Boys will still be boys: Gender differences in trading activity are not due to differences in (over)confidence*

Carlos Cueva* Fundamentos del Análisis Económico (FAE) Universidad de Alicante Iñigo Iturbe-Ormaetxe Fundamentos del Análisis Económico (FAE) Universidad de Alicante

Giovanni Ponti Fundamentos del Análisis Económico (FAE) Universidad de Alicante The University of Chicago and LUISS Guido Carli Roma Josefa Tomás Fundamentos del Análisis Económico (FAE) Universidad de Alicante

January 2019

Abstract

The fact that men trade more than women in financial markets has been attributed to men's overconfidence. However, evidence supporting this view is only indirect. We directly test this conjecture experimentally, by measuring confidence using monetary incentives before participants trade in a simulated market. We find that men are more confident than women in our trading task. Men also trade more, and they hold larger and less diversified portfolios than women. However, we do not find that differences in confidence explain any portion of the gender gap in trading activity. We explore alternative candidate channels such as risk aversion, financial literacy or competitiveness but find that these factors are also unlikely to play a role.

KEYWORDS: Overconfidence; trade volume; gender JEL CLASSIFICATION: C91; D81; D91; G41

^{*} We thank Adam Sanjurjo for helpful comments and Holger Rau for sharing his z-Tree program. We also thank Esther Mata-Pérez, Haihan Yu, and Vita Zhukova for their valuable research assistance. Financial support from Ministerio de Economía y Competitividad (ECO2015-65820-P), Ministerio de Educación, Cultura y Deporte (programa estatal de promoción del talento y su empleabilidad en I+D+i, subprograma estatal de movilidad, plan estatal de investigación científica y técnica y de innovación 2013-16), Generalitat Valenciana (Research Projects Gruposo3/086) and Instituto Valenciano de Investigaciones Económicas (IVIE) is gratefully acknowledged.

[•] Corresponding author: Departamento de Fundamentos del Análisis Económico, Universidad de Alicante, 03071 Alicante (SPAIN), ccuevah@gmail.com

1. Introduction

A well-known result in Finance is that investors trade too much, thereby lowering their average returns. Because of active trading and its associated costs, the average individual investor has been found to pay a performance penalty of up to 3.8 percentage points, compared to passive indexing (Barber et al., 2009; Barber and Odean, 2000; French, 2008; Odean, 1999). In the United States, the annual cost of active trading has been estimated to be around \$100 billion (French, 2008).

The most common explanation for excess trading is *overconfidence* (Benos, 1998; Daniel et al., 1998, 2001; Daniel and Hirshleifer, 2015; Odean, 1998). Overconfident investors believe that the accuracy of their knowledge about asset values is greater than that of the average investor. Such investors believe that they can attain returns above the market average and are, therefore, willing to trade aggressively despite significant transaction costs.

In an influential study, Barber and Odean (2001) indirectly test the hypothesis that overconfidence causes excessive trading by comparing the trading activity of men and women, using gender as a proxy for overconfidence. In a sample of over 35,000 clients at a large discount brokerage, they find that average portfolio turnover is 45% larger for men. Moreover, trading reduces the net returns of both men and women, but the reduction for men is 0.94 percentage points higher. Barber and Odean conclude that their findings provide strong support for the hypothesis that excess trading is caused by overconfidence. However, since they only observe a trader's gender, and not his or her overconfidence, their conclusion relies on the strong assumption that overconfidence is the only dimension relevant to trading frequency along which men and women differ. It is in this sense that we consider indirect their test of the hypothesis that overconfidence causes excessive trading.

Several subsequent studies have replicated the large gap in trading activity between men and women (Agnew et al., 2003; Dorn and Sengmueller, 2009; Grinblatt and Keloharju, 2009). However, the evidence in support of the view that such a gap is due to gender differences in overconfidence is, again, only indirect. Two separate bodies of work motivate this conjecture. The first, originally from psychology and, more recently, economics, documents that men are more confident than women regardless of actual ability in areas such as mathematics (Campbell and Hackett, 1986; Hyde et al., 1990; Niederle and Vesterlund, 2007), academic achievement (Bengtsson et al., 2005; Lundeberg et al., 1994) and finance (Estes and Hosseini, 1988; Prince, 1993; Yablonsky, 1991). The second, coming from behavioral finance, posits overconfidence as a key ingredient in theoretically explaining the high volume of trade observed in financial markets (Benos, 1998; Daniel et al., 1998, 2001; Daniel and Hirshleifer, 2015; Odean, 1998).

