Games and Economic Behavior 156 (2026) 135–148

Contents lists available at ScienceDirect

Games and Economic Behavior

journal homepage: www.elsevier.com/locate/geb

When do prediction markets return average beliefs? Experimental evidence ☆

Marco Mantovani [a,*], Antonio Filippin [b,c]

[a] University of Milan Bicocca, Department of Economics, Piazza dell'Ateneo Nuovo 1, 20126, Milan, Italy
[b] University of Milan, Department of Economics, Via Conservatorio 7, 20122, Milano, Italy
[c] Institute for the Study of Labor (IZA), Schaumburg-Lippe-Str. 5–9, 53113, Bonn, Germany

ARTICLE INFO

JEL classification:
D81
C92
D40

Keywords:
Prediction markets
Information aggregation
Laboratory experiment
Risk preferences
Beliefs

ABSTRACT

In prediction markets, prices can be interpreted as the average belief of the traders under restrictive theoretical assumptions, i.e. specific risk preferences and the Prior Information Equilibrium. The validity of these assumptions depends on the specific market institution and on the composition of the market in terms of risk preferences. In this paper we test in a laboratory experiment the main elements that should affect the distance between prices and average beliefs, manipulating the market institution and the market composition. We do not find that risk preferences significantly affect prices. We find instead that in the double auction– where at least partial information aggregation is expected – prices are closer to the average belief than in the call auction – where, instead, belief aggregation is expected. We show that traders update beliefs in the direction of observed prices, rather than of the true state.

1. Introduction

The paper presents an experiment to assess the conditions under which the price in a prediction market can be interpreted as representing the average belief of the traders. Conveying information through prices is one of the fundamental functions of markets. Prediction markets are exchange systems designed specifically to aggregate dispersed beliefs into market prices (Wolfers and Zitzewitz, 2004; Arrow et al., 2008). They are widely used as forecasting tools across a broad range of domains (see, for example, Luckner et al., 2008, and references therein). Their spread and relevance has grown significantly in recent years. For example, Polymarket, a decentralized platform built on the Polygon blockchain, now hosts around ten thousand markets on topics ranging from politics and sports to cryptocurrency and current events. In January 2025, it recorded approximately 450,000 active traders, with total trading volume exceeding $9 billion in 2024.

A paradigmatic prediction market consists of a set of Arrow-Debreu securities, where each security pays 1$ if a specific state of the world materializes. The price of each security is interpreted as the probability of that state according to the market, i.e. the average belief of the traders. However, the theoretical conditions under which prices accurately reflect average beliefs are quite restrictive.

☆ We would like to thank Peter Bossaerts, Dan Friedman, Anna Gilbert Rivas, Stephanie Heger, Peter Katuščák and Thomas Kittsteiner for useful suggestions. All remaining errors are ours. We gratefully acknowledge financial support from IFREE Small Grant Program and PRIN grant no. P20228SXNF of the Italian Ministry of University and Research.

Declaration of Interest: the Authors declare no competing interests related to the manuscript.

Replication files: https://osf.io/c264t/overview.

* Corresponding author.
E-mail addresses: marco.mantovani@unimib.it (M. Mantovani), antonio.filippin@unimi.it (A. Filippin).

https://doi.org/10.1016/j.geb.2025.12.004
Received 29 January 2025
Available online 25 December 2025

First, price-to-state inferences must not occur: traders must act as price-takers and hold their beliefs constant throughout. Second, traders must be expected utility maximizers with specific and homogeneous risk preferences. Theoretical models of prediction markets typically sidestep the first issue by assuming that traders hold exogenous beliefs – opinions about the state of the world on which they agree to disagree (Wolfers and Zitzewitz, 2004; Ottaviani and Sørensen, 2015). This assumption allows the market to aggregate beliefs rather than information, since traders behave as if there is no additional information to extract from observed (or hypothetical) prices. In a more general framework where beliefs are related to the state of the world through some signal-generating technology, the absence of price-to-state inferences corresponds to a well-known benchmark in information aggregation experiments: the Prior Information Equilibrium (PIE) (Plott and Sunder, 1982). A PIE is a Walrasian equilibrium in which price-taking traders do not update their beliefs based on prices. Conversely, when traders do infer the private information of others from observed or hypothetical prices and update their beliefs accordingly, prices no longer reflect the traders' ex-ante beliefs. In the limiting case, this process leads to the classic benchmark of information aggregation: the Rational Expectations Equilibrium (REE). Importantly, even under a PIE, prices do not necessarily equal the average belief because risk preferences also play a crucial role (Manski, 2006; Wolfers and Zitzewitz, 2006). Risk preferences determine how traders' net demands respond to different prices. This responsiveness influences the extent to which private beliefs are revealed through market behavior. For instance, the demand of a risk-neutral or risk-loving trader only indicates whether the price is above or below their belief, but not how far. In contrast, risk averse traders buy (or sell) more and more as the difference between the price and their belief increases. Moreover, the degree of risk aversion negatively correlates with the net demand. At the aggregate level, the degree of risk aversion influences how prices align with average beliefs (Gjerstad, 2005; He and Treich, 2017; Fountain and Harrison, 2011). Specifically, under CRRA preferences, when an event is deemed likely on average, the PIE price lies below the average belief for risk-neutral or mildly risk-averse traders, matches it when the CRRA coefficient equals 1 (log-utility), and exceeds it as risk aversion strengthens. To summarize, the equivalence between prices and average beliefs, which we adopt throughout the paper as the benchmark because it underpins traditional prediction markets, holds in theory only under a PIE with sufficiently high risk aversion.

We study these conditions – no price-to-state inferences/PIE and level of risk aversion – in a controlled laboratory experiment. In more detail, we manipulate ***i)*** the possibility of drawing inferences from prices by varying the market institution, and ***ii)*** the degree of risk aversion by grouping subjects according to their elicited preferences. As for market institutions, we compare two canonical formats – the single call market and the continuous double auction – as mechanisms for private belief aggregation. These institutions vary in the extent to which they allow or encourage price-to-state inferences. In the call market, inferences must rely on hypothetical reasoning: "If the market-clearing price were x, then others might have different information than I do, and I should revise my belief and demand at price x accordingly." In the double auction, inferences can rely on observed behavior: "If others are willing to buy at price x, then they likely hold different information, and I should revise my belief and demand accordingly." We hypothesize that price-to-state inferences are more viable in the double auction than in the call market. As a result, prices in the call market should approximate a PIE, while prices in the double auction should move closer to the REE due to (partial) information aggregation. We also elicit traders' ex-post beliefs – after the market closes – as a direct measure of belief revision induced by market activity. To manipulate risk preferences, we elicit individual risk aversion through an independent and incentivized task. We use these measures to construct markets composed of traders with systematically different levels of risk aversion (see Crockett et al., 2021, for a similar approach). This exogenous variation in risk preferences should yield clear comparative predictions: ***a)*** prices should diverge more from the uninformed prior in more risk-averse markets, and ***b)*** prices should align more closely with the average belief when the CRRA coefficient is near 1.

Contrary to theoretical predictions, we find no evidence that risk aversion significantly affects market prices under either trading institution. On the other hand, the trading institution plays a significant role, and market prices reflect more closely the true state in the double auction. While this result is not surprising in itself, the comparison in terms of belief aggregation is counterintuitive. Prices in the double auction are closer to the average belief than those in the call auction. In other words, market prices tend to reflect the average information held by traders in the institution where the PIE assumption is likely violated – namely, the double auction – whereas in the call auction, where the PIE assumption is expected to hold, prices remain consistently closer to the uninformed prior. This difference does not emerge at the start of the trading period, but only at its conclusion, as double auction prices converge toward the average belief. We then scrutinize the type of inferences the participants draw from prices, looking at how their beliefs change after the market. Ex-post beliefs show no evidence of sophisticated inference from prices to the underlying state (see similar findings in Choo et al., 2017). Conversely, they appear to be revised naïvely in the direction of observed prices. At the same time, the variance of ex-post beliefs across traders with different information tends to shrink relative to ex-ante beliefs. Traders disagree less after participating in the market than they did before. Therefore the convergence of prices to the average beliefs in the double auction occurs while ex-post beliefs are revised away from the true state of the world, particularly for traders who received the most informative signals. Prices and beliefs co-move in a way that suggests eventual convergence to a point between the uninformed prior and the true state of the world, approximately where the average belief is located.

