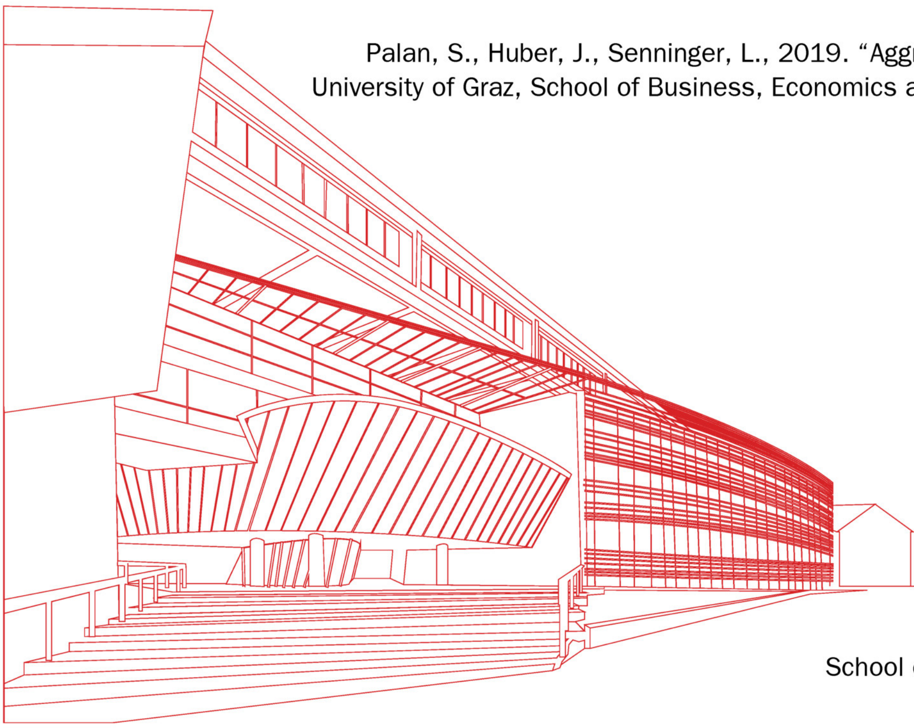
KARL-FRANZENS-UNIVERSITÄT GRAZ
UNIVERSITY OF GRAZ
School of Business, Economics and Social Sciences

Aggregation mechanisms for crowd predictions

Stefan Palan, Jürgen Huber, Larissa Senninger

Working Paper 2019-01
May 15, 2019



Recommended citation:
Palan, S., Huber, J., Senninger, L., 2019. "Aggregation mechanisms for crowd predictions", University of Graz, School of Business, Economics and Social Sciences Working Paper 2019-01.

Working Paper Series
School of Business, Economics and Social Sciences
University of Graz
ISSN 2304-7658
http://bit.ly/graz-working-papers
sowi-wp@uni-graz.at

UNIVERSITY OF GRAZ

Aggregation mechanisms for crowd predictions

Stefan Palan
University of Graz, Department of Banking and Finance
University of Innsbruck, Department of Banking and Finance
Corresponding author: +43(316)380-7306, stefan.palan@uni-graz.at

Jürgen Huber
University of Innsbruck, Department of Banking and Finance

Larissa Senninger
University of Innsbruck, Department of Banking and Finance

Working Paper 2019-01
May 15, 2019

Abstract

When the information of many individuals is pooled, the resulting aggregate often is a good predictor of unknown quantities or facts ("wisdom of crowds"). This aggregate predictor frequently outperforms the forecasts of experts or even the best individual forecast included in the aggregation process. However, an appropriate aggregation mechanism is considered crucial to reaping the benefits of a "wise crowd". Of the many possible ways to aggregate individual forecasts, we compare (uncensored and censored) mean and median, continuous double auction market prices and sealed bid-offer call market prices in a controlled experiment. We use an asymmetric information structure where subjects know different subsets of the total information needed to exactly calculate the asset value to be estimated. We find that prices from continuous double auction markets clearly outperform all alternative approaches for aggregating dispersed information and that information is only useful to the best-informed subjects.

Keywords: information aggregation, asymmetric information, wisdom of crowds
JEL: C53, C83, G14

Working Paper Series
School of Business, Economics and Social Sciences
University of Graz
ISSN 2304-7658
http://bit.ly/graz-working-papers
sowi-wp@uni-graz.at

Aggregation mechanisms for crowd predictions☆

Stefan Palan[a,b,*], Jürgen Huber[b], Larissa Senninger[b]

[a]*Department of Banking and Finance, University of Graz, Universitätsstraße 15, 8010 Graz, Austria*
[b]*Department of Banking and Finance, University of Innsbruck, Universitätsstraße 15, 6020 Innsbruck, Austria*

Abstract

When the information of many individuals is pooled, the resulting aggregate often is a good predictor of unknown quantities or facts ("wisdom of crowds"). This aggregate predictor frequently outperforms the forecasts of experts or even the best individual forecast included in the aggregation process. However, an appropriate aggregation mechanism is considered crucial to reaping the benefits of a "wise crowd". Of the many possible ways to aggregate individual forecasts, we compare (uncensored and censored) mean and median, continuous double auction market prices and sealed bid-offer call market prices in a controlled experiment. We use an asymmetric information structure where subjects know different subsets of the total information needed to exactly calculate the asset value to be estimated. We find that prices from continuous double auction markets clearly outperform all alternative approaches for aggregating dispersed information and that information is only useful to the best-informed subjects.

Keywords: information aggregation, asymmetric information, wisdom of crowds
JEL: C53, C83, G14

"Wisdom of crowds", after Surowiecki's (2004) book of the same name, is a term used to describe the observation that the aggregate of forecasts by multiple people is often a better predictor of actual outcomes than the forecasts of experts or even the best individual forecast included in the aggregation process. A number of studies have set out to document this outperformance (e.g., Gordon, 1924; Bruce, 1935; Sauer, 1998; Berg et al., 2008a,b) and to explore and describe which forecasters and forecasting targets most readily lend themselves to successful crowd prediction (e.g., Lorge, 1958; Brown and Sauer, 1993; Berg and Rietz, 2003; Gruca et al., 2003; Polgreen et al., 2007; Davis-Stober et al., 2014).

☆The authors thank Ben Greiner, Redzo Mujcic, Owen Powell, James Tremewan and seminar participants at the University of Graz and the Vienna University of Economics and Business for valuable comments, as well as Christian Schitter for research assistance. Palan gratefully acknowledges funding by the Austrian Science Fund FWF (START-Grant Y617-G11 Kirchler and SFB F63 Huber).

*Corresponding author.

Email addresses: `stefan.palan@uni-graz.at` (Stefan Palan), `juergen.huber@uibk.ac.at` (Jürgen Huber), `larissa.senninger@gmail.com` (Larissa Senninger)

In the present paper, we instead aim to compare different mechanisms of aggregating crowd predictions regarding predictive accuracy in a setting with asymmetric information. Our experiment includes very simple mechanisms, like calculating the average or median of individual predictions, and more complex ones, like using prices from a continuous double auction market. We aim to answer the question whether simple mechanisms perform equally well or even better than more complex ones and should thus be the instruments of choice, or whether more complex mechanisms yield better predictions, which offset their higher costs in terms of time and infrastructure expenditures. We are of course not the first to ask this question. In work directly related to ours, Clemen (1989) provides a literature review on combining forecasts. He finds that in the majority of cases simple aggregation mechanisms are more effective than more complex ones. This result is supported by the more recent work of Soll et al. (2009), who report that simple averaging is the most effective way of combining individual judgments. Other work in favor of averaging individual estimates is Budescu and Yu (2006) and Lichtendahl Jr et al. (2013) (both comparing it to using Bayes' rule) and Larrick et al. (2012) (in effect comparing it to randomly choosing an individual estimate).[1] One more sophisticated averaging approach is advanced by Budescu and Chen (2014), as e.g., they use a model that identifies experts in the crowd and weights their opinions by relevance when aggregating the individual estimates to a group opinion.

In a more nuanced finding, Malone et al. (2009) argue that averaging is a surprisingly good tool when estimating a certain number, but that in more complex situations more complex mechanisms are needed to aggregate information efficiently. They list "prediction markets" and markets with monetary or non-monetary incentives as being such mechanisms. In line with this view, market-based mechanisms have indeed gained significant attention in recent decades. In prediction markets, the market's organizers create an asset whose value is tied to the outcome to be estimated.[2] Defining such assets thus transforms the estimation of an unknown outcome or its probability into a task that can be accomplished by a market. In markets, prices have the role of aggregating available information. We explore a setting with asymmetric information, as in many relevant predictions (e.g., future stock prices, betting outcomes, etc.) participants will typically have different information and – even more relevant – information of different quality. We mimic this with our experimental design, where we can clearly distinguish better and worse informed subjects.

[1] Larrick et al. (2012) use Jensen's (1906) inequality to prove that the absolute forecast error of the average estimate must be smaller than or equal to the average of the individual estimates' absolute forecast errors. For the task of arriving at a point forecast, this implies that the average over a set of estimates is a (probabilistically) better – i.e., more precise – predictor of the value to be estimated than a randomly chosen element of the set of estimates.

[2] Such a derivative asset may, for example, at a pre-defined maturity date, pay a fixed amount of money conditional on an underlying event having occurred (e.g., a contract that pays \$ 1 if politician X gets elected). Alternatively, the asset may pay an amount that is a linear function of the underlying number to be estimated (e.g., a contract that pays $\$\,x \cdot 100$, where x is the vote share of politician X, in percent).

1. Experimental design

We propose a research design which is simple, easy to understand and allows studying our research question under controlled conditions. Using a laboratory experiment, we first let participants estimate the value of a jar filled with coins. We then provide them with partial information about the coins in the jar and elicit updated estimates. Finally, subjects trade the jars in a market, which aggregates their dispersed and noisy information into market prices. This procedure allows us to analyze – and compare the performance of – multiple mechanisms for aggregating dispersed information. The mechanisms we study are (1) (censored) means and medians of individual, incentivized estimates, (2) mean, median and closing prices as well as closing bid-ask midpoint of a continuous double auction, and (3) the uniform settlement price from a sealed bid-ask call auction.

1.1. Assets and information levels

In preparation for our experiment, we fill four plastic jars with 1-euro and 20-, 5- and 1-cent coins. Figure 1 shows a photo and Table 1 presents information about the value of the coins in each of the jars, designated A through D.

Figure 1: Photo of the four plastic jars employed in the experiment.

Jars contain an average of 25 euros (s.d. 2.58), made up of, on average, 8 euros each in coins of 1 euro (s.d. 1.58 euros), 20 cents (s.d. 1.10 euros), and 5 cents (s.d. 0.74 euros) as well as 1 euro in coins of 1 cent (s.d. 0.24 euros). Subjects are informed that these four types of coins are contained in each of the jars. They can also obtain (imperfect) information about the value of the coins contained in each individual jar[3] from two sources. First, each

[3]We will hereafter use expressions like "value of the coins in the jar" interchangeably with "value of the jar" or BBV (buyback value).

