Wilcoxon signed-rank tests was implemented to determine whether the median per-session allocative efficiency $\mathrm{M}_{\mathrm{AEff}}$ exceeds a benchmark of M_0 ranging from 0.75 to 1 with a step size of 0.05:

$$H_0 : \mathrm{M}_{\mathrm{AEff}} \leq \mathrm{M}_0 \quad \text{versus} \quad H_a : \mathrm{M}_{\mathrm{AEff}} > \mathrm{M}_0.$$

Each test was applied to allocative efficiencies computed relative to the corresponding round for each experimental session separately. As can be seen from the bottom right corner of Figure 3, initial allocative efficiency is significantly below 0.8, but as early as the third round, allocative efficiency significantly surpasses the value of 0.85 and increases further until the final rounds when efficiency is significantly above 0.9.

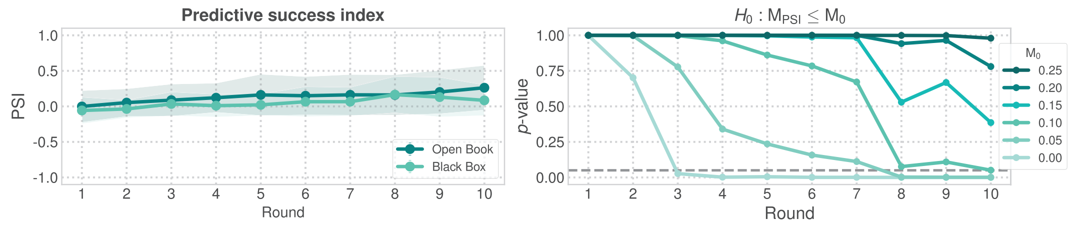
Hence, whereas initial allocative efficiency is significantly lower than 80%, it eventually is significantly higher than 90%. ■

3.2. *Predictive Success.* Selten and Krischker's PSI sheds some more light on the convergence pattern.[28] The results, which are reported in Figure 4, indicate that PSI increases over rounds. Note that the values obtained are in line with Dolgopolov et al. (2019) in spite of our crude estimation of the areas (see footnote). To formally test that there indeed is an increase, we resort again to hypothesis testing. Accordingly, we apply, for each round separately, the one-sample Wilcoxon signed-rank test with the Pratt method to test

$$H_0 : \mathrm{M}_{\mathrm{PSI}} \leq \mathrm{M}_0 \quad \text{versus} \quad H_a : \mathrm{M}_{\mathrm{PSI}} > \mathrm{M}_0,$$

where $\mathrm{M}_{\mathrm{PSI}}$ is the median per-session PSI and M_0 varies from 0 to 0.25 with a step size of 0.05. Similarly as before, for each test, the individual observations correspond to predictive success indices of the individual experimental sessions computed with respect to the given round, whereby we pool across all (Open Book as well as Black Box) treatments.

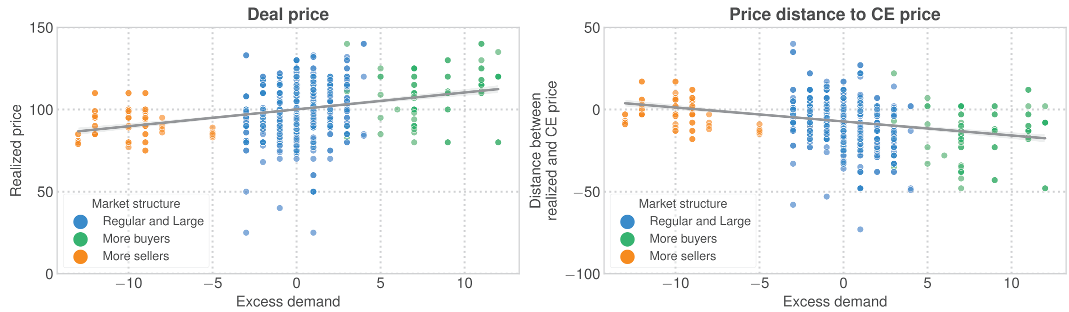
[28] We thank an anonymous reviewer for suggesting the PSI as a criterion.



NOTES: *Left*. PSI with a one-standard-deviation interval around the average per round (turquoise shade) evolving over time. *Right*. Convergence in terms of predictive success. *p*-Values correspond to the hypothesis tests pooling across all (Open Book and Black Box) treatments. The horizontal dashed line indicates the 0.05 significance level.

FIGURE 4

PREDICTIVE SUCCESS INDEX (LEFT) AND HYPOTHESIS TESTING (RIGHT)



NOTES: The data are pooled across all (Open Book and Black Box) treatments. *Left*. Prices attained in the first round as a function of excess demand. *Right*. First-round DCE as a function of excess demand. Fitted linear regression models are displayed as gray lines.

FIGURE 5

FIRST-ROUND DEAL PRICES AS A FUNCTION OF EXCESS DEMAND

Initially, we cannot reject the claim that PSI assumes nonpositive values, as is illustrated in Figure 4. In fact, since the combined first-round PSI roughly equals zero (refer to the LHS of Figure 4), meaning that the hit rate and the area are approximately the same, the initial level of "convergence" might be purely accidental. Nevertheless, PSI significantly exceeds zero after round two (the tests suggest that by the final round $M_{PSI} > 0.10$ at a significance level of 0.05), corroborating the findings from Hypotheses 1–3 that convergence takes a handful rounds.

3.3. *Extra-Marginal Supply and Demand.* As a final step in our macro-level analysis, we take a closer look at how extra-marginal supply and demand affect deal prices, which is a layer of analysis we had not seen previously in the literature.[29] Since the results above show that prices quickly converge to the range of CE prices, it is particularly worth investigating how prices are formed at the very beginning, during the first trading round.

To this end, we examine how first-round deal prices and the corresponding DCE vary as a function of excess demand, that is, the difference between the number of buyers and sellers in the market. The results are plotted in Figure 5. As one would expect, the more buyers enter the market, the more intense competition among them becomes, and the higher the deal prices. Similarly, tilting the market in the opposite direction—toward more sellers—decreases the prices. In fact, the median first-round deal price amounts to 110 in markets with more buyers, to 100 in regular and large markets, and to 95 in markets with more sellers, differences being significant at the 5% level (Mann–Whitney–Wilcoxon [MWW] tests applied to all

[29] We thank an anonymous reviewer for suggesting the analysis of extra-marginal effects.

three pairs of groups of first-round deal prices pooled from the three types of markets, respectively). Curiously, however, increasing excess demand has an opposite effect in terms of DCE. The first-round prices in asymmetric markets with more buyers are much further away from the CE range ($M_{DCE} = -13$) compared to those in regular and large markets ($M_{DCE} = -5$) and asymmetric markets with more sellers ($M_{DCE} = 0$), differences again significant at the 5% level (MWW tests applied to all three pairs of groups of first-round distances to CE prices pooled from the three types of markets, respectively). To understand better why this might be the case, we now turn to the next section, in which we inspect bidding/asking strategies followed by individual buyers and sellers.

4. INDIVIDUAL-LEVEL RESULTS

Our second layer of analysis takes a closer look at how bidding unfolds at the level of individual market participation. To this end, we examine—separately for buyers and sellers—how subjects behave initially, as well as within and in between trading rounds. Our aim is to understand bid and ask behaviors of buyers and sellers better. In particular, we want to check whether there are systematic differences between buyers and sellers related to the observed market-level patterns of asymmetric convergence occurring from below CE.

To make buyers and sellers comparable, we measure their "aggressiveness" in terms of their bids and asks by comparing how much of the surplus a typical buyer $b_i \in B$ requests compared to a typical seller $s_j \in S$ as they bid β_i and ask σ_j. To make these requests directly comparable across all individuals, we quantify them by computing for every buyer $b_i \in B$ and every seller $s_j \in S$ their demands for payoff relative to their corresponding valuations (as used sometimes in finance), as defined below.[30]

DEFINITION 5. *Relative demands* implied by buyer b_i's bid $\beta_{i,k}^{(\cdot)}$ and seller s_j's ask $\sigma_{j,k}^{(\cdot)}$, respectively, are given by $\rho_{b_i,k}^{(\cdot)} = (\overline{\beta}_i - \beta_{i,k}^{(\cdot)})/\overline{\beta}_i$ and $\rho_{s_j,k}^{(\cdot)} = (\sigma_{j,k}^{(\cdot)} - \underline{\sigma}_j)/\underline{\sigma}_j$.

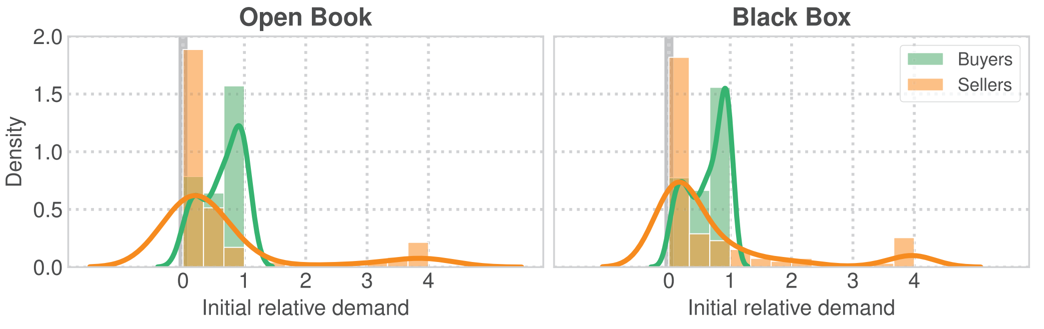
To decouple the influence that different information channels might have on trading behavior, we set Open Book treatments with partial/full feedback against Black Box treatments throughout the analysis. Recall that Black Box treatments provide no feedback to the participants other than their own realized deals. As such, they serve as a useful benchmark to identify the basic mechanisms driving the observed behavior.

4.1. *Initial Bids/Asks.* We first compare buyers' and sellers' very first bids and asks in terms of aggressiveness. Accordingly, in the following, the unit of observation of interest is an individual's initial relative demand, that is, $\rho_{b_i,1}$ for a buyer and $\rho_{s_j,1}$ for a seller, respectively.

In Black Box treatments, all initial relative demands are independent from each other by design since all order-book information is hidden from the participants. By contrast, in Open Book treatments, those experimental subjects who are not the very first to bid/ask are potentially influenced by having observed other participants' actions. Hence, we restrict ourselves in the analyses of those treatments to the initial relative demands that were placed *before* the participants could gain any feedback from anyone else, thus ensuring observation independence.[31] Note that this results in a smaller set of Open Book observations compared with Black Box.

[30] In contrast to our absolute measure of distance to CE prices, we opted for this relative measure of aggressiveness here, again because it is set up "against" the results we find; that is, that buyers are more aggressive. This result would be stronger if other measures were chosen—notice that the relative measure that we use makes a buyer demanding the same absolute amount as a seller appear *less* aggressive.

[31] In the treatments with full access to the order book, this only applies to the very first actions in the first round of the market session. Under the Same-side OB feedback, this holds true more generally for the very first offers made on each side of the market separately. Finally, under the Other-side OB feedback, this holds for all initial offers on the market side to act first which precede the first action on the opposite side.



NOTES: Histograms and Gaussian kernel density estimates of initial relative demands $\rho_{b_i,1}$ and $\rho_{s_j,1}$ for buyers and sellers in Open Book and Black Box treatments. Note that initial relative demands greater than 4 are clipped and redistributed into the rightmost bin (this applies to roughly 4% of values).

FIGURE 6

INITIAL RELATIVE DEMANDS

The resulting distribution of the initial relative demands $\rho_{b_i,1}$ and $\rho_{s_j,1}$ is depicted in Figure 6 for Open Book and Black Box treatments; a formal statistical analysis is provided below (see also Table A.1). The results indicate that buyers initially are more aggressive than sellers.[32]

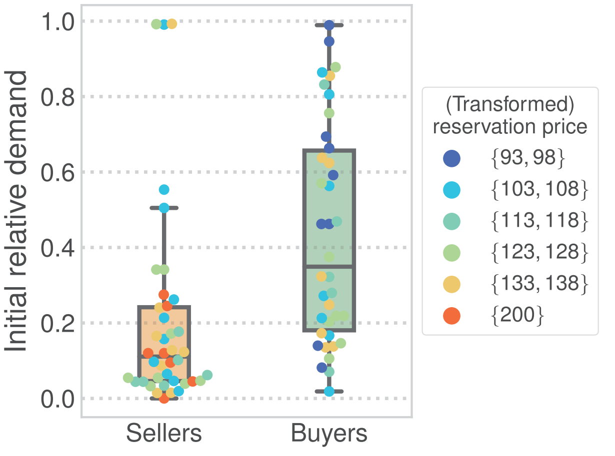
At first sight, one might expect the opposite to hold given that buyers in our experiments face a bounded range of possible bids (between one and their valuation) in comparison to the quasi-unbounded range that sellers may choose from (any number from their valuation up to a maximum input of 99,999 was possible). To investigate the consequences of this asymmetry in the action space, particularly whether it might actually produce the asymmetry in aggressiveness, we conducted a dedicated set of experiments where an upper bound is imposed on sellers' asks such that buyers' and sellers' action spaces are symmetrical to each other about the CE range. More specifically, we ran 10 regularly structured markets where the desired symmetry was achieved by setting the maximum permissible ask to 200 and shifting all reservation prices 10 units down (refer to Figure 2 and Table 1 for more details).[33] Indeed, notice that applying a linear transformation to map seller s_j's action space $[\underline{\sigma}_j, 200]$ onto $[1, 201 - \underline{\sigma}_j]$ makes seller s_j directly comparable to buyer b_i with $\overline{\beta}_i = 201 - \underline{\sigma}_j$. The corresponding initial relative demands (expressed in terms of buyers' action spaces) are illustrated in Figure 7. The results demonstrate that restricting sellers' asks does not imply qualitatively different results—buyers are still substantially more aggressive in their bidding.