Our paper lies at the intersection of two strands of the literature. The first concerns the accuracy of prediction markets. Field evidence provides some support for the ability of prediction markets to reveal the state of the world. The Iowa Electronic Market, for instance, outperformed opinion polls in U.S. presidential elections between 1988 and 2000 (Berg et al., 2008). In a study of three Bundesliga seasons, Spann and Skiera (2009) find that both prediction markets and betting odds yield comparably accurate forecasts, significantly outperforming professional tipsters. Laboratory experiments confirm the accuracy of prediction markets in more controlled settings (Healy et al., 2010; Horn et al., 2014, for reviews).[1] Our study focuses on the accuracy of prediction markets as measured by the extent to which prices represent traders' beliefs. The relationship between prices and beliefs has also been studied outside prediction markets. It has been shown that when the supply of an asset is fixed, prices can exceed the expected value based on traders' beliefs (Miller, 1977); speculative motives can further amplify such mispricing (Harrison and Kreps, 1978; Harris and Raviv, 1993; Morris, 1996). These models rely on two key assumptions: the presence of market frictions, such as short-selling constraints or market incompleteness and disagreement among traders based on subjective opinions. Experimental evidence supports this view: Palfrey and Wang (2012) show that overpricing disappears only when markets are complete and short selling is allowed, an environment similar to ours. Ottaviani and Sørensen (2015) study how liquidity constraints and risk aversion can induce underreaction to public information, when traders start from heterogeneous subjective priors. As mentioned above, this branch of the literature treats beliefs as exogenous, meaning that subjective opinions are fixed or updated from public signals. In contrast, we consider a setting where beliefs result from Bayesian updating of a common prior with private signals. In single call markets, the evidence consistently shows that traders do not anticipate the informational content of hypothetical market-clearing prices (Ngangoué and Weizsäcker, 2021; Biais et al., 2017; Filippin and Mantovani, 2023). Biais et al. (2017) and Filippin and Mantovani (2023) (who use part of the data from this paper) show that the demand schedules in a call auction are indistinguishable from those elicited in a non-strategic random-price mechanism. Conversely, this does not hold in double auctions. This naturally connects our work to the second strand of the literature, that on information aggregation initiated by Plott and Sunder (1982, 1988), who study how dispersed private information is aggregated into prices. The trading process in double auctions helps traders update their beliefs: observing prices and executed trades allows them to infer information and revise their posterior accordingly (Ngangoué and Weizsäcker, 2021). If these revised beliefs assign greater likelihood to the true state, their feedback on trading behavior would push prices further in that direction. Recent evidence shows that this price-to-state inference often fails to achieve the rational expectations equilibrium (REE) (Choo et al., 2017; Corgnet et al., 2023). Despite their imperfections, double auctions tend to outperform call markets in aggregating dispersed information (see, among others, Chen and Plott, 2008; Friedman, 1984; Friedman and Ostroy, 1995; Guarnaschelli et al., 2003; Kagel and Levin, 1986; Kagel, 2004; Ketcham et al., 1984; Palan et al., 2020; Sakurai and Akiyama, 2017; Smith et al., 1982). Overall, this literature supports our assumption that traders behave as if they "agree to disagree" in call markets, but not in double auctions. At first glance, our findings are consistent with this view. However, non-strategic behavior observed in call auctions is accompanied by much less opinion aggregation than the PIE theory would predict. Similarly, inference from prices in the double auction does not lead to learning the true state, but rather to a convergence of beliefs toward the realized price.

The structure of the paper is as follows. Section 2 derives the testable implications. Sections 3 and 4 describe the experimental design and the procedures. Results are reported in Section 5. Section 6 discusses further implications, while Section 7 concludes.

2. Theoretical background and testable implications

A prediction market is a common-value asset market with heterogeneous beliefs.[2] There are two ex-ante equally likely states, $e \in \{Red, Blue\}$. An Arrow-Debreu security pays 100 if $e = Blue$ and 0 otherwise. Its price is denoted by $p \in [0, 100]$. There are n traders, each holding a belief $b_i \in [0, 100]$, representing the subjective probability that $e = Blue$. This belief coincides with the expected value of the asset for trader i. We refer to $b_i = 50$ as the *uninformed belief*, and denote the average belief as $\bar{b} = \frac{1}{n}\sum_i b_i$.

Traders receive an equal endowment m of a numeraire good. The security is in zero net supply, and short positions are settled at the realized value of the security: 100 if $e = Blue$, and 0 otherwise. Therefore, there is no aggregate risk in the market and no constraints on short-selling.

The first condition for prices to match the average belief is that prices reflect but do not influence beliefs. Suppose traders form their beliefs updating a common prior using private signals. If they do not infer others' information from prices, and trade based on their private information only, the corresponding Walrasian equilibrium is a PIE. However, if traders use prices to infer the underlying state (price-to-state inference), beliefs and prices become endogenous and interact. At the extreme, this leads to a Rational Expectations Equilibrium (REE) in which prices are fully revealing and beliefs are mutually consistent.[3]

We argue that whether this inference takes place depend on the market institution. In call auctions, traders must speculate about the beliefs (and thus the signals) that would rationalize hypothetical clearing prices, without observing prices. In contrast, double auctions provide traders with observable price data (bids, asks, transactions), which they may interpret as signals about others' information. Observing relatively high prices (e.g., above 50) may suggest that traders' beliefs lean toward $e = Blue$. If traders interpret this as evidence of informative private signals favoring that state, they may revise their own beliefs accordingly, generating a feedback loop that further pushes prices upward. The converse applies to relatively low prices. If inference occurs in the double auction but not in the call auction, we expect prices in the former to move closer to the true state. This motivates our first testable hypothesis:

- **Hp(Institution):** $\left|p^*_{DA} - 50\right| > \left|p^*_{CA} - 50\right|$.

where p^*_{DA} (p^*_{CA}) is the market-clearing price in the double (call) auction.

[1] The experimental literature has also explored information acquisition (Page and Siemroth, 2017, 2021), incentives for manipulation (Deck and Porter, 2013; Choo et al., 2022), and biases induced by the structure of the state space (Sonnemann et al., 2013).

[2] This section derives our main testable implications by adapting to our experimental setting a standard framework for prediction markets (Gjerstad, 2005; Wolfers and Zitzewitz, 2006; Manski, 2006; He and Treich, 2017).

[3] In this limit, the no-trade theorem applies (Milgrom and Stokey, 1982) and trading based on beliefs requires a violation of common knowledge of the information structure.

Let us now consider a PIE setting in which traders behave as price-takers, treating p as fixed and not informative of others' private information. Under CRRA preferences with relative risk aversion parameter θ, trader i's optimal demand is given by (see Gjerstad, 2005):

$$q_i^*(p, b_i, \theta > 0) = \frac{(1-p)^{\frac{1}{\theta}} b_i^{\frac{1}{\theta}} - p^{\frac{1}{\theta}} \left(1-b_i\right)^{\frac{1}{\theta}}}{(1-p)p^{\frac{1}{\theta}} \left(1-b_i\right)^{\frac{1}{\theta}} + p(1-p)^{\frac{1}{\theta}} b_i^{\frac{1}{\theta}}} m, \tag{1}$$

where a negative demand means being a net (short) seller. Demand is zero when $p = b_i$, and decreases with p. The former statement holds in general for any expected utility maximizer, while the latter holds under non-increasing absolute risk aversion. Risk-neutral and risk-seeking players ($\theta \leq 0$) invest their full endowment in long (short) positions when $p < b_i$ ($p > b_i$).

When $\theta = 1$, Eq. (1) simplifies to:

$$q_i^*(p, b_i, \theta = 1) = \left(b_i - p\right) \frac{m}{p}. \tag{2}$$

Imposing the market-clearing condition $\sum_i q_i^* = 0$ yields the Walrasian equilibrium price $p^*_{\theta=1} = \bar{b}$, consistent with the standard interpretation that a prediction market aggregates individual beliefs such that the equilibrium price reflects the market's collective estimate of the event's probability.

The equality $p^*_{\theta=1} = \bar{b}$ no longer holds in a PIE when $\theta \neq 1$. Under CRRA preferences, as θ increases above 1, the individual demands implied by Eq. (1) drive equilibrium prices closer to the true state. Conversely, when θ falls below 1, equilibrium prices converge toward the uninformed belief ($p = 50$). Similar patterns emerge under CARA preferences (Filippin and Mantovani, 2023) and under symmetric belief distributions with non-increasing risk aversion (He and Treich, 2017). In general, higher degrees of risk aversion tend to shift prices farther from 50 and closer to the true state. This observation motivates our second testable hypothesis:

- **Hp(Risk)** : $\left|p^*_{HIGH} - 50\right| > \left|p^*_{LOW} - 50\right|$,

where p^*_{HIGH} (p^*_{LOW}) denotes the market-clearing price in markets characterized by high (low) risk aversion. The underlying intuition is that risk aversion dampens the responsiveness of traders: for a given belief-price gap, more risk-averse participants trade smaller quantities. Consequently, larger price deviations from 50 are required to clear the market in high-risk-aversion settings.

