Financial Market Instability: Myths and Truths

Carlos Cueva,[1] Aldo Rustichini[2]

[1]Faculty of Economics, University of Cambridge,
Austin Robinson Building, Sidgwick Ave, Cambridge, CB3 9DD, UK
[2]Department of Economics, University of Minnesota
1925 4th Street South 4-101, Hanson Hall, Minneapolis, MN 55455-0462, USA

The hypothesis that gender composition of financial markets affects price stability has been frequently suggested in policy discussions. We test this hypothesis, and compare the effect of gender composition with other group characteristics such as confidence, risk attitude and cognitive skills. We conduct 10 male-only, 10 female-only and 10 mixed-gender experimental asset markets. Male and female markets are comparable in terms of volatility and deviations from fundamentals, mixed gender markets are substantially more stable. Higher average cognitive skills of the group reduce market volatility. Analysis of individual behavior supports the group findings, showing that subjects with higher cognitive skills trade more rationally and thus earn significantly higher profits. Mixed, rather than homogeneous, gender composition and high cognitive skills are equally important for financial market stability.

The nature and causes of financial market instability have long been a topic of utmost importance in economics. The devastating global consequences of the financial crisis of 2008 have brought this topic to the spotlight more than ever, attracting media attention on an unprecedented scale. There are many sides to the emergent public debate, but the underlying theme remains: *what policies can effectively encourage financial stability?*

A variety of responses to this question have focused on the attitude to risk-taking of participants in the finance industry. A possible reason for concern is that the vast majority of professional traders are men: for instance, a recent study found that the percentage of female equity analysts at Wall Street brokerage firms dropped from 16% in 1996 to under 14% in 2005 (*1*). The idea that greater female participation could reduce reckless risk-taking and foster financial stability has received a great deal of attention.[1] In support of this view, there is extensive evidence that, on average, men are more risk-taking than women (see e.g. review (*2*)) and more motivated to compete (*3–5*). Men also tend to be more overconfident in their ability to make investment decisions, which can lead them to trade excessively and incur greater losses than their female counterparts (*6*). More recently, circulating testosterone levels and second-to-fourth digit ratios, which are thought to indicate pre-natal exposure to testosterone, have been found to predict financial risk-taking (*7–10*), performance and endurance in the securities trading business (*11, 12*), and career choices in finance (*9*). This hormone is present in much higher concentrations in men than in women and has been associated with aggression (*13, 14*), reduced sensitivity to punishment (*15*), impaired empathy (*16*) and decreased generosity (*17*).

The complex and non-replicable nature of financial market phenomena makes clear predictions regarding the effect of policy intervention elusive. In this respect, experimental economics has been a very fruitful approach, thanks to its ability to provide replicable data in simplified environments. Experimental asset markets have been used to evaluate the impact of policies such as restrictions on short-selling (*18–20*), futures contracting, margin buying, limit price change rules or brokerage fees (*18, 21*), and have been proposed as a valuable tool for assessing the

[1] see for example New York Times, 2009: http://www.nytimes.com/2009/02/01/business/worldbusiness/01iht-gender.3-420354.html; New York Magazine, 2010: http://nymag.com/news/businessfinance/64950/; The Guardian, 2011: http://www.guardian.co.uk/world/2011/jun/19/neuroeconomics-women-city-financial-crash

effectiveness of antibubble laws (*22*).

We employ established methodology in this field to address the conjecture that gender composition of the market influences price stability, by studying the stability of experimental asset markets with different gender composition. We investigate the size of this hypothetical effect in relation to other relevant characteristics of the market such as cognitive skills, risk aversion and confidence levels. Cognitive skills were measured by a subject's performance in two unrelated tasks: a standard test of nonverbal IQ consisting of a series of pattern matching tasks (*23*), and a backward-induction solvable computer game called "hit-15", designed to test planning ability and strategic thinking (*24*). Risk aversion was measured by a task in which subjects had to choose between a set of 15 pairs of gambles (*25*). Confidence was measured by subjects' guesses about their relative performance compared to the rest of the group immediately before trading. All tasks were rewarded in an incentive compatible way.[2]

In the experimental asset market, a group of volunteers could trade cash and assets in a computerized double-auction mechanism (*26, 27*). A market consisted of 15 trading periods, lasting 2 minutes each. During a trading period, traders could post bids (offers to buy) or asks (offers to sell) for an asset at a price of their choice. All bids and asks were visible to the group, and any trader could accept them and execute the transaction at any time, provided he or she had sufficient funds. Each trader was initially endowed with 10 assets and a loan of 10,000 "francs" cash (the experimental currency). After each trading period, assets paid a random dividend or cost from a known distribution with zero expected value. At the end of the last period, each asset was paid a maturity value of 360 francs. The exchange rate for actual payment was 360 francs $= 1\ GBP$. The typical market size was 10 traders. A total of 284 volunteers took part

[2]A minor exception is the relative confidence measurement: since the score is bounded between 0 and 100, and we rewarded subjects for being within 10% of the true value, a subject who believes to be the best in the group would maximize her chance of winning by selecting 90, rather than 100. The same is true for a subject who believes to be the worst in the group. However, this is unlikely to have had any effect on behavior, since only 3 subjects in our sample selected a number above 89, and these in fact selected 100.

in the experiment. We conducted 10 female-only markets, 10 mixed-gender markets, and 10 male-only markets. The trading room was open-plan, so that traders could observe the gender composition of the market. We did not remark or make this feature salient in any way[3]. Asks and bids were displayed anonymously on traders' computer screens, so that they could not link any offer to the identity of a specific trader.