We used MWW tests to compare the group of buyers' initial relative demands versus the group of sellers' initial relative demands, which were both pooled from the same—depending on the test—(set of) treatment(s) (refer to Table 2 and cf. Table A.1). Recall that the initial relative demands considered are independent both within and across the two groups regardless of the treatment(s) considered.

MWW tests confirm what Figures 6 and 7 suggest, that is, the fact that buyers, on average, initially bid more aggressively in pursuit of higher profits than sellers. The median initial relative demand for buyers in Open Book and Black Box markets is 0.68 and 0.71, respectively, compared with the median initial relative demand for sellers of 0.19 in both, and the differences between buyers' and sellers' initial relative demands in both Black Box and Open Book markets, respectively, are significant with a $p < 0.01$.

[32] Note that the observed unevenness is unaffected if we consider logarithmic demands or bid/ask distance to CE instead (cf. Figure A.10).

[33] Due to technical issues, sellers that should have had the highest reservation prices, 103 and 108, were wrongly assigned a reservation price of 1. Notice that this enabled them to act much more aggressively than our experimental design would permit. Nonetheless, the asymmetry in aggressiveness persists *in spite* of this.



NOTES: Distribution of initial relative demands (expressed in terms of buyers' action spaces) in experimental sessions with restricted asks. Overlaid is a scatterplot corresponding to individual bids/asks color-coded by reservation price. The median relative demand for buyers and sellers is 0.35 and 0.11, respectively, and the differences are significant with a $p < 0.01$ (MWW test applied to the group of all sellers' initial relative demands and the group of all buyers' initial relative demands pooled from the experimental sessions with restricted asks).

FIGURE 7

RESTRICTING SELLERS' ASKS SYMMETRICALLY FROM ABOVE

TABLE 2

INITIAL RELATIVE DEMANDS IMPLIED BY FIRST BIDS AND ASKS

Treatment			Median Initial Relative Demand		MWW
Feedback	Market Structure	Price Rule	Buyers	Sellers	*p*-Value
Open Book			0.68 (42)	0.19 (70)	<0.01
Black Box			0.71 (356)	0.19 (353)	<0.01
	Regular		0.71 (160)	0.2 (161)	<0.01
	Asymmetric	First price	0.71 (118)	0.17 (119)	<0.01
	Large	First price	0.72 (78)	0.19 (73)	<0.01
		All	0.71 (398)	0.19 (423)	<0.01

NOTE: Median initial relative demands comparing buyers and sellers in Open Book and Black Box treatments. Each p-value corresponds to an MWW test applied to the group of all buyers' initial relative demands and the group of all sellers' initial relative demands pooled from the treatments indicated in the first three columns. Numbers in brackets report the number of bids/asks considered in the analysis. See Table A.1 for an overview of the individual market results.

As discussed at the beginning of the section, for the analysis of initial bids and asks, the Black Box treatment comes with a major benefit from the analyst's point of view due to the fact that no participant sees any other's action when placing his/her first bid or ask and has not received any other feedback either—that is, everyone starts off by knowing nothing more than their role and reservation price, both of which are assigned at the outset of the game. In Black Box, this produces a large number of independent individual observations, because we can include every subjects' initial bid or ask in the analysis. Hence, the buyer–seller difference is statistically significant in almost every treatment. By contrast, in Open Book, those who are not the very first to bid/ask are potentially influenced by having observed other participants' actions. Accordingly, there are significantly fewer initial relative demands to analyze. At the aggregate level, however, significance of the same kind of buyer–seller difference in terms of initial aggressiveness is confirmed by the MWW test run on the group of all sellers' initial relative demands and the group of all buyers' initial relative demands pooled across all Open Book treatments combined (with a p-value of less than 0.01), but we could not conduct the same kind of test at individual treatment level due to the small amount of data that can be included. Nevertheless, the fact that Open Book asymmetries are virtually identical to Black Box asymmetries (and qualitatively the same when sellers' asks are also restricted—cf. Table A.1) allows to conclude that the asymmetry is truly inherent, and that it is not the expectation of feedback that drives it.[34] To formally confirm that initial bids and asks do not differ between Black Box and Open Book treatments, we employed the Mann–Whitney test for equivalence as proposed by Wellek (1996). More specifically, we tested, for buyers and sellers separately, the null hypothesis that $|\mathbb{P}(\rho_1^{\mathrm{BB}} > \rho_1^{\mathrm{OB}}) - 1/2| \geq 0.1$ with ρ_1^{BB} and ρ_1^{OB} drawn independently from the distributions underlying buyers'/sellers' initial relative demands in Black Box and Open Book, respectively. In both cases, the null hypothesis can be rejected in favor of the alternative hypothesis $|\mathbb{P}(\rho_1^{\mathrm{BB}} > \rho_1^{\mathrm{OB}}) - 1/2| < 0.1$ at the 5% significance level. To see whether the asymmetry between buyers and sellers persists over time and how *realized*—instead of expected—feedback affects their behavior, we now take a closer look at how bidding and asking unfold within and between consecutive rounds.

4.2. *Within- and Between-Round Adjustments: The Simpler Case of Black Box.* Next we analyze subsequent bid and ask adjustments, to understand traders' basic behavioral patterns and, again, to see whether there are differences between buyers' and sellers' behaviors that could drive the asymmetric convergence patterns observed at the market level. In this section, we shall restrict this analysis to the behavioral patterns identified under Black Box conditions. Black Box treatments exclude several interesting—but complicated—channels of influence of players on one another through order-book feedback, which we shall investigate in the next section. In this section, we focus on the simpler Black Box models and check how Open Book treatments compare with Black Box in terms of these. As far as the statistical analysis is concerned, we would like to stress that the analytical advantages associated with initial demands (which were the most prominent under Black Box) are gone. After initial demands, subsequent adjustments depend on subjects' experience and others' demands or lack thereof, so individual observations are not necessarily independent anymore. Nevertheless, we shall report test statistics as in the prior section, but we would like to stress the fact that the validity of these test statistics does not compare with the analysis of "unspoilt" initial demands from the prior section, where all observations were necessarily independent of each other by design.

4.2.1. *Payoff-based adjustments within rounds.* A subject's strategy following his/her initial relative demand may be expressed via "adjustments" throughout the course of trading in terms of effective relative demands. Adjustments may occur *within* a given round T, and *between* consecutive rounds T and $T+1$.

DEFINITION 6. A *within-round adjustment* is the difference between a player's current $\rho_{\cdot,k}^T$ and previous relative demand $\rho_{\cdot,k-1}^T$ in the same round T, that is, $\delta_{\cdot,k}^T = \rho_{\cdot,k}^T - \rho_{\cdot,k-1}^T$, where $k \geq 2$.

Black Box treatments shut down most feedback channels, leaving only payoff-based learning impulses that may plausibly trigger adjustments: subjects' own prior bids/asks and their (lack of) resulting trading success. Within a given period, by virtue of trading success terminating trading activity for the round, behavior necessarily boils down to a series of failed trading attempts. Accordingly, as one would expect, the vast majority of within-round adjustments $\delta_{\cdot,k}^T$ in Black Box for both sellers and buyers are nonpositive (see Table 3).[35] That is, traders gradually reduce (or repeat) their demand for profit in order to increase their chance of striking a deal and only rarely increase their demand.

[34] We thank an anonymous reviewer for pointing out this interesting observation.
[35] See also Table A.2 for a more detailed analysis.

TABLE 3
WITHIN-ROUND ADJUSTMENTS OF RELATIVE DEMANDS

	Treatment		Buyers		Sellers	
Feedback	*Market Structure*	*Price Rule*	*Median* $\delta^T_{b_i,k}$	$\delta^T_{b_i,k} \leq 0$	*Median* $\delta^T_{s_j,k}$	$\delta^T_{s_j,k} \leq 0$
Open Book			0 (19,098)	86%	0 (20,326)	87%
	Regular		0 (9,269)	86%	0 (7,892)	86%
Full OB	**Asymmetric**	First price	0 (5,112)	85%	0 (6,006)	88%
Full OB	**Large**	First price	−0.01 (4,717)	87%	−0.01 (6,428)	86%
Black Box			−0.03 (14,024)	87%	−0.01 (13,947)	84%
	Regular		−0.03 (6,577)	85%	−0.01 (6,956)	82%
	Asymmetric	First price	−0.03 (3,704)	89%	−0.01 (4,099)	84%
	Large	First price	−0.02 (3,743)	90%	−0.02 (2,892)	88%
		All	−0.01 (33,122)	87%	0 (34,273)	86%

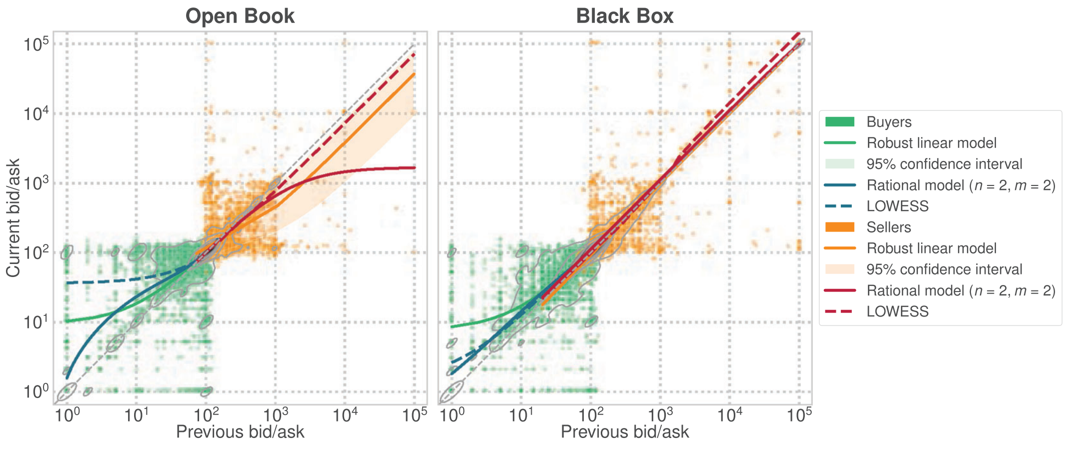
NOTE: Median within-round adjustments $\delta^T_{b_i,k}$ and $\delta^T_{s_j,k}$ and the percentage of nonpositive adjustments, comparing buyers and sellers in Open Book and Black Box treatments across all rounds T. Numbers in brackets indicate the number of bids/asks considered in the analysis. See Table A.2 for an overview of the individual market results.

To analyze within-round adjustment behavior, we set subjects' bids/asks against their previous bids/asks and fit several models including robust linear regressions. The models are fit to a subject's bid/ask as a function of the subject's previous bid/ask, whereby the bids and asks are pooled across all experimental sessions for Open Book and Black Box treatments separately.[36] The results are summarized in Figure 8. Clearly, within a trading round, both buyers and sellers tend to follow up their (unsuccessful) requests for profit with less demanding requests. This occurs with a high degree of "stickiness" with respect to the most recent offers. Note that divergences from the regression lines are driven to a large extent by bids/asks of "round" and focal numbers, in particular asks of 150 and 200 and bids of 50 and 100. Stickiness is also evident from within-round within-subject autocorrelations between bids/asks, with median autocorrelation values for buyers and sellers under Black Box as high as 0.96 and 0.92, respectively (see also Figure A.11).

The analysis suggests that there are no major differences between buyers and sellers, at least not in the direction that would explain the market-level asymmetries: levels of yieldingness (i.e., nonstickiness and stepsize) are comparable, and if anything higher for buyers than for sellers, which does not favor buyers over sellers in terms of bargaining dynamics. Comparison of within-round adjustments across rounds (illustrated in the top row of Figure 9) reveals that buyers yield slightly more only in the first few rounds and that such differences disappear in later rounds. Jointly, with the finding of higher buyer aggressiveness from the last section, these findings help explain why convergence to CE prices from below takes place relatively quickly: buyers tend to be significantly more aggressive at the very beginning of trading—which explains why first deal prices occur below the CE range, but then all traders yield (and buyers do so a bit more quickly), which brings prices closer to the CE range.

We can sharpen how and when buyers and sellers yield differentially through an analysis of each trading round that lasted 120 seconds in terms of four 30-second time windows and comparisons of buyers and sellers within each of these via MWW tests. More specifically, we looked at every 30-second time window of every trading round separately and applied an MWW test to the group of all buyers' within-round adjustments and the group of all sellers'

[36] Robust fits (using Huber M-estimators) are applied to downweight outliers—while 95% of sellers' asks assume a value smaller than 300, outliers far away from the relevant range do occur as asks up to a value of 99,999 were permissible in the standard "unrestricted" setting. Because the linear regression models are, strictly speaking, misspecified, we also investigate, as a sanity check, medians of bids/asks given previous values (see Figure A.12). We thank an anonymous reviewer for this suggestion. Another model we fit is LOWESS (locally weighted scatterplot-smoothing) with a smoothing span of 0.2 of the data. The main motivation behind fitting various models was to highlight better the differences between buyers and sellers.



NOTES: Comparison of buyers' and sellers' within-round behavioral patterns. Buyers' bids $\beta_{i,k}$ for $k \geq 2$ pooled across all Open Book (left) and Black Box (right) treatments and sellers' asks $\sigma_{j,l}$ for $l \geq 2$ pooled across all sellers' asks in Open Book (left) and Black Box (right) treatments are estimated via robust linear regressions (i.e., $b_0 + b_1\beta_{i,k-1}$ and $s_0 + s_1\sigma_{j,l-1}$ for buyers and sellers, respectively) with bootstrap estimates of 95% CI (shaded), via density kernel estimates (encircled by thin gray lines), via robust nonlinear model fits (quadratic/quadratic rational functions, that is, $\frac{c_0^B+c_1^B\beta_{i,k-1}+\cdots+c_n^B(\beta_{i,k-1})^n}{1+d_1^B\beta_{i,k-1}+\cdots+d_m^B(\beta_{i,k-1})^m}$ and $\frac{c_0^S+c_1^S\sigma_{j,l-1}+\cdots+c_n^S(\sigma_{j,l-1})^n}{1+d_1^S\sigma_{j,l-1}+\cdots+d_m^S(\sigma_{j,l-1})^m}$ with $n = m = 2$ for buyers and sellers, respectively), and via nonparametric LOWESS regressions. Note the logarithmic scale on both axes.