Table 1: Value of coins in jars

jar	A	B	C	D	Total
1 euro	9	10	7	6	32
20 cents	7.2	9.6	8.4	6.8	32
5 cents	7.05	9	7.6	8.35	32
1 cent	1.26	0.66	0.88	1.2	4
Total	24.51	29.26	23.88	22.35	100

subject is handed each of the jars for 15 seconds to view, turn, weigh in their hands, etc. Subjects are not allowed to open the jars or use any means other than their senses to analyze the jars' contents. Second, subjects are provided with one of four information levels for each of the jars. More precisely, each subject receives information level I0 for one of the jars, I1 for another jar, I2 for yet another and I3 for the fourth jar. Subjects assigned information level I0 do not receive any additional information about the coins in the jar. Subjects assigned level I1 receive full information about the number (and, separately stated on the computer screen, value) of the 1 euro coins in the jar. Subjects assigned level I2 receive full information about the number (and value) of the 20 cent coins in the jar in addition to the information contained in level I1. Subjects assigned level I3 receive full information about the number (and value) of the 5 cent coins in the jar, in addition to the information contained in level I2. Thus I3 subjects are fully informed about the number and value of 1 euro, 20 cent and 5 cent coins in the jar. No subject receives information about the number (or value) of the 1 cent coins in the jar.

To summarize, all subjects have some, but incomplete information about the value of a jar from viewing and handling the jar for 15 seconds. Information levels I1 through I3 are cumulative, such that subjects with higher information levels have all the information of subjects with lower information levels, plus additional information, and are thus strictly better informed than subjects with lower information levels. Designate as V_1, $V_{0.2}$, $V_{0.05}$ and $V_{0.01}$ the value of 1 euro, 20 cent, 5 cent and 1 cent coins in a jar. Then, depending on information level, subjects have the following information about a non-stochastic lower bound of jar value BBV:[4]

- I0: lower bound equals 0
- I1: lower bound equals V_1
- I2: lower bound equals $V_1 + V_{0.2}$
- I3: lower bound equals $V_1 + V_{0.2} + V_{0.05}$

[4]Of course, since subjects can view and handle jars, they can instantly establish a lower bound above 0 even in I0.

1.2. Procedure

The experiment consists of six sessions with 24 subjects each, conducted on February 22 and 23, 2017, in the Innsbruck EconLab. The 144 subjects were recruited from a standard student subjects pool using hroot (Bock et al., 2014) and the experiment was conducted using GIMS 7.4.16 (Palan, 2015), programmed in z-Tree 3.6.7 (Fischbacher, 2007).

Half of the six sessions employ a call auction (CA), the other half a continuous double auction (CDA) trading protocol. In each session, subjects arrive outside the lab and, after an experimenter has checked their IDs, are randomly assigned to workstations in the lab. An experimenter then reads out aloud the instructions on the respective trading mechanisms, with subjects reading along using personal sets of paper copies of the instructions, which they retain for the entire experiment.[5] Subjects then complete a trial period to get acquainted with the trading interface. Following that, we hand out a second set of instructions that contains information on the asset, on the tasks to perform in the experiment, and on the payoff calculation.

The 24 subjects in each session are split into three groups of eight subjects each. These groups remain fixed throughout the experiment (partner matching). A session consists of 12 trading periods, structured into four blocks of three periods each (one block for each jar). At the beginning of each block, the first subject in each group receives one of the four jars, may view and handle it for 15 seconds and then has to hand it on to the next subject in the group, until all eight subjects have had a chance to inspect the jar. Subjects then submit estimates of the jar value on their respective computers. In each group, two subjects each then receive information levels I0 through I3, such that each information level is represented twice in each group of eight. After having received this information, subjects submit updated estimates of the jar value. They do so again at the beginning of the second and third periods in each block of three periods. The estimates are incentivized as follows: for each estimate that is within $\pm 5\%$ of the true value they receive 20 cents, for each estimate that is within $\pm 15\%$ they receive 10 cents, and for each estimate that is within $\pm 25\%$ they receive 5 cents.

After they have submitted their estimates, subjects are each endowed (virtually, on the computer) with 5 jars and an amount of experimental euros averaging twice the value of the 5 jars, while ensuring that subjects cannot calculate the jar value from their cash endowment.[6] The ratio of outstanding cash to the value of outstanding assets, commonly referred to as

[5] A translation of all instructions, which were originally in German, is included in the online appendix.

[6] For the determination of these euro amounts, we started from two principles. First, there should be no direct correspondence between euro amount and jar value to prevent traders from inferring the latter from the former. Second, the cash-to-asset value ratio should be constant at a value of 2 across all markets. We thus obtained the euro values as follows: We randomly drew (and redrew), for each subject, cash endowments from a uniform distribution over [200,300] experimental euros. We repeated the drawing until the absolute deviation of total cash endowment in the market from total asset endowment value equalled, to two decimal places, 2. We thus obtain individual cash endowments which vary substantially around twice the value of the asset endowment, while ensuring that the cash-to-asset value ratio always equals 2 at the market level. See Table A.8 in the appendix for details. Subjects are symmetrically informed that each subject is endowed with 5 jars and they are told that each subject is endowed with a euro amount that varies across subjects and periods. They are not informed about details of the cash endowment determination algorithm.

the cash-to-asset ratio, thus is 2. This ensures that traders are able to make transactions at reasonable frequencies and prices but it is also reasonably low to avoid biasing our results by cash endowment effects (see Kirchler et al., 2012 and Noussair and Tucker, 2016 and the references therein for evidence on the effect of cash endowments on mispricing). Subjects then trade assets for cash for three minutes both in the CA and in the CDA treatments. Unexecuted orders can be canceled without cost at any time, and are executed according to price followed by time priority. Shorting stocks and borrowing money is not possible. No interest is paid on cash and there are no transaction costs.

Periods within a block are independent in the sense that subjects' endowments are reset to the same starting values at the beginning of every period. Procedures follow the same pattern across blocks, except that traders' information levels and the jar they trade change (every trader receives information level I0 in one block, I1 in one block, I2 in one block and I3 in one block). Subjects are fully and publicly informed about the procedures just outlined.

Finally, we ask subjects the financial literacy questions 2, 3, 4, 7, 10, 12, and 16 of van-Rooij et al. (2011). The computer then randomly chooses one of the questions and subjects earn an additional €1.00 if their answer on this question is correct. The questionnaire is followed by the payment. Subjects' final payoff is determined by randomly drawing one period, summing the value of final asset holdings and cash holdings, dividing by an exchange rate of 30 and adding the earnings from the estimation task. Payment is handed over individually and privately and subjects are asked not to divulge details about the experiment to other students. The experiment lasted approximately 75 minutes and the average payment was €16.02 per subject (s.d. 3.19). Figure 2 illustrates the session structure.

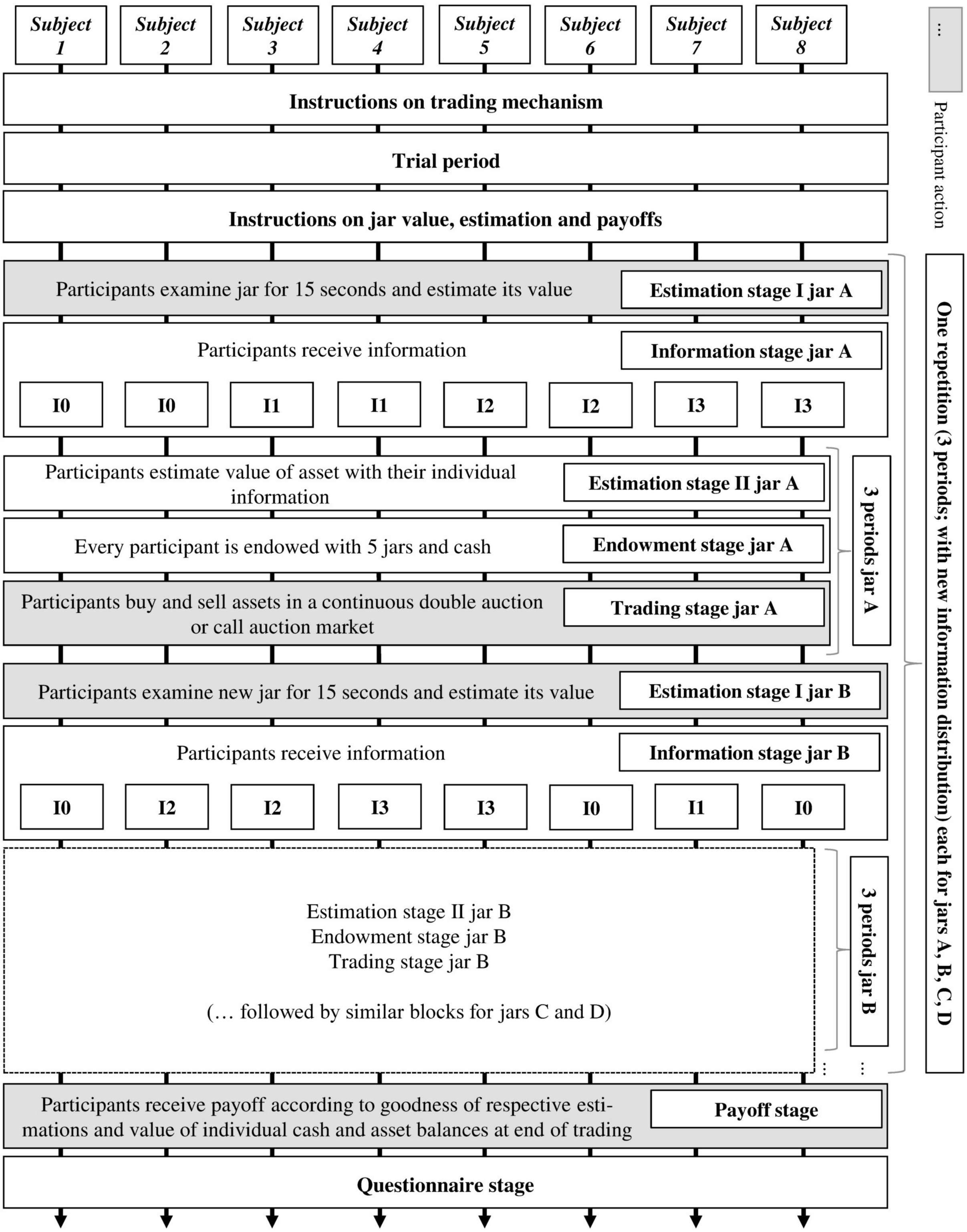


Figure 2: Structure of an experimental session.

1.3. Discussion of design choices

Before continuing to the discussion of our experimental results, we wish to take a moment to discuss some of our design choices. We accordingly structure this section by topic.

Independence of estimates Several authors caution that some conditions have to be met for crowd estimates to outperform other forecasting mechanisms. Surowiecki (2004) for example argues that individuals not only need to have different opinions about the issue in question, they also need to be able to make independent guesses. Similarly, Herzog and Hertwig (2011) recommend mixing participants with different backgrounds and to ask for their opinions independently. They even suggest deliberately perturbing crowd members' original opinions by influencing them in one way or the other. Lorenz et al. (2011) notes that care needs to be taken when sharing information with estimators, since "even mild social influence can undermine the wisdom of the crowd effect in simple estimation tasks" (Lorenz et al., 2011, 9020). We account for these insights by giving subjects no misleading cues regarding jar values and by making them judge the jar values for themselves, privately and independently. We furthermore 'perturb' (in an unbiased sense) their unaided assessments by providing them with differing levels of information.