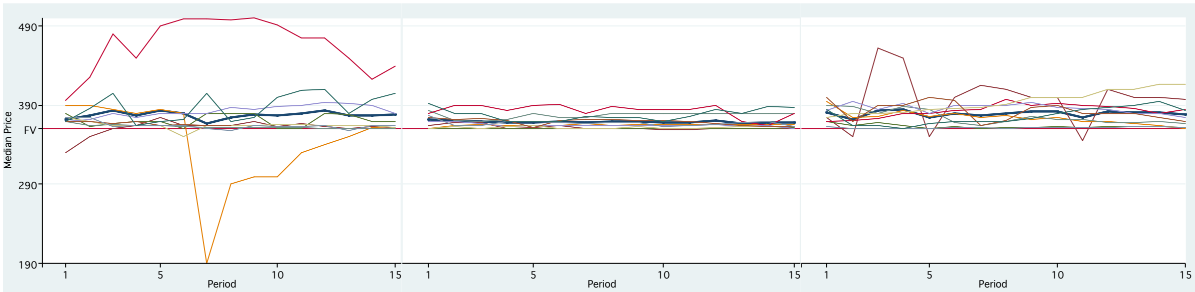
To measure price instability in each market we compute its amplitude: the difference between the maximum and minimum median transaction price observed in a market divided by the fundamental value of the asset, where median transaction prices are calculated for each trading period. Formally: $Amplitude = (max_{t \in T}(P_t) - min_{t \in T}(P_t))/F$; where P_t is the median transaction price at period t, T is the set of 15 periods, and F is the fundamental value of the asset. Since each asset pays on average zero dividends in every period and 360 francs at the end, the fundamental value of the asset is always 360 francs.[4]

Figure 1 shows median transaction prices by period in every market, grouped according to their gender composition. Visual inspection reveals that mixed-gender markets display much greater price stability than female-only or male-only markets. To test the statistical significance of this difference, we compare amplitude across the three groups. Pairwise comparisons using Mann-Whitney U-tests confirm a large and significant difference between single-gender and mixed-gender markets, and no difference between female-only and male-only markets (female-only vs. male-only: $p = 0.910$; female-only vs. mixed-gender: $p = 0.016$; male-only vs. mixed-gender: $p = 0.013$). Grouping together single-gender markets and comparing their amplitude with mixed-gender markets gives an even stronger statistical difference ($p = 0.004$).

We evaluate the possible effect of other relevant characteristics of the market, and compare it with gender composition. This is done by estimating a linear regression model of amplitude

[3]For instance, the invitation email did not mention any gender requirements, instead, we sent identical invitations to female and male volunteers separately

[4]Another measure commonly used to quantify market volatility is normalized absolute deviation. Similar results are obtained when using this alternative measure (see Supplementary text)



(a) Female-only markets (b) Mixed gender markets (c) Male-only markets

Figure 1: Median transaction prices and gender composition of the market. Bold line displays group average.

(log-transformed) as the dependent variable, and cognitive skills (CS), risk aversion, confidence and the indicator variable "mixed" as regressors. Results show that only gender composition and CS are significant predictors of amplitude (see Table 1). The coefficient of CS is six or seven times larger than the coefficient of "mixed" . Since the dependent variable is log-transformed, the size of the effect of the dependent variable should be estimated by taking exponents first. For instance, according to Model 3, mixed sessions have 61% less amplitude than single gender sessions ($e^{-0.947} - 1 = -0.61$). Similarly, a 10% increase from the mean in group CS would lead to a 30% reduction in amplitude ($e^{-5.463 \times 0.1 \times mean(CS)} - 1 = -0.30$). A more substantial increase in average CS score of the group, say, of 30% or 50% from the mean, would lead to a reduction in amplitude of 66% and 83% respectively[5]. Figure 2 displays the key findings with regards to gender composition, CS and price instability.

To understand how these aggregate effects are produced we investigate individual trading behavior. The variable $trade\ gain$ measures a trader's profits from transactions in each period: $trade\ gain_{i,t} = \sum_j (sale\ price^j_{i,t} - F) + \sum_k (F - purchase\ price^k_{i,t})$, where $sale\ price^j_{i,t}$ and $purchase\ price^k_{i,t}$ are the selling and buying prices of the j^{th} and k^{th} transaction executed by trader i in period t. Random-effects panel regression estimation shows that trade gain is strongly

[5]On average, markets had a mean CS score of 0.6620. Since CS is bounded between 0 and 1, the maximum percentage increase possible from the mean is approximately 51%.

Dep. Variable: Amplitude	1 Coefficient (Std. Error)	2 Coefficient (Std. Error)	3 Coefficient (Std. Error)
mixed	-0.842* (0.336)	-0.923** (0.343)	-0.947** (0.350)
CS	-6.032*** (1.620)	-5.636** (1.771)	-5.463** (1.712)
risk aversion	-0.188 (0.558)	-0.191 (0.564)	
confidence	0.035 (0.020)		

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1: OLS Estimates of Amplitude (log-transformed, N = 30). Bootstrapped std. errors using 10000 replications

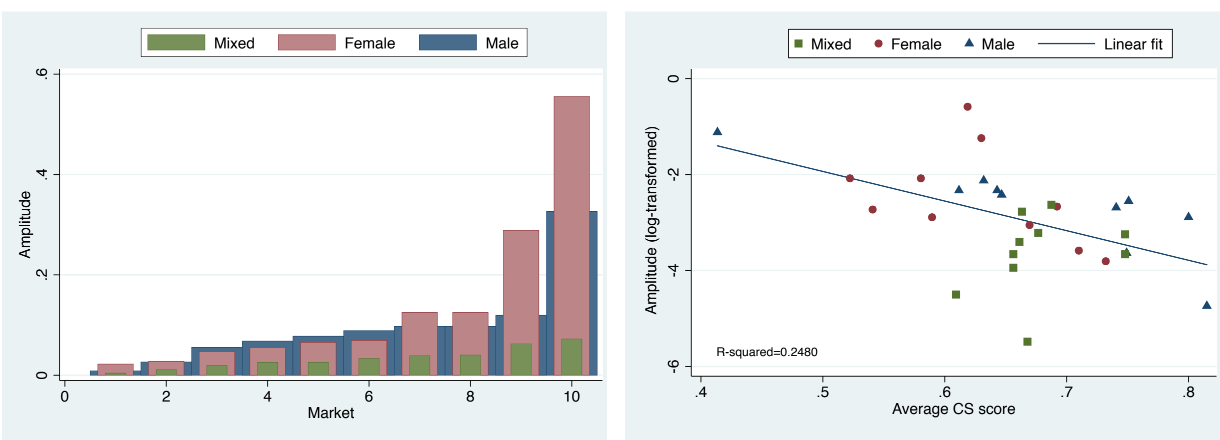


Figure 2: Panel (a) Markets are divided into each gender composition and ranked according to their amplitude. Hence, male-only, female-only and mixed-gender markets are numbered from 1 to 10, according to their ranking in amplitude within markets with similar gender composition. Panel (b) Scatter plot and best linear prediction of amplitude (log-transformed) against average cognitive skills of the market. Observations are also labelled according to the gender composition of the market.