The purpose of this study is to test *directly* whether the gender gap in trading activity can really be ascribed to men's greater overconfidence. We conduct an experiment in which male and female participants can buy or sell risky assets in a simulated market. Subjects are divided into gender-balanced groups who see exactly the same price realizations but do not interact with each other, allowing us to

perform a clean gender comparison.¹ We measure confidence by asking subjects to forecast their own ranking within the group of four in terms of earnings, both before and after trading, with a small financial reward for each correct answer. In turn, we define overconfidence as the difference between subjects' forecasted position and their actual position in the group.

Our findings replicate the key stylized result in the literature, namely, that men trade more, and are more confident than women. Men on average make around 50% more trades than women. However, when we compare men and women with similar levels of confidence or overconfidence we continue to find an equally large gap in the number of trades. Therefore, contrary to Barber and Odean's conjecture, we find that overconfidence does not explain any portion of the gender gap in trading activity.

In our experiment, subjects go through a baseline market, one with transaction costs, one with a competitive payment scheme, and one combining both features. Transaction costs are a crucial reason why overtrading is costly (Barber and Odean 2000). Our experimental design captures this feature of real-world financial markets explicitly. The reason for including a competitive payment scheme is to explore another candidate channel through which gender differences could be driving the observed gap in trading activity, in addition to confidence. To the extent that speculative trading in real-world financial markets has a competitive component, gender differences in willingness to compete are a plausible alternative cause for the gender gap in trading activity.²

We explore two additional possible mechanisms for the gender gap in trade volume: risk aversion and financial literacy. These are estimated from subjects' responses to two additional tasks. A large body of literature documents that women tend to be more risk averse than men (e.g. Charness and Gneezy 2012, Croson and Gneezy 2009, Eckel and Grossman 2008). Therefore, it is important to check whether gender differences in trading activity might be due to differences in risk aversion, rather than confidence. Financial literacy has also been consistently observed to be lower among women and has been shown to correlate with a broad range of behavioral domains, including stock market participation (Almenberg and Dreber, 2015; Lusardi and Mitchell, 2014). Finally, we control for a number of additional socio-demographic variables and standard personality measures by means of a questionnaire administered at the end of the experiment.

As in the case of confidence, however, we find no evidence that any of these alternative channels are responsible for the observed gender differences in trading activity. In this sense, our experimental results suggest that some other underlying mechanisms is at play. We discuss some possible avenues of research in this direction in Section 5.

The remainder of this paper is arranged as follows: Section 2 discusses alternative measures of overconfidence used in the finance literature; Section 3 describes the experimental design and

¹ To minimize experimenter demand effects, subjects are not told anything about the gender composition of the group. However, we ensure that every session has a balanced gender distribution in the laboratory.

² For gender differences in willingness to compete, see e.g. Gneezy *et al.* (2003, 2009), Gneezy and Rustichini (2004), Niederle and Vesterlund (2007), Saccardo *et al.* (2017).

procedures; Section 4 reports our results; Section 5 concludes with a discussion of our findings and of some related works in the literature.

2 Measures of overconfidence in finance

Moore and Healy (2008) review the broad literature on overconfidence and distinguish between three commonly used measures: *overestimation* of one's performance, *overplacement* of one's relative performance, and *overprecision* of one's beliefs. There is no consensus in the finance literature about which of these measures to adopt, with studies typically using either of the last two.

Overprecision is commonly measured as the degree of overestimation of the precision of one's knowledge, and is typically referred to as *miscalibration* in the finance literature. The standard measure of miscalibration requires subjects to provide 90% confidence intervals to a series of questions with a numerical answer, such as the length of the river Nile, the weight of an empty Boeing 747, or the age at death of Martin Luther King. The degree of miscalibration is measured by the percentage of times that the correct answer lies outside the confidence intervals (if this happens more than 10% of the time then the subject exhibits overprecision). Some disadvantages with this measure are that it is difficult to incentivize, it is noisy at the individual level, and it is sensitive to responders' numeracy, general knowledge and ability to conceptualize 90% confidence intervals.