Relation to theory With this paper, we do not wish to challenge theoretical results regarding the aggregation of predictions, nor contribute to the theoretical literature in statistics/econometrics. Such studies usually need to assume some constraints on the prediction target (e.g., which distribution it is drawn from) or on estimator characteristics (e.g., risk-preferences – see Manski, 2006; Gjerstad, 2004; Wolfers and Zitzewitz, 2006; Ottoviani and Sørensen, 2009). We instead conduct an experimental study to see which information aggregation mechanism performs best in an empirical setting, where the distribution the underlying value is drawn from, as well as the distributions of the noise terms in individual estimates, are unknown by subjects, and where subjects are asymmetrically informed.

Incentives In addition to the forecasters differing in their information levels and presumably in how they interpret this information, incentives play a crucial role. In many contexts, incentivizing forecasters to provide their best effort in forecasting is unproblematic, since the forecast solicitors can simply pay forecasters based on the distance between their forecast and the actual outcome, using for example a proper scoring rule. This is less straightforward in market experiments. While in the case of the individual elicitation of forecasts forecasters have no incentives to withhold information, they have such incentives in prediction markets. There, their information is rendered worthless when it becomes publicly known. This argument also applies in our experiment. If forecasts derived from market experiments do a good job of predicting the underlying and unknown value, this is *because* subjects have incentives to perform well in the market, and *despite* them having incentives to withhold their information (particularly when it is superior to others') from other market participants so they alone can profit from it. In any case, we expect subjects in our market experiments to reveal their information only gradually, such that price efficiency improves over time within trading periods, and that later prices are more informative than earlier ones.

2. Results

2.1. Individual behavior

We begin by exploring subjects' estimates. Subjects provide one estimate for the value of the jar they are about to trade prior to receiving information level I0, I1, I2 or I3, and then, after they have received information, provide another estimate at the beginning of every period. We first look at their estimates prior to trading, i.e., in the first period they trade a particular jar, after they have looked at the jar, but before starting to trade the jar (estimation stage I). These estimates are based on the ambiguous information from handling the jar, but not on information they may infer from trading with other subjects.

Overall, subjects underestimate the value of the coins in the jars. After having looked at and handled a jar for 15 seconds, but before receiving explicit information about the coins in the jar, subjects underestimate the true mean jar value of €25 by on average €7.09 ($t(575) = -21.425, p = 0.0000$). After receiving information, this underestimation shrinks to €3.94 ($t(575) = -15.894, p = 0.0000$). Male and female subjects underestimate by €6.66 and €7.43 (gender difference: Welch $t(568.27) = 1.1728, p = 0.2414$) before receiving information, respectively, and by €3.68 and €4.15 after (Welch $t(573.73) = 0.9717, p = 0.3316$).

Result 1. *Participants underestimate jar values. There is no significant gender difference in estimate deviations.*

For our subsequent analyses, we define a subject's jar value estimate deviation Dev as:

$$Dev^{\theta} = ln\left(\frac{Estimate^{\theta}}{BBV}\right) \tag{1}$$

Here, $\theta \in \{pre, post\}$ signifies whether the estimate was made prior to (pre) or after ($post$) revelation of explicit information about the jar value (i.e., I1, I2, I3). Dev^{θ} thus measures the log percentage deviation of estimates from fundamental value.[7] We also define $AbsDev^{\theta} \equiv |Dev^{\theta}|$ as the absolute value of Dev^{θ}.

Table 2 regresses Dev and $AbsDev$ on subjects' experience in judging jars and on their information level (JarNo equals 1 for the first jar a subject sees, 2 for the second, etc.).[8] The table documents that subjects' forecasts improve as subjects gain experience across different jars. If JarNo=2, for example, this implies that this is the second jar a subject has encountered in the experiment. Furthermore, additional information also significantly improves subjects' forecasts, with the coefficients monotonously increasing (decreasing, in the case of $AbsDev^{\theta}$) with the information level.

[7]Due to the log specification, this measure is independent of the choice of numeraire (i.e., whether one expresses prices as taler/jar or jars/taler). See Powell (2016) for details.

[8]Table A.9 in the appendix repeats this analysis but includes subject dummy variables (albeit, to conserve space, not in the output) to give a better indication of the explanatory power of the models ($R^2 > 0.5$ throughout) when accounting for subject heterogeneity.

	Dev^{pre}	$AbsDev^{pre}$	Dev^{post}	Dev^{post}	$AbsDev^{post}$	$AbsDev^{post}$
Intercept	-0.760^{***}	0.832^{***}	-0.350^{***}	-0.501^{***}	0.403^{***}	0.581^{***}
	(0.038)	(0.032)	(0.021)	(0.031)	(0.018)	(0.026)
JarNo	0.141^{***}	-0.151^{***}		0.060^{***}		-0.071^{***}
	(0.014)	(0.012)		(0.009)		(0.008)
I1			0.061^{**}	0.061^{**}	-0.100^{***}	-0.100^{***}
			(0.030)	(0.029)	(0.026)	(0.024)
I2			0.141^{***}	0.141^{***}	-0.155^{***}	-0.155^{***}
			(0.030)	(0.029)	(0.026)	(0.024)
I3			0.362^{***}	0.362^{***}	-0.361^{***}	-0.361^{***}
			(0.030)	(0.029)	(0.026)	(0.024)
R^2	0.151	0.230	0.222	0.276	0.265	0.360
Adj. R^2	0.149	0.228	0.218	0.271	0.261	0.356
RMSE	0.374	0.310	0.257	0.248	0.221	0.206
Num. obs.	576	576	576	576	576	576

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors in parentheses.

Table 2: OLS regressions of jar value estimates, before (*pre*) and after (*post*) information provision and using relative (*Dev*) and absolute (*AbsDev*) log deviations.

Table 3 presents the same picture as Table 2, but includes a variable JarPeriod to account for subjects' learning over consecutive periods of trading the same jar. For JarPeriod=2, for example, subjects' estimates of the jar value reflect their experience in the market in the first period of trading the same asset. We find that observing the market across periods helps subjects forecast better. Nevertheless, gaining experience across different jars continues to significantly improve subjects' estimates.

Result 2. *Participants' estimates improve over time, both within and across jars.*

	Dev^{post}	Dev^{post}	$AbsDev^{post}$	$AbsDev^{post}$
Intercept	-0.335^{***}	-0.469^{***}	0.391^{***}	0.547^{***}
	(0.018)	(0.021)	(0.016)	(0.018)
JarPeriod	0.023^{***}	0.023^{***}	-0.029^{***}	-0.029^{***}
	(0.007)	(0.007)	(0.006)	(0.006)
JarNo		0.053^{***}		-0.062^{***}
		(0.005)		(0.004)
I1	0.036^{**}	0.036^{**}	-0.062^{***}	-0.062^{***}
	(0.016)	(0.015)	(0.014)	(0.013)
I2	0.096^{***}	0.096^{***}	-0.106^{***}	-0.106^{***}
	(0.016)	(0.015)	(0.014)	(0.013)
I3	0.293^{***}	0.293^{***}	-0.296^{***}	-0.296^{***}
	(0.016)	(0.015)	(0.014)	(0.013)
R^2	0.193	0.245	0.233	0.322
Adj. R^2	0.191	0.243	0.231	0.320
RMSE	0.235	0.227	0.206	0.193
Num. obs.	1728	1728	1728	1728

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors in parentheses.

Table 3: OLS regressions of jar value estimates at the beginning of each period of trading a particular jar, using relative (Dev) and absolute ($AbsDev$) log deviations.

2.1.1. Estimate aggregation

We first analyze the best way to aggregate subjects' value estimates. We start by using (1) the average and (2) the median values of subjects' estimates. The three rows in Figure 3 illustrate estimate deviations over jars, periods and information levels, respectively, using both mean and median. Overall, we find that mean and median lead to very similar aggregates for subjects' estimates and that neither is clearly superior to the other.

The first row in Figure 3 shows aggregated estimate deviation for each of the four jars before information is received in the left-hand panel, and after information has been received in the right-hand panel. Clearly, the information provided improves the average estimation quality, as estimation errors decrease by on average about one half. The difference is highly significant for all jars (paired t-tests, $t(143) \leq -3.192, p < 0.0017$).

The second row shows aggregated estimate deviation, pooled over all jars, for each of the three periods that subjects trade the same jar. It provides (weak) evidence for some learning, as absolute estimate deviations decline slightly with experience.

The third row in Figure 3 shows subjects' estimates depending on information level. The right-hand panel documents that higher information levels correspond to lower estimate deviation, but that only I3 subjects come close to estimating jar values correctly. While I1 does not suffice to significantly improve the quality of estimates (I0 vs. I1, Welch two-sample $t(770.75) = -1.739, p = 0.0824$), the information contained in I2 lowers the estimation error