So far, we have assumed that traders update beliefs correctly according to Bayes' rule when tehy receive a private signal. However, in practice, belief updating may deviate from Bayesian norms. As noted by Snowberg and Wolfers (2010), such deviations can confound inference about risk preferences, and may also affect price informativeness when traders revise beliefs based on prices. To assess whether prices deviate from the average Bayesian belief in the direction of incorrectly updated beliefs, we elicit actual beliefs and define our final testable hypothesis:

- **Hp(Beliefs):** $\rho(\Delta p, \Delta b) > 0$,

where $\Delta p = p^* - \bar{b}$, is the vector of deviations of the price in each market from the corresponding average Bayesian belief, $\Delta b = \bar{b}_e - \bar{b}$ is the vector of differences at the market level between the average elicited belief and the Bayesian one, and ρ is the linear correlation coefficient.

We conclude this section with a note on the inference problem faced by traders. The net demands of two traders with different beliefs and risk preferences may intersect at some price–quantity pair. At such a point, an external observer cannot disentangle the combination of beliefs and risk attitudes that jointly determine the observed bid. In a PIE or REE framework, distinguishing between beliefs and risk preferences is inconsequential: in the former, traders behave as price takers, while in the latter, uncertainty is eliminated. Between these two extremes, however, separating beliefs from risk preferences becomes potentially relevant, though exceedingly complex. For tractability, we assume that traders disregard uncertainty regarding others' risk preferences.

3. Design

Risk elicitation task. At the beginning of the experiment, we elicit subjects' risk preferences using the Investment Game (Gneezy and Potters, 1997). In this task, participants decide how to allocate an endowment of 200 Monetary Units (MU) between a safe account and a risky investment, which yields either 2.5 times the invested amount or zero, with equal probability. There is a closed-form mapping between choices in this game and CRRA coefficients. The Investment Game is particularly well-suited for identifying risk preferences near log-utility (Crosetto and Filippin, 2016), which is theoretically relevant in our context. Moreover, this elicitation method aligns with the theoretical structure of prediction markets, in which risk-neutral and risk-seeking agents behave similarly. Due to the overall length of the experiment and concerns about maintaining incentive salience, we opted for a single elicitation of risk preferences, despite the associated drawback of increased measurement error (Gillen et al., 2019).

Matching. Each session is divided into two markets of 11 participants, based on the median level of elicited risk aversion within the session. This procedure maximizes heterogeneity between the Low and High risk aversion markets while minimizing heterogeneity within each market. Participants are informed that each session includes two separate markets and that they will remain in the same market throughout all trading periods. However, they are not told that market assignment is based on their risk preferences. The hypothesis concerning the effect of risk preferences on market prices is derived under the assumption of non-strategic behavior. Traders are assumed not to consider the risk preferences of others.

Table 1
Signals.

	Signal (s)																
	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58
Urn A	–	–	–	–	–	x	–	–	–	–	–						
Urn B			–	–	–	–	–	x	–	–	–	–	–				
Urn C					–	–	–	–	–	x	–	–	–	–	–		
Urn D							–	–	–	–	–	x	–	–	–	–	–
$p(Blue\|s)$	0	0	0	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{2}{3}$	1	1	1	1

Notes: The table reports the distribution of signals conditional on the selected urn. Each signal takes the form: "There are s blue marbles in the urn." The symbols "-" and "x" indicate that the *column* signal can be sent under the *row* urn; "x" denotes the true number of blue marbles in that urn. The last row presents the posterior Bayesian beliefs of a trader who receives the signal shown in the corresponding column.

Asset market. Participants act as traders across 12 trading periods, each following the same structure. In every period, there are four urns, each containing a different number of blue marbles out of 100: urn A contains 47, urn B 49, urn C 51, and urn D 53. One urn is selected at random in each period, with all urns having equal probability of being chosen. This uniform prior is common knowledge among participants.

Traders are endowed with 1000 MU and trade an asset called "Majority Blue." If the selected urn is C or D, the event "the majority of marbles are Blue" occurs (e = Blue), and each unit of the asset pays 100 MU at the end of the trading period. If the urn is A or B, the event does not occur (e = Red), and each unit of the asset pays 0 MU.

Before each period, traders receive a private signal (s) about the composition of the urn, presented as: "There are s blue marbles in the urn." The signal deviates by no more than 5 units from the true number of blue marbles, i.e., $s \in \{x-5, \dots, x+5\}$, where x is the actual number of blue marbles in the selected urn. Each of the 11 possible signals is assigned to a different trader, meaning that the 11 traders each receive one of the 11 signals with equal probability, as shown in Table 1. For example, if the selected urn is A (47 blue marbles), the 11 traders receive signals ranging from 42 to 52. The procedure used to generate and distribute the signals is common knowledge.

Given the signal, Bayesian updating of the prior generates the posterior belief $p(e = \text{Blue}|s)$ as reported in the last row of Table 1. Some signals ($s \leq 45$ and $s \geq 55$) are fully revealing, leading to $p(e = \text{Blue}|s) \in \{0, 1\}$. Other signals ($s = 46, 47, 53, 54$) are partially informative, i.e. $p(e = Blue|s) \in \left\{\frac{1}{3}, \frac{2}{3}\right\}$. Finally, signals in the range $48 \leq s \leq 52$ are uninformative, as they leave the posterior unchanged at $p(e = \text{Blue} \mid s) = \frac{1}{2}$, matching the prior. Aggregating $p(e = \text{Blue} \mid s)$ over all 11 traders in a market yields an average belief about the event e = Blue: 28.8% if the urn is A, 40.9% if the urn is B, 59.1% if the urn is C, and 71.2% if the urn is D. These average beliefs serve as the benchmark against which we compare the equilibrium prices.

Belief elicitation. In each trading period, we elicit twice traders' subjective beliefs about which urn was selected: once after they receive their private signal but before the market opens (ex-ante belief), and again after the trading period concludes (ex-post belief).

Belief reporting is incentivized using the Binarized Scoring Rule (BSR) (Hossain and Okui, 2013). The BSR compares the sum of squared errors of the reported beliefs (normalized between 0 and 1) with a random number $k \in U[0, 1]$. If the sum of squared errors is lower than k (i.e., if the subject's beliefs are sufficiently accurate) he earns a fixed prize (200 MU); otherwise, he receives nothing. Unlike the Quadratic Scoring Rule (QSR), which pays different amounts depending on belief accuracy, the BSR offers a fixed high reward with a probability that increases with accuracy. Since participants cannot reduce the variance of outcomes, their optimal strategy is to maximize the likelihood of receiving the high reward regardless of their risk preferences. This requires reporting the most accurate estimate of the probability distribution. The BSR is isomorphic to the QSR in terms of expected rewards, but it ensures incentive compatibility even for risk-averse subjects. By contrast, the QSR is not incentive-compatible unless participants are risk-neutral. While the BSR also has limitations (Danz et al., 2022), we consider it the most suitable procedure for a study in which both risk preferences and belief accuracy play a central role.

Market institution. In each trading period, traders enter the market with a monetary endowment of 1000 MU but no asset holdings. Sales are therefore implemented via short selling. This setup ensures that the game is zero-sum and that there is no aggregate risk in the market (see, e.g., Bossaerts et al., 2013; Asparouhova et al., 2017). Net short positions at the end of the trading period are settled at the actual value of the asset. Traders who are net sellers must buy back the assets they sold at 0 MU if the urn is A or B, and at 100 MU if the urn is C or D. To prevent bankruptcy, liquidity is frozen for each order based on the worst-case scenario, assuming a final asset value of 0 for net long positions and 100 for net short positions. This "single asset - homogeneous value" framework is isomorphic to a two-state, two-asset environment (see Corgnet et al., 2023, for a comparison of these setups). Given our setting, the equilibrium prices in a PIE where all traders have CRRA preferences with $\theta = 1$ are 29, 41, 59, and 71 in urns A, B, C, and D, respectively. For comparison, a risk-neutral PIE predicts prices of 45, 50, 50, and 55.[4]

[4] The lack of quantity constraints (aside from the no-bankruptcy condition) is crucial. If each trader held one unit and short sales were prohibited, the risk-neutral PIE would yield prices of 34, 50, 50, and 66, resembling the goods markets in Smith and Williams (2000), where markets may not clear.

We implement the prediction market in each session under either a call auction or a double auction, following a between-subjects design:

Call Auction (CA). Traders exchange the asset through a Single Market Call Auction. In each period, they have two minutes to independently submit limit orders to buy or sell the asset in a closed order book. As orders are submitted, a visual representation of each trader's net demand schedule updates in real time. Traders may buy and sell as many units of the asset as they wish, compatibly with the available liquidity. The equilibrium price is determined as the price that equalizes aggregate demand and supply, i.e., where the aggregate net demand equals zero, thereby maximizing the volume of trade. If demand and supply are equal over a range of prices, the equilibrium price is set as the average of those prices. If no price exactly clears the market, some orders may be (partially) unexecuted. Execution priority is given to buy (sell) orders with higher (lower) limit prices.

Double Auction (DA). Traders exchange the asset through a Continuous Double Auction. In each period, the market remains open for three minutes, during which participants may buy or short sell the asset. The longer duration compared to the CA reflects the more interactive nature of this market, as traders typically require less than two minutes to submit their individual demand in the CA, making extended time unnecessary. In the DA, limit orders are posted in an open order book. Trades are executed automatically whenever bid and ask prices match. Traders may buy and sell as many units of the asset as they wish, compatibly with the available liquidity. They also have full real-time information about all executed trades in their market, including prices and quantities.