Unlike overestimation, which is the tendency to overestimate one's performance in *absolute* terms (e.g. the expected number of correct answers in an exam), overplacement is the tendency to overestimate one's *relative* ability or performance. Overplacement is typically referred to as the *better-than-average* effect in the finance literature. Our measure of overconfidence is in line with this approach. In contrast to miscalibration, the better-than-average effect is easy to explain and incentivize: one simply asks subjects to forecast their relative position in the group in terms of their performance in the trading task, and pays them for a correct answer. This makes the better-than-average measure less noisy and less vulnerable to potential gender differences in general knowledge, numeracy, or response styles.

A number of studies have investigated the relationship between trading activity and overconfidence. In a survey of individual investors, Glaser and Weber (2007) find a significant positive correlation between trade volume and better-than-average confidence, but not with miscalibration. Grinblatt and Keloharju (2009) find a significant correlation between trade volume of Finnish investors and a psychological overconfidence assessment performed during the (compulsory) Finnish military service. Dorn and Sengmueller (2009) find no effect of overconfidence on trade except when looking at excess turnover for investors who have high better-than-average confidence. Biais *et al.* (2005) and Fellner-Röhling and Krügel (2014) find no effect of miscalibration on trade volume in experimental asset markets with asymmetric information. Finally, Michailova and Schmidt (2016) conduct experimental asset markets à la Smith *et al.* (1988) with groups of overconfident and underconfident

subjects according to a custom measure of miscalibration and find higher trade volume in overconfident markets.³

To sum up, the better-than-average effect has been partially successful in predicting trade volume while miscalibration has been generally unsuccessful. Additionally, the better-than-average effect has been found to be more pronounced in men than in women (Bengtsson et al., 2005; Beyer, 1990; Beyer and Bowden, 1997; Niederle and Vesterlund, 2007), while miscalibration has not (Biais et al., 2005; Deaves et al., 2008; Fellner-Röhling and Krügel, 2014). Both of these facts suggest that the better-than-average effect is a more appropriate measure than miscalibration to try to explain the gender gap in trade volume.

3 Experimental Design and Procedures

We recruited 192 participants (97 female, 95 male) from the undergraduate population of the University of Alicante. Considering the systematically large gender differences in trading activity that have been observed in empirical studies, we were confident that this sample size would have more than enough power to detect such gender differences. For instance, Fellner-Röhling and Krügel (2014) find strongly significant gender differences in trading activity using a more complicated design and a smaller sample size. All subjects were recruited using ORSEE (Greiner, 2004). Invitations, sent via email, did not provide any information about the experiment, which was simply described as a "decision-making experiment". We conducted 8 sessions with 24 subjects each, lasting approximately two hours. Subjects earned, on average, 19. The experiment was programmed in z-Tree (Fischbacher, 2007).

Our simulated trading task is based on Weber and Camerer (1998) and Weber and Welfens (2007). This task has been widely used in the study of the disposition effect, another well-known bias in behavioral finance. We use the data collected in this experiment to investigate the disposition effect in a separate paper (Cueva et al. 2018).

In the baseline treatment, we give subjects 5,000 units of experimental currency ("pesetas") that they can use to buy and sell units of six risky assets over 9 periods. The exchange rate is 1,000 pesetas $= 1 \in$. Subjects do not trade these assets among themselves and cannot influence prices, which are determined at each period by a stochastic process independent from their actions. Subjects cannot borrow additional cash nor short-sell, but can otherwise buy or sell as many assets as they wish in any period.

In addition to the baseline treatment, we introduce two variations: i) a competitive payment scheme and ii) transaction costs. In the *baseline* treatment (B), subjects earn the value of the portfolio upon liquidation (period 10) plus their remaining cash. In the *competitive* treatment (C), only the winner in each group of four earns the value of her portfolio plus the remaining cash, multiplied by two, while

³ See Duxbury (2015) for further discussion of the experimental evidence regarding overconfidence in behavioral finance.

the remaining three members earn nothing. The *transaction costs* treatment (T) is similar to the baseline, except subjects have to pay a fee for each transaction. The cost is a fixed rate for both purchases and sales, chosen from the set {1%, 4%, 7%}, randomized across groups. Finally, in the *competitive with transaction costs* treatment (CT), we combine the features of treatments C and T.

Subjects participate in all four treatments in the same session, each treatment lasting approximately 15 minutes. Treatment order is counter-balanced between sessions.⁴ No feedback is provided at the end of each treatment. In particular, subjects do not learn their relative position in the group at any time during the experiment until the very end.

