2015

Revisiting Information Aggregation in Asset Markets: Reflective Learning & Market Efficiency

Brice Corgnet
Chapman University, corgnet@chapman.edu

Mark DeSantis
Chapman University, desantis@chapman.edu

David Porter
Chapman University, dporter@chapman.edu

Revisiting Information Aggregation in Asset Markets:
Reflective Learning & Market Efficiency

Brice Corgnet, Mark DeSantis, David Porter[†]
Economic Science Institute & Argyros School of Business and Economics, Chapman University

Abstract

The ability of markets to aggregate disperse information leading to prices that reflect the fundamental value of an asset is key to assessing the often-debated efficiency of markets. We study information aggregation in the experimental environment originally created by Plott and Sunder (1988). Contrary to the current belief, we find that markets do not aggregate information. The model that best describes our data, as well as data on information aggregation subsequent to Plott and Sunder (1988), is prior information (Lintner, 1969). That is, traders use their private information but fail to use market prices to infer other traders' information. We argue that reflecting on asset prices to infer others' information requires specific skills related to the concept of cognitive reflection. We develop a learning model in which only a subset of the traders possess this reflective capacity. We show, using both simulations and laboratory experiments, that information aggregation can only be achieved when the market is populated by highly reflective traders and this high level of cognitive reflection is commonly known to all of the traders.

Keywords: Information aggregation, market efficiency, experimental asset markets, behavioral finance.
JEL CODES: C92, G02, G14.

[†] Email of corresponding author: corgnet@chapman.edu

1. Information Aggregation in Experimental Asset Markets

The extent to which markets aggregate disperse information has been at the center of the heated debate on market efficiency in Finance (Fama, 1970; Fama, 1998; Shleiffer, 2000; Thaler, 2005; Shiller, 2015). The empirical assessment of market efficiency is a daunting task because not only is it impossible for researchers to observe traders' private information, but it is also impossible to test market efficiency independent of a specific equilibrium model for asset prices. As noted by Fama (1991, pages 1575-1576):

> *"Ambiguity about information and trading costs is not, however, the main obstacle to inferences about market efficiency. The joint-hypothesis problem is more serious. Thus, market efficiency per se is not testable. It must be tested jointly with some model of equilibrium, an asset-pricing model."*

An alternative approach to the studies of financial time series is to assess information aggregation, that is the market's ability to consolidate disperse information into clear price signals regarding the asset's true value, in experimental asset markets. In these settings the researcher not only has control over the distribution of private information but also knows the fundamental value of the traded asset. It follows that in an experimental asset market, informational efficiency can be tested separately from asset pricing models. This promising approach was introduced by Plott and Sunder (1988) (PS, henceforth) who designed a laboratory environment to study information aggregation. We use their design to further analyze the market's ability to aggregate disperse information by identifying the critical condition(s) under which aggregation occurs. We also provide an assessment of the early evidence of information aggregation in order to revive the PS methodology. Our empirical analysis builds upon the original design proposed by PS and for which striking evidence in favor of information aggregation was initially reported. This design introduces an experimental asset which can only assume three possible values, 50, 240 or 490. Each trader in the market is then informed of a possible value the asset cannot assume. Because half of the traders are given one clue (e.g., "Not 50") and the other half are given the other possible clue (e.g., "Not 240"), the aggregate information available to all traders in the market is complete. PS posit that under the rational expectations model, traders should only trade at the true value of the asset (e.g., 490). This prediction implies perfect information aggregation in the spirit of Fama's (1970) definition of strong-form efficiency according to which all private information should ultimately be incorporated into prices.

Using new data from 204 markets as well as data from previous research on information aggregation (450 markets), we show that, contrary to the original findings reported in PS, experimental markets repeatedly fail to aggregate disperse information. Instead, asset prices are in line with the predictions of the prior information model (Lintner, 1969; PS). According to this model, traders make decisions based solely on their private information and fail to reflect upon asset prices to uncover other traders' information.

These findings led us to reassess the conditions necessary for markets to aggregate information. As exemplified by the rational expectations model, a market's ability to aggregate information is contingent upon traders' ability to unambiguously infer other's information from trading prices. Given the extensive literature in cognitive psychology (e.g., Tversky and Kahneman, 1974; Kahneman, 2011) documenting the failures of individuals (including experts) to apply Bayesian inference adequately, the rationality assumption may have to be reassessed in the light of behavioral finance models (Daniel, Hirshleifer, and Subrahmanyam, 1998, 2001; Hong and Stein, 1999; Shleiffer, 2000; Hirshleifer, 2001). Thus, our understanding of information aggregation in markets may be improved by considering the cognitive skills traders need to infer other traders' information from asset prices. Following recent findings in cognitive psychology we were able to identify cognitive reflection (and not financial literacy, computation skills or general intelligence) as the best predictor of one's ability to infer others' information from prices. Cognitive reflection is commonly assessed using the cognitive reflection test (CRT, henceforth). This test originally consisted of three questions which all have an appealing and intuitive, yet incorrect, answer. Upon reflection, one can disregard the intuitive answer and identify the correct one (Frederick, 2005). CRT questions are commonly asked in *Wall Street* interviews for trading positions (Zhou, 2008; Crack, 2014). Not surprisingly, professional traders were found to score remarkably high on the CRT (Thoma et al. 2015).

To account for the lack of information aggregation in our asset markets, we developed an asset market model in which traders have different levels of *reflection*. We focused on the stylized model in which perfectly *reflective* traders, who utilize asset prices to infer others' private information, interact with *non-reflective* traders, who do not use asset prices to infer others' information. Our framework bears a resemblance to noisy rational expectations models in which markets are assumed to be populated by both rational and noise traders (e.g., Grossman, 1977; Diamond and Verrecchia, 1981). Using model-based simulations, we established two conjectures which we subsequently test through the use of experiments. First, a higher proportion of *reflective* traders in the market is expected to improve information aggregation. Indeed, *reflective* traders have the ability to form asset prices that are more precise signals of the information available to traders, but this occurs only if *reflective* traders believe that prices are set by *reflective* individuals who trade based on information. If *reflective* traders believe that a large proportion of participants are trading randomly, then they will disregard asset prices as accurate signals of traders' information. This will ultimately limit the degree of information aggregation. Our second conjecture thus suggests that information aggregation should only be successful if the high level of traders' cognitive reflection is common information. This second conjecture stresses that the conditions for information aggregation are much more restrictive than originally envisioned.

We tested our two conjectures by recruiting *reflective* subjects, which we define as those individuals whose CRT score ranked in the top 20% of all scores in our student database. These subjects were highly sophisticated as evidenced by their average score of 2.65 on the three-item CRT. This placed them above the MIT (2.18), Princeton (1.63) and Harvard (1.43) samples reported in Frederick (2005). The CRT scores of our subjects were similar to the professional traders surveyed by Thoma et al. (2015) (average CRT = 2.59, n = 102). Consistent with our first conjecture, we show that the recruitment of individuals who are particularly *reflective* led to asset prices that were closer to the true value of the asset. However, these prices were still closer to the prior information predictions than the rational expectations predictions when traders were not informed of their fellow traders' high level of cognitive reflection.

In line with our second conjecture, information aggregation only occurred when the highly *reflective* traders populating the market were aware of each other's high level of *reflection*. Our findings provide empirical support for the general observation of Guesnerie (2005, preface, page xiv) regarding the rational expectations equilibrium concept:

> *"Coordination on the rational expectations equilibrium does not rely, as some optimistically thought at some time, on the rationality hypothesis, but on the 'common knowledge' of rationality."*

2. Experimental Design

2.1. Asset Markets

Our study uses the design of PS and, in particular, their parameterization of *Market 9* (Treatment C). This design introduces an experimental asset which can only assume three possible values: 50, 240 or 490 francs (each franc was worth \$0.001).[1] Each of the twelve traders in the market was privately informed of which possible value the asset could not take. Because half of the traders were given one clue (e.g., "Not 50") and the other half were given the other possible clue (e.g., "Not 240"),

[1] The exchange rate was chosen so that average subjects' earnings for the experiment were similar to average payments for a three-hour experiment at the lab where the study was conducted (i.e., average subjects' earnings were \$46.45).

the aggregate information available to all traders in the market was complete. Consistent with the rational expectations model (PS), prices should reflect the true value of the asset (e.g., 490). Convergence to the rational expectations prediction in this design constitutes the primary evidence of information aggregation in experimental asset markets.[2] PS report compelling evidence for information aggregation in their *Market 9* (see Figure 1) session conducted at the California Institute of Technology (Caltech).

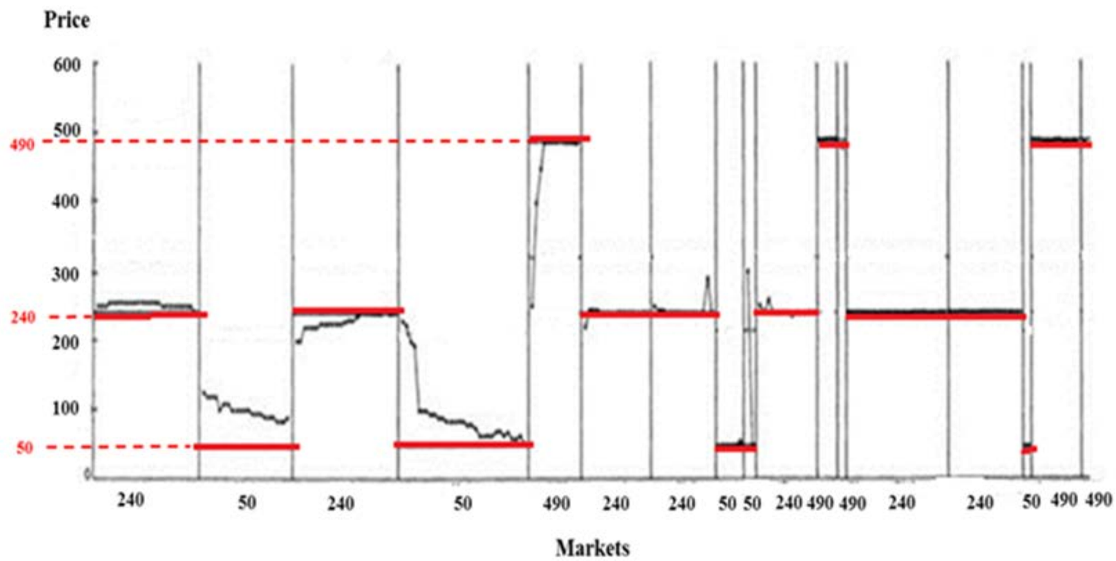


Figure 1. Chart of transaction prices per period taken from PS, *Market 9*. Each transaction is denoted by a black dot. The value predicted by the rational expectations model (i.e., the true value of the asset) is indicated by a horizontal line. The true value of the asset (50, 240 or 490) is displayed below each of the 17 markets.