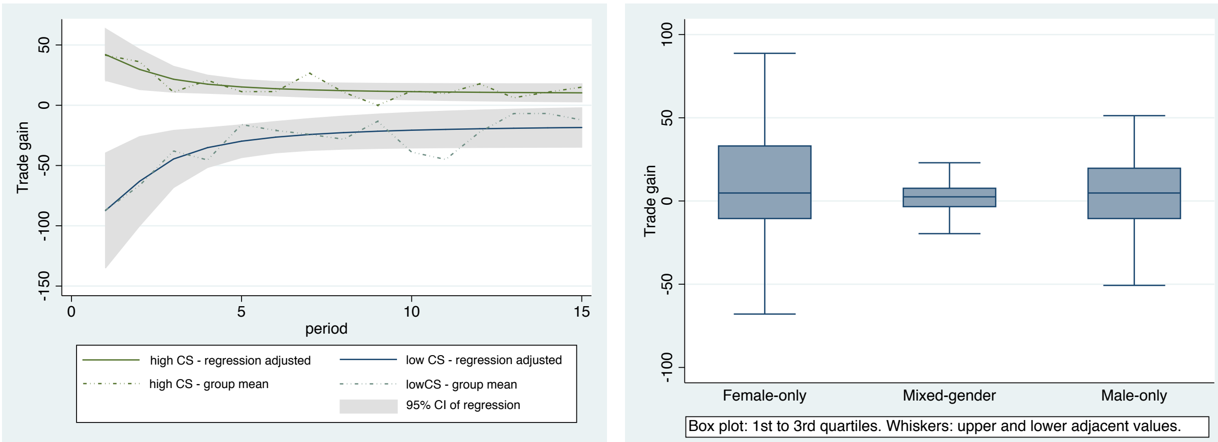


Figure 3: Fig. 3(a) Trade gain over time for traders with CS scores in the top and bottom third of our sample. Curves fitted from fractional polynomial regressions of *trade gain* on *period*. Fig 3(b) Distribution of trade gain in each treatment.

influenced by CS, and not by gender. On the other hand, there are strong learning effects: as the market progresses, high and low CS traders reduce the gap in performance, although not completely (see Fig. 3 (a); regression results can be found in Table 6 in the Supplementary text). The aggregate impact of CS is large: for instance, traders with CS score in the top 10% of our sample earned on average $10.41 GBP$, whereas those in the bottom 10% only earned $5.93 GBP$.

Even though male and female traders do not behave differently at the individual level, the gender composition of the market has a strong effect on the dispersion of trading performance. Specifically, the variance of trade gain across subjects is much larger in single-gender sessions than in mixed gender sessions (Levene's robust test statistic for equality of variances: $p = 0.0086$), and is particularly large in female-only sessions (see Fig. 3 (b)).

In the next stage of analysis, we evaluate more precisely how traders' rationality can affect profitability and price stability. To this end, we develop a model of rational traders facing noisy traders and compare observed behavior to the predictions of the theory[6]. The key assumption of

[6] see Supplementary text for further details on the model

the model is that rational and non-rational traders traders coexist in the market (we refer to non-rational traders as "noisy"). Rational traders can make profits trading with noisy traders who post bids and asks at prices away from the fundamental value of the asset. The model shows that profitable behavior of rational traders does not consist in just trading at fundamental value: in markets with ample liquidity and limited availability of assets such as ours, rational traders may be induced to purchase assets at prices above the fundamental value. This is necessary to avoid running out of assets before the end of the market, which would result in missing potential opportunities to sell at inflated prices. The model therefore predicts some degree of speculation even of rational agents, who can buy at prices higher than the fundamental value to sell later at an even higher price to noisy traders.

In the model, the time pattern of prices depends on the relative weight of rational and noisy traders. With few or no noisy traders, prices are flat at the fundamental value. As the fraction of noisy traders increases, prices deviate more from fundamental value, with rational traders accepting to pay a higher (speculative) price in anticipation of profitable re-sale later.

In our estimates, we take the measure of cognitive skill as an indicator of the degree of rationality of a trader. Results confirm that a trader's cognitive skill predicts trading rationality as described by the model: it strongly decreases a trader's tendency to sell at prices below the fundamental value, but not his or her tendency to buy at prices above the fundamental value. This behavior is indeed consistent with rational trading in a market with ample cash but limited availability of assets (recall that in our experimental markets all traders received a large cash loan of 10,000 francs and a relatively small endowment of 10 assets). Crucially, gender composition of the market has a similar effect: holding cognitive skill constant, a trader is more likely to behave in a rational way if the market has a mixed-gender composition. Finally, we continue to find strong evidence of learning, reflected in a gradual convergence between traders with low and high cognitive skills (see Table 2 and Figure 4).

Dep. variable:	$Y_{i,t} = 1\ if\ min(ask_{i,t}, accepted\ bid_{i,t}) < FV$ $= 0\ otherwise$		$Y_{i,t} = 1\ if max(bid_{i,t}, accepted\ ask_{i,t}) > FV$ $= 0\ otherwise$	
	1	**2**	**3**	**4**
	Coefficient	**Coefficient**	**Coefficient**	**Coefficient**
	(Std. Error)	(Std. Error)	(Std. Error)	
$P_{t-1} - FV$	-0.003	-0.002	0.006***	0.006***
	(0.002)	(0.002)	(0.002)	(0.002)
CS	-3.350***	-2.229***	-1.213*	-0.689
	(0.561)	(0.402)	(0.477)	(0.383)
mixed	-0.716**	-0.707**	0.092	0.094
	(0.257)	(0.241)	(0.201)	(0.200)
$mixed \times female$	0.281	0.307	0.511	0.514
	(0.321)	(0.300)	(0.269)	(0.267)
period	-0.168***		-0.019	
	(0.030)		(0.024)	
$CS \times period$	0.133**		0.064	
	(0.048)		(0.034)	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