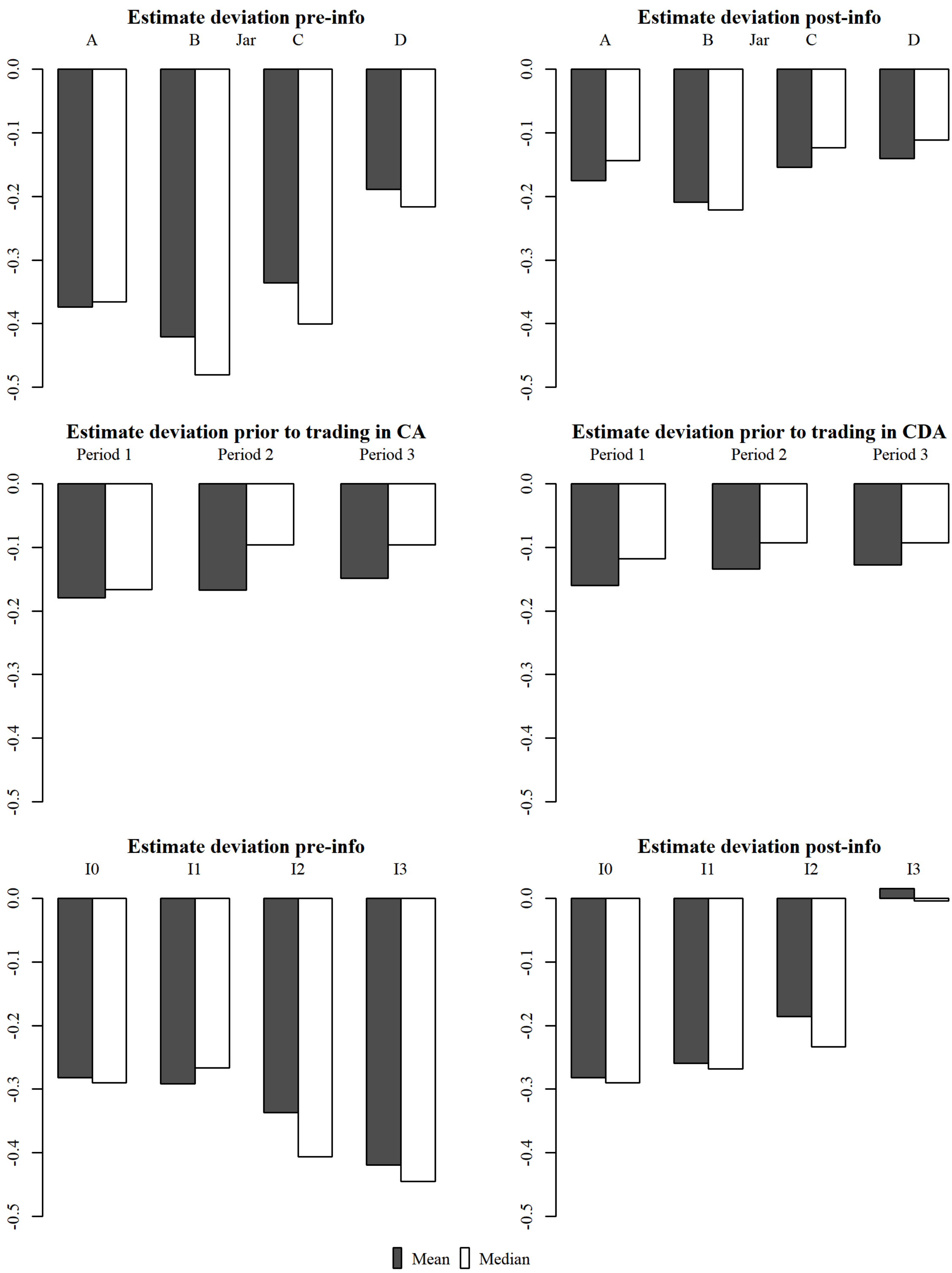


Figure 3: Mean and median log estimate deviation in units of BBV by jar, period and information level.

by about one third (I0 vs. I2, $t(638.67) = -5.066, p = 0.0000$). I3 is the only information level which allows essentially accurate estimates of the coin value in the jars (I0 vs. I3, $t(460.06) = -16.967, p = 0.0000$).

Interestingly, the left-hand panel documents a slight difference in estimate quality, with estimates worsening with increasing subsequently received information. The reason for this picture lies in the design of our experiment. Each subject receives each information level exactly for one jar. The order in which subjects receive the information levels is randomized. However, a subject who is currently estimating the value of the jar for which she will, after the estimate, receive I0 information, may in previous periods already have seen higher information levels. By comparing her estimate prior to receiving this high-quality information with her estimate after receiving the information, she may have learned to avoid large estimation errors prior to information provision. Conversely, a subject who is currently estimating in a period in which, after submitting the estimate, she receives I3 information, has never before received such high-level information and thus has not had the chance to learn from her mistakes in the past to the same degree. We provide evidence supporting this explanation in Appendix C.

Figure 4 displays estimates by InfoLevel and across periods. It documents several noteworthy patterns. First, I3 subjects' estimates lie close to, and are unbiased around, the BBV (Wilcoxon signed rank test cannot reject median differenc equal to 0, $V = 42798$, $p = 0.1705$). Second, all other estimate deviations are significantly different from zero (Wilcoxon signed rank tests for each information separately all yield $p = 0.0000$) and from I3 (Wilcoxon signed rank tests separately comparing I3 to the other information levels all yield $p = 0.0000$). Third, I2 estimates fall short of the true jar value by about 20% (5 euros), with no material learning either across periods within individual jars, or across jars. The quality of I2 estimates also seems to constitute an upper bound on the level of estimate precision subjects with lower information levels can achieve through experience. While the estimates for levels I0 and I1 are below those of I2 until around periods 4 to 6, they are relatively similar to I2 in the final quarter of the experiment, after subjects have gained some experience (Wilcoxon signed rank tests separately comparing I2 to I0 and to I1 yield $p > 0.2$). Thus, information dissemination seems to work to a degree that reflects the second highest information level, but not the highest.

Result 3. *Participants with the highest information level submit significantly more precise estimates than all other information levels. For lower information levels, experience can substitute information, such that experienced subjects with level I0 and level I1 information submit estimates of similar precision as subjects with level I2 information.*

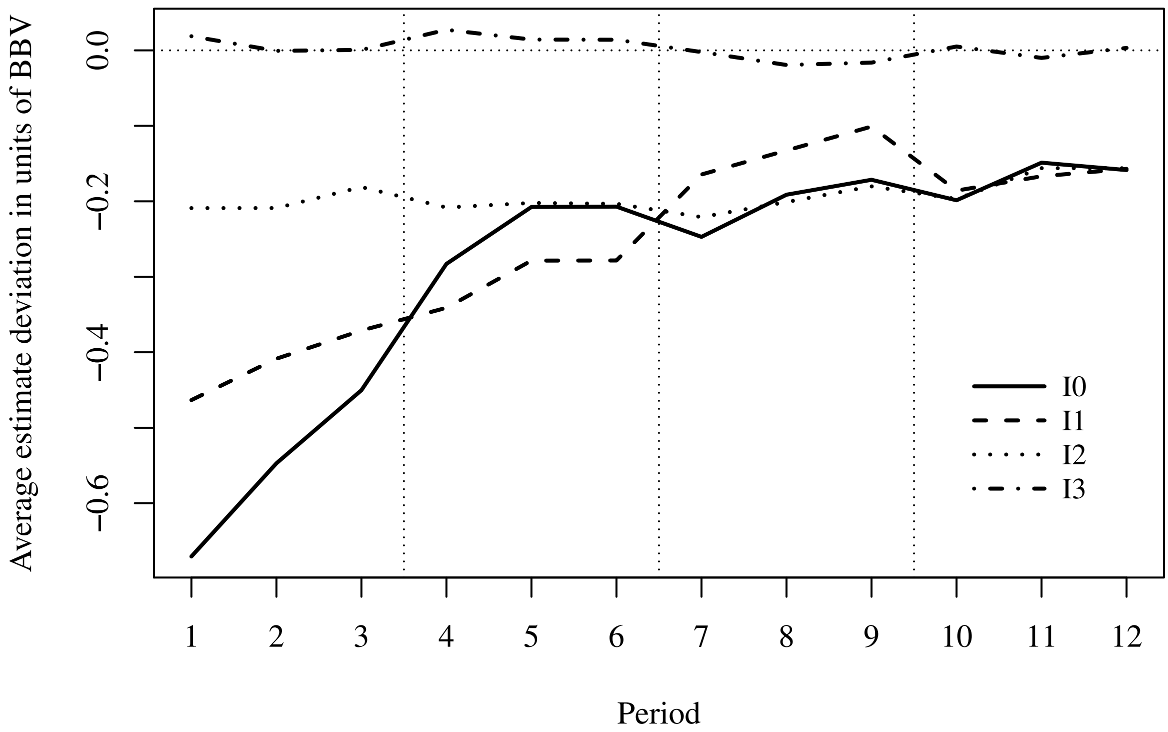


Figure 4: Mean estimate deviation in units of BBV across periods, by InfoLevel. Blocks of periods with different jars are distinguished by vertical dotted lines.

Figure 5 shows the average of Dev^{pre}, the deviation of subjects' estimates from the true jar value before receiving information about the jar value. The p-values stem from t-tests of the hypothesis of equal average deviations when comparing deviations at the beginning of different blocks of periods. The p-values lacking lines to clarify which blocks are being compared compare neighboring blocks (i.e., the block starting in period 1 vs. the block starting in period 4, 4 vs. 7, and 7 vs. 10). The figure suggests that subjects learn and improve their estimates between the first (period 1), second (period 4) and third (period 7) blocks, but not between the third and fourth (period 10).

A way to improve aggregate estimation quality may be to remove outliers before aggregating individual estimates. We find, however, that trimming subjects' estimates by removing a percentage of all observations from each tail of the estimate distribution has negligible effects on the quality of the mean (trimmed) estimate (see Appendix B for details).

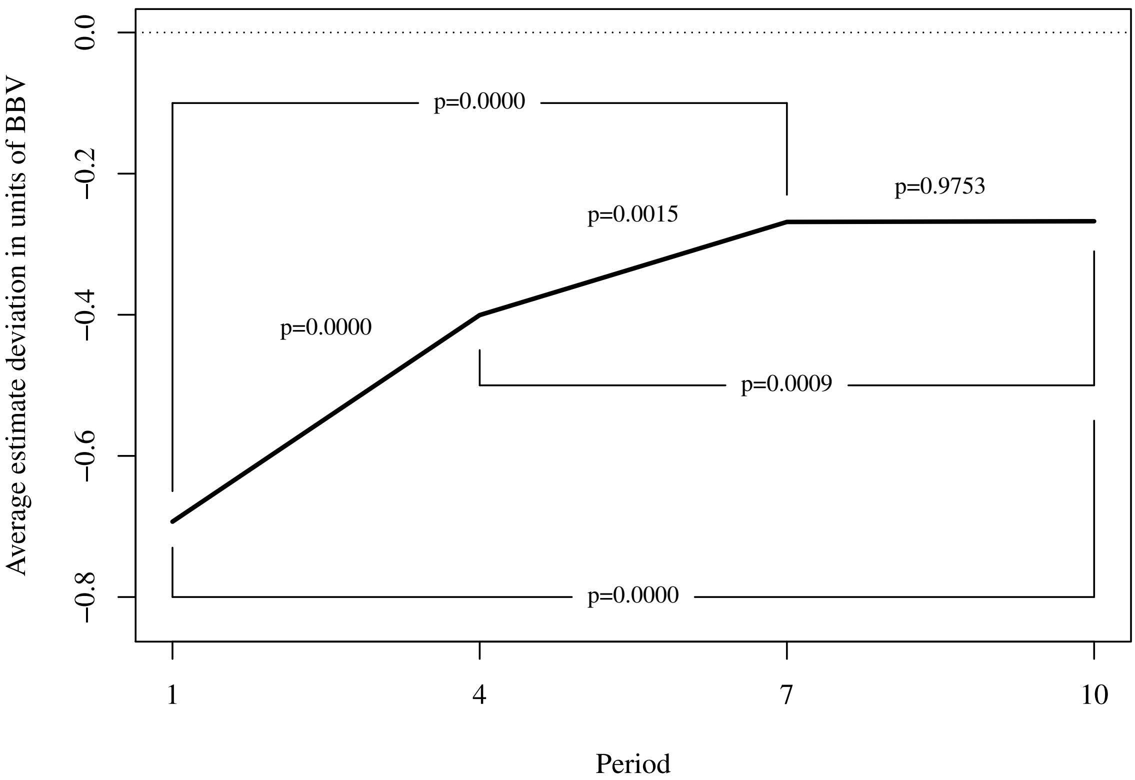


Figure 5: Average estimate deviation in units of BBV across blocks of periods and p-values from t-tests of equality.