4. Procedures and payment scheme

The sessions were conducted between April 2017 and March 2018 at the EELAB of the University of Milan-Bicocca. The experimental software was developed using `z-Tree` (Fischbacher, 2007).

All sessions followed identical procedures. Upon arrival, participants were randomly assigned to cubicles in the lab. We first elicited participants' risk attitudes and then assigned them to either the High or the Low risk aversion market. Participants then received detailed instructions on the rules and functioning of the relevant market institution (*CA* vs. *DA*) and completed a battery of quizzes. Quiz 1 covered urns and signals; Quiz 2, the belief elicitation procedure; Quiz 3, limit orders; Quiz 4, short selling and the monetary consequences of order execution; and Quiz 5 (and Quiz 6 in *DA*) addressed the functioning of the market interface in the corresponding auction format. Reading of the instructions, whose complete translation is available in the Online Appendix, proceeded only once all participants had successfully completed the quizzes. For each quiz, we recorded the number of mistakes made before arriving at the correct answer as a proxy for each participant's comprehension of the task.

Urns were assigned to trading periods using a pseudo-random procedure. We generated a random sequence of 12 periods, in which each of the four urns appeared exactly three times. This sequence was assigned to half of the sessions for each market institution, while the reversed sequence was assigned to the other half.

Within each of the 12 periods, participants first received their private signal and had 30 seconds to report the probability that each urn had been selected. This was followed by the trading period. The software verified that the no-bankruptcy condition held for all possible states of the world before accepting any order, and returned an error message if the condition was violated. In the DA, traders observed the order books and realized trades in real time. In the CA, traders saw only their own individual demand in real time, while the equilibrium price and resulting individual portfolio were revealed after the order submission phase closed. Under both trading institutions, participants were subsequently asked to report again the probability that each urn had been selected, without receiving any feedback about the true state.

At the end of the experiment, the computer randomly selected: (i) the outcome of the Investment Game (whether the risky asset paid 2.5 or 0 for each unit invested) separately for each participant; (ii) one trading period for all participants in the session to determine payments from the trading task; and (iii) either the ex-ante or the ex-post belief measure for each participant in one period. To prevent hedging within periods, participants were informed that the period chosen for (iii) was always different from that in (ii). For the belief task payments computed via the Binarized Scoring Rule (BSR), a separate random number was assigned to each participant. Both the random numbers and selected periods in (iii) varied across participants to avoid social comparison effects. To ensure transparency and credibility, participants received detailed information at the end about the distribution of all random draws within the session, as well as the actual urn selected in each period. Finally, participants were notified of their earnings, completed a brief questionnaire, and received their compensation anonymously. The average duration of the sessions was about two hours, and the average payment was 15.9€.

We collected data from 20 sessions, i.e., 440 experimental subjects, equally split between CA and DA. These numbers correspond to 20 independent observations for each institution – 10 High and 10 Low risk aversion markets – with data from 480 trading periods overall, 120 for each urn. We test for hypotheses *Hp(Institution)* and *Hp(Risk)* using a Mann-Whitney rank-sum test based on 20 Vs. 20 independent observations.[5] This number of independent observations provides a statistical power of 0.8 to detect a large standardized effect size (Cohen's $d \approx 0.9$) at the 0.05 significance level for the two-tailed test.

[5] Hypothesis *Hp(Institution)* compares prices obtained in 20 DA markets and 20 CA markets, jointly considering high and low risk-aversion markets. Hypothesis *Hp(Risk)* compares prices obtained in 20 high risk-aversion markets and 20, low risk-aversion markets, jointly considering DA and CA markets. Therefore, as we use the same test, we have by construction the same power for the two hypotheses.

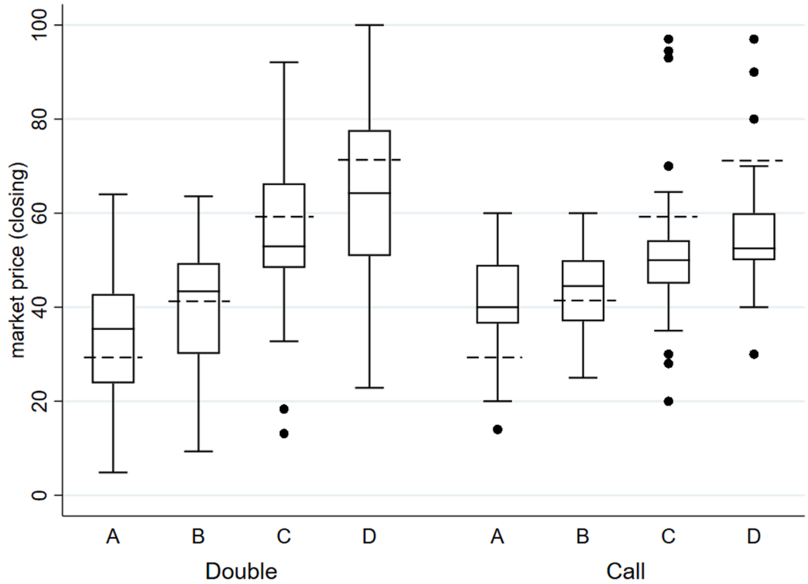

Fig. 1. Prices in the double and in the call auction.
Note: The boxplot shows, for each urn, the distribution of closing prices in the Double Auction and of market-clearing prices in the Call Auction. The dashed lines represent the average Bayesian beliefs conditional on the urn.

5. Results

In this section, we begin by comparing the equilibrium prices with the average beliefs across the two trading institutions. We then examine the effect of risk aversion on prices. Finally, we analyze the role of beliefs and their consistency with the signals received. All non-parametric tests are based on one independent observation per market, i.e., $N = 20$ for each trading institution.

Prices in call and double auction: Informational inefficiency and average beliefs (Hp(Institution))

Fig. 1 summarizes the distributions of closing prices in the two institutions, pooling data from both High and Low risk aversion markets. For the DA, we define the 'closing price' in each period as the average price of the last ten executed trades. These correspond on average to one third of the executed trades in a typical period. In the CA, all trades are executed at the market-clearing price, which we similarly refer to as the closing price. It is common to assess informational efficiency using the absolute differences between prices and REE or PIE (e.g., Plott and Sunder, 1982; Choo et al., 2017; Corgnet et al., 2023). This literature implicitly assumes risk neutrality when computing the PIE. In the context of prediction markets, it is more appropriate to compare prices to the average Bayesian belief, i.e., assuming log-utility (although, in practice, the difference is minimal). Prices exhibit informational inefficiency, as they tend to be closer to the average Bayesian belief (PIE—dashed lines in the figure) than to the REE (0 in Urns A and B; 100 in Urns C and D). The absolute difference between the closing price and the average Bayesian belief is smaller than that between the closing price and the true value of the asset (REE) in 90% of the periods in the DA, and 97% in the CA.

Comparing the two institutions, closing prices are closer to the average Bayesian beliefs in the DA than in the CA, where they instead tend to cluster around the uninformed prior (50). A Mann-Whitney rank-sum test confirms that the absolute difference between prices and the uninformed prior is significantly larger in the DA than in the CA, in line with hypothesis *Hp(Institution)* ($U = 3.354$, $p < .001$). Table 2 corroborates these results. A battery of Wilcoxon signed-rank tests does not reject the equality between closing prices and the average Bayesian belief in the DA, while it finds significant differences for urns A, C, and D in the CA (Panel A). Again with the exception of urn B, a battery of Mann-Whitney rank-sum tests rejects the equality of closing prices between the DA and the CA (Panel B, second row).

Prices in the DA also respond more strongly than in the CA when traders receive additional information that makes the average beliefs more accurate. Specifically,urns A and D distribute more information than urns B and C. A Mann-Whitney rank-sum test confirms that the average absolute difference in prices between high- and low-information urns ($|p_A - p_B|$ and $|p_C - p_D|$) is significantly larger in the DA than in the CA ($U = 2.705$, $p = 0.007$).

- **Result (institution):** Equilibrium prices are significantly closer to the uninformed prior in the CA than in the DA, where they are not significantly different from the average Bayesian belief.

The first part of the main result above is consistent with *Hp(Institution)*. The second part is not, at least at face value: we conjectured that the CA would better replicate the prior information assumption underlying the interpretation of prices as average ex-ante beliefs. This finding may reflect a composition effect, with risk preferences pushing prices toward the uninformed prior (in both the DA and

Table 2
Non-parametric tests on prices.