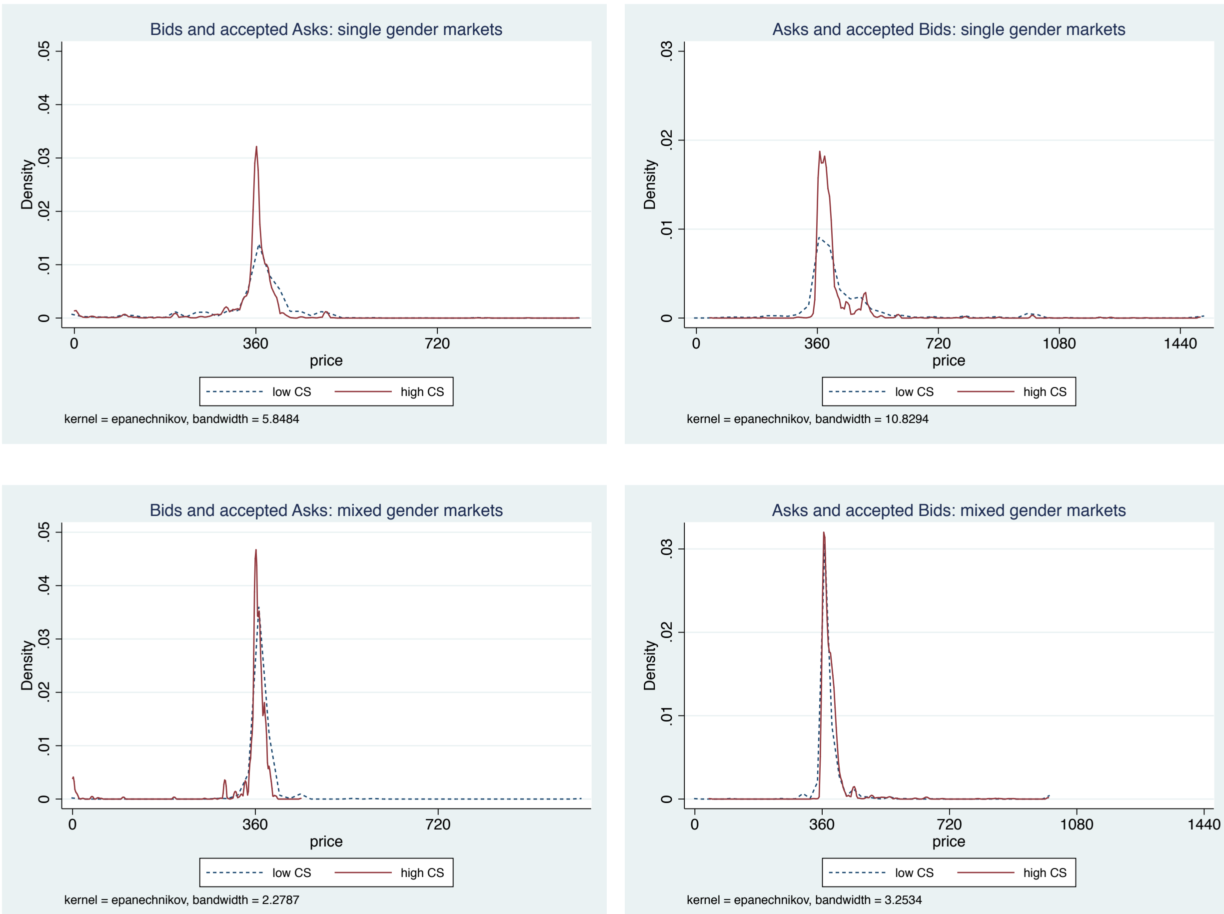
Table 2: Random Effects Probit Regressions. We construct two separate dependent variables tracking the selling behavior and the buying behavior of each subject period by period. Regression 1 and 2 (N = 3888): dependent variable = 1 if trader's minimum ask or accepted bid in a period is less than 360 francs, = 0 otherwise. Regression 3 and 4 (N = 3846): dependent variable = 1 if a trader's maximum bid or accepted ask in a period is greater than 360 francs, = 0 otherwise. A probit model is commonly used for analyzing the relationship between a dichotomous dependent variable and a set of explanatory variables. Since the regressors of interest are time-invariant, a random-effects model is required. According to the predictions of regression 2, a trader with CS score in the top third of our sample attempts a sale at a price below FV with a 1.4% probability in single-gender markets and with 0.3% probability in mixed-gender markets; on the other hand, a trader with CS score in the bottom third of our sample attempts does so with a probability of 11% in single-gender markets and 4% in mixed-gender markets. As regression 1 shows, there is significant learning for traders with lower CS but less so for traders with higher CS.

In conclusion, our results show that experimental mixed gender markets are considerably more stable than markets consisting exclusively of male or female traders. Hence, if one were to follow some policy suggestions to the extreme and "let women run Wall Street", then the outcome might be just as much instability as we have with an all-male market. If our main concern is to guarantee asset price stability, the best policy solution seems to be to encourage equal gender participation in financial markets.

Why is this mixed composition better? Most likely because it puts a brake on excessive competitive behavior that is induced in single gender environments. These results confirm earlier findings in the experimental literature on gender and competition (*3–5*), and seem therefore robust to different experimental setups, extending to experimental asset markets. In our data we also find support for the principle that "boys will be boys" (*6*): confidence is higher in males, and the mixed gender composition enhances the difference (see Figure 5). However, when compared to those of cognitive skills, the size of the destabilizing effects of excessive confidence and attitudes to risk are small.

Cognitive skills are crucial for the dynamics of price and trading in financial markets. They have two effects; they give *both* higher profit for the individual and higher stability for the market. This last effect shows that search for profits does not necessarily induce market instability. The effect of cognitive skills operates to compensate irrational exuberance: smarter traders reduce the destabilizing impact of irrational traders. Earlier results that "explained away" price bubbles in experimental asset markets by appealing to subjects' "active participation bias" (*28*) or misunderstanding of the trading environment (*29*) should be in part re-evaluated. Our model suggests an additional explanation: in a world of rational and noisy traders even rational traders may find it optimal to speculate (that is, buy at prices higher than the fundamental value). The present data validates this theoretical conclusion. In line with recent field evidence on professional traders (*12*), experience also matters: in our data, learning reduces substantially (by about three fourths) the gap in trading profits due to differences in cognitive skills.

Stability of financial markets is desirable, but difficult to achieve, and might require more than financial regulation. The present study shows that stability might be improved by a rich variety of factors. In our data, the size of the effects of gender and cognitive skills are large and comparable, a finding suggesting that cognitive skills are at least as important as the biological and hormonal factors recently investigated in the literature. Our study suggests that the most promising approach is a balanced approach, giving as much weight to the desirable potential of intelligent profit-seeking as to the dangers of reckless competition. Intelligent and balanced gender participation in financial markets appears to be an effective solution.