2.1.2. Learning

We next turn to learning effects within a block of three periods when the same jar was traded, i.e., to whether subjects' estimates improve over these three periods. For each subject and jar, we define $\Delta AbsDev_t$ as the change in absolute log estimate deviation from one period to the next (period 1 to period 2 and period 2 to period 3 for trading the same jar), after subjects have received information, as shown in equation (2):

$$\Delta AbsDev_t \equiv AbsDev_{t+1}^{post} - AbsDev_t^{post} \tag{2}$$

We then regress $\Delta AbsDev_{t\in 1,2}$ on subjects' absolute log estimate deviation after they receive information in the fist period of trading a new jar ($AbsDev_{t=1}^{post}$), interacted with dummy variables for whether the learning took place over the first or over the second period. The coefficients of these regressors can thus be interpreted as the fraction of the initial absolute log estimate deviation that subjects correct due to learning from trading. We report the results in the first content column of Table 4.

	$\Delta AbsDev_t$	$\Delta AbsDev_t$	$\Delta AbsDev_t$
$AbsDev_1^{post} \times P1$	-0.218^{***}	-0.225^{***}	-0.260^{***}
	(0.013)	(0.015)	(0.024)
$AbsDev_1^{post} \times P2$	-0.068^{***}	-0.075^{***}	-0.111^{***}
	(0.013)	(0.015)	(0.024)
CA		0.009	0.004
		(0.006)	(0.007)
Ability		0.002	-0.003
		(0.002)	(0.004)
Jar A			0.012
			(0.011)
Jar B			0.009
			(0.012)
Jar C			0.013
			(0.012)
Jar D			0.009
			(0.011)
Female			0.013^{*}
			(0.007)
R^2	0.205	0.209	0.214
Adj. R^2	0.204	0.206	0.208
RMSE	0.114	0.114	0.113
Num. obs.	1152	1152	1152

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors in parentheses.

Table 4: OLS regressions of $\Delta AbsDev_t$ on initial absolute log estimate deviation after information revelation, interacted with period dummy variables (but no intercept) and other regressors.

The coefficients document that the estimate after the first trading period is about 22 percentage points closer to BBV than the estimate before the first trading period, and that the second trading period yields another improvement of about 7 percentage points. In the second column we add a dummy variable for the call auction sessions and a measure of subjects' estimation ability. We define the latter as:

$$Ability \equiv ln\left(\frac{AbsDev_{t=1}^{post}}{\overline{AbsDev_{t=1}^{post}}}\right), \tag{3}$$

where $AbsDev_{t=1}^{post}$ is the subject's absolute log estimate deviation for a particular jar, after receiving information and before trading in the first period of trading this jar, and

$\overline{AbsDev^{post}_{t=1}}$ is the average of the same variable over all subjects. *Ability* thus is the (log) percentage outperformance of the subject's estimate relative to the mean estimate by all subjects. Adding these two variables to our regression model does not affect the discovered learning effects. In the third column, we also add dummy variables for the four different jars, as well as gender, which shows that learning does not differ much (by approximately 1%) between female and male subjects. Overall, none of the more complex specifications materially improves the explanatory power (i.e., R^2) of the first regression model.

Result 4. *Participants' estimates of jar value improve by about 22 percentage points over the first period of trading, and by another about 7 percentage points over the second. Neither the trading mechanism nor participants' estimation ability or gender materially moderates this learning process.*

2.2. Market-level results

We now turn to the comparison of the two market mechanisms Call Auction vs. Continuous Double Auction. Figure 6 plots mean transaction price deviations over time, measured in periods. The top panel displays data from the CDA, the bottom from the CA treatment. If subjects learned across jars and over time, we would expect a monotonous upward trend. There is no clear evidence for such learning, except for CA in the first two periods.

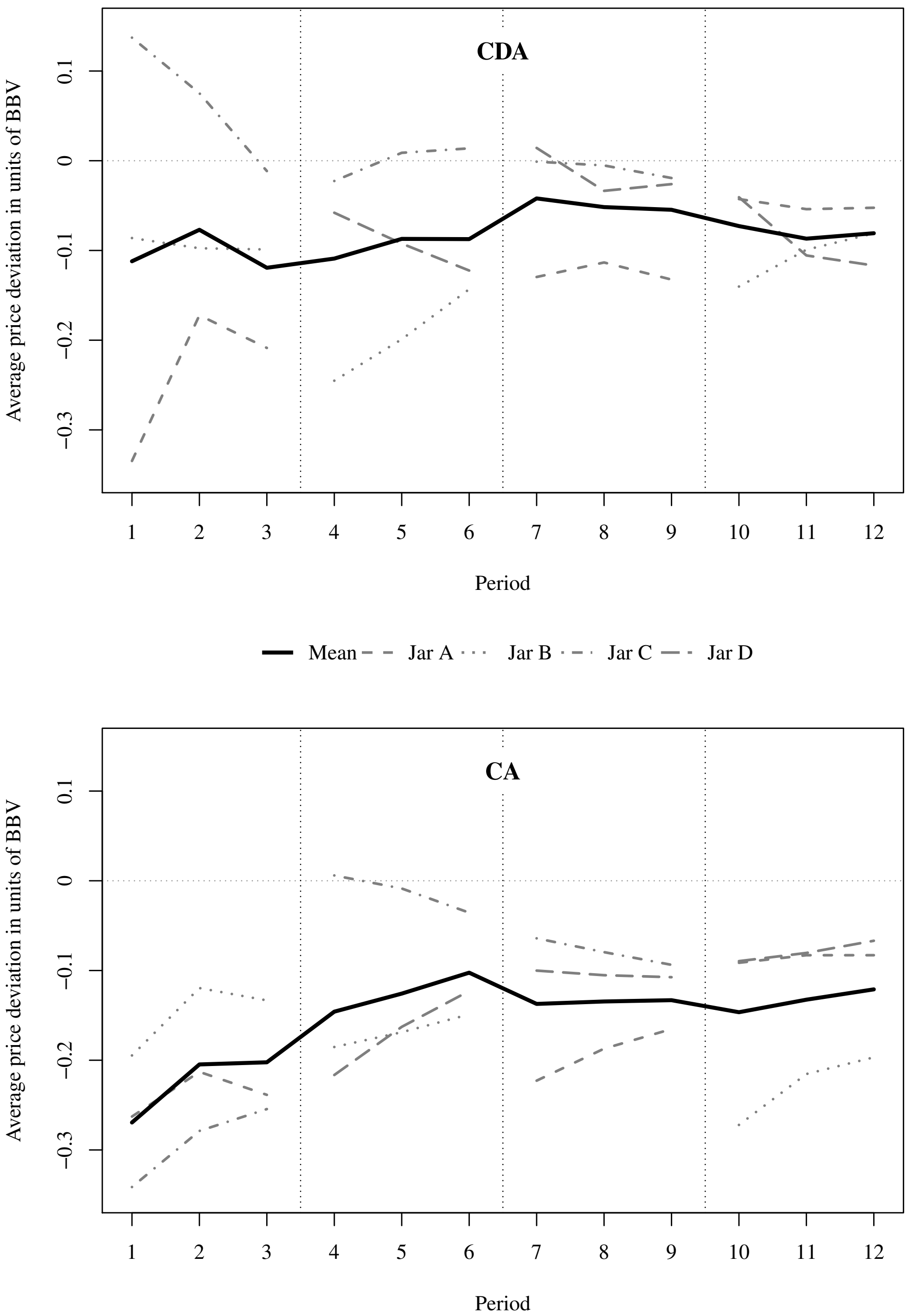


Figure 6: Deviation of mean transaction prices from BBV (in units of BBV) over trading periods. Continuous double auction data in top, call auction data in bottom panel. The solid black lines give the overall average, while the dashed gray lines give deviations for individual jars.

Figure 7 plots individual jars' and the mean's price development over the three periods each jar is traded for, separately for CA and CDA. In neither treatment do we observe learning across periods within a jar, but we see that prices deviate less within CDA than CA.

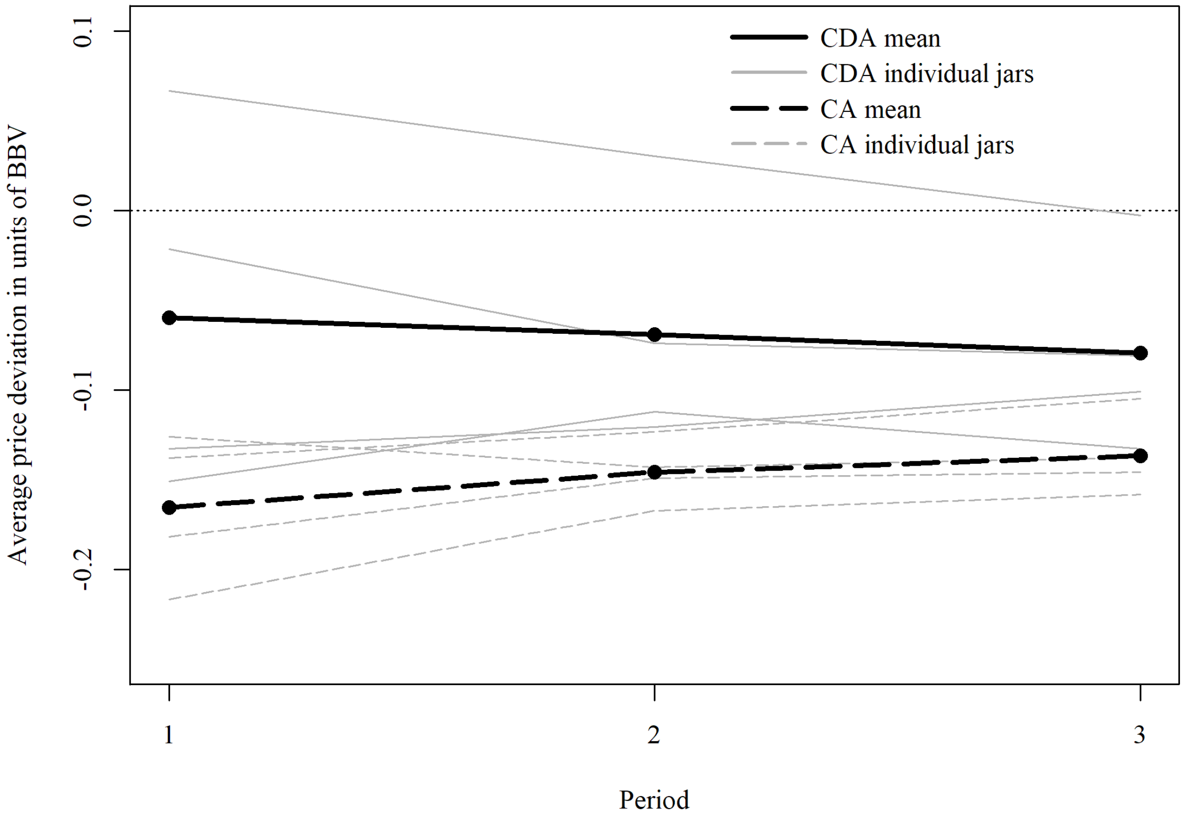


Figure 7: Mean jar price deviations from BBV (in units of BBV) in continuous double auction and call auction markets.