FIGURE 8

WITHIN-ROUND BID/ASK ADJUSTMENTS

within-round adjustments that occurred in the time window under observation, whereby we pooled across all Black Box and Open Book treatments, respectively. For Black Box, we find that buyers concede more than sellers during the first three 30-second time windows of the three initial rounds.[37] Subsequently, the differences disappear throughout most of the trading rounds and only appear early in each round.[38]

We repeat these analyses for the Open Book treatments. Overall, we find that, as was visible in Figures 8 and 9, buyers and sellers adjust their demands even more similarly than in Black Box.[39] A direct comparison of buyers/sellers in Black Box versus Open Book treatments reveals that during the first 30 seconds of the first round, both buyers and sellers lessen their demands in Open Book treatments more than their counterparts in Black Box treatments. This might drive the phenomenon that deal prices in the former converge faster to the CE range. Afterward, demand reductions tend to be larger in Black Box as opposed to Open Book treatments, and the differences are more prevalent for buyers than for sellers, which goes hand in hand with markedly higher volatility in Black Box at the individual level (refer to Figure 9), as well as slower convergence to CE. These phenomena are also visible in Figure 10, which depicts how distances between bids/asks and CE prices evolve over time at the macro level. Indeed, the buyer–seller differences stand out the most at the beginning of the first trading round, and they are more pronounced in Black Box treatments. These differences become smaller in subsequent trading rounds. Figure 10 also illustrates that all traders—especially buyers in Black Box treatments—"reset" their demands for profit and begin a new trading round with renewed aggressiveness. We shall investigate this phenomenon next in our analysis of between-round adjustments.

[37] We have $p < 0.01$ indicating greater buyer yieldingness during the [60, 90] time window of the first round, the [30, 60] and [60, 90] time windows of the second round, and the [0, 30] and [30, 60] time windows of the third round.

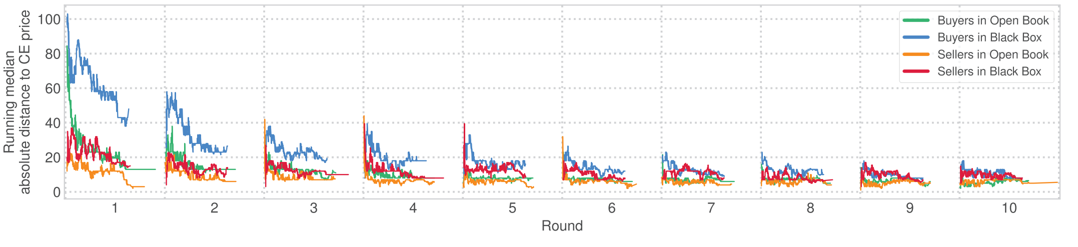
[38] We have $p < 0.01$ indicating greater buyer yieldingness during the [0, 30] time window in rounds 4, 5, 7, 8, and 10.

[39] There are hardly any instances of differential yielding evidence ($p < 0.01$): buyers are more yielding in the [60, 90] time window of the first round, in the [0, 30] time window of the third round, and in the [30, 60] time window of the seventh round; sellers in the [30, 60] time window of the eighth round.

NOTES: Distribution of bid/ask within-round adjustments $\delta^T_{b_i,k}$ and $\delta^T_{s_j,k}$ (top) and between-round adjustments $\Delta^T_{b_i,1}$ and $\Delta^T_{s_j,1}$ (bottom) over the first 10 rounds. The overlaid lines highlight the median values. The bid/ask within-round and between-round adjustments are pooled across all Open Book (left) and all Black Box (right) treatments, respectively.

FIGURE 9

WITHIN-ROUND (TOP ROW) AND BETWEEN-ROUND (BOTTOM ROW) BID/ASK ADJUSTMENTS OVER TIME



NOTES: The median absolute bid/ask distances to CE are calculated across all experimental sessions using sliding windows of size 100 (at the outset of each round, the minimum window size is set to 5). Specifically, all Open Book/Black Box bid/ask distances to CE are sorted by trading round T and time t_T, and, for time t_T, the median distance is computed by taking into account the distance at time t_T and the most recent 99 distances preceding it in the trading round T under observation. If there are strictly less than 99 but at least four distances preceding it, all of these are taken into account. Otherwise, the median value is not computed.

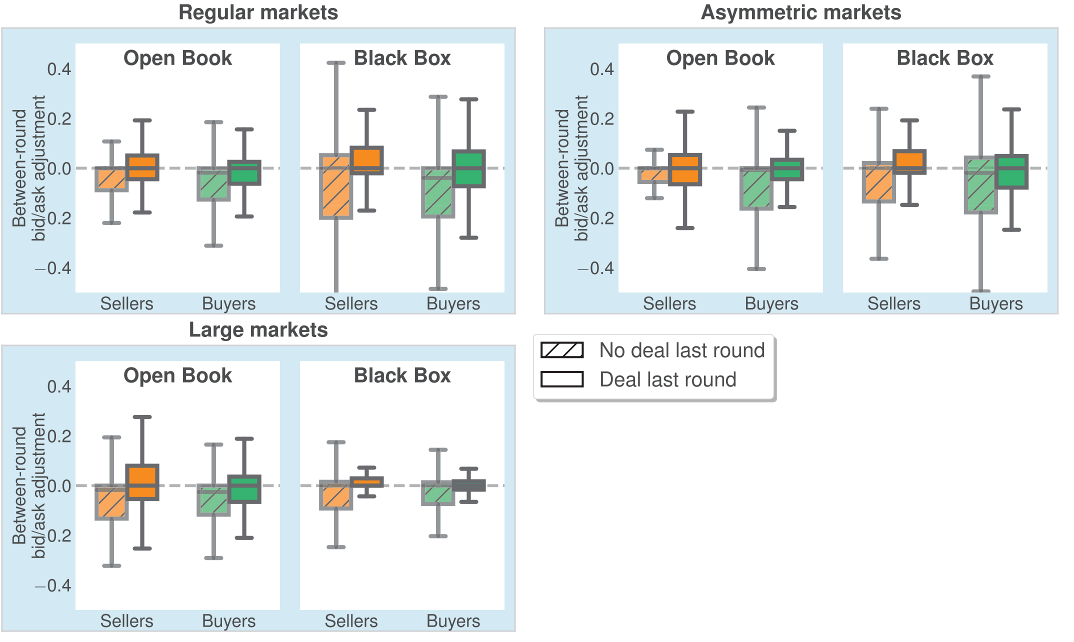
FIGURE 10

RUNNING MEDIAN ABSOLUTE DISTANCE TO CE

4.2.2. *Payoff-based adjustments between rounds.* We now analyze adjustments that occur from one round to another. Concretely, we analyze first bids $\beta^T_{i,1}$ and asks $\sigma^T_{j,1}$ submitted in a given trading round $T \geq 2$ and how these depend on (not) having made a deal in the previous round $T-1$.

DEFINITION 7. A *between-round adjustment* is the difference between a player's first relative demand $\rho^T_{\cdot,1}$ in round $T \geq 2$ and first relative demand $\rho^{T-1}_{\cdot,1}$ in the previous round $T-1$, that is, $\Delta^T_{\cdot,1} = \rho^T_{\cdot,1} - \rho^{T-1}_{\cdot,1}$, where $T \geq 2$.

In contrast to within-round adjustments, between-round adjustments are not necessarily preceded by a failed trading attempt, and we can differentiate between between-round ad-



NOTES: Distribution of between-round adjustments $\Delta^T_{b_i,1}$ and $\Delta^T_{s_j,1}$ in response to (not) having made a deal at the end of round $T-1$ for all trading rounds $T \geq 2$. The depicted between-round adjustments are pooled across all Open Book/Black Box regular, asymmetric, and large markets, respectively. See Figure A.13 for an overview of the individual market results.

FIGURE 11

BETWEEN-ROUND ADJUSTMENTS

justments following a deal versus no deal. As Figure 11 indicates, this distinction is important, because traders who did not deal reduce demand much more than those who did, as is confirmed by a series of MWW tests.[40] In fact, as is visible from the figure, the median adjustment following a deal is zero across treatments for both market sides, and period-to-period price adjustment dynamics are driven by those who failed to trade, who become less aggressive in order to increase their chances of trading next period. These findings complement our prior analysis of within-round adjustments, which identified patterns of demand reduction following failure to trade. Whereas the within-round pattern has been investigated, this between-round dynamic is a novel finding.[41] Jointly, the two dynamics create the well-known patterns of "market jaws" (Bossaerts and Plott, 2008), which get smaller over time—see Figure 10.

Beyond the deal versus no-deal differences in adjustments, we again find differences between buyers and sellers, especially in early rounds. As is visible in the bottom row of Figure 9, this analysis reveals a similar story to what we found regarding within-round adjustments (see the top row): early on, buyers make larger between-round adjustments (becoming less aggressive) than sellers, especially after failure to trade, but these differences vanish over time.[42] There are no notable differences between subjects in Black Box and Open Book treatments in terms of the levels of between-round adjustments, but adjustments are more volatile

[40] MWW tests, which were applied to the group of all buyers'/sellers' between-round adjustments following a deal and the group of all buyers'/sellers' between-round adjustments not following a deal pooled from the same—depending on the test—(set of) treatment(s), have a $p < 0.01$ for both buyers and sellers in almost all of the scenarios (see Table A.3 and Figure A.13).

[41] Since contributions discussed in Friedman and Rust (1993), some more recent work has been done on rationalizing bid/ask adjustments within period (Hollifield et al., 2004; Rostek and Weretka, 2012), but not between periods.

[42] MWW tests confirm this (with $p < 0.01$) in Black Box after the first, third, and eighth round, and in Open Book after the first and second rounds. For each trading round separately, an MWW test was applied to the group of all

in Black Box (see Figure 9). How between-round adjustments relate to within-round adjustments at the macro level is visualized in Figure 10.

4.2.3. *Two simple rules.* Taken together, payoff-based adjustments from the analysis of Black Box can be summarized by two simple rules:

- *Within-round adjustments to secure a deal*: Buyers/sellers gradually become less aggressive until they make a deal or trading ends for the round.
- *Between-round adjustments to manifest or improve the current deal status*: After making a deal in a prior round, buyers/sellers start the new trading round with similarly (or more) aggressive demands. Otherwise, after failing to make a deal in a prior round, buyers/sellers make less aggressive demands in the next round.

Qualitatively, these rules apply for buyers and sellers alike.[43] Figure 10 illustrates them at an aggregate level across all experimental subjects. Quantitatively, buyers are more aggressive than sellers, with difference being most pronounced in the first few rounds and standing out most in Black Box markets. Note that this buyer-seller discrepancy in terms of aggressiveness also explains why high excess demand leads to deal prices much more distant (below) from the CE range in comparison to the prices implied by equally high excesses of supply that results in prices barely above CE. Importantly, by virtue of obtaining qualitatively (and initially even quantitatively) similar results for Open Book and Black Box, our findings also suggest that the asymmetric convergence to CE that is observed at the market level is not driven by (the expectation of) the feedback from the order book but is rather inherent to the trading process. One difference that comes out over time between Black Box and Open Book is that traders in Black Box are less yielding, presumably due to the lack of order-book feedback, as is evident from Figure 10. Before we get into a formal analysis, inspection of Figure 10 already visualizes very clearly that the role of available order-book feedback is largest during the very first round of trading: whereas trading in Black Box and Open Book treatments starts off with comparable demands, incoming feedback in Open Book treatments leads to much quicker reductions (especially by buyers). How order-book feedback and price realizations affect the trading dynamics will be our final layer of analysis.

5. ORDER-BOOK EFFECTS

5.1. *Within- and Between-Round Dynamics in Open Book.* In the three Open Book treatments (i.e., Same-side OB, Other-side OB, and Full OB), more available information potentially influences bid/ask behavior of traders on both sides, not just trader's own history and own past deals as was the case for Black Box. This means that the analysis that we have performed for Black Box does not capture some of the potentially relevant aspects, even though—as was confirmed in the previous section—there was qualitatively positive evidence for the same ingredients of payoff-based learning rules in Open Book treatments as well: demands were adjusted depending on trade success, and buyers were more aggressive than sellers.[44]

Quantitatively, however, estimates were different, effect sizes of different magnitudes, and models generally fit worse, which indicates that access to the order book indeed changes bidding/asking behaviors. To investigate how, we perform a second set of analyses that makes use

buyers' between-round adjustments and the group of all sellers' between-round adjustments corresponding to the round under observation, which were pooled across all Black Box and Open Book treatments, respectively.

[43] Note that they are exactly the kinds of patterns depending on positive/negative impulse implied by "aspiration adjustment" theory (Sauermann and Selten, 1962), as were experimentally observed in bargaining games in early experimental work by Tietz and Weber (1972).

[44] Recall that we qualitatively identified the same behavioral patterns in Open Book as in Black Box, in particular as concerns the differential aggressiveness of buyers and sellers (refer to Figures 8, 9, and 10; see also Tables A.2 and A.3 and Figures A.11, A.12, and A.13).

of the partial–versus–full order-book access design of our experiments. The variation in terms of plausible and available impulses permits us to identify how prior bids, asks, and realized prices trigger behavioral adjustments. In sum, this analysis will identify realized deal prices as the central source of influence, and we shall find no further evidence of buyer–seller differences related to usage of the order book.

A disclaimer is opportune for our analysis, because our aim is not to identify the true underlying cognitive processes that drive these adjustments (candidates for which include Bayesian reasoning, imitation, trial-and-error, momentum, ZI trading, etc.). Instead, our analytical approach is aimed at identifying in which direction and how feedback impulses drive bid/ask adjustments, and it is agnostic to the cognitive foundations in the sense that we shall pursue an econometric analysis of which impulses correlate with what kinds of responses without getting into a deeper quest concerning the true psychological model for these adjustments.