(a) Buying behavior (b) Selling behavior

Figure 4: Buying and selling behavior for traders with CS scores in the top third and bottom third of our sample, in markets with different gender composition. Panel (a) buying behavior: kernel density estimates of posted bids and accepted asks. The distribution of prices has a significantly higher variance, mean and median for traders with low CS than traders with high CS, regardless of gender composition of the market (Levene's robust test for equality of variance: $p < 0.0001$, Mann-Whitney U-test: $p < 0.0001$); the distribution of prices displays significantly less volatility in mixed-gender markets for both types of traders, but especially for lower CS traders (std. deviation: low CS, single-gender: 97.4; low CS, mixed-gender: 48.0; $p < 0.0001$; high CS, single-gender: 77.1; high CS, mixed-gender: 70.7; $p < 0.0001$). Panel (b) selling behavior: kernel density estimates of posted asks and accepted bids. The distribution has higher variance for traders with low CS than for traders with high CS in single-gender markets ($p < 0.0001$) but not in mixed-gender markets ($p = 0.335$). The distribution displays a sharper drop in prices below 360 (the fundamental value of the asset) in high CS traders than in low CS traders.

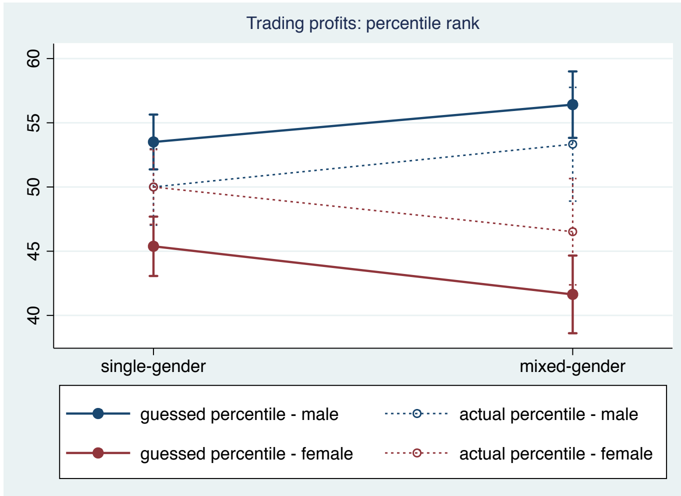


Figure 5: Confidence about relative trading performance. This figure displays traders' mean guesses about their percentile rank in the market with respect to their trading profits. Guesses are elicited before the start of the market. Hollow circles connected by dashed lines display actual mean ranks. Whiskers indicate standard errors for each estimate. Confidence levels are significantly higher in men than in women. The difference is significant even in single gender markets ($p = 0.0132$) and particularly large in mixed gender markets ($p = 0.0004$). Men make marginally higher profits in mixed markets, although the difference is not significant ($p = 0.194$).

References and Notes

1. C. Green, N. Jegadeesh, Y. Tang, *Financial Analysts Journal* **65**, 65 (2009).

2. R. Croson, U. Gneezy, *Journal of Economic Literature* **47**, 448 (2009).

3. U. Gneezy, M. Niederle, a. Rustichini, *The Quarterly Journal of Economics* **118**, 1049 (2003).

4. U. Gneezy, A. Rustichini, *American Economic Review* **94**, 377 (2004).

5. M. Niederle, L. Vesterlund, *The Quarterly Journal of Economics* **122**, 1067 (2007).

6. B. M. Barber, T. Odean, *The Quarterly Journal of Economics* **116**, 261 (2001).

7. H. Vermeersch, G. T'Sjoen, J.-M. Kaufman, J. Vincke, *Hormones and behavior* **53**, 463 (2008).

8. C. Apicella, *et al.*, *Evolution and Human Behavior* **29**, 384 (2008).

9. P. Sapienza, L. Zingales, D. Maestripieri, *Proceedings of the National Academy of Sciences of the United States of America* **106**, 15268 (2009).

10. P. Brañas Garza, A. Rustichini, *PloS one* pp. 1–17 (2011).

11. J. M. Coates, J. Herbert, *Proceedings of the National Academy of Sciences of the United States of America* **105**, 6167 (2008).

12. J. M. Coates, M. Gurnell, A. Rustichini, *Proceedings of the National Academy of Sciences of the United States of America* **106**, 623 (2009).

13. J. Archer, *Neuroscience and biobehavioral reviews* **30**, 319 (2006).

14. J. M. Carré, S. K. Putnam, C. M. McCormick, *Psychoneuroendocrinology* **34**, 561 (2009).

15. J. van Honk, *et al.*, *Psychoneuroendocrinology* **29**, 937 (2004).

16. J. van Honk, *et al.*, *Proceedings of the National Academy of Sciences of the United States of America* **108**, 3448 (2011).

17. P. J. Zak, *et al.*, *PloS one* **4**, e8330 (2009).

18. R. R. King, V. L. Smith, A. W. Williams, M. Van Boening, *Nonlinear Dynamics and Evolutionary Economics*, P. I, D. R. H, C. P, eds. (Oxford University Press, 1992), pp. 445–495.

19. L. Ackert, F. R. B. of Atlanta, *Bubbles in Experimental Asset Markets: Irrational Exuberance No More*, Working paper series (Federal Reserve Bank of Atlanta, 2002).

20. E. Haruvy, C. N. Noussair, *The Journal of Finance* **61**, 1119 (2006).

21. D. P. Porter, V. L. Smith, *Journal of Behavioral Finance* **4**, 7 (2003).

22. E. F. Gerding, *Wisconsin Law Review* (2007).

23. J. C. Raven, *Raven's Standard Progressive Matrices (SPM) and Ravens Standard Progressive Matrices Plus (SPM Plus)* (Pearson, 2000).

24. S. V. Burks, J. P. Carpenter, L. Goette, A. Rustichini, *Proceedings of the National Academy of Sciences of the United States of America* **106**, 7745 (2009).