Figure 8 plots the standard deviation of transaction price deviations from BBV over the trading periods for each jar. It shows a downward trend, indicating harmonization of subjects' estimates in light of their observations in the market. It also contains a line showing estimate deviations, which follow a similar pattern, yet remain at a higher level. This documents that market prices offer more precise predictions of jar value than individual estimates.

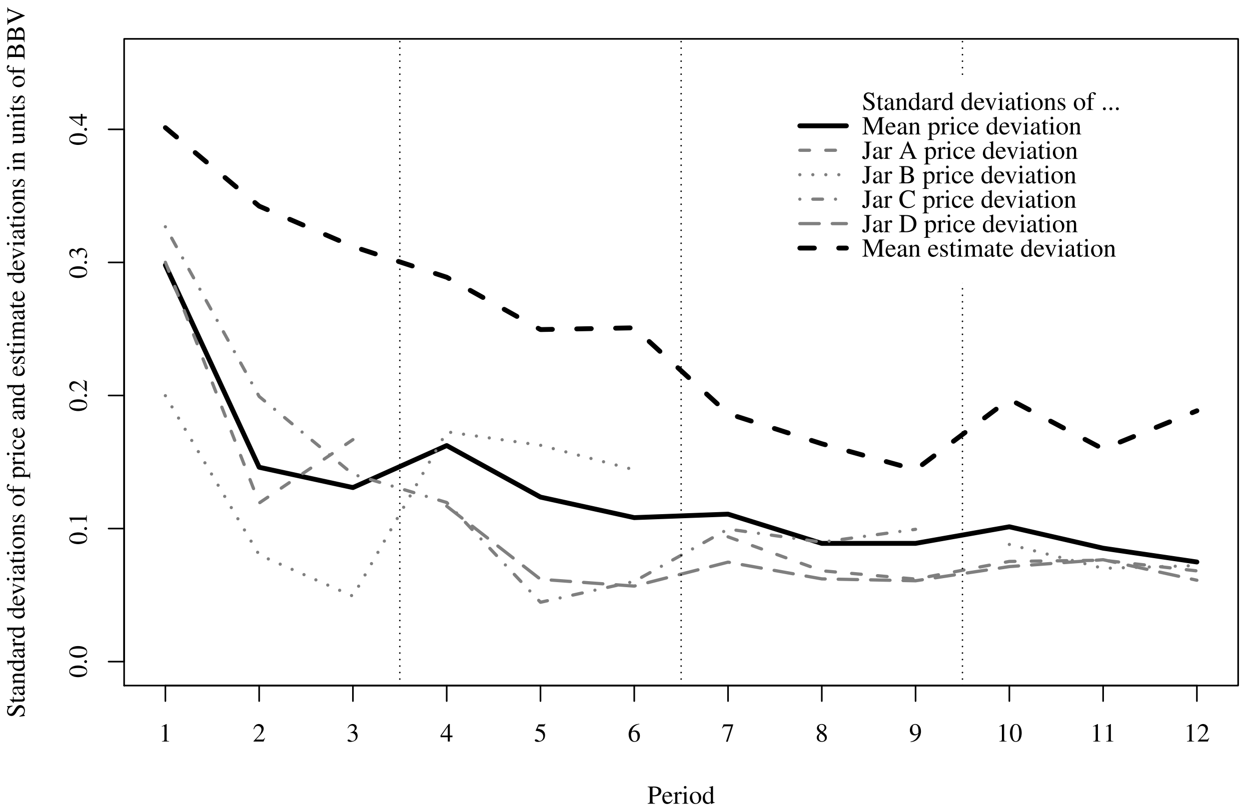


Figure 8: Standard deviation of the deviations of transaction prices from BBV (in units of BBV) over trading periods.

2.3. Subject earnings

Table 5 reports a regression analyzing the percentage change in subjects' wealth, $\Delta Wealth$:

$$\Delta Wealth = \left(\frac{FinalCash + FinalAssets \cdot BBV}{InitialCash + InitialAssets \cdot BBV} - 1 \right) \cdot 100 \tag{4}$$

We multiply by 100 to scale up the regression coefficients for better legibility. Regarding the regressors, we again use $AbsDev^{post}$, the log deviation from BBV of subjects' estimates of the jar value after information provision in period 1 and at the beginning of the period in periods 2 and 3 of trading each jar. This variable constitutes an inverse measure of subjects' precision in estimating jar values, incorporating the information provided by the experimenter and the information gathered by observing (and participating in) trading. $AbsDev^{pre}$, a similar measure as $AbsDev^{post}$, is calculated only once for each jar (when subjects first estimate the jar value) and is kept constant within the three periods of trading of each jar. It thus measures a subject's estimation ability, bar explicit information about the jar value (from information levels I1 through I3) and bar learning effects from trading. *Female* is a dummy variable for subject gender (using the obvious coding), and I1 through I3 are dummy variables denoting a subject's information level in any given period.

	$\Delta Wealth$	$\Delta Wealth$	$\Delta Wealth$
$Intercept$	2.295^{***}	1.541^{***}	1.444^{***}
	(0.290)	(0.317)	(0.320)
$AbsDev^{post}$	-6.837^{***}	-5.140^{***}	-6.395^{***}
	(0.766)	(0.810)	(0.894)
$AbsDev^{pre}$			1.313^{***}
			(0.460)
$Female$	-1.449^{***}	-1.547^{***}	-1.574^{***}
	(0.352)	(0.353)	(0.354)
$I1$		−0.177	−0.250
		(0.270)	(0.272)
$I2$		−0.042	−0.319
		(0.259)	(0.276)
$I3$		1.981^{***}	1.481^{***}
		(0.384)	(0.396)
R^2	0.171	0.204	0.211
Adj. R^2	0.170	0.202	0.208
RMSE	4.038	3.961	3.945
Num. obs.	1728	1728	1728

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors adjusted for 144 clusters at the subject level in parentheses.

Table 5: Regressions of percentage change in subject's wealth (evaluated at BBV) over the course of a period.

The highly significant coefficients for $AbsDev^{post}$ in Table 5 show that subjects with greater $AbsDev^{post}$, and thus a relatively low-quality estimate of BBV, end up with lower wealth than subjects who are more successful in estimating jar value. When we add information level dummy variables, it is interesting to see that the coefficient for I3 is significant even after controlling for $AbsDev^{post}$. This is caused by the non-linearity of the relationship between information level and final profits. I3 subjects earn 3.29% more than I0-I2 subjects on average (see Figure A.9 in the appendix). The lower earnings of female subjects may stem from the fact that female subjects end each period holding on average 4.47 jars, while male subjects average 5.66. Remember that jars are on average undervalued. Male traders thus tend to be net buyers of jars, paying less than BBV, but earning BBV for each jar bought in this way.[9] The rightmost column finally adds $AbsDev^{pre}$ to explore whether sub-

[9] Including a measure of subjects' risk-preferences (following Dohmen et al., 2011) does not materially affect these findings.

jects' innate estimation ability ("pure" ability, without information or prior experience with the jar being traded) helps them outperform. This seems not to be the case. The effect of the positive coefficient of $AbsDev^{pre}$ is in fact nearly entirely compensated by the larger (in absolute terms) negative coefficient of $AbsDev^{post}$ in this model.[10]

Table 6 shows the average log deviations from BBV when using different mechanisms to aggregate predictions of the true jar value. Period refers to the period within a block of trading a single jar and the table lists averages across all jars. The aggregation methods summarized in the table are the midpoint of the bid-ask spread at the end of the trading period in the CDA (CDA mid), the closing, median and average prices in the CDA (CDA close, CDA mean, CDA median), the median and mean jar value estimates after receiving information (Est. median, Est. mean), and the price in the CA (CA). In this and the following paragraphs, we focus on the first period, as in many situations outside of the lab where good estimates of an unknown quantity are required, it is impractical to let subjects trade/estimate for multiple periods. The table shows that the absolute deviation is lowest when using the midpoint of the bid-ask spread in the continuous double auction.

Table 7 displays p-values when comparing the average deviations resulting from the use of the aggregation mechanisms listed in Table 6. Table 7 uses only data from the first period within a block. Furthermore, the rows and columns in the table are sorted by increasing absolute deviation in the first period.[11,12] The data documents that predictions based on CDA data clearly outperform the CA and mean and median estimates. The differences within the CDA are not significant. When relying only on estimates, the median outperforms the mean as an aggregation mechanism.

Result 5. *CDA prices are closest to true jar values. CA prices and individual estimates perform significantly worse. Limiting the analysis to the simple estimates, aggregation using the median outperforms the mean.*

[10] However, we caution against placing too much weight on this final column, since the results may to some degree be driven by collinearity. The Pearson correlation between Dev^{pre} and Dev^{post} is 0.527 (using only the first periods of trading a jar, to isolate the pure effect of the information levels, without influence from trading experience).

[11] We discard 7 out of our 1746 (0.4%) offers outstanding at the end of a period because they have prices of 1000 or above, which are likely not meant to be serious and even if so, would bias our results without adding valuable insights. Furthermore, we are most interested in Period 1 data and these outliers only ocur in Period 3 data.

[12] We end up with 4 out of 108 (3.7%) periods where we cannot calculate a bid-ask midpoint due to missing best bid or best ask values.

Period	CDA mid	CDA close	CDA median	CDA mean	Est. median	CA	Est. mean
1	-4.314	-5.092	-6.067	-7.443	-16.114	-16.651	-20.875
2	-6.250	-8.813	-6.823	-7.569	-13.909	-13.671	-17.907
3	-4.031	-8.245	-7.622	-8.204	-11.888	-13.216	-16.338
All	-5.177	-7.384	-6.837	-7.739	-13.970	-14.513	-18.373

Table 6: Log Deviation from BBV (in %) resulting from different aggregation mechanisms. Columns are sorted in ascending order by absolute deviation in the first period of each block.

	CDA mid	CDA close	CDA median	CDA mean	Est. median	CA	Est. mean
CDA mid		0.9777	0.8734	0.8692	0.0013	0.0008	0.0000
CDA close	0.9777		0.8255	0.8264	0.0001	0.0000	0.0000
CDA median	0.8734	0.8255		0.9862	0.0005	0.0003	0.0000
CDA mean	0.8692	0.8264	0.9862		0.0016	0.0009	0.0000
Est. median	0.0013	0.0001	0.0005	0.0016		0.8445	0.0453
CA	0.0008	0.0000	0.0003	0.0009	0.8445		0.0704
Est. mean	0.0000	0.0000	0.0000	0.0000	0.0453	0.0704	

Table 7: p-values from pairwise t-tests comparing the log deviations from BBV resulting from different aggregation mechanisms using only data from the first period within a block.

3. Conclusion

The present paper reports on a lab experiment studying different mechanisms for aggregating dispersed information. We use the controlled conditions of the experimental laboratory to compare the quality of estimates of an unknown quantity stemming from (1) subjects' estimates, (2) continuous double auction, or CDA, market prices and (3) call auction, or CA, market prices. We find that prices in a CDA constitute the best aggregation mechanism, characterized by the lowest prediction error.