Perhaps because of this early compelling evidence, only a few works have utilized the original setting of PS to assess the robustness of information aggregation in experimental asset markets. Exceptions include the works of Biais et al. (2005), Hanson, Oprea and Porter (2006), and Veiga and Vorsatz (2010).

Biais et al. (2005) focus their study on the role of personal characteristics such as self-monitoring and overconfidence in explaining trading profits whereas Hanson, Oprea and Porter (2006) and Veiga and Vorsatz (2010) study the manipulability of experimental asset markets. These works, however, do not follow the PS approach for assessing information aggregation. According to PS information aggregation occurs when rational expectations predictions regarding prices, allocations and profits

[2] Indeed, the markets with contingent-claim assets (Treatment A) for which PS report market prices close to the rational expectations predictions are evidence of information dissemination (Plott and Sunder, 1982) rather than information aggregation. Following Plott and Sunder (1982), we consider information dissemination to characterize the scenario in which a subset of the traders have perfect information (i.e., they know the true value of the asset). Information aggregation is arguably a more challenging task as no trader has sufficient information to definitively know the value of the asset (see e.g., Desgranges and Guesnerie, 2005 for a theoretical argument).

outperform those of alternative models such as, for example, the prior information model of Lintner (1969). Importantly, none of these works replicate the original PS design differing in relevant dimensions such as the number of markets, the fundamental value of the asset or the traders' endowments (see Table A1 in Appendix A). In our study, we use an experimental setup designed to replicate (as closely as possible) the original PS design (*Market 9*). Not only do we use the same parameters as in their original study with respect to possible asset values and the number of periods, but we also use their original instructions, training and terminology (see Appendix B). The only notable difference with the original PS design is that our study uses a computerized double auction instead of an oral double auction.[3] As a large majority of trading now operates electronically, this choice was made intentionally. The use of electronic double auctions also facilitates the replicability of our study.

2.2. Procedures

We recruited a total of 144 individuals from a subject pool of more than 1,500 individuals at a major Western US University. We conducted a total of 12 sessions with 12 traders in each. In ten of the sessions traders were endowed with 1,200 francs in cash (baseline sessions). To ensure our results are not dependent upon this endowment and to mirror the design of PS, in the remaining two sessions each subject's cash endowment was a 25,000 franc loan that had to be repaid at the end of each period (loan sessions). Each session consisted of 17 independent markets. Before each market started, subjects completed a training quiz regarding the random device (a spinning wheel) which was to be used during the experiment to draw the actual value of the asset (50, 240 or 490 francs) at the end of each market period. During the training, subjects had to predict the outcome of the spinning wheel over 10 trials (see Appendix B, Instructions Part 1). Each correct prediction was rewarded 25 cents, and each incorrect answer incurred a 10 cent penalty as in the original design of PS. Before the experiment started, subjects also completed a 7-question quiz on the mechanics of the market (see Appendix B, Instructions Part 3).

[3] There exist two other minor differences between our design and that of PS. First, in 10 out of 12 (baseline) sessions we did not use the loan procedure of PS to determine endowments. Second, we used 5-minute instead of 7-minute periods as in PS. However, because our trading mechanism was computerized, subjects could undertake at least as many trades as in the original 7-minute periods with oral auctions. The average (median) number of trades in PS was 13.5 (15.0) in Market 9 compared to 32.5 (28.0) in our study (*p-value* < 0.001).

2.3. Tests

At the end of each session, subjects completed a series of tests and a demographic survey for a total of 25 minutes. All of these tasks were computerized. Subjects were paid a flat fee of $3 to complete the tests. As is common practice in the literature, no pay-for-performance was used for the tests. We chose to administer tests which have been found to correlate with trading behavior in a series of recent works using both experimental and archival data (e.g., Grinblatt, Keloharju and Linnainmaa, 2011, 2012; Biais et al. 2005). This includes a financial literacy test, the CRT, a general intelligence test (Raven test) and a self-monitoring questionnaire. We describe these tests and the related literature in Appendix C.

3. Results

3.1. Information Aggregation: Competing Models

Following PS we consider three competing models of trading behavior in our information aggregation experiments: rational expectations (RE), prior information (PI), and maximin (MM).

3.1.1. Rational expectations (RE)

Under rational expectations, all subjects trade in equilibrium as if they knew the private information of the other traders in the market. Given the PS design, the pooled information of all traders identifies the value of the asset with certainty. It follows that, under RE, price predictions are equal to the actual value of the asset. Moreover, in this scenario if any trades occur in equilibrium, then they must take place at a price equal to the value of the asset.

3.1.2. Prior information (PI)

In this model traders do not infer other traders' information from market prices but apply Bayes' rule to compute the expected value of the asset given their own information (Lintner, 1969). That is, subjects base their trades solely on the information received at the beginning of the market and fail to reflect on asset prices to uncover other traders' information. As a result, PS considers the PI asset price prediction to equal the expected value of the trader with the most positive prior information (i.e., the trader whose prior information about the asset leads to the highest expected value across traders). A more general version of the prior information model implies that a transaction can occur at any price between the expected value of the asset for the trader with the least positive prior information and the expected value of the asset for the trader with the most positive prior information. PS define the PI price prediction as a single number (210 when the value of the asset is 50 and 316.9 otherwise). However, any price in the range (156.9, 210) <156.9, 316.9> {210, 316.9} is actually consistent with

PI when the value of the asset is (50) <240> {490}, where 156.9, 210 and 316.9 correspond to the expected value of the asset for traders who hold the clue "Not 490", "Not 240" and "Not 50" (see Table 1, Panel A). The average price for each of these price intervals is equal to 183.5, 236.9 or 263.5 when the value of the asset is 50, 240 or 490. We refer to these average prices as the predictions of the generalized prior information model (GPI, henceforth).

3.1.3. Maximin (MM)

Similar to PI, traders do not aggregate information in this model. Instead, subjects are expected to buy (sell) the asset only when they are certain that the price is below (above) the minimum (maximum) value they could possibly receive given their prior information. In our design, this implies that a trader with the clue "Not 50" will be willing to buy for any price below 240 while being willing to sell the asset for any price above 490.

The predictions of the three models as well as the average transaction prices across the 10 baseline sessions are detailed in Table 1.

Table 1. Price predictions for each model.

Panel A. Expected value of asset given the subject's clue.			
Clue	Not 50	Not 240	Not 490
Expected value	316.9	210	156.9
Panel B. Predicted value of asset according to each model.			
Model/True Value	50	240	490
RE (as in PS)	50	240	490
PI (as in PS)	210	316.9	316.9
GPI (This paper)	183.5	236.9	263.5
MM (as in PS)	50	240	240
Average Transaction Price	219.5	240.8	263.1
(SD)	(227.5)	(241.7)	(251.8)

3.2. Statistical Analysis of Information Aggregation

3.2.1. Our data

We start by presenting the average price per minute over the 10 baseline sessions for each of the 17 markets (see Appendix D for graphs of average prices for each session separately).

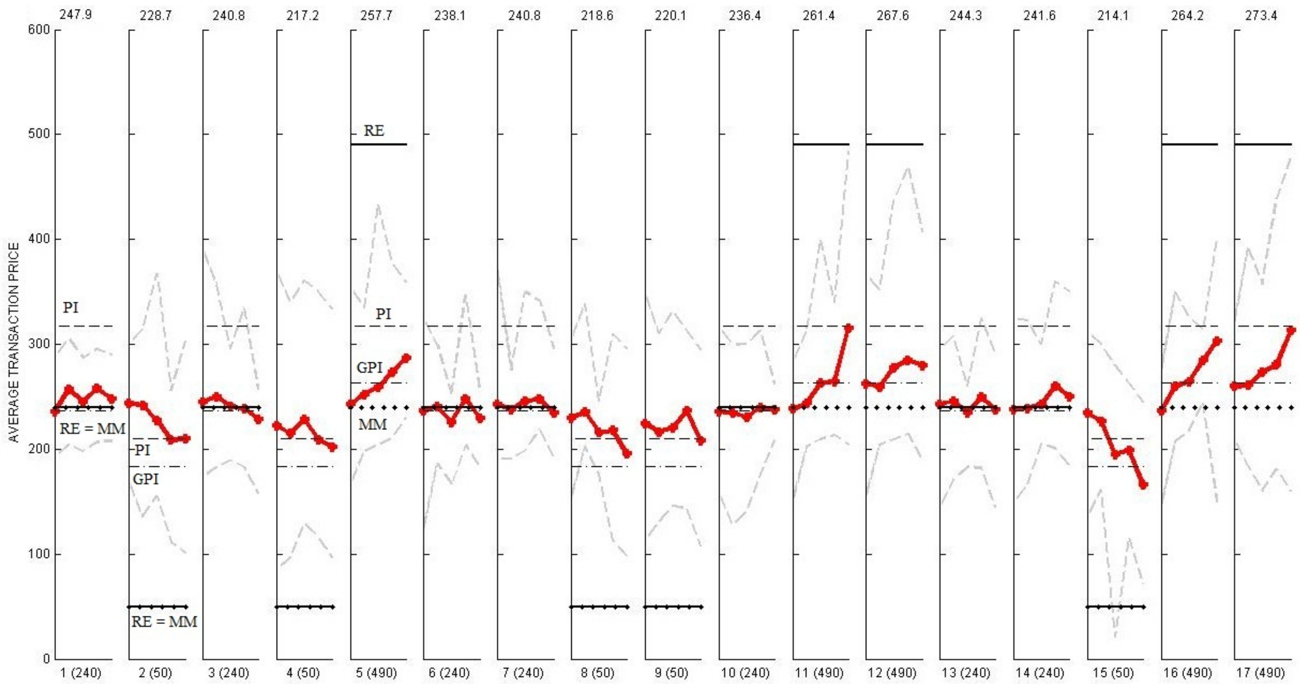


Figure 2. Average price per minute over the 10 baseline sessions for each of the seventeen markets (thick, red curve). The dashed, light grey curves represent the minimum and maximum average price per market per minute for the 10 baseline sessions. The average price per market period is listed at the top of each subfigure, and the true value of the asset is denoted at the bottom of each subfigure. The rational expectations value (RE) is indicated by a solid horizontal line; the prior information value is indicated by a dashed horizontal line (PI); the maximin value is indicated by a dotted horizontal line (MM); and the generalized prior information value is indicated by a dash-dot horizontal line (GPI).