25. C. A. Holt, S. K. Laury, *American Economic Review* **92**, 1644 (2002).

26. V. L. Smith, G. L. . Suchanek, A. W. . Williams, *Econometrica* **56**, 1119 (1988).

27. C. Noussair, S. Robin, B. Ruffieux, *Experimental Economics* **4**, 87 (2001).

28. V. Lei, C. N. Noussair, C. R. Plott, *Econometrica* **69**, 831 (2001).

29. M. Kirchler, J. Huber, T. Stöckl, *American Economic Review* **102**, 865 (2012).

30. J. Huber, M. Kirchler, *Experimental Economics* **15**, 89 (2012).

31. U. Fischbacher, *Experimental Economics* **10**, 171 (2007).

32. C. Cueva, E. Roberts, A. Rustichini, Hormonal factors in financial decision making (2012). Work in progress.

33. G. Caginalp, D. Porter, V. Smith, *Proceedings of the National Academy of Sciences* **95**, 756 (1998).

Materials and Methods

The trading task

We implement a computerized double auction market inspired by the classic design developed in (*26*). During a trading period, participants can submit any number of bids and asks, provided they have sufficient funds to complete the transaction. A bid (ask) consists of an offer to buy (sell) a single asset at a specified price. For simplicity, we do not allow block trading (i.e. submitting bids or asks for multiple assets at a time). The market operates with an "open book": all bids and asks submitted during a trading period are listed on subjects' screens, anonymously and ordered by price. A subject can accept any number of bids or asks provided he or she has sufficient funds to complete the transaction.

The design in (*26*) has been shown to be sensitive to particular kinds of subject misunderstanding. Specifically, (*29*) point out that declining fundamental values are an important source of confusion and miss-pricing. These authors find that such source of misunderstanding can be addressed by providing intuitive examples in the instructions (*29*), showing subjects a plot of the fundamental value process (*30*), or using constant fundamental values.

With this in mind, we depart from (*26*) in three ways: task instructions are written in a simpler, more intuitive way designed to be easily understood by non-economists; the trading environment is similarly intuitive, implemented with z-tree software (*31*); finally, assets have a constant fundamental value.

The use of a constant fundamental value follows (*27*). In this setup, each asset pays either -24, -16, 4 or 36 "francs" with equal probability at the end of every period , plus a maturity value of 360 francs at the end of period 15. Francs have a conversion value of 360 francs = 1 GBP. Since dividends every period have zero expected value, the fundamental value of the asset is constant at 360 francs.

At the start of the trading task, each subject receives 10 assets and a 10,000 francs loan. Payoffs at the end of the trading task (in pounds) are given by

$$(C + 360 \times A + \sum_{t=1}^{15} d_t - 10,000)/360$$

where C is final cash balance, A is final asset holdings, and d_t is total dividends or costs at period t. In order to maintain total available cash in the market constant, dividends were not added to subjects' payoffs until the end.[7]

Each trading period lasted 2 minutes. During trading, subjects could see all outstanding bids and asks in the market, all concluded transaction prices for that period, their current cash and asset holdings, and a plot of average transaction prices in every past period. At the end of a trading period dividends for that period were announced. These were the same for every asset in the market. Subjects were also provided with a summary of their total cash, assets and dividends up to that period. Before the new trading period began, subjects were asked to make a guess about the average transaction price in the next period. Each accurate guess was rewarded with an extra 10 pence at the end of the session. There were 15 trading periods in total, plus an additional practice period at the beginning.

To check that participants had understood the task correctly, they were asked to complete a 6-item questionnaire about the trading task. Subjects making any mistakes in the questionnaire were approached one by one so that any misunderstandings could be verbally addressed.

Risk aversion measurement

We implemented a simple lottery task based on (*9, 25*). Subjects were faced with a list of 15 choices. Each choice was between a risky lottery and a safe option. The risky lottery was always

[7] Another feature of classic design in (*26*) shown by (*29*) to favor overpricing is that in their markets the cash to asset ratio is increasing. Since their assets pay positive dividends in every period, the amount of cash available in the market increases over time, while the value of assets steadily declines. This would not be an issue in our design since asset values are constant and dividends are on average zero. However, we avoid any potential influence of changes in available cash in the market by only adding dividend payouts or holding costs at the end.

a payoff of 0 or 10 with equal probability. The safe option gradually increased in value for each of the 15 items. The task was designed in a way such that an individual with reasonable risk preferences would choose the risky option for the first few items and at some point switch to the safe option when the latter was high enough. For example, a risk neutral individual would select the risky option for the first 10 items and the safe option for the last four items (see Table 3).

Table 3: The lottery task

	Safe option	**Risky option**
1.	2.50	0 or 10 with equal probability
2.	2.75	0 or 10 with equal probability
3.	3.00	0 or 10 with equal probability
4.	3.25	0 or 10 with equal probability
5.	3.50	0 or 10 with equal probability
6.	3.75	0 or 10 with equal probability
7.	4.00	0 or 10 with equal probability
8.	4.25	0 or 10 with equal probability
9.	4.50	0 or 10 with equal probability
10.	4.75	0 or 10 with equal probability
11.	5.00	0 or 10 with equal probability
12.	5.25	0 or 10 with equal probability
13.	5.50	0 or 10 with equal probability
14.	5.75	0 or 10 with equal probability
15.	6.00	0 or 10 with equal probability

All 15 items were simultaneously presented on their screen and subjects could take as much time as they wished to complete and review all 15 decisions. For each subject, one of their choices would be randomly selected with equal probability at the end of the experiment for payment.

Risk aversion was measured as a subject's risk premium based on his or her choices of lotteries. That is, the difference between the expected value of the risky option ($5\ GBP$) and the lowest safe option chosen over the risky option. Our measurement is therefore bounded between 2.50 if a subject always chooses the safe option, to -1.25 if a subject always chooses the risky option.

Cognitive skills measurements

Hit 15

This task is based on (*24*). Subjects were presented with a row of 15 squares and had to try to reach the last square before the computer. The game starts from the leftmost square. Subject and computer make their moves in turns from the square selected by the previous mover and can only advance a maximum of 3 squares at a time. This game is solvable by backward induction. The unique winning strategy is to select the 3^{rd}, 7^{th}, 11^{th}, or 15^{th} square whenever possible. Subjects had 30 seconds to make each move and played up to 6 full rounds with the computer. They earned 50 pence each time they hit the 15^{th} square. The computer's strategy was designed to make a mistake in the first move of each round and to play optimally thereafter. This meant that a subject who has discovered the winning strategy could win every round of the game.