Panel A: Within-treatment differences of prices Vs. average Bayesian beliefs by urn	Urn A		Urn B		Urn C		Urn D	
	z	p-value	z	p-value	z	p-value	z	p-value
CA	3.883	< .001	1.792	0.073	−3.025	0.003	-3.771	< .001
DA	1.717	0.086	0.149	0.881	−1.045	0.296	−1.531	.126
Panel B: Across-treatment differences of prices by urn	Urn A		Urn B		Urn C		Urn D	
	U	p-value	U	p-value	U	p-value	U	p-value
Opening price	1.271	0.204	1.65	0.099	0.676	0.499	.555	0.579
Closing price	−2.083	0.037	−.595	0.552	1.956	0.046	2.137	0.033

Notes: Panel A reports the Wilcoxon signed-rank test statistic and the corresponding p-value on the (urn-by-urn) difference between prices and average Bayesian beliefs, separately for the Call (CA) and the Double (DA) auction. Panel B reports, for each urn, the Mann-Whitney U test statistic and the corresponding p-value on the (urn-by-urn) difference in prices between the Double and the Call auction. The relevant price is always the market-clearing one in the Call auction. For the Double auction, the first row reports the opening price (average of the first ten trades, or around the first 1/3 of the executed trades in a typical period), while the second row reports the closing price (average of the last ten trades, or around the last 1/3 of the executed trades in a typical period). A positive U statistic indicates a higher value in the Double auction. All statistics are computed using one observation per market (20 independent observations per institution).

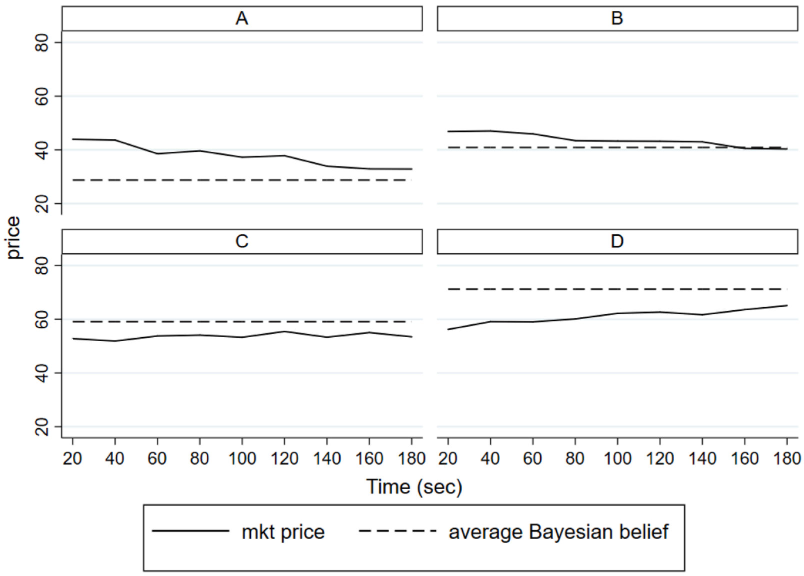

Fig. 2. Convergence of prices within the double auction.
Notes: The figure shows the evolution of prices in the double auction: the moving average across all markets over time within a trading period (3 minutes). The dashed lines correspond to the average Bayesian beliefs given the urn.

the CA), and belief revision occurring in the DA only. In the following sections, we examine the roles of risk preferences and beliefs more closely. Before doing so, two points are worth noting.

First, prices appear more volatile in the DA, as shown in Fig. 1. This larger variance may foster the occurrence of so-called 'mirages' (Camerer and Weigelt, 1991), situations where prices conflict with the information held by the market. In our context, a mirage occurs when the true state of the world is deemed less likely than an incorrect one, i.e., $p_A, p_B > 50$ and $p_C, p_D < 50$. Mirages are slightly more frequent in the DA (22% vs. 16%), but this difference is not statistically significant according to a Fisher's exact test ($p = 0.297$).

Second, the difference between the two institutions emerges over time. Fig. 2 shows that, in the DA, prices converge towards the average Bayesian beliefs during the trading period. By contrast, opening prices in the DA (mean over the first ten trades) do not significantly differ from the market-clearing prices in the CA, as confirmed in Table 2 (Panel B, first row).

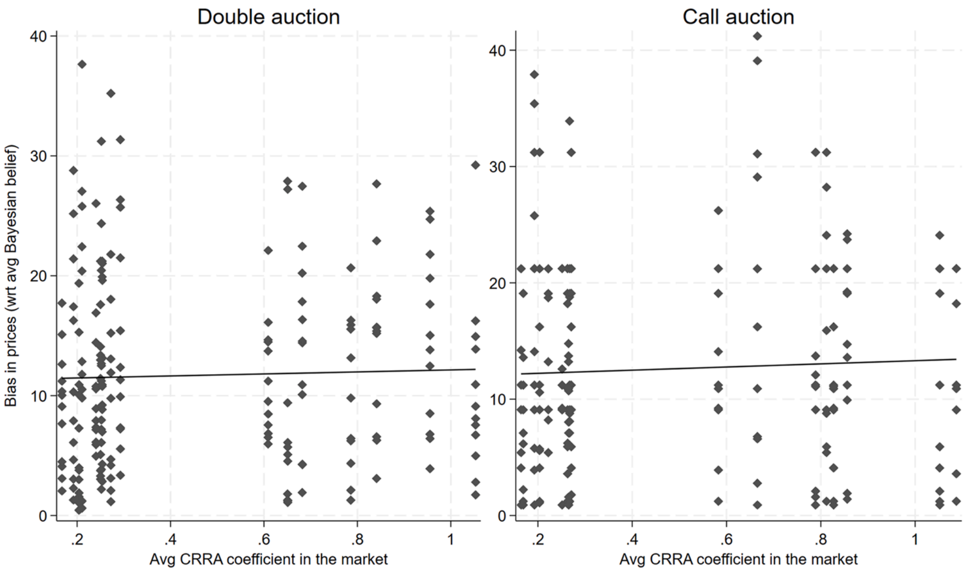

Fig. 3. Risk aversion and difference between prices and average beliefs.
Notes: The figure shows, for the double auction, the distance of prices in each market/period from the average Bayesian belief, plotted against the average CRRA coefficient in the market. Gaps between around 0.3 and 0.6 reflect the matching procedure. A linear fit is superimposed. The panel for the CA is reproduced from Filippin and Mantovani (2023).

Risk aversion (Hp(Risk))

Fig. 3 plots the average degree of risk aversion (CRRA) in each market against the absolute difference between the closing price and the average Bayesian belief, for both the CA and the DA. The CRRA coefficient is estimated for each participant using choices from the Investment Game. Under CRRA and PIE, this difference should be zero (price equals average beliefs) when the CRRA coefficient is 1, and should increase as risk aversion decreases. Thus, a negative correlation is expected, but this is not observed: the difference between price and Bayesian belief is orthogonal to the average degree of risk aversion, and the (non-significant) effect of risk aversion does not differ across institutions. In particular, closing prices are not closer to the average Bayesian belief when the average CRRA approaches log-utility ($\theta = 1$).

The matching protocol exogenously manipulates the degree of risk aversion across markets. We compare the absolute difference between the closing price and the uninformed prior between high and low risk-aversion markets. The corresponding observed average differences are 2.1, 1.2 for urns A and B in the CA, while they have the wrong sign for urns C and D. The same differences are 3.9, 3.7, 0.8 and -0.2 (wrong sign) in the DA. A rank-sum test fails to reject the null of no difference ($U = 0.487$, $p = 0.626$). Fig. 4 confirms that the distribution of closing prices is very similar between high and low risk-aversion markets in both the DA and the CA (institution-specific tests yield similar results: DA, $U = 0.832$, $p = 0.406$; CA, $U = -.227$, $p = 0.821$).

- **Result (Risk):** Prices are not closer to the uninformed prior in low risk-aversion markets.

We do not find support for hypothesis *Hp(Risk)*. To put the null results into perspective, it is useful to construct a qualitative benchmark based on the observed risk preferences of participants. In the low risk-aversion markets, the average estimated CRRA coefficient is 0.23, which implies PIE prices of approximately 40, 47, 53, and 60 for urns A, B, C, and D, respectively. In the high risk-aversion markets, the average estimated CRRA coefficient is 1.41, yielding PIE prices of approximately 25, 38, 62, and 75. Taking these figures at face value, the (absolute) price difference between low and high risk-aversion markets would be about 15 in urns A and D, and about 9 in urns B and C.

Three caveats apply to the result above. First, despite the large differences expected given the observed degree of risk aversion (as shown above), we cannot entirely rule out the possibility that the null result would not replicate under a different experimental design. Second, the absence of a price response to market-level risk aversion does not imply that individual behavior is unaffected by risk preferences. Data from the CA, where individual demand schedules can be directly analyzed, are examined in detail in Filippin and Mantovani (2023), and show that risk preferences do influence individual behavior in terms of orders and market exposure (see also Fellner and Maciejovsky, 2007). Third, the null result may reflect limited power to detect small or medium-sized effects. However, the consistency of the result across institutions, the detailed analysis in Fig. 3 showing no visible relationship between prices and risk aversion, and the fact that our matching protocol should generate relatively large effects argue against this explanation.