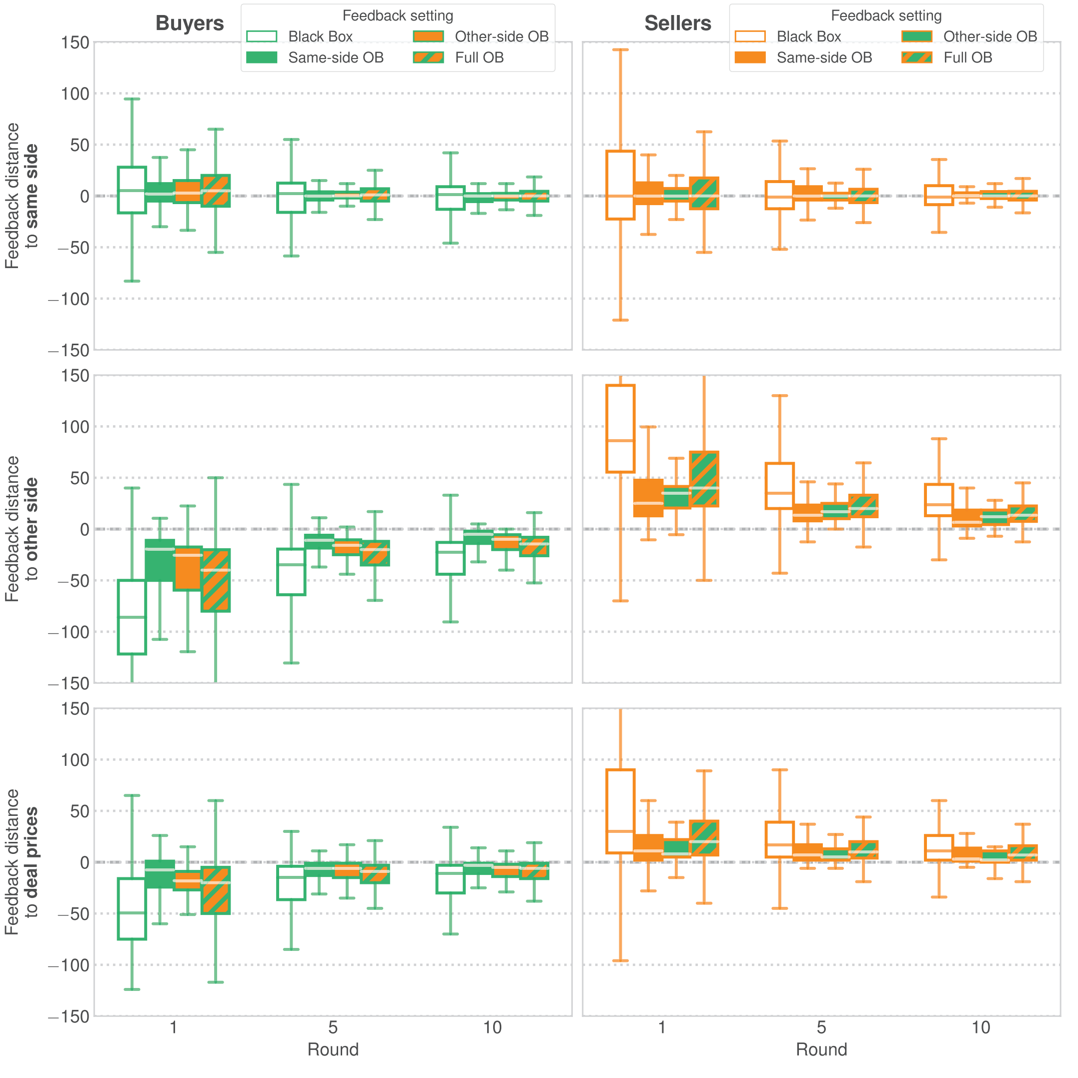
To perform our analysis of feedback effects, we benchmark the Open Book treatments, which offer feedback, against the Black Box treatments, which do not, by imputing the missing feedback with what it would have been had it been provided. We begin our analysis with the effect of the same-side feedback, then turn to the other-side feedback, and finally take a look at the impact of deal information. In all cases, we evaluate the impact of a feedback impulse on a bid/ask by computing the distance between the two.

DEFINITION 8 (FEEDBACK DISTANCE). *Feedback distance* is the distance between a player's action $\alpha(t_T)$, which may be a bid $\beta_i^T(t_T)$ or an ask $\sigma_j^T(t_T)$, and a feedback impulse F, that is, $d(\alpha(t_T), F) = \alpha(t_T) - F$, where F, in principle, may be any statistic that can be extracted or computed from information accessible to the player at any time $t'_T < t_T$ during round T. Our candidates will be others' bids and asks and prior realized prices.

5.2. *Same- and Other-Side Feedback.* As the amount of information that is available to traders grows over time t_T, it becomes increasingly more difficult to disentangle the influences of the different (changing) impulses. For this reason, we begin our analysis by focusing on trading up to the time when the first deal is struck. Interestingly, this window of time is similar in duration across all the treatments, regardless of the (lack of) order-book feedback. It lasts for roughly 10 seconds, during which time approximately 10 market participants place their first bids and asks. A total of 10 seconds is probably too short for feedback to properly kick in and to allow subjects to learn anything important, which might explain why we found no major differences between Black Box and Open Book treatments in our analysis of buyers' and sellers' initial relative demands (refer to Subsection 4.1).

As trading continues, differences between the different feedback regimes do arise. To assess the role that the (lack of) real-time access to others' bids and asks plays, we measure the feedback distance between every action $\alpha(t_T)$ and its same-side and other-side impulses. We take the *same-side/other-side impulse* to be the median of the last 10 actions preceding $\alpha(t_T)$ in round T as submitted by players from the same/opposite side of the market as the focal player. A first indirect piece of evidence that the feedback regime matters is the fact that the corresponding feedback distances attain significantly different values for Black Box compared with Open Book (see Figure 12). We present our direct analysis of feedback effects for the first trading round $T = 1$, where the differences are the most significant (see Figure 10). However, the findings we shall present are qualitatively consistent across all rounds T.

5.2.1. *Same-side feedback.* Feedback distances to the same-side impulses tend to be more concentrated around 0 when such feedback is provided compared to Black Box, which indicates that subjects do pay attention to the actions of their trade rivals from the same market side and act more similarly to them (refer to Figure 12). Since the participants are more or less equally likely to submit a bid/ask higher or lower than the median of the past bids/asks, we compare these deviations across the different feedback settings by considering their absolute values instead. The median absolute distance in the first round thus amounts to 22.5 and



NOTES: Difference between a subject's bid/ask and the median of the last 10 bids/asks on the same side of the market (top), the median of the last 10 bids/asks on the other side of the market (middle), and the last realized price (bottom) during the first, fifth, and tenth trading round for all four feedback variations.

FIGURE 12

FEEDBACK DISTANCE FROM SAME-SIDE BID/ASK ACTIVITY (TOP ROW), OTHER-SIDE BID/ASK ACTIVITY (MIDDLE ROW), AND REALIZED PRICES (BOTTOM ROW) ACROSS ALL FEEDBACK REGIMES

30 for the buy side and the sell side under Black Box and, in the same order of the market sides, to 16 and 15 for Full OB, to 9.75 and 9.25 for Same-side OB, and to 12 and 7 for Other-side OB. For every trading round and every market side separately, we applied an MWW test to the group of all corresponding absolute values of the same-side feedback distances pooled across all Black Box treatments and the group of all corresponding absolute values of the same-side feedback distances pooled across all Same-side/Other-side/Full Open Book treatments. MWW tests confirm that the absolute values of the same-side feedback distances are significantly larger in Black Box (with $p < 0.01$), but differentiating between the other treatments is more difficult because other feedback impulses may be at play and interact with one another (in addition to/instead of the access to the same side of the market). Hence, we refine our findings regarding the role that the same-side feedback plays by performing an additional analysis checking for imitative patterns, which are most plausible with regard to same-

side feedback. The details of this analysis are in the Appendix (refer to Figure A.14 and the accompanying text), but the finding is quite intuitive: imitative behavior is indeed most prevalent in Same-side OB followed by Full OB and not consistently identified in Other-side OB and Black Box.

Notice that buyers and sellers, as concerns reacting to same-side feedback, behave in a way that is indistinguishable in terms of aggressiveness. We confirm this separately for Same-side OB and Full OB treatments (i.e., the ones that permit access to these impulses) via Mann–Whitney tests for equivalence with the absolute values of individual subjects' same-side feedback distances as the units of observation: results indicate that (at the 5% significance level and with the equivalence margin set to 0.1) the absolute values of same-side feedback distances do not differ markedly between buyers and sellers.

As a final note, notice that the distributions of the same-side feedback distances for buyers and sellers in Black Box—where same-side impulses are absent—are more left-skewed and right-skewed, respectively, than in the other treatments. Hence, in Black Box, that is, without any order-book feedback whatsoever, players are more aggressive and less yielding, which is in line with our previous findings.

5.2.2. *Other-side feedback.* We now turn to other-side feedback. As before with same-side feedback, again visible in Figure 12, buyers and sellers do not react too differently to other-side impulses. Mann–Whitney tests for equivalence applied to the absolute values of individual subjects' other-side feedback distances confirm that in the Other-side OB and Full OB treatments, the absolute values of other-side feedback distances are similar for buyers and sellers (at the 5% significance level and with the equivalence margin set to 0.13). Comparing treatments, however, differences in line with the previous analysis are identified, as subjects in Black Box are significantly more aggressive. The median absolute distance in the first round is 86 for buy and sell side in Black Box, whereas, in the same order of the market sides, this reduces to 40 for Full OB, to 19.5 and 25.75 for Same-side OB, and to 25.5 and 35 for Other-side OB.[45] Because subjects in Black Box do not have access to other-side impulses, these discrepancies suggest that other-side order-book feedback acts as a "pull" force bringing both sides of the market closer together and reducing traders' aggressiveness overall.[46]

The analysis of same-side and other-side bid/ask feedback impulses can be summarized by stating two complementary forces:

- *Imitation of others*: Same-side feedback makes traders bid/ask more similarly to what others did recently.
- *Bid–ask contraction*: Other-side feedback speeds up bid/ask adjustments toward the bids and asks of the other market side.

The interplay of these two forces increases the speed at which the bid–ask spread contracts compared with Black Box and reduces volatility as all adjustments point in the same direction.[47] As a result, more feedback speeds up convergence.[48] In terms of the number and timing of realized deals, the resulting efficiency gains over Black Box are largest in the first round (median number of deals: Black Box: 3.5, Open Book: 6; median time between deals: Black Box: 17, Open Book: 8), but present across all rounds (see also Figure A.15).[49]

[45] MWW tests comparing Black Box with Same-side/Other-side/Full Open Book treatments, which were conducted in a similar manner as in the same-side feedback analysis, put Black Box (with $p < 0.01$) deviations above the others.

[46] At first glance it appears as if this contraction is strongest in Same-side OB, but this is the case because bids and asks were closer together in those market sessions than in others already in the first place.

[47] This is very visible in Figure A.16.

[48] See efficiency over time in Figures 3 and A.7.

[49] MWW tests reveal that (at the significance level of 1%) the greater the amount of feedback given, the more deals are struck every round (without significant difference between Same-side OB and Other-side OB). The unit of observation in the MWW tests was the number of deals struck in a trading round of an individual experimental session, and the tests compared the number of deals struck for sessions pooled across all possible feedback treatment (Black Box, Same-side OB, Other-side OB, and Full OB) pairs.

5.3. *Realized Deals.* Our final layer of analysis investigates the effects of deal prices on bid/ask behavior. We follow a similar analysis as in the previous section and compare every subjects' action $\alpha(t_T)$ with the latest *deal-price impulse* (i.e., the price of the most recent deal in that market). The corresponding feedback distances are illustrated in Figure 12. Notice that deal prices have a pull akin to the one observed for other-side feedback. Comparing treatments, Black Box again results in the most aggressive behavior (with the median absolute feedback distance equal to 50 and 32 for buyers and sellers, respectively, as opposed to, in the same order of the market sides, 21 and 20 for Full OB, 10 and 12 for Same-side OB, and 18.5 and 10 for Other-side OB).[50] Whereas there are no differences between buyers and sellers in terms of individual subjects' absolute feedback distances in Full OB (as confirmed by the Mann–Whitney test for equivalence at the 5% significance level with the equivalence margin set to 0.1), this is not the case for Same-side OB and Other-side OB treatments. In the former, it is buyers who bid closer to past realized prices, whereas in the latter, it is sellers who are drawn closer to the deal prices (indicated by MWW tests with $p < 0.01$). Notice that this always applies to the side of the market which only has access to buyers' trading activity. Given that there is a bias toward the buy side of the market and that initial prices tend to be significantly in buyers' favor, this should not come as a surprise. As trading goes on, these differences disappear.

6. CONCLUSION

We conducted a large number of market experiments within a uniform setting to stress-test convergence predictions. Our results are that convergence occurs after a handful of trading rounds via equilibration from initially below-CE prices, as in Smith (1962) and similar studies. The evidence is collected across markets where such an asymmetry is not "built in" at all. Indeed, asymmetric convergence occurs across a rich variety of market settings, all of the single-item continuous DA variety. They were typically of medium size, with a dozen traders on each side. More work is needed to understand convergence patterns in smaller and larger markets, under different frames, with other subject pools, and under different DA market institutions to understand why recent contributions (Lin et al., 2020) may have found different patterns.

Our experimental design did not favor buyers structurally, so the asymmetric dynamics that we observed are rather indicative of behavioral asymmetries between buyers and sellers. The most notable difference between the two that we identified is that buyers are initially more aggressive than sellers. This difference is present and qualitatively similar across all treatments. In terms of subsequent adjustments, we do not observe further notable differences in terms of adjustments of buyers versus sellers that would reinforce the asymmetric price convergence that begins with the initial bids and asks. Sellers are less aggressive initially but not more yielding subsequently. These kinds of individual-level results open up rich avenues for future work related to further developing, fitting, and testing individual learning models such as individual evolutionary learning (Arifovic and Ledyard, 2011), or building on variants of zero/minimal intelligence (Gode and Sunder, 1993). It would also be interesting to see if our observed buyer–seller asymmetry might change when using other frames such as a more neutral one with less trade-inducing language, or in an explicit employer–worker market.