Subjects are randomly assigned into groups of two men and two women. They are told that they are in a group of four but they do not know the identity, or gender, of their group mates. To ensure that subjects perceive a similar gender-balanced environment in every session, men and women are placed in alternating cubicles. This is done by asking men and women to randomly draw a number from different envelopes: one envelope contains odd numbers and the other contains even numbers. Subjects are then seated in the cubicle corresponding to their drawn number. In the three sessions in which a perfect gender balance was not achieved, the remaining subjects from the more numerous gender were asked to draw a number from the envelope assigned to the other gender.

Subjects are also told that each group member faces the same price path of the six assets. Unlike Weber and Camerer (1998) and Weber and Welfens (2007), who use the same asset prices for all experimental subjects, we determine different random price paths for each group of four. Otherwise, our markets are similar to the neutral markets in Weber and Welfens (2007). All markets begin in period -3, in which all assets have the same initial price of 100 pesetas. From period -2 onwards, the price of each asset follows an independent path, going up by 6% or down by 5%. Subjects are told that each asset has a different probability of a price increase, but they are not told the actual probabilities. They are also informed that the probability of a price increase is constant for each asset over the whole treatment and that price changes are independent of previous prices and subjects' actions. The actual probabilities are chosen randomly without replacement from the set {0.4, 0.45, 0.5, 0.5, 0.6}.

Following Weber and Welfens (2007), we do not inform subjects of the set of probabilities that assets might take. The main advantage of this is that, without further information, subjects are more likely to evaluate each asset independently from the other assets. If instead, we gave subjects full information, then a Bayesian updater would need to take into account the prices of all assets to form beliefs about each asset's underlying probability (due to the fact that probabilities are drawn without replacement). Of course, an alternative would have been to inform subjects about the set of probabilities, and draw them with replacement, but that would have resulted in very heterogeneous markets, which

⁴ Sessions 1 and 2: B, C, T, CT; sessions 3 and 4: T, CT, B, C; sessions 5 and 6: C, B, CT, T; sessions 7 and 8: CT, T, C, B. Subjects played a short trial period before each treatment.

would have made comparisons across groups harder. Also, this environment would be somewhat less representative of real financial markets, in which probabilities are generally unknown.

Subjects observe prices from period -3 to 10, but can only trade in periods 1 to 9. Periods -3 to 0 provide some information about the assets before trading, while period 10 is used to liquidate portfolios. Each trading period lasts one minute. At the beginning of periods 1, 6 and 10 we ask subjects to rank each asset in terms of their underlying probability of a price increase. They receive 100 additional pesetas if they guess the position of all six stocks correctly. The expected reward in this task was intentionally very small (0.10€ if the full ranking of stocks is guessed correctly, zero otherwise), to minimize potential hedging considerations.

To measure confidence, we ask subjects at the beginning and at the end of each treatment to forecast their own ranking within their group of four in terms of total earnings (final cash + portfolio value). Subjects are reminded that they have been assigned to a group of four participants at random and that they will be paid an additional 100 pesetas for guessing their position correctly. Again, we intentionally chose a small reward in this task in order to minimize potential hedging considerations. Subjects need to select one out of four options: "First", "Second", "Third" or "Fourth". We recode this information so that confidence = 1 when her guess is "Fourth" and confidence = 4 when it is "First".

We define overconfidence as the difference between ex-ante confidence and actual rank. Similar to ex-ante confidence, the variable rank takes values $\{1, 2, 3, 4\}$ and equals four for the winner in the group. Our measure of overconfidence can take values from -3 to +3, where a positive value means the initial guess was too confident. Even though we also elicit ex-post confidence, we do not use this measure in our analysis because it is affected by actual performance in the task, which makes it endogenous to trade volume.

To measure risk aversion, subjects complete a Multiple Price List task (MPL, Holt and Laury, 2002) at the beginning of the session. The task consists of a sequence of 21 binary decisions between a fixed lottery and an increasing safe option. The lottery pays a prize of 5,000 pesetas with 50% probability and nothing otherwise. The safe option ranges from 0 to 5,000 pesetas. At the end of the experiment, we randomly select one of the 21 decisions for payment. We measure risk aversion as the difference between the expected value of the lottery (2,500 pesetas) and the elicited certainty equivalent (the safe amount first preferred to the lottery along the sequence) divided by 2,500 so that we normalize the measure between -1 and +1, with 0 representing risk neutrality.