PS performed several analyses to gauge whether or not their subjects aggregated their private information through trading. For consistency, we apply the PS analytical techniques to our data. Following their lead, we give information aggregation its "best chance" by only considering the last occurrence of each of the possible asset values: 50, 240, and 490 (i.e., markets 15, 14, and 17, respectively). To assess information aggregation, we report for each session the mean absolute price deviation between the price and the value predicted by each model (PI, GPI, RE, and MM) in Table 2. For each session this value is calculated as:

$$\underset{i}{\text{average}}|p_i - m|$$

where i represents a transaction, p_i corresponds to the transaction price, and m is the predicted price based upon the appropriate model. Thus, the mean absolute price deviation is computed as the average over all transactions in markets 14, 15, and 17 for each session. Note that for each of the 10 baseline sessions, the mean absolute price deviation is smallest for one of the prior information models (either PI or GPI). Given the tendency of an influx of cash to boost prices in asset markets

(Caginalp, Porter and Smith, 1998; 2001), it is not surprising that the deviations from rational expectations predictions are even larger for our loan sessions than for our baseline sessions.

Table 2. Comparison of actual prices to model predictions at the end of each market. As in PS, only markets 14, 15, and 17 are considered.

		Mean Absolute Price Deviation				*Percentage of Convergent Price Changes*			
Sessions	Session	PI	GPI	RE	MM	PI	GPI	RE	MM
Baseline	1	45.30	68.26	136.33	127.33	58.1%	55.8%	51.2%	53.5%
	2	103.47	83.63	122.76	58.16	52.3%	52.3%	54.7%	54.7%
	3	59.83	42.48	121.92	98.89	57.1%	52.4%	60.3%	54.0%
	4	53.56	43.69	118.37	57.25	63.3%	56.7%	65.0%	61.7%
	5	78.08	68.71	145.85	89.93	50.5%	49.5%	51.5%	53.6%
	6	54.96	22.28	150.10	77.10	61.5%	60.0%	61.5%	63.1%
	7	77.77	65.28	142.26	71.30	61.9%	64.9%	60.8%	63.9%
	8	58.28	37.33	115.93	77.52	59.1%	57.6%	54.5%	54.5%
	9	59.20	26.36	117.45	60.24	55.2%	50.9%	62.0%	61.3%
	10	56.48	18.41	141.86	74.35	69.4%	69.4%	67.7%	61.3%
Loan	11	159.37	177.04	229.14	263.87	54.4%	54.4%	55.4%	54.4%
	12	116.81	135.21	204.72	219.14	57.4%	57.4%	54.1%	56.6%
Total average		76.92	**65.72**	145.56	106.26	**58.4%**	56.8%	58.2%	57.7%
PS (1988)	*Market 9*	136.00	–	**0.00**	83.00	–	–	–	–

In our statistical analysis and following PS we use the Wilcoxon Rank Sum test and the Sign Rank test, if not stated otherwise. We use data from the baseline sessions for the tests, although similar findings are obtained when the loan sessions are included. Using the Wilcoxon Rank Sum test we thus confirm that the mean absolute price deviation of asset prices for all of the sessions is significantly lower when computed with respect to PI or GPI than RE (both *p-values* < 0.001). Similarly, the mean absolute price deviations computed with respect to MM are significantly smaller than the corresponding RE deviations (*p-value* < 0.001). Moreover, the mean absolute price deviation is also lower for PI {GPI} than MM (*p-value* = 0.112 {0.010}). Thus, our data demonstrate that both the PI and the GPI models better describe market prices in the information aggregation experiments than the RE or MM models. We illustrate this finding in Figure 3 which displays the histogram of average prices across all of our markets for each of the three possible asset values (50, 240 or 490) for the 10 baseline sessions. In each figure, we represent the range of prices that is consistent with the GPI model. Interestingly, the rational equilibrium only falls in the range of possible prices predicted by GPI when the value of the asset is 240. This is actually the only case where we observe evidence of information aggregation. However, this is mostly due to the fact that average prices (regardless of the true value of the asset) tend to be concentrated around 240 regardless of the value of the asset. The range of prices consistent with GPI includes a large proportion (71.2%) of market average prices.

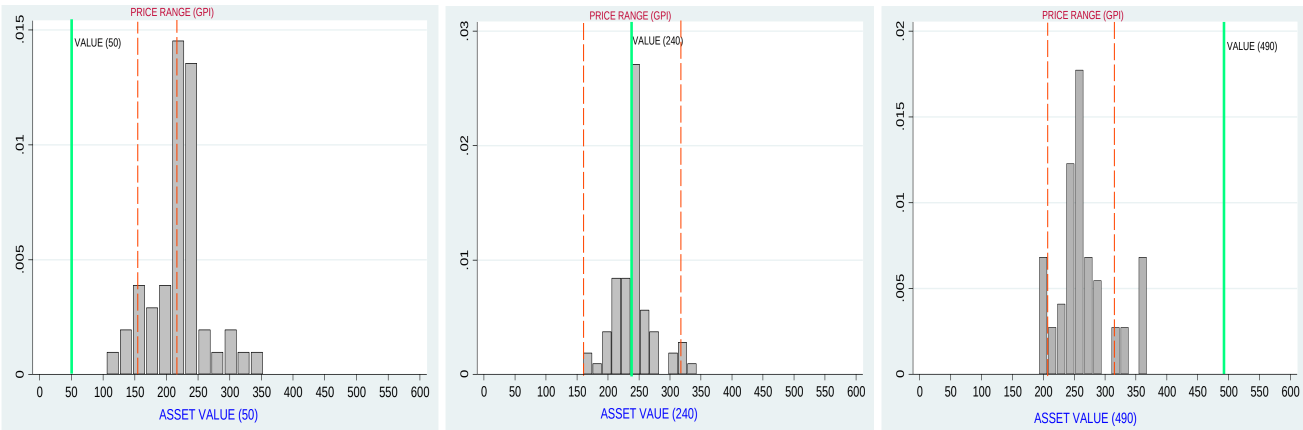


Figure 3. Histogram of average prices for asset values 50, 240 and 490 for all markets in the 10 baseline sessions. Dashed lines represent the range of prices that are in line with the predictions of the GPI model. The upper bounds of the GPI price range corresponds to the PI model prediction in PS.

Another measure of information aggregation developed in PS is the percentage of convergent price changes. The $N+1^{st}$ transaction is considered to be convergent if its price is closer to the selected model's prediction than the price of the previous transaction, N. That is, the $N+1^{st}$ transaction is deemed convergent if:

$$|p_{N+1} - m| \leq |p_N - m|$$

where p_N is the market price of the N^{th} transaction and m is the model's predicted value. For each model the ratio of convergent price changes to the total number of transactions per session is reported in Table 2.[4] On average, the PI model performs best with respect to this metric, although differences across models are not statistically significant (all *p-values* > 0.1 for Normal Proportion tests).

3.2.2. Previous research data

We perform a similar analysis with data provided by Hanson, Oprea and Porter (2006) (HOP, henceforth) and Veiga and Vorsatz (2010) (VV, henceforth). These are the only studies of which we are aware that use (for their baseline treatment) the same type of market design as PS (see Table A1

[4] As an initial transaction is required to determine if the subsequent transaction is convergent, we utilize the total number of transactions minus three as the denominator in this ratio.

in Appendix A). Biais et al. (2005) differ from the original PS study by introducing a call auction either before or after the double auction market starts.[5]

Consistent with our data, the PI and GPI models perform best in all ten sessions across both studies in terms of mean absolute price deviation (see Table A2 in Appendix A). The mean absolute price deviation is significantly lower when computed for PI {GPI} than RE for both HOP (*p-value* = 0.006, {0.016}) and VV (*p-value* = 0.021, {0.043}). Mean absolute price deviation is also significantly lower for PI {GPI} than MM for both HOP (*p-value* = 0.010 {0.025}) and VV (*p-value* = 0.006 {0.025}). We report no significant differences in mean absolute price deviation between RE and MM for both HOP (*p-value* = 0.885) and VV (*p-value* = 0.873).

In line with the analysis of our data, the PI, MM and RE models do not differ in the proportion of convergent prices across sessions in HOP and VV (All *p-values* > 0.1 for Normal Proportion tests). In summary, data from previous studies that utilized a design similar to PS confirm that PI is a better model than RE for describing market prices in information aggregation experiments. This suggests that although traders utilize their private information to make trading decisions, they are typically unable to use market prices to update their beliefs regarding other traders' information.

Although Bias et al. (2005) did not implement the same auction mechanism as the original PS study, to the best of our knowledge this paper in conjunction with the present one are the only studies that used the same asset values as *Market 9* in PS (50, 240 and 490). In Figure A1 in Appendix A, we display average transactions prices for each asset value for Biais et al. (2005), our study, and the original PS session (*Market 9*). The Bias et al. (2005) results demonstrate a weak relationship between asset prices and asset value. Their findings are remarkably consistent with ours as asset prices appear to be much more in line with GPI than with RE.

Although asset prices do not appear to be in line with rational expectations predictions, it is still possible that prices converged to rational expectations predictions over time. In Appendix E, we provide a detailed analysis of learning across market periods in our baseline sessions. We show that asset prices are closer to rational expectations predictions in the last three market periods than in the first three market periods. Asset prices are also closer to rational expectations predictions for the last three transactions of a given period than the first three transactions. However, prices are still closer to the predictions of the prior information models (PI / GPI) than the rational expectations predictions even when considering the last three transactions of each of the last three markets (see Table E1).

[5] Also, in Krahnen and Weber (2001) some agents are completely uninformed in the baseline treatment. This represents a significant deviation from the original PS design.

Our findings shed light on the inherent difficulty of experimental asset markets to aggregate disperse information even in the simplest environment in which the asset can only assume three possible values. Using the same data set, Corgnet, DeSantis, and Porter (2015, CDP henceforth) perform an extensive analysis of the determinants of traders' earnings. They find that individuals with high cognitive skills (i.e., Financial literacy, IQ, and cognitive reflection) earn more. This highlights the fact that information aggregation may require specific cognitive skills that, despite their high level of academic education, our subjects may not possess. This observation echoes the comment of Radner (1982) regarding the outstanding traders' abilities the rational expectation model may require in a situation in which traders hold disperse information.

Given the heterogeneity in cognitive skills, even across college populations (e.g., Frederick, 2005), one must reassess the potential of the prior information model to explain information aggregation. We first highlight some of the shortcomings of the prior information model before developing and testing a new model of information aggregation based on traders' cognitive limitations.