Finally, the informational variations of our treatments allow investigating in more detail which parts of the order-book feedback trigger what kinds of behavioral responses at the individual level, and how these differences aggregate to differences in performance at the market level. Connecting our findings with related efforts such as focusing on complete versus incomplete information (Kimbrough and Smyth, 2018) or on what information traders actively seek (Kirchsteiger et al., 2005) is useful to inform optimal information design in DAs, especially as availability of information may also determine what kinds of algorithms may be used by

[50] Again, these differences are significant at the significance level of 1% based on MWW tests, which were conducted in a similar manner as in the same-side and other-side feedback analysis.

nonhuman traders. These kinds of questions are becoming increasingly relevant as the digitization and automation of markets proceeds further (Bao et al., 2021).

APPENDIX A

A.1 *Experimental Details.* This Appendix contains details including data, registrations, and analyses, which can also be found at our Open Science Framework (OSF) registry under https://osf.io/gu62n/. Here, we provide some summary of the trading platform and the experimental sessions.

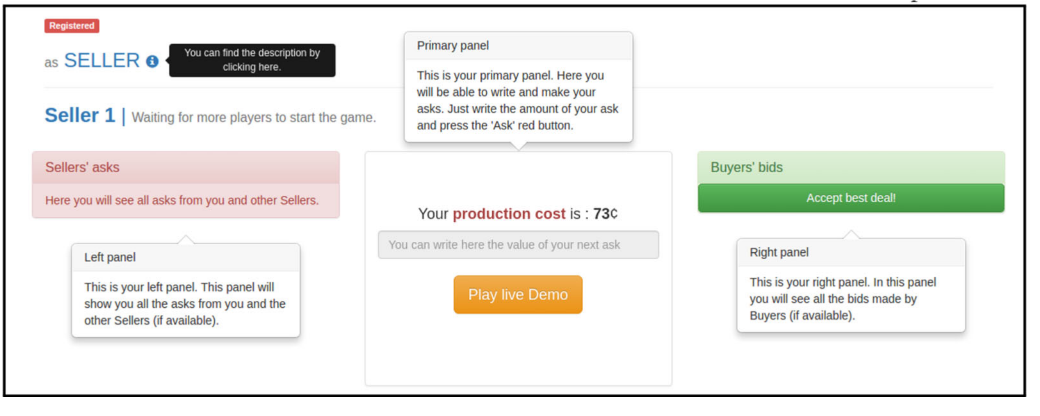
A.1.1 *Trading platform.* SciOn (https://scienceexperiment.online/demo scienceexperiment.online/demo) is an experimental software programmed in PHP and designed to handle multiple large real-time trading experiments in parallel. The kind of simultaneous and real-time market interaction that we are after requires instantaneous synchronization, which we ensure via a WebSocket communication protocol. This protocol allows interaction between client and server with lower overheads, thus facilitating real-time message exchange. Importantly, it allows to quasi-constantly "refresh" an experimental subject's page, thus providing every user with real-time information about other users' bids and/or asks and realized prices (see Figure A.1 for an impression of the front end). In addition, the efficiency of the WebSocket protocol allows running multiple trading rooms separately at the same time, each of which may be large.[51] To possess live control over the current status of the whole system (message dispatching and queuing, transaction count, database, query, etc.) we developed a server monitor, which also contains an emergency module that enables communication with the participants in case of failure in the main server.

The software is easy to handle, with a preprogrammed input mask for market components such as bids, valuations, market size, etc., and available for experimentation and real-time classroom use.[52]

The main experiments were performed on Amazon Mechanical Turk (as in Paolacci et al., 2010; Horton et al., 2011). The task could only be selected by workers in the United States

[51] We tested our platform successfully with over 10 simultaneous trading rooms with at least 20 participants each and for single trading rooms of up to 400 participants.

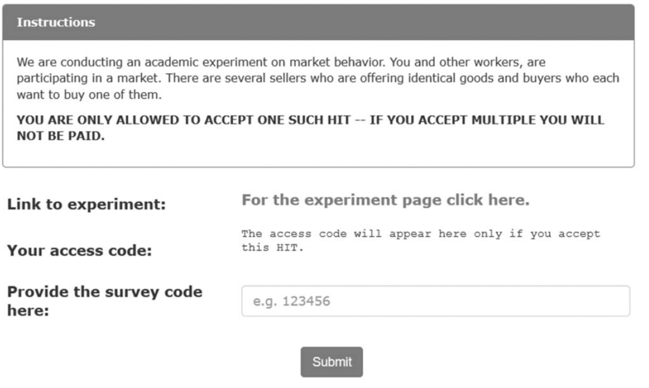
[52] We thank Christoph Kuzmics for being our test pilot for the classroom app.



NOTES: Subjects could familiarize themselves with the trading environment during the instruction phase. Depending on feedback treatment, the different components (e.g., Buyers' bids) were either disabled or enabled. Under Sellers' asks and Buyers' bids, own and other subjects' bids and asks appear in real time, and subjects could click on them to make a deal. The central field is where own bids/asks can be placed.

FIGURE A.1

TRADING INTERFACE (FOR A SELLER)



Instructions

We are conducting an academic experiment on market behavior. You and other workers, are participating in a market. There are several sellers who are offering identical goods and buyers who each want to buy one of them.

YOU ARE ONLY ALLOWED TO ACCEPT ONE SUCH HIT -- IF YOU ACCEPT MULTIPLE YOU WILL NOT BE PAID.

Link to experiment: For the experiment page click here.

Your access code: The access code will appear here only if you accept this HIT.

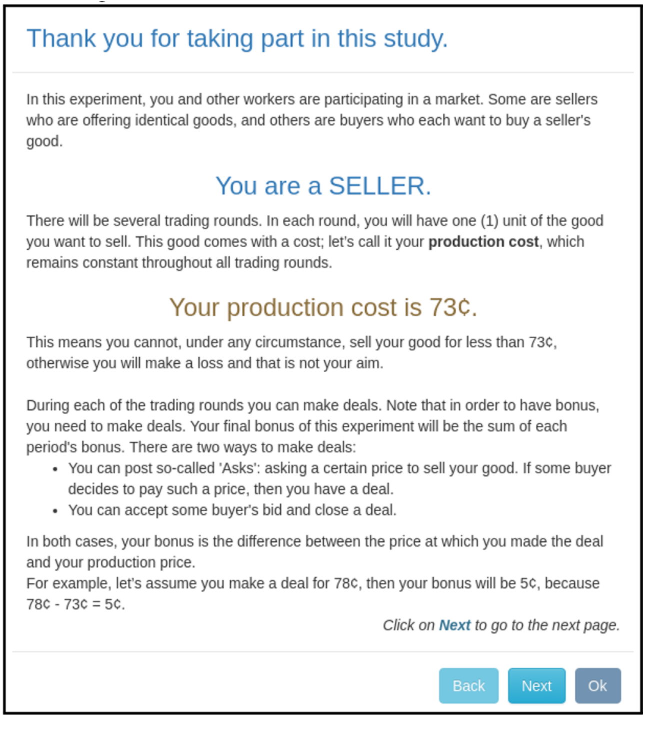
Provide the survey code here: e.g. 123456

Submit

NOTES: This is how our experiments were advertised on Amazon Mechanical Turk during the recruitment period.

FIGURE A.2

EXPERIMENTAL ADVERTISEMENT ON THE ONLINE RECRUITMENT PLATFORM



Thank you for taking part in this study.

In this experiment, you and other workers are participating in a market. Some are sellers who are offering identical goods, and others are buyers who each want to buy a seller's good.

You are a SELLER.

There will be several trading rounds. In each round, you will have one (1) unit of the good you want to sell. This good comes with a cost; let's call it your **production cost**, which remains constant throughout all trading rounds.

Your production cost is 73¢.

This means you cannot, under any circumstance, sell your good for less than 73¢, otherwise you will make a loss and that is not your aim.

During each of the trading rounds you can make deals. Note that in order to have bonus, you need to make deals. Your final bonus of this experiment will be the sum of each period's bonus. There are two ways to make deals:

- You can post so-called 'Asks': asking a certain price to sell your good. If some buyer decides to pay such a price, then you have a deal.
- You can accept some buyer's bid and close a deal.

In both cases, your bonus is the difference between the price at which you made the deal and your production price.
For example, let's assume you make a deal for 78¢, then your bonus will be 5¢, because 78¢ - 73¢ = 5¢.

*Click on **Next** to go to the next page.*

Back Next Ok

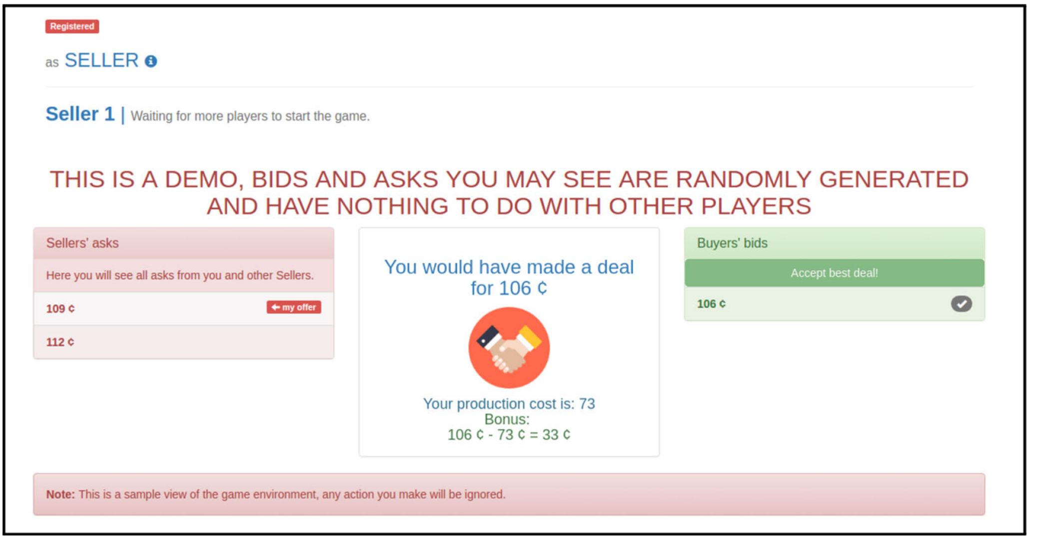
NOTES: These are the instructions provided to a seller prior to the demo phase (see Figure A.4).

FIGURE A.3

AN EXAMPLE OF EXPERIMENTAL INSTRUCTIONS (THE CASE OF A SELLER)

who had not participated in any of the account holder's (ETH Descil) previous experiments. The experimental instructions contained a link to the trading Web page (see below) and a personalized access code. The flat rate participation fee was $1.00 with a possible bonus as high as $7.50 per 10 rounds of play, depending on treatment. The task overview read as follows:

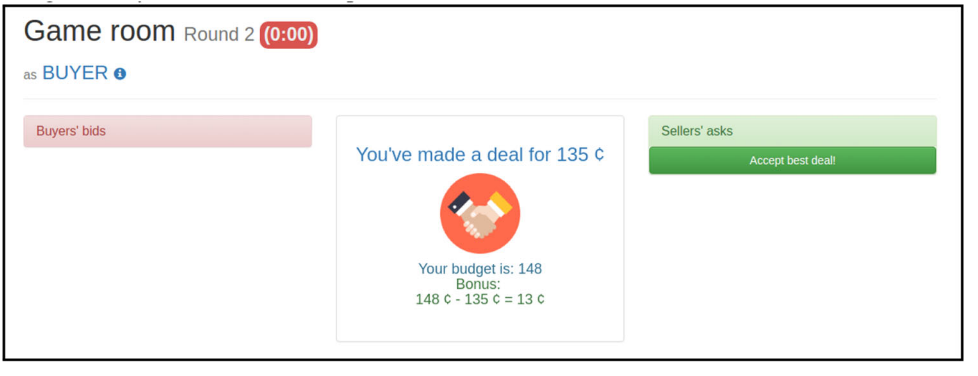
- **Title**: Take part in a market experiment where you are trading goods, WITH BONUS
- **Description**: You will trade in a market and will receive your profits as a BONUS at the end of the game



NOTES: Subjects were able to play around with a noninteractive demo, where (depending on treatment) random bids and/or asks arrived on the two market sides.

FIGURE A.4

HOW THE TRADING DEMO LOOKS



NOTES: Subjects saw their own deals as displayed below in all treatments. In addition, they saw a deal being made by others and at what price in all treatments other than Black Box.

FIGURE A.5

WHEN A DEAL IS MADE

- **Keywords**: game, interactive, market
- **Reward**: $1.00

Workers were then able to open a description of the task that contained instructions, including a hyperlink to the trading Web page used for the experiment. Figure A.2 shows a screenshot of the description pane.

If the worker accepted the HIT and followed the instructions, he would next receive his personal access code to enter the trading experiment on the trading Web page. He would first get instructions as detailed in Figure A.3. In particular he would be informed whether he is a buyer or a seller and his respective budget or production cost.