At the end of the experiment, subjects complete a questionnaire. The questionnaire contains standard socio-demographic questions such as age and field of studies, a number of self-assessed psychological measures using a Likert-type scale, and a three-item Financial Literacy Test (FLT, Lusardi and Mitchell, 2008). Among the various measures collected, we primarily focus on risk aversion and financial literacy, since these are most likely to have a chance of explaining some portion of the gender gap in trade volume *a priori*. The other psychological measures include a short version of the Big Five test and some items aimed at exploring the sources of the disposition effect, investigated

in a separate study (Cueva *et al.* 2018).⁵ These variables are used as additional controls in the last part of our analysis (see Appendix C for details).

There are advantages and disadvantages to an experimental design in which subjects do not trade among themselves. The main disadvantage is that our environment departs from the standard setup in finance models of overconfidence (e.g. Odean, 1998; Daniel and Hirshleifer 2015). In this sense, our experiment does not provide a good test of these models. However, the aim of our study is to test whether gender differences in trade volume can be explained by overconfidence, not to test a specific model of overconfidence and trade volume.

A criticism might be that overconfidence plays no role in our experiment from a theoretical standpoint. However, this is not the case. In Appendix A, we present a simple model which formalizes the link between overconfidence and trade volume in our environment. We follow the usual approach by which overconfident traders are assumed to overestimate the precision of their private signals. In our experiment, private signals are simply the assets' prices. Traders who are overconfident in our experiment will overestimate the informativeness of assets' prices when making inferences about their underlying probabilities, which will lead to negatively biased estimates of the return volatility of the optimal portfolio. As a result, risk averse traders who are overconfident will allocate a larger share of their endowment to the optimal portfolio than risk averse traders who are not. As traders update their optimal portfolio over time, this generates a positive relationship between trade volume and overconfidence.

The main advantage of our design is that subjects are unable to affect market prices. We see this as a more realistic way of modelling the trading problem faced by small private investors, who are typically price-takers. An experiment in which subjects trade among themselves but in which each trader has a negligible effect on prices would require a large market, and would therefore need to be conducted online rather than in the lab. This would entail losing a large amount of experimenter control. Another problem of multiplayer markets is that the behavior of each trader is highly dependent on the behavior of other traders. This makes a reliable statistical analysis of individual differences much more difficult, since observations are not independent. Finally, relying on a small number of large markets with endogenously determined prices entails losing almost all experimenter control over the variable which is likely to have the biggest effect on trade volume – prices.

4 Results

We first discuss gender differences in aggregate behavior, disregarding treatment effects. For each individual we first compute the average number of transactions per period in each treatment, and then

⁵ The Big Five test is a widely used psychological questionnaire which classifies human personality into five factors: openness, conscientiousness, extraversion, agreeableness and neuroticism (McCrae and John, 1992). We use a Spanish translation of the reduced form of the Big Five test (Benet-Martínez and John, 1998; John and Srivastava, 1999). The other items are self-control (Tangney et al., 2004), difficulty recognizing one's mistakes, optimism, willingness to take risks, confidence, competitiveness and enjoyment of winning.

we compute the average across the four treatments. P-values are obtained from Mann-Whitney U-tests unless otherwise stated. Panel (a) of Figure 1 displays mean number of trades per period for men and women. Men trade significantly more in all periods. The difference is highest in the first period (5.1, p=0.0001), and lowest in the last one (1.6, p=0.0175). Pooling all periods together, men make an average of 9.1 transactions per period (N=94), while women only make 5.9 (N=97).⁶ This difference is highly significant (p<0.0001, N=191) and of similar magnitude to previous empirical work (Agnew *et al.* 2003, Barber and Odean 2001).

[FIGURE 1 HERE]

Figure 1. Trading activity and confidence of men and women over all treatments. Panel a) Mean number of trades per period. Panel b) Histograms of median ex-ante confidence across the four treatments (1=lowest, 4=highest). Panel c) Mean number of trades (± SE bars) per period, disaggregated by confidence level (data for confidence levels below two are omitted due to insufficient observations).

A common alternative measure of trading activity used in finance is portfolio turnover, which is defined as the total value of transactions divided by average portfolio value. Since in our case all subjects have the same initial endowment, we define turnover as the total value of transactions, because we do not need to normalize. We compute again averages across the four treatments. Men buy and sell an average of 8,969 pesetas worth of assets per treatment, or 179% of their initial endowment, while women trade 5,853 pesetas worth of assets, or 117% of their endowment. Again, this difference is highly significant (p<0.0001). Note that, since all assets have an initial price of 100, turnover is approximately the number of transactions times 100, and the two measures are very highly correlated. For this reason, even though we focus on the number of transactions throughout the paper, all our results apply equally to turnover.