4. Reflective Learning Model

4.1. Limitations of the Prior Information Model

4.1.1. Allocations and profit distributions

The relative success of the prior information model in predicting asset prices should be tempered by its inability to predict allocations or profit distributions across market subjects (see Appendix F). Following PS, we consider the allocation predictions of each model. While the RE model predicts that subjects should not trade in these markets (except at the true value of the asset), the PI and MM models suggest trading should occur. Moreover, these models indicate that all shares should be held by a specific subset of investors at the end of each market period. The model predictions are summarized in Table 3.

Table 3. Clue of the traders predicted by each model to hold all of the shares.

Asset Value	50	240	490
RE	No trading (except at 50)	No trading (except at 240)	No trading (except at 490)
PI / GPI	Not 240	Not 50	Not 50
MM	No trading (except at 50)	Not 50	Not 50

According to the PS statistical measure of allocation efficiency (see Appendix F1), the allocation predictions of the PI model are no better than the predictions of the MM model (see Table F1 in Appendix F1). In Table 4, we report significant differences between the predictions of the PI model and the actual holdings in our experimental sessions.

Table 4. Predicted and actual holdings of shares for the traders who should hold all of the shares according to the prior information models (PI / GPI). All traders start with 4 shares.[+]

Asset Value	50	240	490
PI / GPI predicted holding	8	8	8
Actual holding	4.57	5.81	4.78
Sign Rank Test (*p-value*)	**<0.001**	**<0.001**	**<0.001**

[+]Baseline data were used. Inclusion of data from the two loan sessions yields similar results. Tests are completed using the average holding for a given session in a given state.

Not surprisingly, as the PI model fails to explain final allocations, it also struggles to explain the distribution of profits across trader types. Using the PS analysis, we do not find a significant difference in the ability of the PI or the MM model to explain profit distributions. Moreover, RE outperforms both PI and MM (see Appendix F2), although our results significantly differ from the original PS analysis.

One issue for all of the models presented thus far is that they do not account for heterogeneity across traders. This is a serious limitation given the findings of Biais et al. (2005) and more recently of Corgnet, DeSantis and Porter (2015). CDP show that cognitive skills (and, in particular, cognitive reflection) play a central role in explaining trader earnings and trading behavior in information aggregation markets, even after controlling for general intelligence, financial literacy, computation skills and risk attitudes. We thus focus our analysis of traders' heterogeneity on cognitive reflection (as measured by the CRT).

4.1.2. Heterogeneity in cognitive skills

A trader who is able to use prices to update his or her own beliefs about the true value of the asset should ultimately be better informed than the trader who does not learn from prices.[6] Thus, traders who use prices to update their prior information (we will refer to these traders as *reflective*) should trade more consistently with the true value of the asset than those who disregard prices as a signal of the true value of the asset (*non-reflective* traders). Formally, we say that a trade is *consistent* with asset value w in {50, 240 or 490}, as long as it implies buying (selling) the asset for a price below (above) w. Because *reflective* traders are better able to learn the true value of the asset in information aggregation markets than *non-reflective* traders they will also tend to obtain higher earnings.

CDP show that cognitive reflection, as measured by the CRT, is the best predictor of trading *consistently* with the true asset value (as well as trader earnings) even after controlling for general intelligence, financial literacy, computation skills and risk attitudes. The *reflective* capacity, which

[6] This is the case as long as all other traders are not trading randomly.

allows traders to infer others' clues from asset prices, is closely related to an individual's ability to apply Bayes' rule adequately and refrain from using simple heuristics. Thus, it is not surprising that CRT, which is our best predictor of *reflective* capacity, has also been shown to explain an individual's ability to avoid known heuristics and behavioral biases (e.g., Cokely and Kelley, 2009; Oechssler, Roider and Schmitz, 2009; Campitelli and Labollita, 2010; Toplak, West and Stanovich, 2011). We illustrate the positive relationship between CRT and trading *consistently* with the true value of the asset in Table 5 (see CDP for a detailed statistical analysis).

Table 5. Trading *consistent* with the true value of the asset for all individual-level data across CRT scores.[+]

CRT scores		0-1	2-3	4-5	6-7
Proportion of *consistent* trades	Average	53.6%	56.8%	57.6%	62.1%
	Median	50.0%	60.0%	60.0%	70.0%
Proportion of subjects		35.0%	35.0%	20.8%	9.2%

[+]Baseline data were used. Inclusion of the data from the loan sessions yields similar results.

Thus, even in the context of a pool of college students who are homogenous in terms of educational background, we uncover significant heterogeneity in cognitive skills which ultimately affects trading behavior (see CDP for detailed analyses). To further investigate the effect of heterogeneity in cognitive skills on information aggregation in asset markets, we develop a model in which *reflective* and *non-reflective* traders interact.

4.2. Reflective Learning

We propose a model in which highly *reflective* and *non-reflective* traders interact. Our model resembles noisy rational expectations models (e.g., the theoretical works of Grossman, 1977 and Diamond and Verrecchia, 1981, as well as the experimental work of Bloomfield, O'Hara and Saar 2009) because it allows for different levels of traders' sophistication (*reflection* in our model). However, our work contrasts with the noisy rational expectations literature because it is a learning model that does not rely on an equilibrium approach to asset pricing.

In our model, *reflective* traders are able to use observed prices to infer other traders' information by properly applying Bayes' rule. The *non-reflective* traders are assumed to trade randomly, deciding to either buy or sell the asset at a given price with probability 50%.

4.2.1 Description of the model

Let α represent the proportion of *reflective* (REF) traders and $(1-\alpha)$ the proportion of *non-reflective* (or *noise*) traders in the market. The collective belief regarding the proportion of *reflective* traders in the market is denoted α_b. Note that α_b is only equal to α under the assumption of "common information" of traders' level of *reflection*, i.e., that all traders share the correct belief of α.

Trading

We follow our experimental design in modeling the trading process. In particular, we consider that prices occur as a result of a continuous flow of bids and asks posted by traders (see Appendix B). At the beginning of each period, a trader (selected at random) posts a bid-ask spread to the market. Then, another trader is selected at random to post a bid-ask spread. In line with our experimental design, the bid (ask) of the newly selected trader will replace the current bid (ask) if and only if it improves the current spread.[7] Traders (selected at random) continue to post bid-ask spreads to the market until a trade occurs. A trade occurs if a newly selected trader's bid is greater than the current best ask or if the trader's ask is less than the current best bid. If a trade occurs, then the price, p, is set at the current best bid (or ask) and the trading book is erased.[8]

For a *reflective* trader, the bid-ask spread is determined as a function of the trader's belief regarding the true asset value. Specifically, for a *reflective* trader, the bid (ask) is a randomly drawn integer from a uniform distribution on the interval $[1, REF_w]$ ($[REF_w, 600]$), where REF_w:= E[W|"Not w"] denotes the belief of the true value of the asset (W) for the *reflective* trader with clue "Not w", w $\in \{50, 240, 490\}$. Note that the interval [0,600] corresponds to the range of permissible values in the experimental design. Moreover, a trader's ask must be greater than his or her bid similarly to our experimental design. For a *non-reflective* trader, the bid and ask are randomly drawn from a uniform distribution on the interval $[1{,}600]$.

Updating rules

After each transaction, the *reflective* traders update their beliefs of the true value of the asset by applying Bayes' rule to infer other traders' information from market prices.[9] *Noise* traders do not update their beliefs and simply trade randomly buying or selling the asset with probability 50% regardless of the price.[10] To understand the basic dynamics of the model, consider the following

[7] Note, however, that the bid-ask spread is only updated when the newly selected trader has enough cash (shares) to cover the bid (ask) position.

[8] The book only consists of the best bid and the best ask.

[9] *Reflective* traders do not update their beliefs if a trade fails to occur in a given period. This is a simplifying assumption because *reflective* traders could learn others' information from the absence of a trade.

[10] Our framework can be extended by considering the general case in which *non-reflective* traders are not necessarily *noise* traders. For example, one could consider that traders are reflective-type η if they need to observe η prices consistent with a certain value of the asset before updating their belief accordingly. This extension of our model would relate to recent works stressing the prominent role of inattention in financial decisions (e.g., Agnew, Balduzzi, and Sunden, 2003; Andersen et al. 2015).

example. Suppose that the true value of the asset is 50. Based on their prior information ("Not 240", "Not 490" or "Not 50"), *reflective* traders' expectations regarding the value of the asset would be: $REF_{240} = 210$, $REF_{490} = 156.9$ and $REF_{50} = 316.9$.[11] Let us consider traders with the clue "Not 240".[12] Assume the first-transaction price is p = 180. The traders with the clue "Not 240" know that any transaction between *reflective* traders at this price must involve a trader with the clue "Not 490" as the seller (because they are the only *reflective* traders who value the asset below 180) and a trader with either the clue "Not 50" or "Not 240" as the buyer. It follows that a transaction occurring at p = 180 is more likely to involve a *reflective* trader with clue "Not 490" than a *reflective* trader with the clue "Not 50". Indeed, a trader with the clue "Not 490" may transact with a trader who received either the "Not 50" or the "Not 240" clue whereas traders with the "Not 50" and "Not 240" clues will not trade with each other at this price.

It follows that for p = 180 (the same argument holds for any price in (156.9, 210)), traders with the clue "Not 240" will update their beliefs downward giving less weight to the asset value equaling 490 and more weight to the asset value being 50. After observing p = 180, the traders with the "Not 240" clue will update their belief of the true value of the asset using Bayes' rule as follows:

$$\mathrm{E}[\mathrm{W} \mid \text{“Not 240”}, p = 180] = \frac{50\times\mathrm{P}[p|50]\times\mathrm{P}[50]+490\times\mathrm{P}[p|490]\times\mathrm{P}[490]}{\mathrm{P}[p|50]\times\mathrm{P}[50]+\mathrm{P}[p|490]\times\mathrm{P}[490]}$$

where the prior probabilities for the asset value being 50 or 490 are respectively $\mathrm{P}[50] = 0.35$ and $\mathrm{P}[490] = 0.20$ given the parameters of our experimental design (see Table A1 in Appendix A). Also, if the value of the asset is 50 then the *reflective* traders in the market either hold the clue "Not 240" (in which case they value the asset at $REF_{240} = 210$) or "Not 490" (in which case they value the asset at $REF_{490} = 156.9$). Since the price (p = 180) is between the valuation of both types of *reflective* traders in the market, a trade will occur between two *reflective* traders as long as they hold a different clue. In our model, this occurs with probability $\frac{\alpha^2}{2}$. The rest of the trades will involve at least one *non-reflective* trader. These trades will occur with probability $\frac{1-\alpha^2}{2}$ in our model. It follows that the probability of a trade occurring at price p = 180 when the value of the asset is 50 is given by $\mathrm{P}[p|50] = \frac{\alpha^2}{2} + \frac{1-\alpha^2}{2} = \frac{1}{2}$. A similar reasoning applies to show that $\mathrm{P}[p|490] = \frac{1-\alpha^2}{2}$ (see Appendix G1).