In the next step, the agent was able to see a short video explaining the pane seen during the experiment and how bids, asks, and trades are displayed. Figure A.4 illustrates the display as seen during this explanation video. In the center, each agent can enter his bid or ask. On the

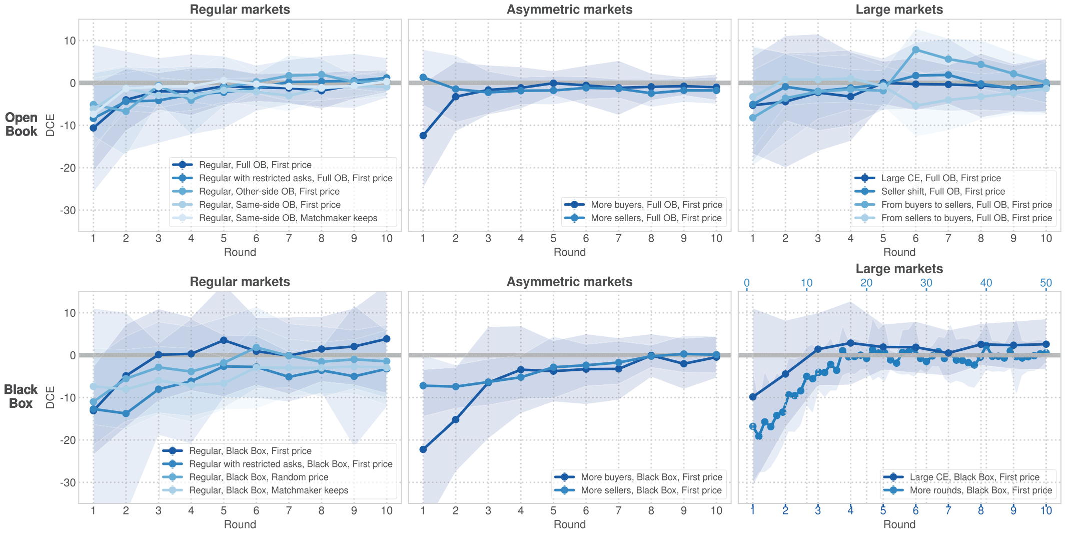


FIGURE A.6

CONVERGENCE TO CE PRICES IN TERMS OF DCE

right-hand side agents can see the bids/asks from the other side of the market (in the treatments where this information is available). On the left-hand side agents can see the bids/asks from their side of the market (in the treatments where this information is available).

If an agent made a deal the pane changed as shown in Figure A.5. In particular, the agent was informed about the monetary profit he would receive from the deal (price minus production cost for sellers and budget minus price for buyers).

Upon completion of the experiment each worker received a personalized survey code to submit on Amazon Mechanical Turk (AMT). This allowed us to track the bonus each player had to be paid.

A.2 *Data Analysis.*

A.2.1 *Convergence patterns.*

A.2.2 *Predictive success index (PSI).* Although subject inactivity did not pose a significant issue in our experiments, (temporary) absence of any number of subjects from a trading round may nevertheless potentially alter the corresponding competitive equilibrium (CE) range and, consequently, result in a PSI which then would not properly reflect how good the CE range is as a predictor for the prices (realized in the long term).[53] The left-hand side of Figure 4 illustrates the PSI for the data treated without attempting to correct for such issues, which is broken up as per the individual market types in Figure A.8.

Taking a closer look at the raw data, which includes 1,238 individual rounds in total, it turns out that the effective CE ranges vary comprising between a minimum of one and a maximum of 66 prices with a median of 6.[54] The absence of subjects thus may create two problems. On the one hand, it may shrink the CE range to a single price (which occurs in 26% of the cases), possibly resulting in values of the hit rate close to zero and hence low values of PSI even though the prices may well be in close proximity to the CE. On the other hand, the CE range may become very wide, potentially resulting in values of the area close to one, although a closer inspection of the data would hint at misspecification issues. Indeed, when this

[53] Note that the CE ranges are calculated at the level of individual rounds for each of the experimental sessions separately.

[54] No deal was struck in seven of these rounds.

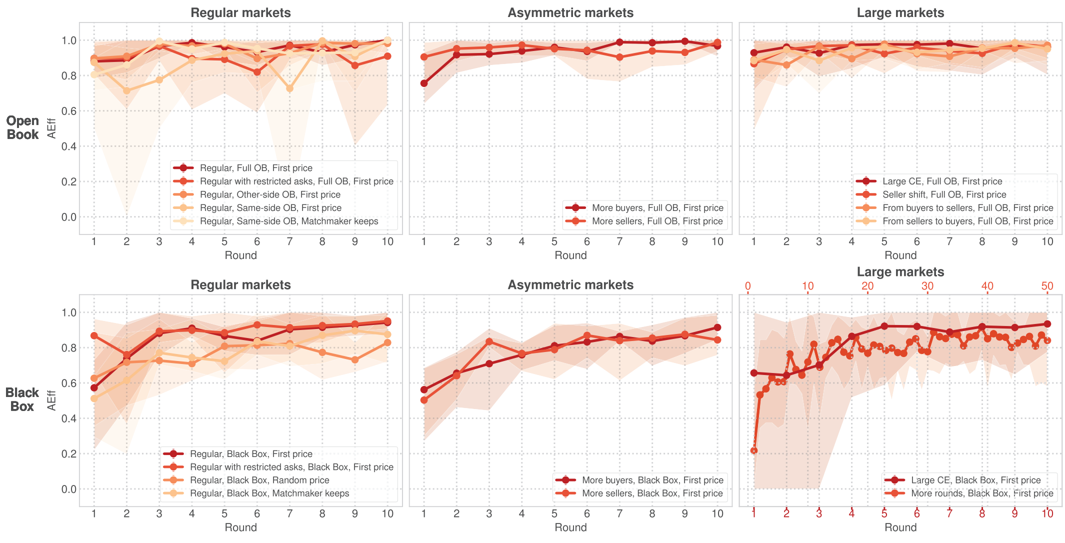


FIGURE A.7

REALIZED ALLOCATIVE EFFICIENCY (AEff)

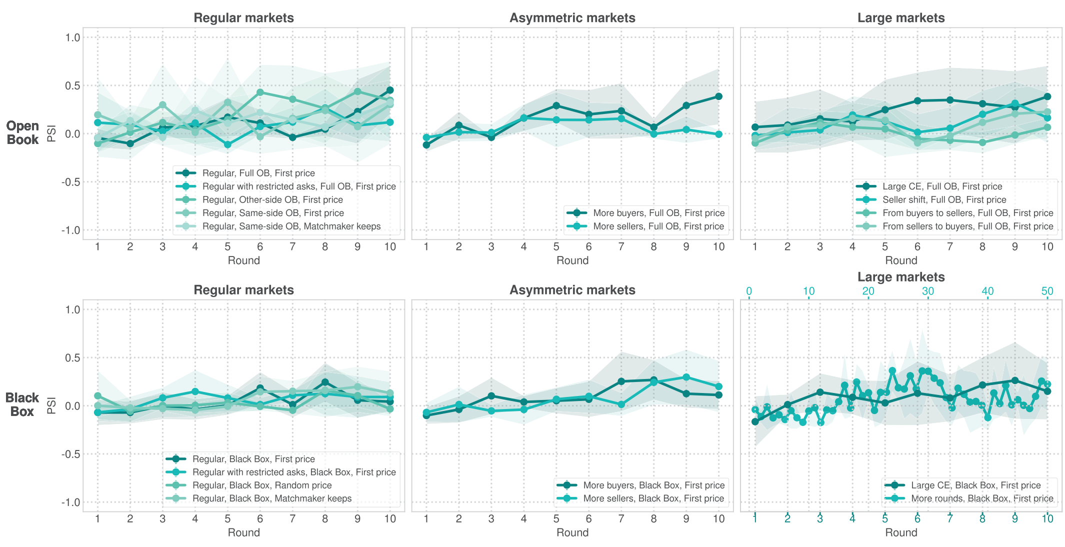
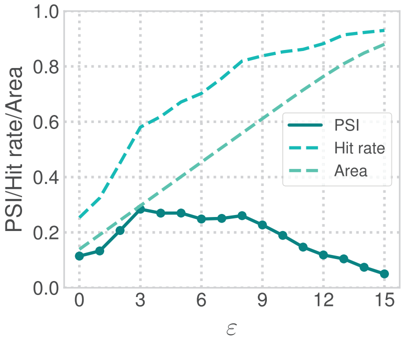


FIGURE A.8

PREDICTIVE SUCCESS INDEX (PSI)

is the case, the prices can typically be observed to converge to a much narrower range of values within the CE range. Note that for our data, the values for the area range from as low as 0.02 to as high as 0.87, with a median of 0.14.

All in all, to ameliorate the problems associated with disproportionate CE ranges without interfering with the data too much, we follow a simple procedure. Notice first that shrinking the length of the CE range to zero causes the hit rate and the area to tend to zero as well. In contrast, expanding the CE range until it covers the entire range of possible price realizations yields a hit rate and an area of one. In both cases, the PSI amounts to zero, so neither of the two extremes is a good choice. What can be done instead is to systematically probe all the variations in between, and to pinpoint the one with the highest PSI, and compare it with the original CE range. Specifically, we recompute the predictive success indices by substituting



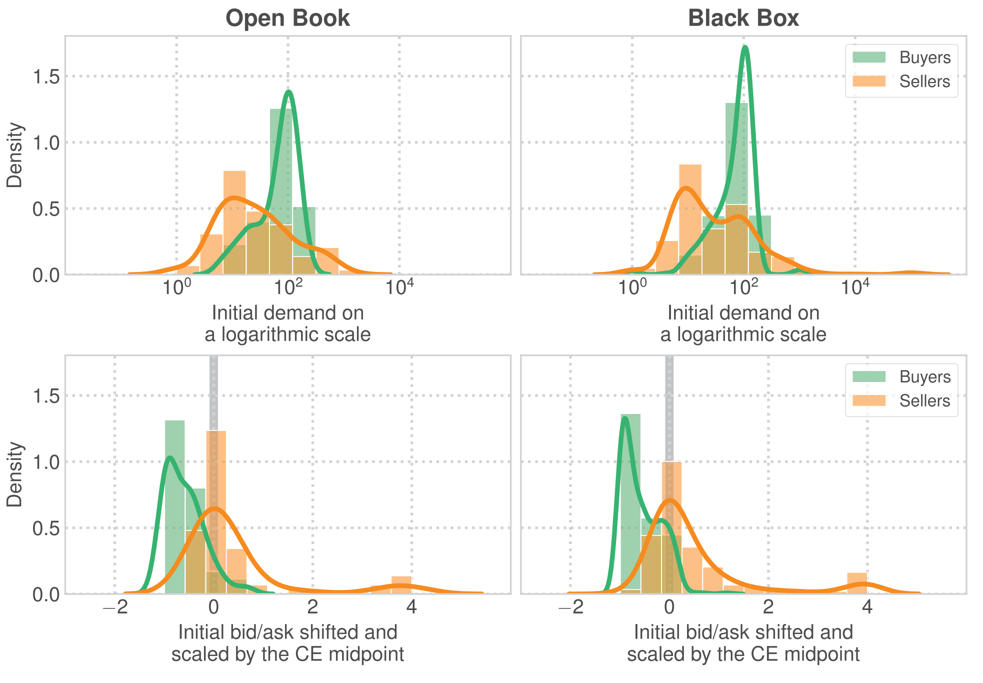
Treatments	PSI	*Hit rate*	*Area*
Open Book	0.31	0.62	0.30
Black Box	0.25	0.54	0.29
All	0.28	0.58	0.30

Note that the numbers do not add up due to rounding.

NOTES: *Left*. Hit rate, area, and PSI across all experimental treatments, sessions, and trading rounds as a function of ε modulating the widths of the original CE ranges $[\underline{P}^*, \overline{P}^*]$. *Right*. Overall PSI for $\varepsilon = 3$. The reported values are obtained by taking into account all prices realized in the relevant experimental sessions across all trading rounds. Note that the area is computed as the arithmetic mean of the per-session per-round areas, whereby every realized price is given the same weight.

FIGURE A.9

PREDICTION SUCCESS INDEX OF $[\underline{P}^* - \varepsilon, \overline{P}^* + \varepsilon]$



NOTES: Histograms and Gaussian kernel density estimates of the initial demands for buyers and sellers in Open Book and Black Box treatments. *Top*. Absolute demands on a logarithmic scale. *Bottom*. Initial bids/asks shifted and scaled by the CE midpoint, that is, $(\beta_{i,1} - P^*_M)/P^*_M$, $(P^*_M - \sigma_{j,1})/P^*_M$ for $P^*_M = (\underline{P}^* + \overline{P}^*)/2$ (values greater than 4 are clipped and redistributed into the rightmost bin, which applies to roughly 4% of values).

FIGURE A.10

INITIAL DEMANDS

the original CE ranges $[\underline{P}^*, \overline{P}^*]$ with ranges of the form $[\underline{P}^* - \varepsilon, \overline{P}^* + \varepsilon]$ for various values of $\varepsilon \in \mathbb{R}$, whereby we use the same value of ε for all trading rounds and all sessions taken into account. The results are reported in Figure A.9 and indicate that the best trade-off between precision and accuracy can be achieved by extending the CE ranges by 3 units upward and downward, which is not too far off from the actual CE ranges of the underlying experimental markets.[55]

A.2.3 *Initial bids/asks.*

TABLE A.1
INITIAL RELATIVE DEMANDS IMPLIED BY FIRST BIDS AND ASKS

	Treatment		*Median Initial Relative Demand*		*MWW*
Feedback	*Market Structure*	*Price Rule*	*Buyers*	*Sellers*	*p-Value*
Open Book			0.68 (42)	0.19 (70)	<0.01
	Regular		0.73 (20)	0.14 (33)	<0.01
Full OB		First price	0.96 (3)	0.14 (5)	
	With restricted asks	First price	0.59 (3)	0.21 (4)	
Other-side OB		First price	0.99 (3)	0.09 (10)	
Same-side OB		First price	0.32 (7)	0.11 (6)	
		Matchmaker keeps	0.5 (4)	0.3 (8)	
Full OB	**Asymmetric**	First price	0.84 (7)	0.37 (6)	(0.42)
	More buyers		0.63 (6)	0.26 (2)	
	More sellers		0.93 (1)	2.68 (4)	
Full OB	**Large**	First price	0.67 (15)	0.19 (31)	<0.01
	Large CE		0.84 (6)	0.43 (16)	
	Seller shift		0.14 (3)	0.05 (5)	
	From buyers to sellers		0.47 (2)	0.06 (4)	
	From sellers to buyers		0.82 (4)	0.33 (6)	
Black Box			0.71 (356)	0.19 (353)	<0.01
	Regular		0.71 (160)	0.2 (161)	<0.01
		First price	0.85 (42)	0.46 (42)	(0.23)
	With restricted asks		0.32 (35)	*0.24 (36)	*(0.16)
		Random price	0.68 (40)	0.13 (38)	<0.01
		Matchmaker keeps	0.84 (43)	0.2 (45)	<0.01
	Asymmetric	First price	0.71 (118)	0.17 (119)	<0.01
	More buyers		0.65 (74)	0.14 (39)	<0.01
	More sellers		0.76 (44)	0.24 (80)	<0.01
	Large	First price	0.72 (78)	0.19 (73)	<0.01
	Large CE		0.65 (38)	0.28 (36)	(0.31)
	More rounds		0.77 (40)	0.15 (37)	<0.01
		All	0.71 (398)	0.19 (423)	<0.01

NOTE: Median initial relative demands comparing buyers and sellers in Open Book and Black Box treatments. *p*-Values correspond to MWW tests applied to the group of all buyers' initial relative demands and the group of all sellers' initial relative demands pooled from the treatments indicated in the first three columns. Numbers in brackets indicate the number of bids/asks considered in the analysis. $*$Due to technical issues, seven sellers were assigned a reservation price of 1 (instead of 103 and 108). Excluding their first asks, the median sellers' initial relative demand amounts to 0.14, and the MWW test yields a *p*-value of less than 0.01.