Panel (b) of Figure 1 displays histograms of the median ex-ante confidence for men and women. Recall that ex-ante confidence goes from 1 (lowest) to 4 (highest). Men are significantly more confident than women (p<0.0001, N=191).⁷ While women typically believe they will be between second and third in their group of four, most men believe that they will end up between first and second. Even though

 $^{^{6}}$ 24 observations in the baseline treatment were lost due to a server crash in one session. This session had the same number of male and female participants (as did most). For these subjects, we compute the mean number of trades per period using the other 3 treatments. Additionally, we exclude one (male) subject as a clear outlier from all of our analysis. This subject averaged 76 transactions per period in the baseline treatment, which is over 14 standard deviations above the mean of the rest of our sample (8.5). No other outliers were detected. For instance, the second most active subject in our sample averaged 29 transactions per period.

⁷ This is a Mann-Whitney U-test on median ex-ante confidence. The number of subjects is 94 men and 97 women. Recall that subjects go through 4 treatments, so we take an observation as the median over these four values, as is done in panel (b) of Figure 1.

we provide no feedback about relative performance, subjects might update their beliefs with experience as they go through the four treatments. We do not observe significant patterns in ex-ante confidence suggestive of learning from treatment to treatment, except a slight increase from the first to the second treatment that subjects played. To rule out any potential learning, we check whether the gender gap in confidence persists using only the first confidence measurement, before subjects have any experience in trading. Results are similar in this case, with men displaying significantly higher confidence than women (p=0.001, N=191).

Panel (c) of Figure 1 illustrates our main result. Namely, that gender differences in trading activity do not disappear when we control for confidence. In particular, except for subjects in the lowest confidence range (confidence ≤ 2.5), men trade significantly more than women at every other confidence level (if confidence = 3, p=0.002, N=60; if confidence = 3.5, p=0.0005, N=31; if confidence = 4, p=0.001, N=43). Surprisingly, Figure 1 shows a negative correlation between median confidence and mean number of trades for women (Spearman's $\rho = -0.204$, p = 0.045, N=97). However, this result does not carry through when we disaggregate behavior treatment by treatment. Finally, to rule out learning, we repeat the analysis using only the first confidence measurement and find similar results (see Figure O1 in the Online Appendix).⁸

We now repeat the analysis using overconfidence instead of confidence. Although both men and women are significantly overconfident (we reject the null hypothesis that overconfidence is zero for both genders), men's average overconfidence (0.663) is 68% higher than women's (0.395). The gender gap in overconfidence is significant (p=0.032), but less stark than in the case of confidence. On the other hand, the difference in overconfidence is no longer significant if we focus on the first treatment subjects participate in (p=0.53). The reason for this is that men tend to rank slightly higher than women (although not significantly overall, p=0.23). Interestingly, the distribution of ranks for men is U-shaped, with modes at the top and bottom ranks, while the distribution for women has modes at ranks 2 and 3.

In Panel (a) of Figure 2 we partition subjects into four levels of overconfidence and compare the number of trades of men and women within each level. The partition is as follows: -1 if median overconfidence < 0, 0 if median overconfidence = 0, 1 if median overconfidence in (0, 1], 2 if median overconfidence >1. The distribution of subjects into these levels is as follows. Level -1, 17.8% (15 men, 19 women), level 0, 25.6% (23 men, 26 women), level 1, 32.5% (28 men, 34 women), level 2, 24.1% (28 men, 18 women). Again, we find that men trade significantly more than women in every group (level -1, p=0.016; level 0, p=0.0037; level 1, p=0.0014; level 2, p=0.0033). We obtain similar results if, instead of using four levels of overconfidence, we use a median split (see Figure O3 in the Online Appendix). To rule out learning and to ensure that our measure of overconfidence is exogenous to trade volume, in Panel (b) of Figure 2 we redo the analysis partitioning subjects according to their

⁸ In Figure O2 in the Online Appendix, we also investigate order effects and find only a moderate trend towards lower trade volume in later treatments, which is not statistically significant.

overconfidence level in the first treatment and compare mean trading volume in the three remaining treatments. Results do not change, with p-values at the four levels always below 0.05.