[11] As the true value of the asset is assumed to be 50, there are no traders with the clue "Not 50" in the market. However, this fact is not known by the *reflective* traders who must therefore act as if traders with the clue "Not 50" existed.
[12] The updating rules for other traders and other possible asset values are similar. Refer to Appendix G1 for more details.

It is worth noting that traders' collective belief regarding the proportion of *reflective* traders in the market (α_b) does not necessarily coincide in our model with the actual proportion of *reflective* traders (α). In that case ($\alpha_b \neq \alpha$), *reflective* traders will use the following formulas:

$$\mathrm{P}[p|50] = \frac{\alpha_b^2}{2} + \frac{1-\alpha_b^2}{2} \text{ and } \mathrm{P}[p|490] = \frac{1-\alpha_b^2}{2} \text{ to update their beliefs.}$$

If the first-transaction price is in the range (210, 316.9), then the opposite logic would apply and traders with the clue "Not 240" would update their beliefs upward instead of downward. For first-transaction prices below 156.9 (above 316.9), *reflective* traders would not update their beliefs because any transaction in that range would imply all *reflective* traders, regardless of their prior information, buying (selling) the asset. In the figure below, we illustrate the updating rules for the traders with the clue "Not 240" after the first transaction.

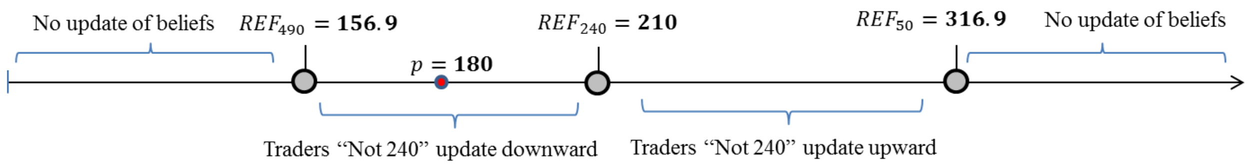


Figure 4. Representation of *reflective* traders' beliefs before the first transaction and updating rules for the trader with the clue "Not 240" for different ranges of first-transaction prices.

After observing the second-transaction price, *reflective* traders with the clue "Not 240"will update their beliefs using E[W | "Not 240", p = 180] as their prior belief. That is, their prior belief after the second transaction is their updated belief after the first transaction. This updating process occurs after each transaction. A detailed description of the model, including the general updating procedure for all possible values of the asset and all possible clues, is included in Appendix G1.

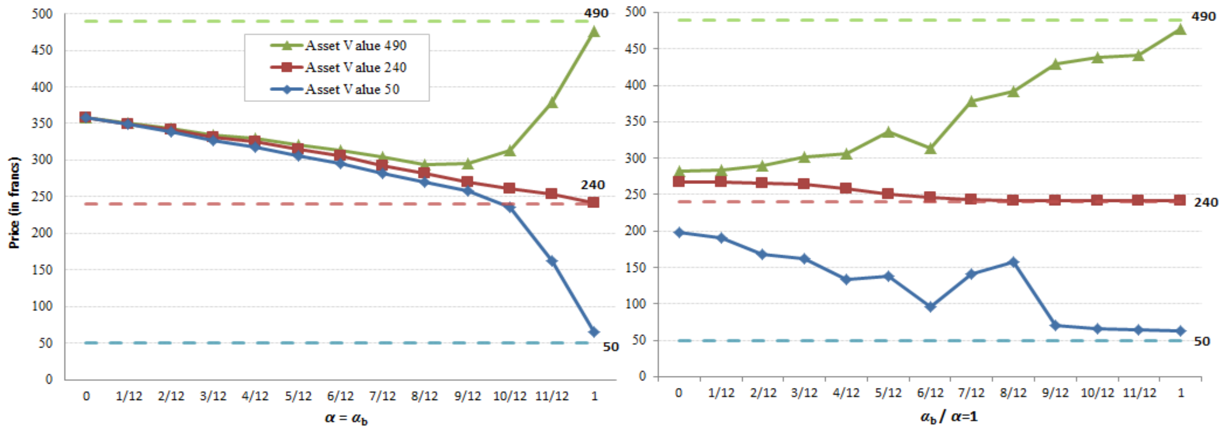
Simulation results and conjectures

We conduct simulations of this model to motivate conjectures which may be tested experimentally. The number of *reflective* traders is determined by the parameter α, while the *reflective* traders' belief of the proportion of *reflective* traders in the market is given by α_b. For each combination of (α, α_b) we perform 1,000 simulations. Each simulation is conducted with 12 traders each endowed with 4 shares of the asset and 1,200 francs similarly to our experimental design. The simulation runs until 30 trades have been made where 30 corresponds to the average number of trades in our baseline sessions. In Table 6, we report our main measure of information aggregation (the average absolute deviation) for the 1,000 simulations of different (α, α_b) combinations. We show that when traders are aware of the actual proportion of *reflective* traders in the market, the average absolute deviation between the price and the true value of the asset decreases as the proportion of *reflective* traders in the market increases ($\alpha = \alpha_b$). Moreover, even a small proportion of *noise* traders (e.g., $\alpha = 11/12$) leads to substantial deviations from the true value. Indeed, $\alpha = \alpha_b = 1$ yields an average absolute deviation (as defined in PS) of 9.41, while $\alpha = \alpha_b = 11/12$ produces an average absolute deviation of 84.85. In the last column of Table 6, we consider the case in which all traders are *reflective* ($\alpha = 1$) but they do not necessarily know that all traders in the market are *reflective* (i.e., $\alpha \neq \alpha_b$). In this case we still find substantial deviations from true value unless the proportion of *reflective* traders is common information ($\alpha = \alpha_b = 1$). Indeed, for $\alpha = 1$ and $\alpha_b = 11/12$ the average absolute deviation is 22.54 compared to 9.41 when the actual proportion of *reflective* traders is common information.

Table 6. Average mean absolute price deviation from true value calculated across all three possible values of the asset (50, 240 and 490)

	Common Information	No Common Information
α_b	$\alpha = \alpha_b$	$\alpha = 1$
0	201.18	135.71
1/12	197.73	131.95
2/12	193.98	121.77
3/12	189.22	113.64
4/12	184.55	99.54
5/12	180.04	87.68
6/12	174.99	78.97
7/12	169.17	70.90
8/12	163.53	70.49
9/12	154.01	28.74
10/12	135.78	23.84
11/12	84.85	22.54
1	**9.41**	**9.41**

We illustrate our findings graphically in Figure 5. When the proportion of *reflective* traders is common information, the average prices converge to the true value of the asset as α approaches 1 (see Panel (a)). Moreover, when all traders in the market are *reflective* ($\alpha = 1$), the average prices again converge to the true value of the asset as α_b approaches 1 (see Panel (b)).



(a) Common information (b) No common information

Figure 5. These figures represent the average price across simulations for each (α, α_b)-combination. In Panel (a), we assume common information that is $\alpha = \alpha_b$. In Panel (b), we assume all traders are *reflective* (α=1) and we display the evolution of average prices as the traders' belief of the proportion of *reflective* traders (α_b) increases from 0 to 1.

In Appendix A (Table A3), we also report the percentage of convergent price changes, which is the other primary measure of information aggregation in PS, for the 1,000 simulations. When the level of trader *reflection* is common information ($\alpha = \alpha_b$), this percentage is greatest when all traders are *reflective* ($\alpha = 1$). If all traders are *reflective*, then the percentage is greatest when there is common information (i.e., $\alpha = \alpha_b$).

The results of our simulations allow us to derive two testable conjectures. The first conjecture states that a higher proportion of *reflective* traders in the market should improve information aggregation.

Conjecture 1. *As the proportion of reflective traders in the market increases, transaction prices will be closer to the true value of the asset.*

The intuition supporting Conjecture 1 follows from the fact that an increase in the proportion of *reflective* traders in the market increases the number of trades that are based on private information. As a result of this increase in the number of informed trades, asset prices will more likely reflect traders' available information. In addition, *reflective* traders will be able to infer others' information by observing asset prices and will subsequently trade based on their updated beliefs of the value of the asset. These subsequent trades will transmit information to the market leading prices to reflect the aggregate information.

There exists, however, one issue with this argument. *Reflective* traders will infer information from asset prices only if they believe that prices are set by *reflective* individuals who trade based on information. If *reflective* traders believe that a large proportion of individuals are not trading based on information (e.g., *noise* traders), then they will disregard asset prices as accurate signals of traders' information. This, in turn, will ultimately preclude information aggregation. This leads to our second conjecture which establishes the crucial role that common information of traders' *reflective* capacity serves in achieving information aggregation in markets. This conjecture states that information aggregation will only occur if traders' high level of *reflection* is common information.[13]

Conjecture 2. *A necessary condition for a market entirely populated with reflective traders to aggregate information is for the number of reflective traders in the market to be common information.*

4.3. Testing the Reflective Learning Model

4.3.1. Recruiting on CRT

To test Conjecture 1, we need to be able to exogenously manipulate the proportion of *reflective* traders in the market. To do so, we rely on the work of CDP who identify CRT as predictive of the ability of traders to reflect on prices and update their beliefs accordingly.[14] We decided to recruit subjects in the top 20% of the CRT scores distribution in order to increase the proportion of *reflective* traders in the market. This subset of our population has an average score of 2.65 on the original three-item CRT, which places them in the top 20% of the distribution of the original CRT scores of 3,428 students surveyed in Frederick (2005).[15] The scores of our high-CRT subjects were significantly higher than the CRT scores of 592 US individual traders who averaged 1.28 (see Krische, 2015) and professional workers in the Finance and Banking sectors with an average score of 1.62 (see Thoma et al. 2015). The only groups that match the CRT scores of our top 20% sample are the 102 professional traders surveyed in Thoma et al. (2015) and the 24 Caltech students who participated in the study of Brocas et al. (2014) (see Figure A2 in Appendix A for a summary of CRT scores across a wide range of samples).

To recruit by CRT scores, we used the results of an extensive survey which was conducted at our home institution one month after our original study. All of the subjects registered in the laboratory's database (n = 1,963) were invited to complete a comprehensive one-hour survey which included the

[13] In Appendix G2, we provide preliminary evidence for two additional conjectures derived from the Reflective Learning Model regarding allocations and trading earnings.