[55] Note that $\varepsilon = 3$ also turned out to be the best choice if we restricted ourselves to (non-)Black Box markets only. Moreover, shrinking the CE ranges instead of extending them, that is, using $[\underline{P}^* - \varepsilon, \overline{P}^* + \varepsilon]$ for $\varepsilon < 0$, led to a decrease in PSI.

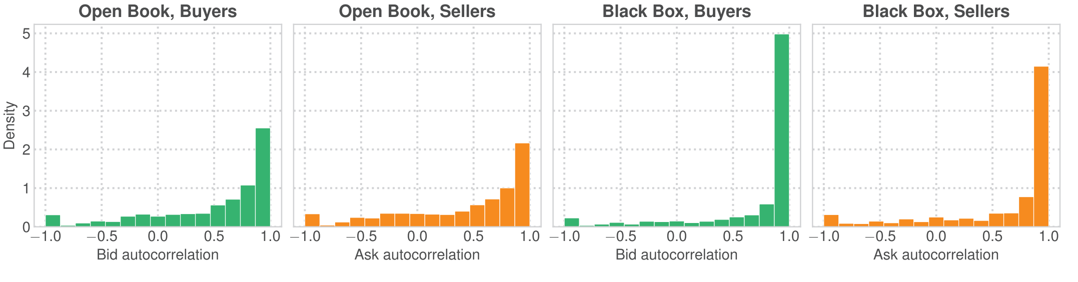
A.2.4 *Within-round adjustments.*

TABLE A.2

WITHIN-ROUND ADJUSTMENTS OF RELATIVE DEMANDS

	Treatment		Buyers		Sellers	
Feedback	*Market Structure*	*Price Rule*	*Median* $\delta^T_{b_i,k}$	$\delta^T_{b_i,k} \leq 0$	*Median* $\delta^T_{s_j,k}$	$\delta^T_{s_j,k} \leq 0$
Open Book			0 (19,098)	86%	0 (20,326)	87%
	Regular		0 (9,269)	86%	0 (7,892)	86%
Full OB		First price	0 (2,092)	87%	0 (2,019)	90%
	With restricted asks	First price	−0.01 (1,521)	86%	−0.01 (589)	80%
Other-side OB		First price	0 (1,830)	89%	0 (1,735)	87%
Same-side OB		First price	0 (1,976)	83%	0 (1,420)	77%
		Matchmaker keeps	0 (1,850)	87%	0 (2,129)	90%
Full OB	**Asymmetric**	First price	0 (5,112)	85%	0 (6,006)	88%
	More buyers		0 (3,639)	82%	−0.03 (824)	92%
	More sellers		−0.01 (1,473)	94%	0 (5,182)	87%
Full OB	**Large**	First price	−0.01 (4,717)	87%	−0.01 (6,428)	86%
	Large CE		−0.01 (2,182)	87%	−0.01 (3,081)	87%
	Seller shift		−0.01 (809)	90%	−0.03 (895)	89%
	From buyers to sellers		−0.01 (692)	86%	−0.01 (1,063)	89%
	From sellers to buyers		−0.02 (1,034)	85%	0 (1,389)	82%
Black Box			−0.03 (14,024)	87%	−0.01 (13,947)	84%
	Regular		−0.03 (6,577)	85%	−0.01 (6,956)	82%
		First price	−0.02 (1,792)	83%	−0.02 (1,938)	86%
	With restricted asks		−0.04 (917)	87%	−0.03 (879)	89%
		Random price	−0.04 (1,765)	88%	−0.01 (1,743)	80%
		Matchmaker keeps	−0.02 (2,103)	84%	−0.01 (2,396)	78%
	Asymmetric	First price	−0.03 (3,704)	89%	−0.01 (4,099)	84%
	More buyers		−0.02 (2,581)	87%	−0.04 (1,158)	84%
	More sellers		−0.03 (1,123)	93%	−0.01 (2,941)	84%
	Large	First price	−0.02 (3,743)	90%	−0.02 (2,892)	88%
	Large CE		−0.02 (1,198)	86%	−0.02 (1,360)	85%
	More rounds		−0.03 (2,545)	91%	−0.02 (1,532)	91%
		All	−0.01 (33,122)	87%	0 (34,273)	86%

NOTE: Median within-round adjustments $\delta^T_{b_i,k}$ and $\delta^T_{s_j,k}$ and the percentage of nonpositive adjustments, comparing buyers and sellers in Open Book and Black Box treatments across all rounds T. Numbers in brackets indicate the number of bids/asks considered in the analysis.

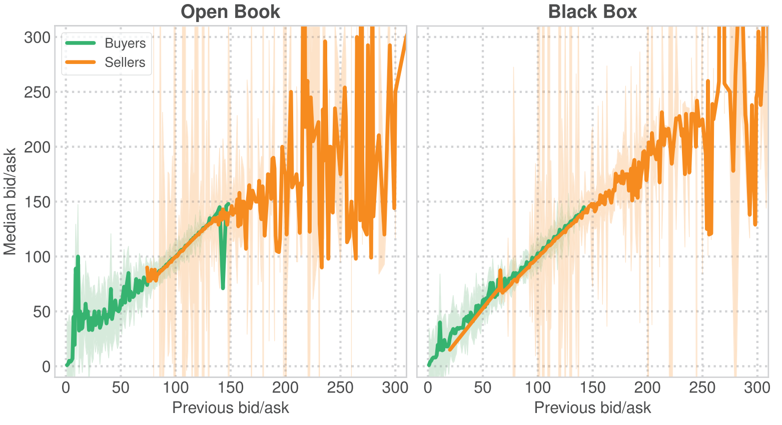


NOTES: Distributions of within-round within-subject correlation between bids/asks submitted by a player in a trading round. The corresponding median values are: 0.71 (buyers in Open Book), 0.63 (sellers in Open Book), 0.96 (buyers in Black Box), and 0.92 (sellers in Black Box).

FIGURE A.11

WITHIN-ROUND WITHIN-SUBJECT BID/ASK AUTOCORRELATIONS

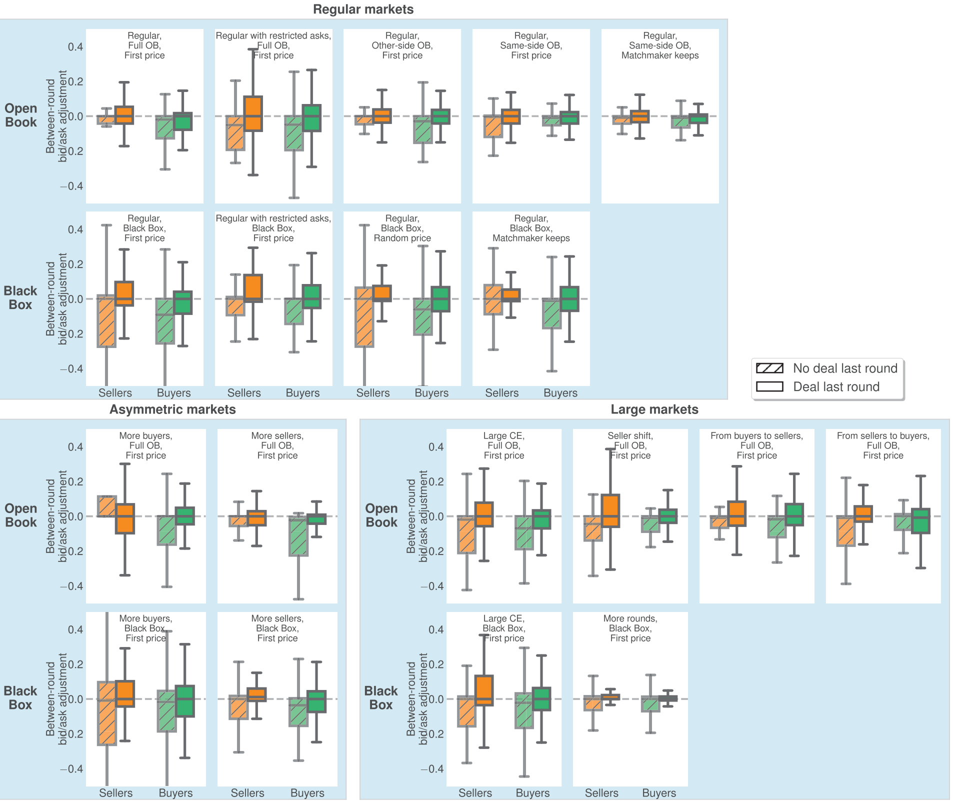
A.2.5 *Between-round adjustments.*



Notes: The shaded areas correspond to standard deviations of the data. Note that we zoomed in on asks smaller than 300 (which applies to 95% of sellers' asks) to make the trends clearer. Bids and asks are pooled across all Open Book and Black Box treatments, respectively.

Figure A.12

median bid/ask following a given bid/ask



NOTES: Distribution of between-round adjustments $\Delta^T_{b_i,1}$ and $\Delta^T_{s_j,1}$ in response to (not) having made a deal at the end of round $T-1$ for $T \geq 2$.

FIGURE A.13

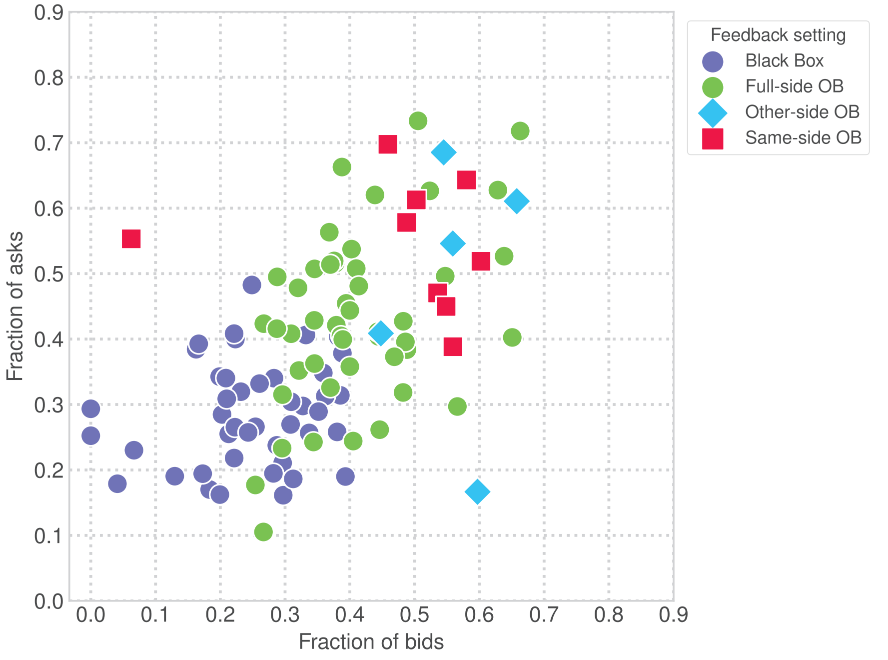
BETWEEN-ROUND ADJUSTMENTS

A.2.6 *Order-book effects.* To investigate to what extent the subjects rely on the feedback concerning their own side of the market when choosing their actions, we take a closer look at imitative behavior and benchmark markets that provide such feedback against those that do not. To this end, we consider each individual market session subject to one of our four feedback treatments separately and simply compute the observed fraction of bids and asks that mirror one of the 10 most recent bids and asks, respectively. The results are displayed in Figure A.14 and clearly show that three distinct clusters form. In the lower-left corner, there is a high concentration of Black Box market sessions. As expected, the corresponding fractions tend to be low, since the participants do not have access to the same-side trading activity, and hence any instance of "imitation" is merely due to chance. Moving toward the center of the plot, one can find a condensed group of markets with full access to the order book, implying that the subjects in these markets respond to the same-side cues more than expected under the Black Box baseline treatment, with the copying rate roughly two times higher. Finally, the third cluster in the upper-right corner consists mostly of markers where Same-side OB feedback is supplied, which again is reasonable because the participants do not have access to other confounding information to rely on to and to guide their decision processes. For these markets, the copying rate can be even three times as high as expected under the Black Box regime.