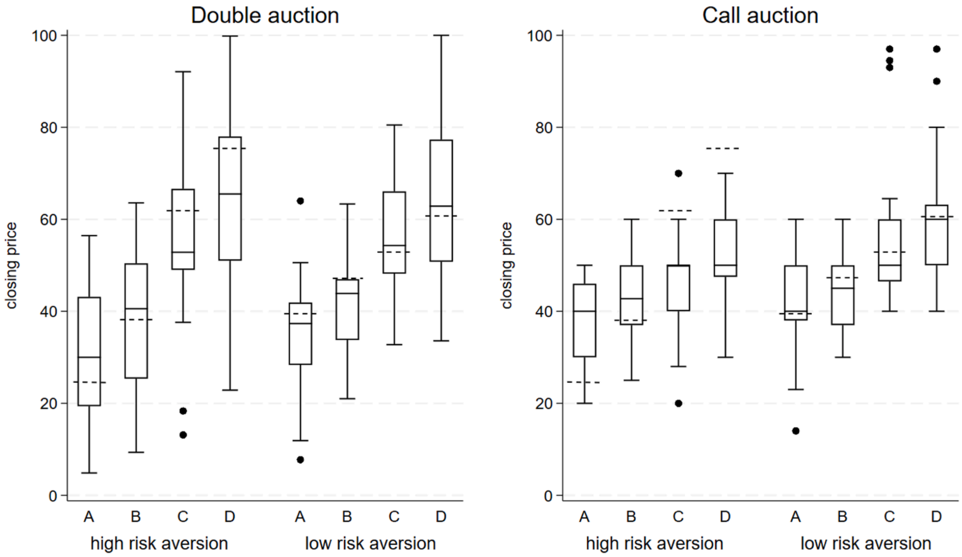

Fig. 4. Risk aversion and prices in the double auction.
Notes: The boxplot shows the distribution of closing prices for each urn in the double auction, separately for high and low risk-aversion markets. The panel for the CA is reproduced from Filippin and Mantovani (2023). The dashed lines report the predicted prices in a PIE where all traders have CRRA preferences as elicited in the Investment Game for high and low risk-aversion markets.

Table 3
Beliefs.

	Signal					Urn			
	42–45	46–47	48–52	53–54	55–58	A	B	C	D
Bayesian	0	33.3	50	66.6	100	28.8	40.9	59.1	71.2
DA ex-ante	4.02	28.25	51.43	76.34	97.60	30.11	42.74	60.11	73.22
CA ex-ante	3.76	27.26	52.54	75.61	97.10	29.46	41.44	61.30	74.29
DA ex-post	13.45	33.02	52.43	69.90	90.33	34.11	44.88	59.84	69.73
CA ex-post	15.99	31.08	49.40	61.35	83.84	33.41	41.19	55.48	64.36

Notes: The 'Bayesian' row reports the posterior following each signal (averaged over urns) and the posterior for each urn (averaged over signals). Subsequent rows show average stated beliefs at the beginning ('ex-ante') and end ('ex-post') of the trading period for each institution (DA: Double Auction, CA: Call Auction).

Bayesian updating and beliefs (Hp(Beliefs))

Another possible source of price deviations from average Bayesian belief is the misperception of the probability that the asset pays off given the signal received, i.e., $p(e = \text{Blue}|s)$. Beliefs are elicited as the probability assigned to urns C and D. Table 3 reports ex-ante and ex-post beliefs across institutions, signals, and urns.

Ex-ante beliefs closely track the Bayesian benchmarks under both institutions. Accordingly, opening beliefs do not differ significantly between institutions. Mann–Whitney tests detect no significant differences between CA and DA at any signal level (signals $[42, 45]$: $U = 0.123$, $p = 0.902$; $[46, 47]$: $U = -.243$, $p = 0.808$; $[48, 52]$: $U = -.947$, $p = 0.344$; $[53, 54]$: $U = 0.974$, $p = 0.330$; $[55, 58]$: $U = 0.950$, $p = 0.342$). Thus, differences in prices between the CA and DA cannot be attributed to differences in individual beliefs. Moreover, the distance between ex-ante beliefs and Bayesian beliefs at the market level does not correlate with deviations of prices from average Bayesian belief ($\rho = 0.082$, $p = 0.615$). We therefore find no support for hypothesis *Hp(Beliefs)* that incorrect Bayesian updating distorts prices.

- **Result (Beliefs):** Deviations of prices from average Bayesian beliefs are not explained by incorrect updating of beliefs.

6. Discussion

The prices we observe in the CA are relatively close, on average, to those predicted in a risk-neutral PIE. Despite the fact that 90% of traders show positive levels of risk aversion in the investment game, one could think that they behave as risk-neutral traders in the market. In the DA, prices start very close to those in the CA, but they converge toward the average Bayesian beliefs during trading. Consistent with our hypothesis regarding the effect of the market institution, one could interpret these closing prices as reflecting partial information aggregation: price-to-state inference induce the revision of ex-ante beliefs toward to the true state. We can verify

Table 4
Predicted and observed quantities.

	Prediction (Avg θ)		Observed DA		CA	
	High RA	Low RA	High RA	Low RA	High RA	Low RA
Urn A	53.1	85.3	38.0	46.1	32.8	34.1
Urn B	35.2	67.1	31.5	40.4	34.2	37.6
Urn C	35.2	67.1	34.2	39.6	34.1	48.3
Urn D	53.1	85.3	33.0	42.1	36.4	43.0

Notes: On the righ-hand side, the Table reports the observed average stock holdings, separately for each Urn, market institution, and for high and low risk-aversion markets. On the left-hand side, the Table reports the predicted traded quantities. Predictions correspond to a PIE where all traders are charaterized by a CRRA coefficient equal to that elicited on average in the Investment Game for high and low risk-aversion markets, respectively.

if prices represent a risk-neutral PIE in the CA and partial information aggregation starting from that benchmark in the DA looking at traded quantities and ex-post beliefs.

We start by assessing some implications of a risk-neutral PIE. First, if traders behave as if they are risk neutral in the market, elicited risk preferences in the investment game should be uncorrelated with the traded volumes, both at the individual and at the market level. Second, risk-neutral traders invest their entire endowment in either long or short positions, and their level of exposure does not depend on their beliefs. Consequently, trade volumes are on aggregate large in a risk-neutral PIE: about 109 assets in urns A and D, and 78 assets in urns B and C. We find evidence against each of these implications in the CA. Individual stock holdings depend negatively on the trader's elicited risk preferences (Spearman $\rho = -0.200$, p-value < 0.01). Traders invest about half of their endowment, and their asset holdings positively depend on the difference between their beliefs and the price (Spearman $\rho = 0.160$, p-value < 0.01).[6] Overall, the volume of traded assets is smaller than in a risk-neutral PIE: 37.6 assets on average.

Table 4 reports more in detail on the traded quantities. For both the DA and the CA, we report the average traded quantity for each urn, separately for high and low risk-aversion markets. For comparison, and in parallel with the exercise in Fig. 4, we predict the quantities that would be traded in a PIE where all traders are charaterized by a CRRA coefficient equal to that elicited on average in the Investment Game for high and low risk-aversion markets, respectively. Two results stand out. First, predicted traded quantities are small not only relative to the risk-neutral benchmark, but also relative to what one would expect given the elicited degree of risk aversion. Second, consistent with the comparative statics reported above at the individual level, predicted quantities are larger in markets populated by traders with lower risk aversion.

We now turn to price-to-state inference. Table 3 shows that ex-post beliefs are indeed revised during the trading period. However, subjects fail to infer the true state of the world. In fact, ex-post beliefs tend to move away from the true state, particularly for the most informative signals. Fully informed traders (those receiving signals 42–45 and 55–58) become more uncertain about the state of the world. Belief updating is similar in both institutions, although ex-post beliefs deviate slightly less from the Bayesian benchmark in the DA. More formally, we measure the precision of individual beliefs as the sum of squared errors relative to the true state of the world. For each signal type, we test the difference between ex-ante and ex-post errors using the Wilcoxon signed-rank test. There is no evidence that beliefs become more precise after trading. On the contrary, a randomly selected individual is significantly more informed before observing market activity than after, in both institutions (DA: $z = -2.800$, $p = 0.005$; CA: $z = -3.733$, $p < 0.001$). Conditioning on the signal, ex-post beliefs significantly worsen in the CA for almost every signal type. Even in the DA, beliefs generally worsen, and significantly so for the most informative signals.

Rather than through sophisticated inference from prices to the state, belief updating appears to follow naive revisions toward the observed price. Fig. 5 shows, for each urn, signal, and institution, the revision from average ex-ante to ex-post beliefs, along with the closing price and the true value of the asset. Ex-ante beliefs move toward the price in almost all cases, while the true state plays little role. Prices are relatively more informative about the state in the DA than in the CA, and correspondingly, ex-post beliefs are more accurate in the DA. Indeed, the sum of squared errors of ex-post beliefs is significantly lower in the DA than in the CA (Mann–Whitney rank-sum test: $U = 2.894$, $p = 0.003$). This type of belief revision is not the price-to-state inference that would lead to information aggregation. It is, instead, similar to the type of updating considered in Manski (2006).