[14] CDP also show that, on average, high-CRT individuals earn more than low-CRT individuals. This is in line with a series of experimental works in the bubble literature (Noussair, Tucker and Xu, 2014; Corgnet et al. 2015) that have shown that high-CRT subjects outperform low-CRT subjects.

[15] This is consistent with the fact that our 144 subjects (who participated in the baseline and loan experiments) obtained remarkably similar CRT scores (1.23) to the sample of Frederick (2005) (1.24).

extended, seven-item CRT developed by Toplak, West and Stanovich (2014).[16] Using the extended CRT, we recruited subjects in the top 20% of the distribution of scores of the 885 students who participated in the survey. We thus recruited subjects who scored 5, 6 or 7 on the extended CRT (see Table A4 in Appendix A for the distribution of the extended CRT scores in the student population). To test Conjectures 1 and 2, we conducted two treatments: one in which CRT scores of all traders in the market were common information and one in which the scores were not common information.

We conducted a total of 4 sessions per treatment with a total of 96 subjects. As intended, the CRT scores of the high-CRT sessions are significantly higher than for either the baseline or loan sessions (all *p-values* < 0.001) (see Table 7). There are no significant differences in CRT scores between the high-CRT sessions with and without common information (*p-value* < 0.873) and between the baseline and loan sessions (*p-value* < 0.587).

TABLE 7. Seven-item CRT scores by treatments.

Treatment	Average (median) -Stand. Dev- CRT score
Baseline (n = 120)	3.31 (3.00) -1.86-
Loan (n = 24)	3.17 (3.00) -1.90-
Top 20% CRT (n = 48) (*No common information*)	5.70 (6.00) -0.94-
Top 20% CRT (n = 48) (*Common information*)	5.72 (6.00) -0.80-

4.3.2. Results

It is clear from Figure 6 that high-CRT sessions led to prices which were closer to the true value of the asset (i.e, RE prediction) than baseline sessions. This observation is in line with Conjecture 1. In line with Conjecture 2, we observe that average prices converge to the true value of the asset, which corresponds to the rational expectations predictions, only for the high-CRT sessions with common information.[17]

[16] See CDP for a detailed description of the measures used in the survey.
[17] In addition to our graphical illustration of conjectures, we direct the reader to video links showing examples of the differences in information aggregation across treatments. In the following links, one can replay Market 17 (last period of the experiment where the true value of the asset is 490) for one baseline session, one high-CRT session without common information and one high-CRT session with common information: (https://sites.google.com/site/financecognitive/videos).

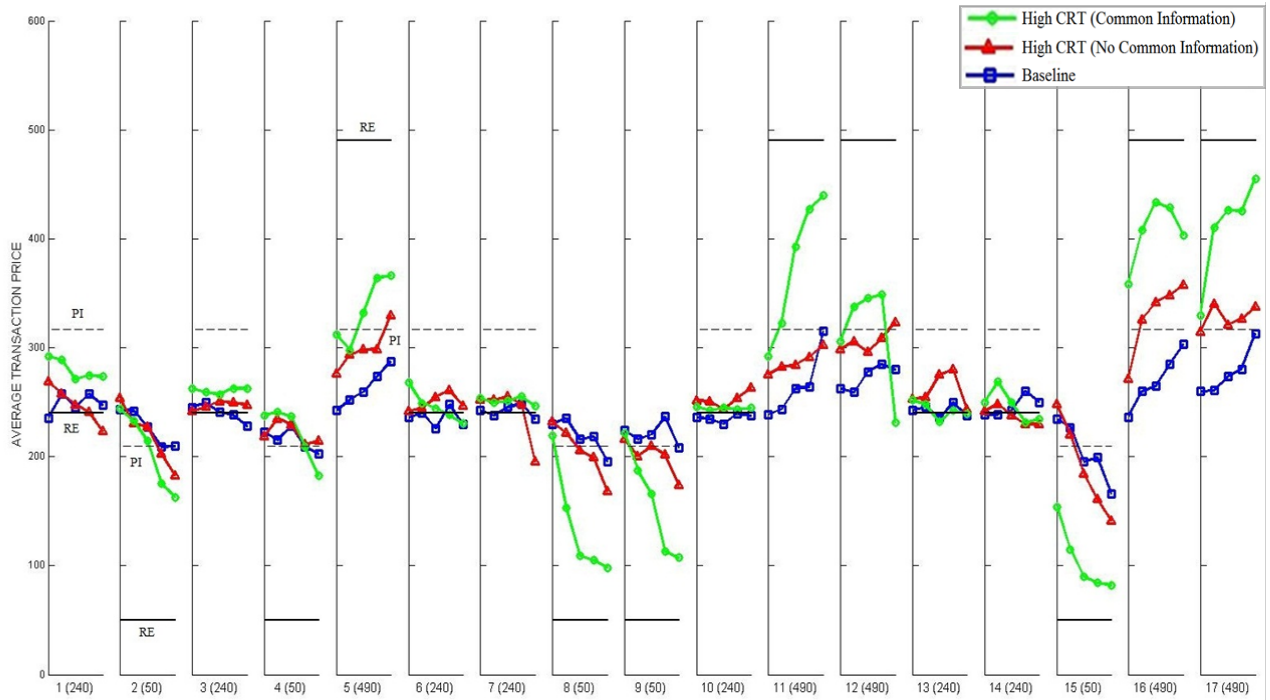


Figure 6. Average price per minute over the 4 high-CRT with common information (green circle markers), the 4 high-CRT without common information (red triangle markers), and the 10 baseline (blue square markers) sessions for each of the 17 markets. The true value of the asset is denoted at the bottom of each subfigure, i.e., 50, 240, and 490. The rational expectations value (RE) is indicated by a horizontal line, and the prior information value (PI) is indicated by a dashed line.

In Table 8, we provide statistical support for our conjectures by computing the mean absolute price deviations with respect to each of the competing models for each treatment. We compute this variable for all the transactions in markets 14, 15 and 17 (as in PS) as well as for the last three transactions in these markets. In our statistical analysis, we will refer to both measures.

We first observe that the mean absolute price deviation computed with respect to rational expectations predictions for the high-CRT sessions with common information (28.28 for the last three transactions) corresponds to a proportion of *reflective* traders greater than $\frac{11}{12}$ (that is, at most one subject per market is a *noise* trader) in our simulations (see Table 6). This suggests that our recruitment of *reflective* traders was effective in limiting the number of *noise* traders in the market.[18]

In line with Conjecture 1, we show that the mean absolute price deviation with respect to rational expectations is smaller in the high-CRT sessions than in the baseline sessions whether we consider common information of trader's level of *reflection* (*p-value* = 0.005 for all transactions as well as for

[18] Applying the same mean absolute price deviation metric as in the baseline sessions implies that, on average, approximately two traders per market are *noise* traders. This suggests that our recruitment of high-CRT subjects had the intended effect of reducing the number of *noise* traders in the market.

the last three transactions) or not (*p-value* = 0.005 for all transactions and *p-value* = 0.157 for the last three transactions) (see Table 8).

Table 8. Comparison of mean absolute price deviations across models for each asset value and for all [last three] transactions in markets 14, 15 and 17.[19]

Treatments	*Mean absolute price deviation*			
	PI	GPI	RE	MM
Baseline	64.69	**47.64**	131.28	79.21
	[71.40]	[**65.79**]	[115.78]	[81.73]
Loan	**138.09**	156.12	216.93	241.50
	[**151.11**]	[163.00]	[226.67]	[247.22]
Top 20% CRT (*No common information*)	57.98	**51.10**	106.71	85.69
	[**70.19**]	[73.08]	[90.08]	[71.14]
Top 20% CRT (*Common information*)	88.87	98.89	**59.89**	72.64
	[114.50]	[132.33]	[**28.28**]	[85.83]

In line with Conjecture 2, the rational expectations model is a better predictor of asset prices in the high-CRT sessions with common information than in the high-CRT sessions without common information (*p-value* = 0.083 for all transactions and *p-value* = 0.043 for the last three transactions). Moreover, the rational expectations price predictions outperform those of the other models only in the sessions with high-CRT traders and common information regarding the level of trader *reflection*. Indeed, when comparing the mean absolute price deviations of RE to PI {GPI} <MM> we obtain the following *p-values* 0.083 {0.043} <0.563> for all transactions [all *p-values* = 0.021 for the last three transactions]. For the high-CRT treatment without common information the RE model's predictions are not as good as those provided by PI {GPI} <MM> (*p-value* =0.021 {0.021} <0.248> for all transactions and all *p-values* > 0.1 for the last three transactions).

The fact that common information of the traders' level of *reflection* is essential for information aggregation is consistent with recent works in the asset bubble experiment literature. These studies show that uncertainty regarding other traders' level of sophistication (*reflection*) can partly account for the emergence of bubbles. For example, Cheung, Hedegaard and Palan (2014) show that mispricing in bubbles experiments is least pronounced when subjects are trained to compute the fundamental value of the experimental asset and when this training is common information to traders. Making training common information reduces the uncertainty regarding market subjects'

[19] As before, we use the PS analysis to compare treatments but alternative analyses would yield the same qualitative results. For example, we obtain similar results using all markets in the analysis or using only the last three transactions of all markets.

understanding of the fundamental value of the asset. To control for strategic uncertainty in asset markets, Akiyama, Hanaki and Ishikawa (2013) designed experimental markets with one human trader and five computer traders which were programmed to follow a fundamentalist strategy. In that context, the authors showed that price forecasts of human traders were substantially closer to the fundamental value of the asset than in markets composed only of human traders. They interpret this finding as evidence that strategic uncertainty regarding other traders' rationality may explain a large part of asset market mispricing.

5. Conclusions

Our findings stress that information aggregation is difficult to achieve in asset markets, even in the simplest case in which the asset may only assume one of three possible values and ambiguity is absent. To account for these findings, we developed a model in which traders may or may not have the ability to reflect on asset prices to infer other traders' information. In this model *reflective* and *non-reflective* traders interact in the spirit of noisy rational expectations models. This model helped us derive two testable conjectures. First, a higher proportion of *reflective* traders in the market should lead prices to reflect the true value of the asset more closely. Second, information aggregation should only be achieved if it is common information to *reflective* traders that the market is populated by *reflective* traders. We tested these conjectures by recruiting *reflective* individuals defined as having scored in the top 20% of all individuals in our subjects' database on the CRT. Consistent with our first conjecture, we show that recruiting *reflective* individuals led to asset prices that more accurately reflected the true value of the asset. In line with our second conjecture, information aggregation only occurred when the *reflective* traders populating the market were aware of each other's high level of *reflection*. Our findings are thus consistent with the original PS results (Market 9) showing evidence of information aggregation in a setting (Caltech) where subjects were *reflective* (e.g., Brocas et al. 2014) and aware of each other's *reflective* capacity.