TABLE A.3
BETWEEN-ROUND ADJUSTMENTS OF RELATIVE DEMANDS

Treatment			*Buyers' Median $\Delta^T_{b_i,1}$ after*			*Sellers' Median $\Delta^T_{s_j,1}$ after*		
Feedback	*Market Structure*	*Price Rule*	*No deal*	*Deal*	*p-Value*	*No deal*	*Deal*	*p-Value*
Open Book			−0.02 (973)	0 (3,292)	<0.01	0 (1,080)	0 (3,296)	<0.01
	Regular		−0.02 (383)	0 (1,181)	<0.01	0 (241)	0 (1,175)	<0.01
Full OB		First price	−0.02 (69)	0 (248)	(0.01)	0 (51)	0 (249)	(0.03)
	With restricted asks	First price	−0.05 (91)	0 (249)	<0.01	−0.05 (22)	0 (246)	(0.02)
Other-side OB		First price	−0.03 (59)	0 (249)	<0.01	0 (41)	0 (247)	(0.03)
Same-side OB		First price	−0.01 (99)	0 (221)	<0.01	0 (60)	0 (219)	(0.02)
		Matchmaker keeps	−0.01 (65)	0 (214)	(0.07)	−0.01 (67)	0 (214)	(0.02)
Full OB	**Asymmetric**	First price	−0.01 (404)	0 (690)	<0.01	0 (494)	0 (693)	(0.06)
	More buyers		0 (381)	0 (343)	<0.01	0 (9)	0 (345)	(0.23)
	More sellers		−0.02 (23)	0 (347)	<0.01	0 (485)	0 (348)	(0.1)
Full OB	**Large**	First price	−0.03 (186)	0 (1,421)	<0.01	−0.02 (345)	0 (1,428)	<0.01
	Large CE		−0.07 (62)	0 (677)	<0.01	−0.02 (130)	0 (675)	<0.01
	Seller shift		−0.01 (41)	0 (291)	(0.02)	−0.04 (54)	0 (294)	<0.01
	From buyers to sellers		−0.02 (40)	0 (213)	<0.01	−0.01 (68)	0 (226)	(0.01)
	From sellers to buyers		0 (43)	−0.01 (240)	(0.38)	−0.01 (93)	0 (233)	<0.01
Black Box			−0.01 (1,632)	0 (2,643)	<0.01	0 (1,519)	0 (2,627)	<0.01
	Regular		−0.04 (434)	0 (883)	<0.01	0 (447)	0 (881)	<0.01
		First price	−0.09 (92)	0 (239)	<0.01	0 (102)	0 (238)	<0.01
	With restricted asks		0 (72)	0 (216)	<0.01	0 (68)	0 (217)	<0.01
		Random price	−0.06 (132)	0 (204)	<0.01	0 (121)	0 (202)	<0.01
		Matchmaker keeps	−0.01 (138)	0 (224)	<0.01	0 (156)	0 (224)	(0.04)
	Asymmetric	First price	−0.02 (448)	0 (530)	<0.01	0 (464)	0 (527)	<0.01
	More buyers		−0.02 (349)	0 (254)	(0.01)	−0.01 (75)	0 (253)	<0.01
	More sellers		−0.04 (99)	0 (276)	<0.01	0 (389)	0.01 (274)	<0.01
	Large	First price	0 (750)	0 (1,230)	<0.01	0 (608)	0 (1,219)	<0.01
	Large CE		−0.02 (112)	0 (193)	<0.01	0 (126)	0 (191)	<0.01
	More rounds		0 (638)	0 (1,037)	<0.01	0 (482)	0 (1,028)	<0.01
		All	−0.01 (2,605)	0 (5,935)	<0.01	0 (2,599)	0 (5,923)	<0.01

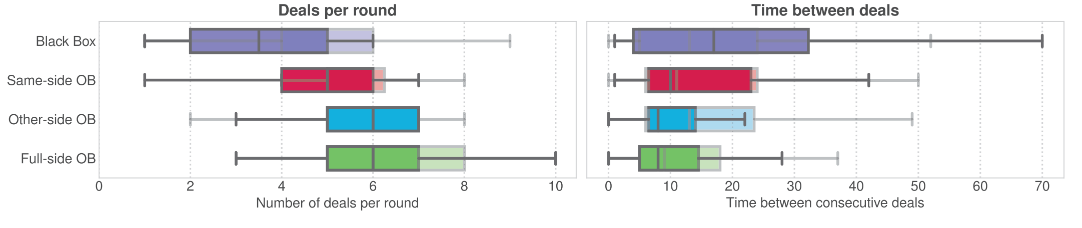
NOTE: Median between-round adjustments $\Delta^T_{b_i,1}$ and $\Delta^T_{s_j,1}$ in response to (not) having made a deal in the previous round $T-1$, comparing buyers and sellers in Open Book and Black Box treatments across all rounds $T \geq 2$ (the corresponding distributions are depicted in Figure A.13). Numbers in brackets indicate the number of bids/asks considered in the analysis. p-Values correspond to MWW tests, which were applied to the group of all buyers'/sellers' between-round adjustments following a deal and the group of all buyers'/sellers' between-round adjustments not following a deal pooled from the treatments indicated in the first three columns.



NOTES: The fraction of bids coinciding with one of the 10 most recent bids versus the fraction of asks coinciding with one of the 10 most recent asks across all 104 market sessions subject to one of our four feedback treatments (over the course of all trading rounds). Overall, there are 28% such bids and asks in the Black Box treatments, 46% when Full OB feedback is provided, 55% in the case of Same-side OB feedback, and 58% under the Other-side OB feedback treatment.

FIGURE A.14

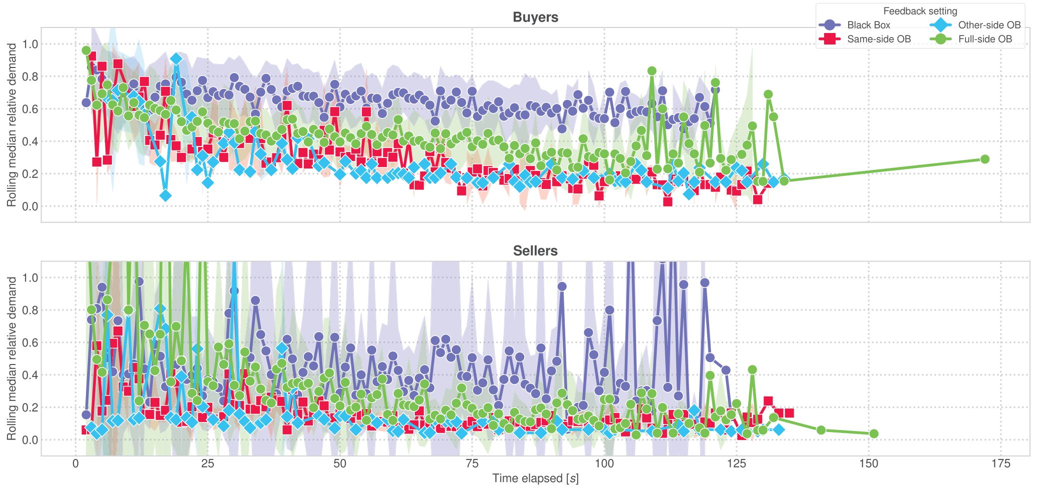
COPYING RECENT BIDDING BEHAVIOR



NOTES: *Left*. Distribution of the number of deals struck in a given round. *Right*. Distribution of the time elapsed between two consecutive deals in a given round. The box plots in the front correspond to the first rounds, and the shaded box plots in the background to all rounds, whereby all experimental sessions were taken into account.

FIGURE A.15

DEALS STRUCK ACROSS THE FOUR FEEDBACK SETTINGS



Notes: At any given time, the median of the last 20 buyers' relative demands $\rho^1_{b_i,k}$ (top) and the median of the last 20 sellers' relative demands $\rho^1_{s_j,l}$ (bottom) are shown. Note that each dot represents the mean over all market sessions per feedback treatment combined and the shaded areas surrounding them the one-standard-deviation intervals.

Figure A.16

EVOLUTION OF RELATIVE DEMANDS OVER THE COURSE OF THE FIRST ROUND

REFERENCES

Alós-Ferrer, C., J. Buckenmaier, and G. Kirchsteiger, "Do Traders Learn to Select Efficient Market Institutions?" *Experimental Economics* 25 (2022), 203–28.

Arechar, A. A., S. Gächter, and L. Molleman, "Conducting Interactive Experiments Online," *Experimental Economics* 21 (2018), 99–131.

Arifovic, J., and J. Ledyard, "A Behavioral Model for Mechanism Design: Individual Evolutionary Learning," *Journal of Economic Behavior and Organization* 78 (2011), 374–95.

Azrieli, Y., C. P. Chambers, and P. J. Healy, "Incentives in Experiments: A Theoretical Analysis," *Journal of Political Economy* 126 (2018), 1472–503.

Bao, T., E. Nekrasova, T. Neugebauer, and Y. E. Riyanto, "Algorithmic Trading in Experimental Markets with Human Traders: A Literature Survey," SSRN Working Paper 3908065, 2021.

Bossaerts, P., and C. R. Plott, "From Market Jaws to the Newton Method: The Geometry of How a Market Can Solve Systems of Equations," Volume 1, Part 1, 1st edition, Chapter 2 (Amsterdam, The Netherlands: Elsevier, 2008), 22–24.

Chamberlin, E. H., "An Experimental Imperfect Market," *Journal of Political Economy* 56 (1948), 95–108.

Charness, G., U. Gneezy, and B. Halladay, "Experimental Methods: Pay One or Pay All," *Journal of Economic Behavior and Organization* 131 (2016), 141–50.

Davis, D. D., and C. A. Holt, "Market Power and Mergers in Laboratory Markets with Posted Prices," *RAND Journal of Economics* 25 (1994), 467–87.

Dolgopolov, A., D. Houser, C. Martinelli, and T. Stratmann, "Assignment Markets: Theory and Experiments," George Mason University Interdisciplinary Center for Economic Science Work WP 1075, 2019.

Farmer, J. D., P. Patelli, and I. I. Zovko, "The Predictive Power of Zero Intelligence in Financial Markets," *Proceedings of the National Academy of Sciences* 102 (2005), 2254–59.

Friedman, D., and J. Rust, *The Double Auction Market: Institutions, Theories, and Evidence* (Westview Press, 1993).

Gode, D. K., and S. Sunder, "Allocative Efficiency of Markets with Zero-Intelligence Traders: Market as a Partial Substitute for Individual Rationality," *Journal of Political Economy* 101 (1993), 119–37.

Hauser, D. J., and N. Schwarz, "Attentive Turkers: MTurk Participants Perform Better on Online Attention Checks than Do Subject Pool Participants," *Behavior Research Methods* 48 (2016), 400–7.

Hollifield, B., R. A. Miller, and P. Sandås, "Empirical Analysis of Limit Order Markets," *Review of Economic Studies* 71 (2004), 1027–63.

Holt, C. A., L. W. Langan, and A. P. Villamil, "Market Power in Oral Double Auctions," *Economic Inquiry* 24 (1986), 107–23.

Horton, J. J., D. G. Rand, and R. J. Zeckhauser, "The Online Laboratory: Conducting Experiments in a Real Labor Market," *Experimental Economics* 14 (2011), 399–425.

Jantschgi, S., H. H. Nax, B. Pradelski, and M. Pycia, "On Market Prices in Double Auctions," CEPR Discussion Paper 17040, 2022.

Kimbrough, E. O., and A. Smyth, "Testing the Boundaries of the Double Auction: The Effects of Complete Information and Market Power," *Journal of Economic Behavior and Organization* 150 (2018), 372–96.

Kirchsteiger, G., M. Niederle, and J. Potters, "Endogenizing Market Institutions: An Experimental Approach," *European Economic Review* 49 (2005), 1827–53.

Lin, P.-H., A. L. Brown, T. Imai, J. T.-Y. Wang, S. Wang, and C. Camerer, "Evidence of General Economic Principles of Bargaining and Trade from 2,000 Classroom Experiments," *Nature Human Behavior* 4 (2020), 917–27.

Paolacci, G., J. Chandler, and P. G. Ipeirotis, "Running Experiments on Amazon Mechanical Turk," *Judgment and Decision making* 5 (2010), 411–19.

Plott, C. R., and V. L. Smith, "An Experimental Examination of Two Exchange Institutions," *Review of Economic Studies* 45 (1978), 133–53.

Pratt, J. W., "Remarks on Zeros and Ties in the Wilcoxon Signed Rank Procedures," *Journal of the American Statistical Association* 54 (1959), 655–67.

Rostek, M., and M. Weretka, "Price Inference in Small Markets," *Econometrica* 80 (2012), 687–711.

Sauermann, H., and R. Selten, "Anspruchsanpassungstheorie der Unternehmung," *Zeitschrift für die gesamte Staatswissenschaft* 4 (1962), 577–97.

Selten, R., "Properties of a Measure of Predictive Success," *Mathematical Social Sciences* 21 (1991), 153–67.

———, and W. Krischker, "Comparison of two Theories for Characteristic Function Experiments," in R. Tietz, ed., *Aspiration Levels in Bargaining and Economic Decision Making* (Berlin, Heidelberg: Springer, 1983), 259–64.

Smith, V. L., "An Experimental Study of Competitive Market Behavior," *Journal of Political Economy* 70 (1962), 111–37.

———, "Effect of Market Organization on Competitive Equilibrium," *Quarterly Journal of Economics* 78 (1964), 181–201.

———, "Experimental Economics: Induced Value Theory," *American Economic Review* 66 (1976), 274–79.

———, *Papers in Experimental Economics* (Cambridge University Press, 1991).

———, G. L. Suchanek, and A. W. Williams, "Bubbles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets," *Econometrica* 56 (1988), 1119–51.

———, and A. W. Williams, "The Effects of Rent Asymmetries in Experimental Auction Markets," *Journal of Economic Behavior and Organization* 3 (1982), 99–116.

Snowberg, E., and L. Yariv, "Testing the Waters: Behavior across Participant Pools," *American Economic Review* 111 (2021), 687–719.

Tietz, A., and H. J. Weber, "On the Nature of the Bargaining Process in the Kresko-Game," in H. Sauermann, ed., *Contributions to Experimental Economics* 7 Tübingen, Germany: J.C.B. Mohr, (1972), 305–34.

Walker, J. M., and A. W. Williams, "Market and Behavior in Bid, Offer, and Double Auctions: A Reexamination," *Journal of Economic Behavior and Organization* 9 (1988), 301–14.

Walras, L., *Principe d'une Theorie Mathematique de l'echange, Memoire lu a l'Academie des Sciences Morales et Politiques* (Guillaumin, 1874).

———, *Theorie mathematique de la richesse sociale* (Corbaz, 1883).

Wellek, S., "A New Approach to Equivalence Assessment in Standard Comparative Bioavailability Trials by Means of the Mann-Whitney Statistic," *Biometrical Journal* 38 (1996), 695–710.