Knowledge of the true state does not improve in the DA despite trading activity. This aligns with literature showing that beliefs do not converge to the REE (Choo et al., 2017). At the same time, the fact that prices improve while beliefs worsen is counterintuitive and suggests that these differences arise from how traders behave given their beliefs, rather than from changes in the beliefs themselves. The question is what mechanism generates these effects.

In the CA, traders fail to translate their information into trading activity. Their demands pivot (switch from positive to negative) around a price closer to the uninformed prior than to their belief, contrary to what expected utility would predict. Filippin and Mantovani (2023) analyze individual demands in the CA and label this pattern *operational conservatism*: traders act as if they were

[6] See Filippin and Mantovani (2023) for an in-depth analysis of individual demands and risk aversion in the CA.

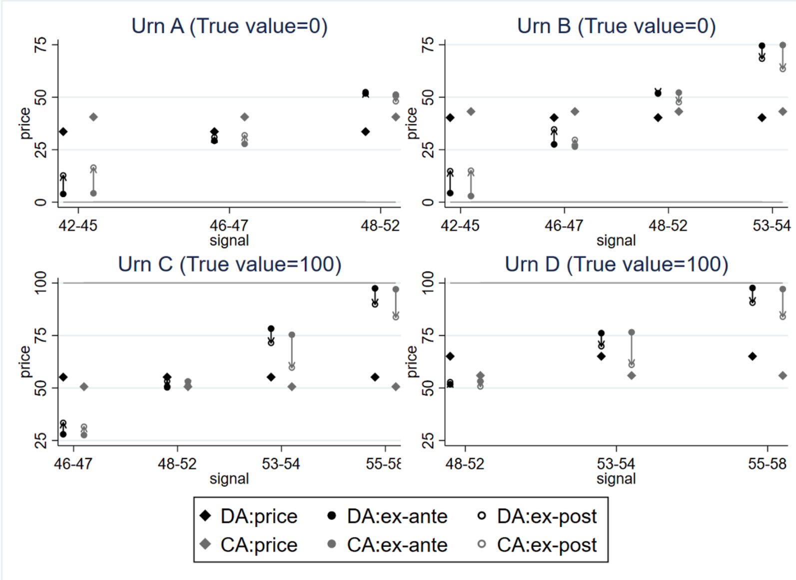

Fig. 5. Revision of beliefs after observing market prices.
Notes: Each plot corresponds to an urn and shows ex-ante (solid circles) and ex-post (hollow circles) average beliefs in both the DA (black) and the CA (gray), by signal type. Arrows connect ex-ante to ex-post beliefs. Diamond symbols indicate the average price under the corresponding urn and institution (prices do not vary across signals for a given urn and institution). The gray line shows the actual value of the asset (0 or 100).

less informed than they actually are. In the DA, traders' activity seems to incorporate information at least in part, although it is less straightforward to formally recover net demands in the DA for comparison with the CA.

This explanation may clarify why prices move toward the true state even while beliefs move toward the price and away from the truth. Since traders act as if their ex-ante belief is closer to 50 than their Bayesian belief, their demand can become more consistent with the Bayesian belief even if their belief worsens during trading. For example, a trader with $b_i = 0.04$ (average ex-ante belief in the DA for $s_i \in [42, 45]$) may initially bid as if $b_i = 0.25$. Even if her belief later revises to $b_i = 0.13$ (average ex-post belief in the DA for the same signals), her final bids may still better align with the information she possesses (which implies $b_i = 0$) than at the start.

Several factors may explain why traders transfer more information into their trading activity in the DA. First, failure to translate information into action may stem from perceived ambiguity or complexity in the environment (Bernheim and Sprenger, 2020; Trautmann and Van De Kuilen, 2015; Dimmock et al., 2016; Fattinger, 2018; Halevy, 2007; Asparouhova et al., 2015; Huber et al., 2019), which may feel less pronounced in the DA. Second, strategic interaction may prompt more aggressive bidding: traders confident that the asset will pay may initially bid at low prices to secure high profits, but observing opponents' orders can lead to increasingly competitive bids, resulting in demands that better reflect private information and higher market prices.

While our experiment cannot disentangle these explanations, one implication of the observed dynamics of prices and beliefs is worth highlighting. As beliefs move closer to the price, their variance across traders with different information decreases. Specifically, the variance of beliefs across signals declines by about 25% and 21% in urns A, D and B, C, respectively. In other words, disagreement among traders diminishes after trading. This makes it unlikely that prices will eventually fully reveal the true state of the world. Rather, the joint movement of prices and beliefs suggests that they converge somewhere between the uninformed prior and the true state. By construction, the average ex-ante belief falls within that range, although there is no guarantee that it represents the precise point of convergence.

7. Conclusion

Prediction markets promise a mechanism for aggregating dispersed beliefs that provides proper incentives while being virtually costless. They have grown and spread across several fields despite the lack of solid empirical validation of the underlying theory. This theory relies on the assumption of price-taking behavior (Prior Information Equilibrium, PIE) and further implies that risk preferences influence the equilibrium price, requiring a specific level (log utility) for prices to accurately reflect average beliefs.

We design an experiment in which markets differ in the degree of risk aversion among traders. We manipulate the market institution (call vs. double auction) to compare settings where the PIE benchmark is more or less likely to hold. We also elicit traders' beliefs before the market, to control for incorrect Bayesian updating, and after the market, to assess the extent to which subjects infer information from observed prices.

Our results show that prices in the double auction are relatively close to the average beliefs, whereas in the call auction they remain closer to an uninformed prior. Although higher informational efficiency in the double auction is expected, this result is surprising given that the institution most consistent with PIE is the call auction. Our design allows us to rule out risk aversion as an explanation for the greater informativeness of prices. We do not find evidence that traders' risk preferences substantially affect market prices.

The crucial element that distinguishes the double auction appears to be the feedback between observed prices, beliefs, and trading behavior. While this mechanism might be expected to facilitate inference about the true state of the world, our data suggest otherwise: beliefs are revised toward observed prices rather than toward the true state after the trading period, as one would expect under rational expectations. The influence of rational expectations equilibria (REE) on economics is hard to overstate. Yet research on the strategic foundations of decentralized markets has long highlighted their limited applicability for information aggregation (Wolinsky, 1990; Blouin and Serrano, 2001). Recent experimental evidence confirms that REE rarely emerge, under institutional arrangements designed to favor them as the double auction (Corgnet et al., 2023).

The answer to the question posed in our title – when do prediction markets reflect average beliefs? – is not clear-cut. On one hand, the convergence of prices and beliefs in the double auction suggests that prediction markets can serve as a costless mechanism for belief aggregation. On the other, call markets fail at this task. At the same time, in both institutions the average trader ends up being less informed after the market than before. Directly eliciting beliefs and adopting alternative mechanisms, such as logarithmic market scoring rules, may complement double auctions in generating predictions while avoiding undesired spillovers on traders' beliefs.

Data availability

Data and replication files are available at https://osf.io/c264t.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary material

Supplementary material associated with this article can be found in the online version at 10.1016/j.geb.2025.12.004.

References

Arrow, K.J., Forsythe, R., Gorham, M., Hahn, R., Hanson, R., Ledyard, J.O., Levmore, S., Litan, R., Milgrom, P., Nelson, F.D., et al., 2008. The promise of prediction markets. Science 320 (5878), 877–878.

Asparouhova, E., Bossaerts, P., Eguia, J., Zame, W., 2015. Asset pricing and asymmetric reasoning. J. Polit. Econ. 123 (1), 66–122.

Asparouhova, E., Bossaerts, P., et al., 2017. Experiments on percolation of information in dark markets. Econ. J. 127 (605), F518–F544.

Berg, J., Forsythe, R., Nelson, F., Rietz, T., 2008. Results from a dozen years of election futures markets research. Handbook. exp. econ. results 1, 742–751.

Bernheim, B.D., Sprenger, C., 2020. On the empirical validity of cumulative prospect theory: experimental evidence of rank-independent probability weighting. Econometrica 88 (4), 1363–1409.

Biais, B., Mariotti, T., Moinas, S., Pouget, S., 2017. Asset Pricing and Risk Sharing in a Complete Market: An Experimental Investigation. Technical Report. Toulouse School of Economics (TSE).

Blouin, M.R., Serrano, R., 2001. A decentralized market with common values uncertainty: non-steady states. Rev. Econ Stud. 68 (2), 323–346.

Bossaerts, P., Frydman, C., Ledyard, J., 2013. The speed of information revelation and eventual price quality in markets with insiders: comparing two theories. Rev Financ. 18 (1), 1–22.