Using experimental asset markets allowed us to directly control important elements of the environment which enabled us to test competing models of information aggregation. Controlling the flow of information into the market, the traders' level of *reflection* as well as the common information regarding traders' *reflective* capacity would not have been possible in naturally occurring markets. Our study thus exemplifies the many benefits of the experimental approach for the study of market efficiency and information aggregation initiated by Plott and Sunder (1982, 1988). By highlighting the complexity of information aggregation in asset markets, our findings will hopefully revive experimental works on the topic. Recognizing the difficulty of experimental markets to aggregate information stresses the need for further experimental research focused on assessing the behavioral and institutional roots of market efficiency. For example, it remains to be seen whether information aggregation can be achieved in more complex environments (with more states of the world and ambiguous information) even in the most favorable case in which traders are highly *reflective* and this high level of *reflection* is common information.

Policy implications are also mounting. For example, our work would validate the growing concerns regarding the lack of sophistication of the new wave of investors in the Chinese stock markets and its role in current asset mispricing (e.g., Lopez, 2015). Also, the advent of online brokers has facilitated stock market access for individual investors (Bogan, 2008). Many such investors lack sophistication and training, which may not only affect their wealth but could also induce significant mispricing in the market. A natural question for regulators is whether individuals opening trading accounts should attest some degree of financial training and sophistication. This would not only increase the degree of sophistication (*reflection*) of traders but also make traders' sophistication common information, which, as we have seen, is a crucial ingredient of informational efficiency of markets.

6. References

Agnew, J., P. Balduzzi, and A. Sunden (2003) "Portfolio Choice and Trading in a Large 401(k) Plan," *American Economic Review* 93, 193-215.

Akiyama, E., N. Hanaki and R. Ishikawa (2013) "It is not Just Confusion! Strategic Uncertainty in an Experimental Asset Market," Aix-Marseille School of Economics Working Paper 2013-40.

Andersen S., J. Campbell, K. Meisner-Nielsen and T. Ramadorai (2015) "Inattention and Inertia in Household Finance: Evidence from the Danish Mortgage Market," NBER Working Paper 21386.

Biais, B., D. Hilton, K. Mazurier and S. Pouget (2005) "Judgmental Overconfidence, Self-Monitoring and Trading Performance in an Experimental Financial Market," *The Review of Economic Studies* 72(2), 287-312.

Bloomfield, R., M. O'Hara and G. Saar (2009) "How Noise Trading Affects Markets: An Experimental Analysis," *Review of Financial Studies* 22(6), 2275-2302.

Bogan, V. (2008) "Stock Market Participation and the Internet," *Journal of Financial and Quantitative Analysis* 43, 191-211.

Brocas, I., C. Camerer, J. Carrillo and S. Wang (2014) "Imperfect Choice or Imperfect Attention? Understanding Strategic Thinking in Private Information Games," *Review of Economic Studies* 81, 944-970.

Caginalp, G., D. Porter and V. Smith (1998) "Initial Cash/Stock Ratio and Stock Prices: An Experimental Study," *Proceedings of the National Academy of Sciences* 95(2), 756-61.

Caginalp, G., D. Porter and V. Smith (2001) "Financial Bubbles: Excess Cash, Momentum, and Incomplete Information," *Journal of Behavioral Finance* 2(2), 80-99.

Campitelli, G. and M. Labollita (2010) "Correlations of Cognitive Reflection with Judgments and Choices," *Judgment and Decision Making* 5, 182-191.

Cheung, S., M. Hedegaard and S. Palan (2014) "To See is to Believe: Common Expectations in Experimental Asset Markets," *European Economic Review* 66, 84-96.

Cokely, E.T. and C.M. Kelley (2009) "Cognitive Abilities and Superior Decision Making Under Risk: A Protocol Analysis and Process Model Evaluation," *Judgment and Decision Making* 4, 20-33.

Corgnet, B., M. DeSantis and D. Porter (2015) "What Makes a Good Trader? On the Role of Quant Skills, Behavioral Biases and Intuition on Trading Performance" Economic Science Institute working paper 15-17.

Corgnet, B., R. Hernan-Gonzalez, P. Kujal and D. Porter (2015) "The Effect of Earned Versus House Money on Price Bubble Formation in Experimental Asset Markets," *Review of Finance* 19 (4), 1455-1488.

Crack, T. (2014). "Heard on the Street: Quantitative Questions from Wall Street Job Interviews," Dunedin, New Zealand.

Daniel, K., D. Hirshleifer and A. Subrahmanyam (1998) "Investor Psychology and Security Market Under- and Overreactions," *The Journal of Finance* 53, 1839-1885.

Daniel, K., D. Hirshleifer and A. Subrahmanyam (2001) "Overconfidence, Arbitrage and Equilibrium Asset Pricing," *The Journal of Finance* 56, 921-965.

Desgranges, G. and R. Guesnerie (2005) "Common Knowledge and the Information Revealed Through Prices: Some Conjectures," in *Assessing Rational Expectations 2: Eductive Stability in Economics*," *MIT Press*.

Diamond, D. and R. Verrecchia (1981) "Information Aggregation in a Noisy Rational Expectations Economy," *Journal of Financial Economics* 9(3), 221-235.

Fama, E. (1970) "Efficient Capital Markets: A Review of Theory and Empirical Work," *The Journal of Finance* 25(2), 383-417.

Fama, E. (1991) "Efficient Capital Markets: II," *The Journal of Finance* 46(5), 1575-1617.

Fama, E. (2008) "A Brief History of the Efficient Market Hypothesis," *American Finance Association* speech.

Frederick, S. (2005) "Cognitive Reflection and Decision Making," *Journal of Economic Perspectives* 19(4), 25-42.

Grinblatt, M., Keloharju, M. and J. Linnainmaa (2011) "IQ and Stock Market Participation," Journal of Finance 66:2121-2164.

Grinblatt, M., Keloharju, M. and J. Linnainmaa (2012) "IQ, Trading Behavior, and Performance," Journal of Financial Economics, 104(2):339-362.

Grossman, S. (1977) "The Existence of Future Markets, Noisy Rational Expectations and Informational Externalities," *Review of Economic Studies* 64(3), 431-449.

Guesnerie, R. (2005) "Assessing Rational Expectations 2: Eductive Stability in Economics," *MIT Press*.

Hanson, R., R. Oprea and D. Porter (2006) "Information Aggregation and Manipulation in an Experimental Market," *Journal of Economic Behavior and Organization* 60, 449-459.

Hong, H. and J. Stein (1999) "A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets," *The Journal of Finance* 54, 2143-2184.

Kahneman, D. (2011) "Thinking, fast and slow," *Macmillan*.

Krahnen, J. and M. Weber (2001) "Marketmaking in the Laboratory: Does Competition Matter?," *Experimental Economics* 4(1) 55-85.

Krische, S. (2015) "Who is the Average Individual Investor? Numerical Skills and Implications for Accounting Research" SSRN working paper.

Lintner, J. (1969) "The Aggregation of Investor's Diverse Judgments and Preferences in Purely Competitive Security Markets," *Journal of Financial and Quantitative Analysis* 4(4), 347-400.

Lopez, L. (2015) "6% of China's Newest Stock Market Investors Can't Read," *Business Insider*. http://www.businessinsider.com/6-of-chinas-newest-stock-market-investors-cant-read-2015-3

Noussair, C., S. Tucker and Y. Xu (2014) "A Futures Market Reduces Bubbles but Allows Greater Profit for More Sophisticated Traders," Working Papers in Economics 14/12, University of Waikato, Department of Economics.

Oechssler, J., A. Roider and P.W. Schmitz (2009) "Cognitive Abilities and Behavioral Biases," *Journal of Economic Behavior & Organization* 72(1), 147-152.

Plott, C. and S. Sunder (1982) "Efficiency of Experimental Security Markets with Insider Information: An Application to Rational expectations Models," *Journal of Political Economy* 90, 663-698.

Plott, C. and S. Sunder (1988) "Rational Expectations and the Aggregation of Diverse Information In Laboratory Security Markets," *Econometrica* 56, 1085-1118.

Radner, R. (1982): "Equilibrium under Uncertainty," in *Handbook of Mathematical Economics* II, *North-Holland Publishing Company*.

Shiller, R. (2015) "Irrational Exuberance," 3rd edition, *Princeton University Press*.

Shleifer, A. (2000) "Clarendon Lectures: Inefficient Markets," *Oxford University Press.*

Sinayev A. and E. Peters (2015) "Cognitive Reflection vs. Calculation in Decision Making," *Frontiers in Psychology* 6, 532.

Thaler, R. (2005) "Advances in Behavioral Finance, Volume II (Roundtable Series in Behavioral Economics)," *Princeton University Press.*

Thoma, V, E. White, A. Panigrahi, V. Strowger and I. Anderson (2015) "Good Thinking or Gut Feeling? Cognitive Reflection and Intuition in Traders, Bankers and Financial Non-Experts," *PLoS One* 10(4), e0123202.

Toplak, M., R. West and K. Stanovich (2011) "The Cognitive Reflection Test as a Predictor of Performance on Heuristics and Biases Tasks," *Memory & Cognition* 39, 1275-1289.

Toplak, M., R. West and K. Stanovich (2014) "Assessing miserly information processing: An expansion of the Cognitive Reflection Test," *Thinking & Reasoning* 20(2), 147-168.

Tversky, A. and D. Kahneman (1974) "Judgment under Uncertainty: Heuristics and Biases," Science 185, 1124-1131.

Veiga, H. and M. Vorsatz (2010) "Information Aggregation in Experimental Asset Markets in The Presence Of A Manipulator," *Experimental Economics* 13(4), 379-398.

Zhou, X. (2008) "A Practical Guide To Quantitative Finance Interviews," CreateSpace Independent Publishing Platform; 14th edition.

6. Appendices

Appendix A. Tables and figures

Table A1: Summary of experimental designs comparing our study with previous related works.