Raven's progressive matrices

We administered a subset of Raven's advanced progressive matrices (*23*). These matrices are widely used as a standard non-verbal IQ test. Subjects had 30 seconds per matrix which were presented on their computer screens and were rewarded 50 pence per correct answer. They were shown 10 matrices in total.

We computed CS as a subject's average score in *hit 15* and *Raven's matrices*, normalized between 0 and 1.

Confidence measurements

Before and after every task except the lottery, participants were asked to make a guess about their own performance. They were rewarded an extra 50 pence every time their guesses were within 10% of the true answer.

Two different measures of confidence are used. We refer to them as *absolute* and *relative* confidence. *Absolute confidence* is a subject's guess about the score obtained in a given task. For the trading task, this would be final total balance, as a percentage of initial endowment (we limited the range of possible guesses from 50% to 150% of initial endowment); for hit 15 and Raven's matrices, this would be the number of wins or correct answers respectively. *Relative confidence* is a subject's guess about his or her performance relative to the rest of the group. This was obtained by asking subjects to guess their percentile rank by selecting the position of a slider in a horizontal line, where the left end of the line is labelled "worst" and the right end is labelled "best". This response was coded on a linear scale from 0 to 100.

Hormonal measurements

In order to measure hormonal levels, subjects were asked to provide a saliva sample at the start of the experiment, after the trading task and at the end of the experiment. A scanned picture of their right hand was also taken at the start of the experiment.

This data is being examined in a different paper (*32*).

Supplementary text

Model and Equilibrium behavior

We first describe a stylized version of the typical double auction asset market experiment, where a set of traders choose sequentially in a market with a precise protocol. We then characterize the equilibrium.

A Model of the market

We first provide the description of the environment in a discrete time framework. The length of the time interval is h, which will later be sent to 0. Trading begins at time 0. At time t a trader has a current portfolio $(\theta(t), x(t))$ of assets and cash. In every instant, a rational trader is observing the choices of a set of noisy traders, who can post asks and bids. Specifically in every period the trader can see an ask, which is the offer to sell an asset at a specified price, or a bid, an offer to buy. A pair (a, r) indicates an ask at price r, and (b, r) a bid at price r.

The trader knows his current portfolio, observes the offer (ask or bid) for that period and decides whether to accept or reject. If he rejects, the portfolio is unchanged and he proceeds to the next period. If he accepts an ask (a, r) his portfolio changes to $(\theta(t) + h, x(t) - rh)$; accepting a bid (b, r) takes it to $(\theta(t) - h, x(t) + rh)$. Trading goes until a set termination date T. The final portfolio $(\theta(T), x(T))$ is paid a total cash amount $\theta(T)F + x(T)$ where F is the fundamental value of the assets, and the trading ends.

Noisy traders behave in a fixed way. The probability of observing an ask in the period is $P(a)$ and of a bid is $P(b)$. Given that a is offered, the probability of an ask at price r is given by a fixed density $A(r)$; of a bid by a density $B(r)$.

Traders face the no short-sales constraint and no borrowing constraint:

$$\text{For every } t \in [0, T], \theta(t) \geq 0, x(t) \geq 0. \tag{1}$$

A strategy for a rational trader is a function that for every portfolio (θ, x), time t and standing offer (a, r) or (b, r) prescribes whether to accept or reject. This function is denoted by $\rho(\theta, x, t, (i, r))$, where $i \in \{a, b\}$, $t \in [0, T]$, and takes value in the set $\{accept, reject\}$.

We will see that the optimal strategy of a rational trader can be reduced to a pair of cutoff prices (p, q), each component also depending on the vector $(\theta, x, t, (i, r))$, where p indicates acceptance of any ask with price $r \leq p$, and q acceptance of any bid with price $r \geq q$. We call the policy determining the cutoff prices $\pi(\theta, x, t)$.

The game in continuous time with continuous asset endowments is defined as the limit in h tending to zero of the discrete game. The equation describing the rate of change of the portfolio is given explicitly in the equation 6 below.

Equilibrium behavior

Let

$$F_A^D(p) \equiv \int_0^p A(r)dr, F_B^U(q) \equiv \int_q^{+\infty} B(r)dr, \tag{2}$$

the probability of an acceptable ask (with cutoff price p) and acceptable bid (with cutoff price q); and

$$G_A^D(p) \equiv \int_0^p rA(r)dr, G_B^U(q) \equiv \int_q^{+\infty} rB(r)dr, \tag{3}$$

the expected cash decrease and increase given the price cutoff pair (p, q).

Finally, let

$$f(p, q) \equiv P(a)F_A^D(p) - P(b)F_B^U(q); g(p, q) \equiv -P(a)G_A^D(p) + P(b)G_B^U(q) \tag{4}$$

which describe the expected changes in the two components of the portfolio, summarized by $h(p, q) \equiv (f(p, q), g(p, q))$.

The trader has to choose a policy π that solves:

$$\max_{\pi}(\theta(T)F + x(T)) \tag{5}$$

subject to

$$\text{for all } t \in [s, T], (\dot{\theta}(t), \dot{x}(t)) = h(\pi(\theta(t), x(t), t)) \tag{6}$$

the no short-sales and no borrowing constraints:

$$\text{for all } t \in [s, T], \theta(t) \geq 0, x(t) \geq 0 \tag{7}$$

and a given initial portfolio at time s:

$$(\theta(s), x(s)) = (\theta_s, x_s). \tag{8}$$

For every (θ, x, s) this defines the value $V(\theta, x, s)$ of the program at the optimal policy.

We can now characterize the equilibrium behavior of rational traders. First we define a static problem that will be used to describe the characterization. At portfolio and time (θ, x, s),

$$\max_{(p,q)\in R^2}(f(p, q)F + g(p, q)) \tag{9}$$

subject to:

$$f(p, q)(T - s) + \theta(s) \geq 0,\ \ g(p, q)(T - s) + x(s) \geq 0. \tag{10}$$

Call $\sigma(\theta, x, s)$ this optimal pair. Then one can prove that following characterizes the optimal policy:

1. The optimal policy for the rational trader at (θ, x, s) is the static choice $\sigma(\theta, x, s)$.

2. At the optimal policy, $p = q$.

3. If in the static problem at the optimal solution neither of the constraints 10 is binding, then $p = q = F$.

4. The value function is

$$V(\theta, x, s) = \theta F + x + (T - s)[f(\sigma(\theta, x, s))F + g(\sigma(\theta, x, s))] \tag{11}$$

The value function has the obvious interpretation that the value of the portfolio is the current value (at terminal prices) plus the future gain from trade over the residual time $T - s$ at rate $\dot{\theta}(t)F + \dot{x}(t)$. Note that there is no bid-ask spread: this simplicity is a virtue of the continuous asset model. The optimal policy can be defined in two steps. First consider the price pair setting F as the cutoff for bids and asks. If at this price the final holdings of assets and cash are both positive, then the optimal policy is to hold the price at F.