The outperformance of the CDA is in line with the recent successes of prediction markets and it supports the use of market mechanisms for information aggregation. However, while the CDA outperforms the other aggregation mechanisms, it is at the same time the most complex of the mechanisms employed in our study. A simple estimate (even with incentivization) can be elicited very quickly and using any medium (verbal, paper, online). Conducting a continuous auction market requires considerable investment both in terms of the solicitor's infrastructure and participants' time. Furthermore, the possibility of observing no or only few trades – and the potential cost of guarding against this eventuality – should also be taken into consideration. Whether these additional monetary and non-monetary costs are justified cannot be answered in general. Instead, this question needs to be answered on a case-by-case basis, weighing the CDA's greater costs against the benefits that can be derived from the greater forecast precision it offers.

We hope that in addition to our results per se, our methodology may also help future researchers. Having subjects handle and estimate the value of multiple types of coins in a jar and providing them with varying levels of information about the coins in the jar allows for studying both ambiguity and risk, and for implementing a number of valuable treatment variations. For future research, it would for example be interesting to apply the approach of Budescu and Chen (2014) to our setting. They compare individual subjects' performance with the group and then let only above-average subjects (i.e., 'experts') interact with each other in a second round.

Appendix A. Additional figures and tables

Figure A.9 plots percentage change in subject wealth over the four information levels. The advantage of obtaining information level I3 is evident. None of the pairwise differences between $\Delta Wealth$ among I0, I1 and I2 are significant (Wilcoxon rank sum test, $p > 0.6$ for all comparisons), while all differences are significant when comparing to I3 ($p = 0.0000$ for all comparisons).

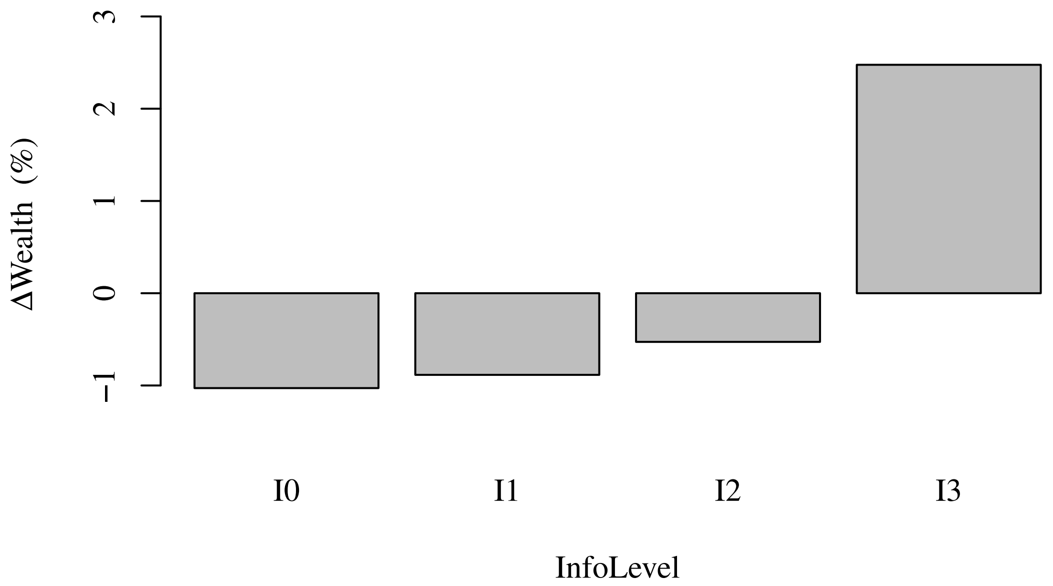


Figure A.9: Percentage change in wealth per subject, over information levels; means over all sessions.

Table A.8 lists the cash endowments subjects received at the beginning of each period of trading a particular jar.

Table A.8: Cash endowments

jar	A	B	C	D	Average
Subject 1	221	316	259	204	250
Subject 2	238	306	215	241	250
Subject 3	232	270	276	222	250
Subject 4	282	287	237	194	250
Subject 5	261	300	221	218	250
Subject 6	213	307	232	248	250
Subject 7	231	267	250	252	250
Subject 8	282	288	220	210	250
Cash-Asset-Ratio	1.99918	2.00017	1.99958	2.00112	

Table A.9 repeats the analysis of Table 2, yet includes subject dummy variables (albeit, to conserve space, not in the output) to give a better indication of the explanatory power of the models ($R^2 > 0.5$ throughout) when accounting for subject heterogeneity.

	Dev^{pre}	$AbsDev^{pre}$	Dev^{post}	Dev^{post}	$AbsDev^{post}$	$AbsDev^{post}$
Intercept	-0.579^{***}	0.736^{***}	-0.592^{***}	-0.743^{***}	0.616^{***}	0.793^{***}
	(0.207)	(0.180)	(0.114)	(0.108)	(0.103)	(0.093)
JarNo	0.135^{***}	-0.147^{***}		0.060^{***}		-0.071^{***}
	(0.012)	(0.010)		(0.008)		(0.007)
InfoLevel1			0.061^{**}	0.061^{**}	-0.100^{***}	-0.100^{***}
			(0.027)	(0.025)	(0.024)	(0.021)
InfoLevel2			0.141^{***}	0.141^{***}	-0.155^{***}	-0.155^{***}
			(0.027)	(0.025)	(0.024)	(0.021)
InfoLevel3			0.362^{***}	0.362^{***}	-0.361^{***}	-0.361^{***}
			(0.027)	(0.025)	(0.024)	(0.021)
R^2	0.551	0.552	0.554	0.607	0.530	0.626
Adj. R^2	0.400	0.400	0.402	0.473	0.370	0.497
RMSE	0.314	0.273	0.225	0.211	0.204	0.182
Num. obs.	576	576	576	576	576	576

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors in parentheses.

Table A.9: OLS regressions of jar value estimates, before (*pre*) and after (*post*) information provision and using relative (*Dev*) and absolute (*AbsDev*) log deviations. Dummy variables for individual subjects were included in the estimations but omitted in the output.

Appendix B. Analysis of trimmed mean estimates

For Figure B.10, we calculate mean estimate deviations after removing outliers from the data. Specifically, we trim subjects' estimates by removing a percentage of all observations from each tail of the estimate distribution. We then calculate the mean estimate deviation for the trimmed data and plot it over different trim levels.[13] The shading in the background indicates ±1 standard deviation around the trimmed mean estimates. The figure suggests that, for our data, the effect of removing outlying observations before averaging has negligible effects on the mean estimate, particularly in light of the wide standard deviation bands.

[13] Note that, at the extremes, a two-tailed trim percentage of zero implies no removal of outliers, while a percentage of 50 implies using the median estimate only.

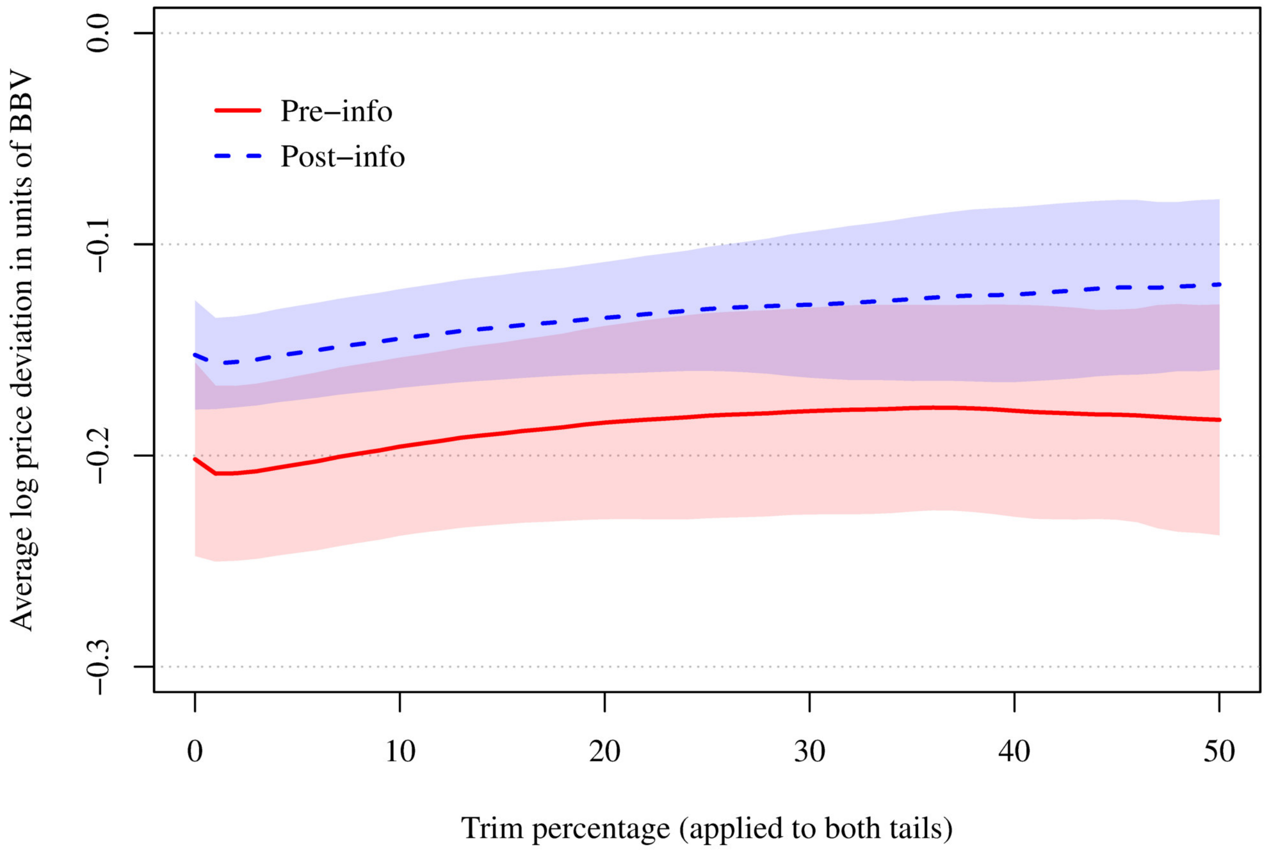


Figure B.10: Log deviation of trimmed mean estimates from BBV (in units of BBV) over different percentages of trimming. Trimming entails removing a set percentage of all observations from each tail of the distribution prior to calculating the mean of the remaining observations. At the two extremes, a percentage of zero implies no removal of outliers, while a percentage of 50 implies using the median estimate only. The shaded areas depict ±1 standard deviation (calculated over the four jar means).

Appendix C. Analysis of estimates before information revelation

Section 2.1.1 mentions (and Figure 3 shows) that subjects' jar value estimates before receiving high-quality information (e.g., I3) are worse than their estimates before receiving low-quality information (e.g., I0). As Table C.10 shows, this is due to subjects learning from the information received in previous blocks of trading different jars. The Table uses data from periods 4, 7 and 10 (i.e., the first periods of trading each jar after the first) and analyzes as dependent variable subjects' absolute log estimate deviation before receiving information. This variable, $AbsDev_t^{pre}$, is regressed on (1) dummy variables for the four jars, (2) dummy variables for whether the subject's history contains periods with information levels I1, I2 or I3, (3) a dummy variable for whether the observations stem from a call auction market, and (4) a dummy variable for whether the subject in question was female. Note that the regressions in Table C.10 forego the use of an intercept in favor of using all four jar dummy variables.