Authors	Number of traders	Number of markets (market length in minutes) - Sessions -	Loan	Endowment [cents] (Assets)	Asset values [cents] (Probabilities)	Training on the random process	Trading mechanism
Plott and Sunder (1988) (*market 9*)	12	17 (7) - 1 -	Yes	2,500 (4)	5, 24, 49 (0.35,0.45,0.20)	Yes	Oral continuous double auction
Our study (2015) Baseline / Loan	12	17 (5) - 12 -	No/Yes	120 / 2,500 (4)	5, 24, 49 (0.35,0.45,0.20)	Yes	Computerized continuous double auction
Biais et al. (2005)	8 to 18	4 (7) - 26 -	No	250 (4)	5, 24, 49 (1/3, 1/3, 1/3) No monetary incentives	No	Oral continuous double auction and call auction
Hanson, Oprea and Porter (2006) Baseline	12	8 (5) - 4 -	No	125 (2)	0, 25, 62.5 (1/3, 1/3, 1/3)	No	Computerized continuous double auction
Veiga and Vorsatz[20] (2010) Baseline	12	7 (5) - 6 -	Yes	250 (4)	12.5, 37.5, 52.5 (1/3, 1/3, 1/3)	No	Computerized continuous double auction

[20] This study was conducted in Europe and the values refer to euro cents.

Table A2. Comparison of actual prices to model predictions at the end of each market for previous research data. As in PS, only markets 14, 15, and 17 are considered.

		Mean Absolute Price Deviation				*Percentage of Convergent Price Changes*			
Treatment	Session	PI	GPI	RE	MM	PI	GPI	RE	MM
HOP (2006)	1	8.38	21.37	40.62	43.96	57%	61%	69%	61%
	2	17.45	4.17	25.05	19.82	61%	65%	70%	67%
	3	14.12	9.42	31.86	26.68	65%	61%	63%	62%
	4	8.97	25.43	38.04	44.29	68%	67%	61%	67%
VV (2010)	1	45.41	47.16	95.69	88.55	69%	62%	68%	62%
	2	51.63	60.34	102.25	103.09	69%	68%	79%	68%
	3	61.04	92.65	153.07	189.09	82%	78%	82%	80%
	4	86.08	96.01	162.45	156.78	59%	57%	62%	55%
	5	56.90	55.64	106.93	102.52	73%	68%	76%	70%
	6	41.52	51.33	68.62	70.92	75%	65%	78%	63%
Total average		**39.15**	46.35	82.46	84.57	68%	65%	**71%**	66%
PS (1988)	*Market 9*	136.00	–	**0.00**	83.00	–	–	–	–

Table A3. Average percentage of convergent price changes with respect to the true value calculated across all three possible values (50, 240 and 490) across 1,000 simulations.

α_b	Common Information $\alpha = \alpha_b$	No Common Information $\alpha = 1$
0	50%	61%
1/12	50%	62%
2/12	50%	66%
3/12	50%	71%
4/12	50%	72%
5/12	50%	79%
6/12	50%	77%
7/12	50%	83%
8/12	51%	84%
9/12	52%	84%
10/12	54%	84%
11/12	61%	83%
1	**99%**	**99%**

Table A4. Distribution of extended CRT scores
for the 885 students in the lab database who took the survey.

CRT score	% of students
0	12.77
1	19.32
2	18.42
3	14.01
4	13.33
5	9.83
6	7.23
7	5.08

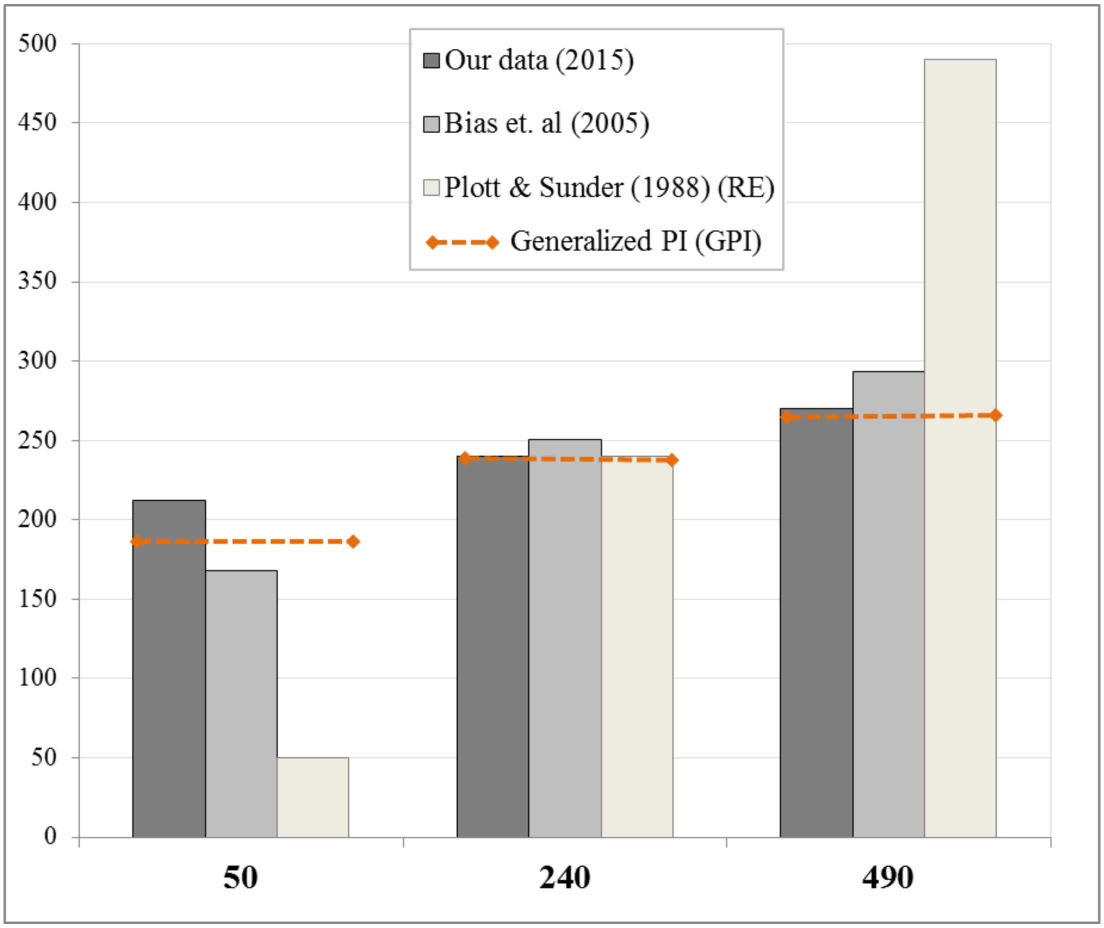


Figure A1. Average prices for asset values 50, 240 and 490 across studies.

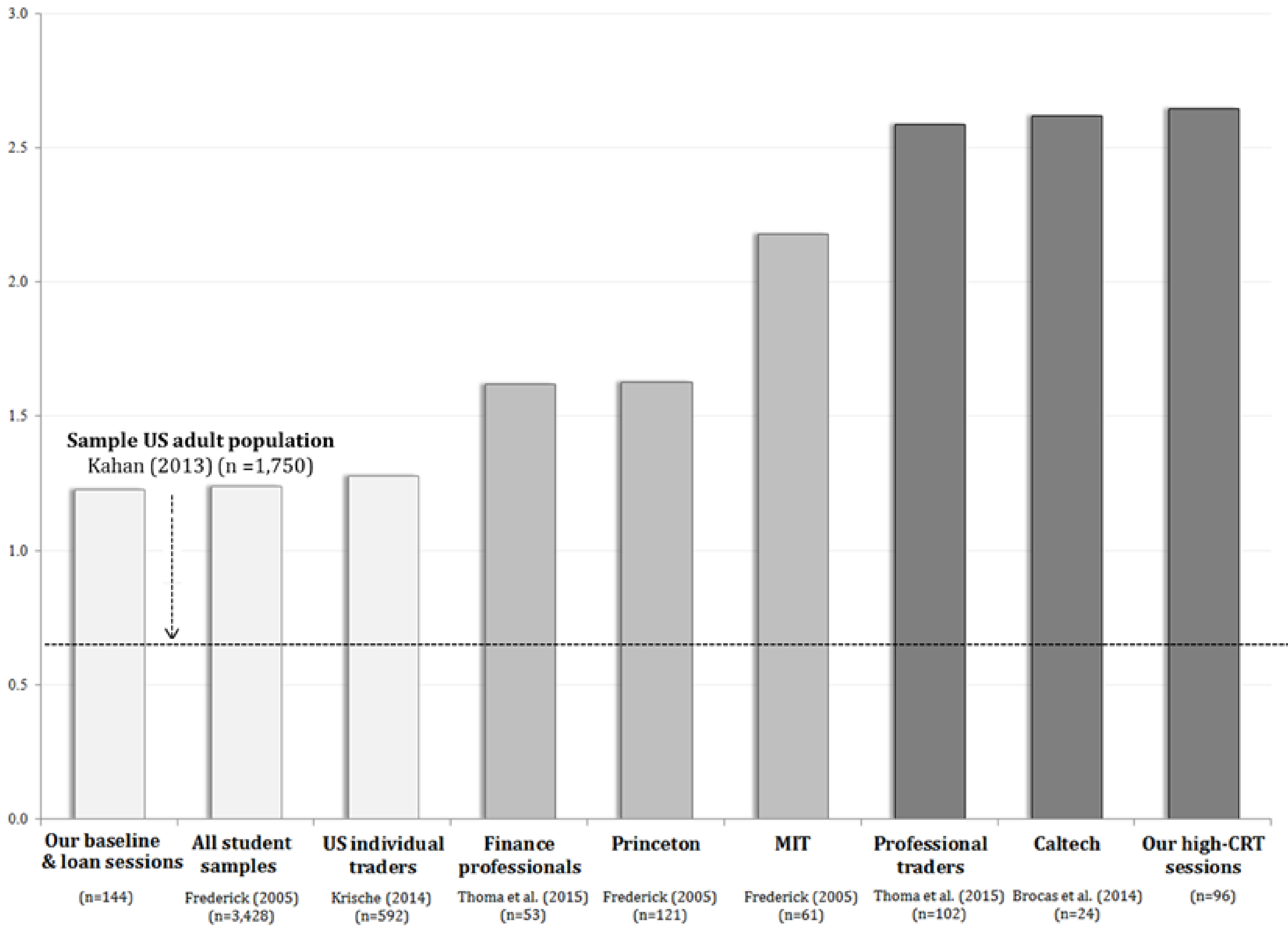


Figure A2. Average three-item CRT scores for a wide range of samples.

Sinayev and Peters (2015) also suggest that the average three-item CRT is below 1 in the general US population (n = 2,703).

<u>Online Appendices</u>

Appendix B. Instructions (click here)

Appendices C to G. (click here)