Suppose instead that the final holding of the assets with that policy is negative (we may say that the initial endowment of assets is small). Then the cutoff price has to be increased to make the selling less likely and the buying more likely. The increase is just enough that the holdings of assets at the final time T is zero. Raising the cutoff price reduces the instantaneous rate of gain from trading to a smaller value.

When $p = q$ we can write the functions f and g simply as $f(p)$ and $g(p)$. This function f is increasing and g is decreasing, so for example the optimal price at (θ, x, s) when the assets constraint is binding is

$$f^{-1}\left(\frac{-\theta}{T - s}\right). \tag{12}$$

Further tables, and figures

We provide results on further measures of market instability for each gender composition of the market in Table 4. Each variable is formally defined below:

$$Turnover = \sum_{t=1}^{t=T} transactions_t / A; \tag{13}$$

$$Amplitude = (max_t(P_t) - min_t(P_t))/F; \tag{14}$$

$$Normalized\ (absolute)\ deviation = \sum_{t=1}^{T}\sum_{i}^{I}(price_{i,t} - F)/(100 \times A); \tag{15}$$

$$Dispersion = \sum_{t=1}^{t=T}|P_t - F|/(F \times T); \tag{16}$$

$$Bias = \sum_{t=1}^{T}(P_t - F)/(F \times T); \tag{17}$$

where $transactions_t$ is the total number of transactions executed in period t, T is the total number of trading periods in the market ($T = 15$), A is the total number of assets in the market, P_t is the median transaction price at period t, F is the fundamental value of the asset ($F = 360$), I is the turnover in period t, and $price_{i,t}$ is the price of the i^{th} transaction executed in period t[8].

Table 4 displays average market measures for each gender composition and p-values from pairwise comparisons using the non-parametric Mann-Whitney U-test.

Table 5 displays average market measures for each gender composition and for data reported by (*27*). Comparison between our data and (*27*) is important because their experimental asset markets are the closest in design to ours. As it can be seen, their markets displayed higher instability than ours. However there are a few important differences that may help explain this difference: in 4 out of their 7 markets, (*27*) endowed traders with a loan of 100,000 cash, and in the remaining 3 markets, they gave traders 10,000 cash that they could keep; trading periods in their markets lasted between 3 and 5 minutes, whereas they only lasted 2 minutes in our markets; finally, their software and instructions were different to ours. We took extra care in minimizing

[8]We have divided normalized absolute deviation by 100 to keep it comparable to other studies, based on the design with $F = 3.60$ rather than $F = 360$

the chances of confusion by writing the instructions in an easier, less technical language, and by asking subjects to complete an instructions quiz.

As shown by (*33*) and by our model of rational traders facing noisy traders, markets with higher liquidity tend to induce greater overpricing. Finally, their longer trading periods will have naturally led to a larger number of transactions.

Table 6 displays results from a random-effects linear regression of *trade gain* against time, gender and cognitive skills (referred to in the body of the article).

Market measure	Female-only	Male-only	Mixed-gender	*female-only* vs. *male-only* (p-value)*	*single-gender* vs. *mixed-gender* (p-value)
Turnover	3.26	2.54	2.56	0.598	0.059
	(0.92)	(0.75)	(0.87)		
Amplitude	0.14	0.10	0.03	0.910	0.004
	(0.17)	(0.09)	(0.02)		
Normalized deviation	1.01	0.78	0.32	0.880	0.020
	(1.21)	(0.63)	(0.37)		
Dispersion	0.07	0.06	0.03	0.597	0.059
	(0.09)	(0.04)	(0.02)		
Bias	0.05	0.05	0.03	0.174	0.253
	(0.09)	(0.03)	(0.02)		

*p-values from pairwise Mann-Whitney U-tests

Table 4: Market measures, averaged for female-only, male-only and mixed-gender markets. Between-sessions standard deviations shown in parentheses

Market measure	Female-only	Male-only	Mixed-gender	Noussair et al 2001
Turnover	3.26	2.54	2.56	4.19
	(0.92)	(0.75)	(0.87)	(1.18)
Amplitude	0.14	0.10	0.03	0.52
	(0.17)	(0.09)	(0.02)	(0.54)
Normalized deviation	1.01	0.78	0.32	2.24
	(1.21)	(0.63)	(0.37)	(0.98)
Dispersion	0.07	0.06	0.03	
	(0.09)	(0.04)	(0.02)	
Bias	0.05	0.05	0.03	
	(0.09)	(0.03)	(0.02)	

Table 5: Market measures, averaged for female-only, male-only, mixed-gender markets and for experiments by *Noussair et al, 2001* (dispersion and bias not reported by the authors). Between-sessions standard deviations shown in parentheses.

Dep. var: $trade\ gain$	**Coef.** (std. error)	**Coef.** (std. error)	**Coef.** (std. error)
CS	235.003**	235.003**	221.341**
	(80.593)	(80.583)	(68.513)
$period$	7.058***	7.058***	7.062***
	(2.042)	(2.042)	(2.041)
$CS \times period$	-10.672***	-10.671***	-10.678***
	(3.135)	(3.134)	(3.132)
$mixed$	29.080	30.707	-2.418
	(64.947)	(65.788)	(3.602)
$CS \times mixed$	-48.118	-49.445	
	(98.650)	(99.448)	
$mixed \times female$	1.497		
	(5.398)		
$constant$	-154.810**	-154.810**	-145.836***
	(52.101)	(52.095)	(44.294)

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: Random Effects Panel Regression Estimation. Std. errors adjusted for 30 clusters (one for each market). $trade\ gain_{i,t} = \sum_j (sale\ price^j_{i,t} - FV) + \sum_k (FV - purchase\ price^k_{i,t})$, where $sale\ price^j_{i,t}$ and $purchase\ price^k_{i,t}$ are the selling and buying prices of the j^{th} and k^{th} transaction executed by trader i in period t.