	$AbsDev_t^{pre}$	$AbsDev_t^{pre}$	$AbsDev_t^{pre}$	$AbsDev_t^{pre}$
Jar A	0.852^{***}	0.753^{***}	0.745^{***}	0.721^{***}
	(0.041)	(0.046)	(0.047)	(0.048)
Jar B	0.861^{***}	0.763^{***}	0.755^{***}	0.731^{***}
	(0.038)	(0.043)	(0.045)	(0.046)
Jar C	0.806^{***}	0.707^{***}	0.699^{***}	0.675^{***}
	(0.035)	(0.041)	(0.043)	(0.044)
Jar D	0.749^{***}	0.650^{***}	0.642^{***}	0.618^{***}
	(0.045)	(0.048)	(0.050)	(0.051)
JarNo	-0.145^{***}	-0.046^{*}	-0.046^{*}	-0.053^{**}
	(0.012)	(0.026)	(0.026)	(0.026)
$I1 \in$ history		-0.031	-0.031	-0.024
		(0.037)	(0.037)	(0.037)
$I2 \in$ history		-0.086^{**}	-0.086^{**}	-0.077^{*}
		(0.040)	(0.040)	(0.040)
$I3 \in$ history		-0.278^{***}	-0.278^{***}	-0.269^{***}
		(0.037)	(0.037)	(0.037)
CA			0.016	0.009
			(0.024)	(0.024)
Female				0.061^{**}
				(0.025)
R^2	0.716	0.745	0.746	0.748
Adj. R^2	0.714	0.742	0.741	0.744
RMSE	0.307	0.292	0.292	0.291
Num. obs.	576	576	576	576

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors in parentheses.

Table C.10: OLS regressions of $|Dev_t^{pre}|$ on dummy variables for the jars, for having previously received information levels I1 through I3, for the trading mechanism and subject gender.

The first content column in Table C.10 documents that subjects learn across jars. The more jars they have previously traded, the lower their absolute log estimate deviation. Content column 2 refines this finding by showing that much of the learning stems from having previously received high-quality information about a jar's value. Subjects who have, for previous jars, received I3 information, submit jar value estimates which are about 30% more accurate than subjects who have yet to experience I3 information. Content column 3 shows that the trading mechanism does not affect the dependent variable, while the final column suggests that male subjects' estimates are about 6% more accurate than female subjects' after controlling for all other variables. The R^2 values indicate an excellent model fit.

Appendix D. Analysis of estimation ability on market prices

Table D.11 regresses the absolute log deviation of period mean and median price, closing price and closing bid-ask spread in the CDA as well as the absolute log deviation of the CA price on the period average absolute log deviation of subjects' estimates before information revelation, controlling for individual jar effects. The regressions use only the first period of trading for each jar in order to isolate, as far as possible, the pure ability effect from learning across periods. The table shows that estimate quality, i.e., subjects' average ability in forming estimates of jar value, improves market prices in the CA, but not in the CDA. Table D.12 repeates the analysis with the even more stringent specification that it uses only the very first period in a session. It confirms the findings from Table D.11.

	CDA mean	CDA median	CDA close	CDA mid	CA
Intercept	0.092	0.097^{*}	0.110^{**}	0.135^{*}	0.029
	(0.063)	(0.049)	(0.046)	(0.066)	(0.036)
$AbsDev^{pre}$	0.152	0.117	0.014	−0.016	0.342^{***}
	(0.107)	(0.084)	(0.079)	(0.113)	(0.056)
Jar 2	−0.002	−0.006	0.011	0.054	0.014
	(0.062)	(0.049)	(0.046)	(0.066)	(0.036)
Jar 3	−0.051	−0.049	−0.029	−0.050	−0.048
	(0.062)	(0.049)	(0.046)	(0.065)	(0.036)
Jar 4	−0.056	−0.050	−0.065	−0.067	0.008
	(0.066)	(0.051)	(0.048)	(0.069)	(0.038)
R^2	0.141	0.150	0.105	0.115	0.585
Adj. R^2	0.030	0.040	−0.010	0.001	0.531
RMSE	0.132	0.103	0.097	0.139	0.077
Num. obs.	36	36	36	36	36

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors in parentheses.

Table D.11: OLS regressions of period averages of absolute log market price deviations on absolute log deviations of estimates before information revelation and jar dummies, using only the first periods of trading each jar.

	CDA mean	CDA median	CDA close	CDA mid	CA
Intercept	0.335	0.291	0.401^{**}	0.184	−0.145
	(0.246)	(0.152)	(0.149)	(0.130)	(0.159)
$AbsDev^{pre}$	−0.063	−0.060	−0.299	−0.058	0.530^{**}
	(0.285)	(0.176)	(0.173)	(0.150)	(0.203)
Jar 2	−0.209	−0.184	−0.154	−0.118	0.020
	(0.184)	(0.114)	(0.112)	(0.097)	(0.070)
Jar 3	−0.112	−0.095	−0.039	−0.010	0.001
	(0.180)	(0.111)	(0.109)	(0.095)	(0.070)
R^2	0.206	0.344	0.446	0.273	0.726
Adj. R^2	−0.270	−0.049	0.114	−0.164	0.561
RMSE	0.215	0.133	0.130	0.113	0.075
Num. obs.	9	9	9	9	9

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors in parentheses.

Table D.12: OLS regressions of period averages of absolute log market price deviations on absolute log deviations of estimates before information revelation and jar dummies, using only the first periods of trading per session.

References

Berg, J., Forsythe, R., Nelson, F., Rietz, T., 2008a. Results from a dozen years of election futures markets research. In: Plott, C. R., Smith, V. L. (Eds.), Handbook of Experimental Economics Results. Elsevier, Ch. 80, pp. 742–751.

Berg, J. E., Nelson, F. D., Rietz, T. A., 2008b. Prediction market accuracy in the long run. International Journal of Forecasting 24 (2), 283–298.

Berg, J. E., Rietz, T. A., 2003. Prediction markets as decision support systems. Information Systems Frontiers 5 (1), 79–93.

Bock, O., Baetge, I., Nicklisch, A., 2014. hroot: Hamburg registration and organization online tool. European Economic Review 71, 117–120.

Brown, W. O., Sauer, R. D., 1993. Fundamentals or noise? evidence from the professional basketball betting market. The Journal of Finance 48 (4), 1193–1209.

Bruce, R. S., 1935. Group judgments in the fields of lifted weights and visual discrimination. The Journal of Psychology 1 (1), 117–121.

Budescu, D. V., Chen, E., 2014. Identifying expertise to extract the wisdom of crowds. Management Science 61 (2), 267–280.

Budescu, D. V., Yu, H.-T., 2006. To bayes or not to bayes? a comparison of two classes of models of information aggregation. Decision Analysis 3(3), 145–162.

Clemen, R. T., 1989. Combining forecasts: A review and annotated bibliography. International journal of forecasting 5 (4), 559–583.

Davis-Stober, C. P., Budescu, D. V., Dana, J., Broomell, S. B., 2014. When is a crowd wise? Decision 1 (2), 79–101.

Dohmen, T. J., Falk, A., Huffman, D., Schupp, J., Sunde, U., Wagner, G., 2011. Individual risk attitudes: Measurement, determinants, and behavioral consequences. Journal of the European Economic Association 9 (3), 522–550.

Fischbacher, U., 2007. z-Tree: Zurich Toolbox for Ready-made Economic Experiments. Experimental Economics 10 (2), 171–178.

Gjerstad, S., Oct. 2004. Risk aversion, beliefs, and prediction market equilibrium, university of Arizona, Department of Economics Working Paper 04-17.

Gordon, K. H., 1924. Group judgments in the field of lifted weights. Journal of Experimental Psychology 7 (5), 398–400.

Gruca, T., Berg, J. E., Cipriano, M., 2003. The effect of electronic markets on forecasts of new product success. Information Systems Frontiers 5 (1), 95–105.

Herzog, S., Hertwig, R., 2011. The wisdom of ignorant crowds: Predicting sport outcomes by mere recognition. Judgment and Decision Making 6 (1), 58.

Jensen, J. L. W. V., 1906. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. Acta Mathematica 30 (1), 175–193.

Kirchler, M., Huber, J., Stöckl, T., 2012. Thar she bursts - reducing confusion reduces bubbles. The American Economic Review 102 (2), 865–883.

Larrick, R. P., Mannes, A. E., Soll, J. B., 2012. The social psychology of the wisdom of crowds. Psychology Press, Ch. 13, pp. 227–242.

Lichtendahl Jr, K., Grushka-Cockayne, Y., Pfeifer, P., 2013. The wisdom of competitive crowds. Operations Research 61 (6), 1383–1398.

Lorenz, J., Rauhut, H., Schweitzer, F., Helbing, D., 2011. How social influence can undermine the wisdom of crowd effect. PNAS 108 (22), 9020–9025.

Lorge, Irving; Fox, D. D. J. B. M., 1958. A survey of studies contrasting the quality of group performance and individual performance. Psychological Bulletin 55 (6), 337–372.

Malone, T., Laubacher, R., Dellarocas, C., 2009. Harnessing crowds: Mapping the genome of collective intelligence. SSRN working paper 1381502.

Manski, C. F., 2006. Interpreting the predictions of prediction markets. Economics Letters 91 (3), 425–429, revision of NBER Working Paper 10359, March 2004.

Noussair, C. N., Tucker, S., 2016. Cash inflows and bubbles in asset markets with constant fundamental values. Economic Inquiry 54 (3).

Ottoviani, M., Sørensen, P. N., 2009. Aggregation of information and beliefs: Asset pricing lessons from prediction markets. SSRN discussion paper 1447369.

Palan, S., 2015. GIMS - software for asset market experiments. Journal of Behavioral and Experimental Finance 5, 1–14.

Polgreen, P. M., Nelson, F. D., Neumann, G. R., 2007. Use of prediction markets to forecast infectious disease activity. Clinical Infectious Diseases 44 (2), 272–279.

Powell, O., 2016. Numeraire independence and the measurement of mispricing in experimental asset markets. Journal of Behavioral and Experimental Finance 9, 56–62.

Sauer, R. D., 1998. The economics of wagering markets. Journal of Economic Literature 36 (4), 2021–2064.

Soll, J., Richard, B., Larrick, P., 2009. Strategies for revising judgment: how (and how well) people use others' opinions. Journal of Experimental Psychology: Learning, Memory and Cognition 35 (3), 780–805.

Surowiecki, J., 2004. The Wisdom of Crowds. Abacus, London.

vanRooij, M., Lusardi, A., Alessie, R., 2011. Financial literacy and stock market participation. Journal of Financial Economics 101, 449–472.

Wolfers, J., Zitzewitz, E., 2006. Interpreting prediction market prices as probabilities. SSRN discussion paper 898597.