Camerer, C., Weigelt, K., 1991. Information mirages in experimental asset markets. J. Business, 463–493.

Chen, K.-Y., Plott, C.R., 2008. Markets and information aggregation mechanisms. Handb. Exp. Econ. Results 1, 344–352.

Choo, L., Kaplan, T.R., Zultan, R., 2019. Information aggregation in Arrow–Debreu markets: an experiment. Exp. Econ. 22 (3), 625–652.

Choo, L., Kaplan, T.R., Zultan, R., 2022. Manipulation and (MIS) trust in prediction markets. Manage. Sci. .

Corgnet, B., Deck, C., DeSantis, M., Hampton, K., Kimbrough, E.O., 2023. When do security markets aggregate dispersed information? Manage. Sci. 69 (6), 3697–3729.

Crockett, S., Friedman, D., Oprea, R., 2021. Naturally occurring preferences and general equilibrium: a laboratory study. Int. Econ. Rev. 62 (2), 831–859.

Crosetto, P., Filippin, A., 2016. A theoretical and experimental appraisal of four risk elicitation methods. Exp. Econ. 613–641. https://doi.org/10.1007/s10683-015-9457-9

Danz, D., Vesterlund, L., Wilson, A.J., 2022. Belief elicitation and behavioral incentive compatibility. Am. Econ. Rev. 112 (9), 2851–2883.

Deck, C., Porter, D., 2013. Prediction markets in the laboratory. J. Econ. Surv. 27 (3), 589–603.

Dimmock, S.G., Kouwenberg, R., Wakker, P.P., 2016. Ambiguity attitudes in a large representative sample. Manage. Sci. 62 (5), 1363–1380.

Fattinger, F., 2018. Trading complex risks. Available at SSRN 3086358 .

Fellner, G., Maciejovsky, B., 2007. Risk attitude and market behavior: evidence from experimental asset markets. J. Econ. Psychol. 28 (3), 338–350. https://doi.org/https://doi.org/10.1016/j.joep.2007.01.006

Filippin, A., Mantovani, M., 2023. Risk aversion and information aggregation in binary asset markets. Quant. Econ. 14 (2), 753–798.

Fischbacher, U., 2007. z-Tree: zurich toolbox for ready-made economic experiments. Exp. Econ. 10 (2), 171–178.

Fountain, J., Harrison, G., 2011. What do prediction markets predict? Appl. Econ. Lett. 18 (3), 267–272. https://doi.org/10.1080/13504850903559575

Friedman, D., 1984. On the efficiency of experimental double auction markets. Am. Econ. Rev. 74 (1), 60–72.

Friedman, D., Ostroy, J., 1995. Competitivity in auction markets: an experimental and theoretical investigation. Econ. J. 105 (428), 22–53.

Gillen, B., Snowberg, E., Yariv, L., 2019. Experimenting with measurement error: techniques with applications to the caltech cohort study. J. Polit. Econ. 127 (4), 1826–1863.

Gjerstad, S., 2005. Risk Aversion, Beliefs, and Prediction Market Equilibrium. Microeconomics 0411002. EconWPA.

Gneezy, U., Potters, J., 1997. An experiment on risk taking and evaluation periods. Q. J. Econ. 112 (2), 631–45.

Guarnaschelli, S., Kwasnica, A.M., Plott, C.R., 2003. Information aggregation in double auctions: rational expectations and the winner's curse. Inf.n Syst. Front. 5 (1), 63–77.

Halevy, Y., 2007. Ellsberg revisited: an experimental study. Econometrica 75 (2), 503–536.

Harris, M., Raviv, A., 1993. Differences of opinion make a horse race. Rev. Financ. Stud. 6 (3), 473–506.

Harrison, J.M., Kreps, D.M., 1978. Speculative investor behavior in a stock market with heterogeneous expectations. Q. J. Econ. 92 (2), 323–336.

He, X.-Z., Treich, N., 2017. Prediction market prices under risk aversion and heterogeneous beliefs. J. Math. Econ. 70, 105–114.

Healy, P.J., Linardi, S., Lowery, J.R., Ledyard, J.O., 2010. Prediction markets: alternative mechanisms for complex environments with few traders. Manage. Sci. 56 (11), 1977–1996.
Horn, C.F., Ivens, B.S., Ohneberg, M., Brem, A., 2014. Prediction markets–a literature review 2014. J. Prediction Mark. 8 (2), 89–126.
Hossain, T., Okui, R., 2013. The binarized scoring rule. Rev. Econ. Stud. 80 (3), 984. https://doi.org/10.1093/restud/rdt006
Huber, J., Palan, S., Zeisberger, S., 2019. Does investor risk perception drive asset prices in markets? experimental evidence. J. Banking Finance. 108, 105635.
Kagel, J.H., 2004. Double auction markets with stochastic supply and demand schedules: call markets and continuous auction trading mechanisms. In: Advances in Understanding Strategic Behaviour. Springer, pp. 181–208.
Kagel, J.H., Levin, D., 1986. The winner's curse and public information in common value auctions. Am. Econ. Rev. 76 (5), 894–920.
Ketcham, J., Smith, V.L., Williams, A.W., 1984. A comparison of posted-offer and double-auction pricing institutions. Rev. Econ. Stud. 51 (4), 595–614.
Luckner, S., Schröder, J., Slamka, C., 2008. On the forecast accuracy of sports prediction markets. In: Negotiation, Auctions, and Market Engineering. Springer, pp. 227–234.
Manski, C.F., 2006. Interpreting the predictions of prediction markets. Econ. Lett. 91 (3), 425–429.
Milgrom, P., Stokey, N., 1982. Information, trade and common knowledge. J. Econ. Theory 26 (1), 17–27.
Miller, E.M., 1977. Risk, uncertainty, and divergence of opinion. J. Finance 32 (4), 1151–1168.
Morris, S., 1996. Speculative investor behavior and learning. Q. J. Econ. 111 (4), 1111–1133.
Ngangoué, M.K., Weizsäcker, G., 2021. Learning from unrealized versus realized prices. Am. Econ. J. Microeconomics 13 (2), 174–201.
Ottaviani, M., Sørensen, P.N., 2015. Price reaction to information with heterogeneous beliefs and wealth effects: underreaction, momentum, and reversal. Am. Econ. Rev. 105 (1), 1–34.
Page, L., Siemroth, C., 2017. An experimental analysis of information acquisition in prediction markets. Games. Econ. Behav. 101, 354–378.
Page, L., Siemroth, C., 2021. How much information is incorporated into financial asset prices? experimental evidence. Rev. Financ Stud. 34 (9), 4412–4449.
Palan, S., Huber, J., Senninger, L., 2020. Aggregation mechanisms for crowd predictions. Exp. Econ. 23 (3), 788–814.
Palfrey, T.R., Wang, S.W., 2012. Speculative overpricing in asset markets with information flows. Econometrica 80 (5), 1937–1976.
Plott, C.R., Sunder, S., 1982. Efficiency of experimental security markets with insider information: an application of rational-expectations models. J. Polit. Econ. 90 (4), 663–698.
Plott, C.R., Sunder, S., 1988. Rational expectations and the aggregation of diverse information in laboratory security markets. Econometrica J. Econometric Soc. 56 (5), 1085–1118.
Sakurai, H., Akiyama, E., 2017. Bubbles and information in continuous double auction and call market: an experiment. Available at SSRN 2930083 .
Smith, V.L., Williams, A.W., 2000. The boundaries of competitive price theory: convergence, expectations, and transaction costs. In: Bargaining and Market Behavior: Essays in Experimental Economics, pp. 286–319.
Smith, V.L., Williams, A.W., Bratton, W.K., Vannoni, M.G., 1982. Competitive market institutions: double auctions vs. sealed bid-offer auctions. Am. Econ. Rev. 72 (1), 58–77.
Snowberg, E., Wolfers, J., 2010. Explaining the favorite–long shot bias: is it risk-love or misperceptions? J. Polit. Econ. 118 (4), 723–746.
Sonnemann, U., Camerer, C.F., Fox, C.R., Langer, T., 2013. How psychological framing affects economic market prices in the lab and field. Proc. Natl. Acad. Sci. 110 (29), 11779–11784.
Spann, M., Skiera, B., 2009. Sports forecasting: a comparison of the forecast accuracy of prediction markets, betting odds and tipsters. J. Forecast. 28 (1), 55–72.
Trautmann, S.T., Van De Kuilen, G., 2015. Ambiguity attitudes. The Wiley Blackwell handbook of judgment and decision making 1, 89–116.
Wolfers, J., Zitzewitz, E., 2004. Prediction markets. J. econ. perspect 18 (2), 107–126.
Wolfers, J., Zitzewitz, E., 2006. Interpreting Prediction Market Prices as Probabilities. Technical Report. National Bureau of Economic Research.
Wolinsky, A., 1990. Information revelation in a market with pairwise meetings. Econometrica: J. Econometric Soc. , 1–23.