[FIGURE 2 HERE]

Figure 2. Trading activity and overconfidence of men and women over all treatments. Panel a) Mean number of trades (\pm SE bars) per period, disaggregated by overconfidence level: -1 if median overconfidence < 0; 0 if median overconfidence is 0; 1 if median overconfidence in (0, 1]; 2 if median overconfidence > 1. Panel b) Mean number of trades (\pm SE bars) per period in the last three treatments, disaggregated by overconfidence level in the first treatment: -1 if overconfidence ≤ -1 ; 2 if overconfidence ≥ 2).

We next check whether these results hold in each treatment. Table 1 displays the mean number of trades, confidence, profit, rank and overconfidence, disaggregated by gender and treatment. Indeed, trading activity is significantly larger for men than for women in all treatments. On average, the difference between men and women is 61, 62, 43 and 39 percent in treatments B, C, T and CT, respectively. Confidence is also significantly higher for men in every treatment, and the gap is stable across treatments. There are no significant differences in trading profits between men and women, although men tend to come slightly above women, particularly in the treatments without transaction costs. On the other hand, male profits display significantly higher variance (see Figure A1 in the Appendix), which is to be expected given that men tend to hold more assets than women (2,813 pesetas worth of assets vs 1,932, p<0.0001, see Section 4.1).

		В	С	Т	СТ
	Male	10.614 (5.724)	11.981 (7.326)	6.669 (4.214)	7.350 (4.960)
# Trades	Female	6.571 (3.713)	7.389 (4.920)	4.661 (2.764)	5.293 (3.163)
	p-value	< 0.0001	< 0.0001	0.0003	0.0002
	Male	3.293 (0.711)	3.362 (0.801)	3.106 (0.861)	3.181 (0.879)
Confidence	Female	2.812 (0.764)	2.990 (0.823)	2.691 (0.834)	2.835 (0.850)
	p-value	0.0001	0.0008	0.0006	0.0045
	Male	5245 (481)	5208* (491)	4942 (579)	5025* (626)
Profit	Female	5165 (388)	5078* (322)	4915 (377)	4956* (381)
	p-value	0.094	0.116	0.932	0.811
	Male	2.561 (1.177)	2.702 (1.181)	2.447 (1.232)	2.542 (1.241)
Rank	Female	2.447 (1.075)	2.289 (1.020)	2.536 (1.001)	2.474 (0.991)

Table 1. Means and standard deviations (in parentheses) of number of trades, confidence, profit, rank and overconfidence, disaggregated by treatment and gender.

	p-value	0.511	0.011	0.576	0.666
	Male	0.732 (1.248)	0.660 (1.308)	0.660 (1.492)	0.638 (1.537)
Overconfidence	Female	0.365 (1.335)	0.701 (1.308)	0.155 (1.357)	0.361 (1.386)
_	p-value	0.106	0.900	0.025	0.193
Observations		82 (M), 85 (F)	94 (M), 97 (F)	94 (M), 97 (F)	94 (M), 97 (F)

Note: # Trades = mean number of transactions per period; Confidence = pre-task guess of position in the group of 4 (1 lowest, 4 highest); Rank = final position in the group of 4 (1 worst, 4 best); p-values are computed using Mann-Whitney U-tests. *Potential (rather than realized) profits, since only the winner in each group earns double his/her profits.

An alternative measure of trading performance which allows for a more reliable comparison of gender differences is the variable rank. Rank equals 1 if a subject makes the lowest profits in the group, and equals 4 if she makes the highest profits in the group (recall that each group is composed of two men and two women who face the same prices). There are no significant gender differences in rank except in treatment C, where men tend to rank significantly higher than women. This is consistent with previous results by Cueva and Rustichini (2015) and Eckel and Füllbrunn (2015) who observe somewhat higher earnings for men in mixed-gender experimental asset markets. More broadly, our evidence is in line with known gender differences in performance in competitive environments (Gneezy et al., 2003; Gneezy and Rustichini, 2004; Niederle and Vesterlund, 2007; Saccardo et al., 2017). Finally, men tend to be more overconfident than women in all treatments except C, but the difference is only statistically significant in T. Therefore, despite the large differences in confidence, our results only lend partial support to the hypothesis that men are more overconfident than women in their relative trading performance.

We now estimate the gender gap using random effects regressions that control for confidence in each treatment. The dependent variable is the natural logarithm of average transactions per period. This transformation is standard and helps to improve the normality of the data. Results do not change if we use the number of trades instead. Adding one to the number of trades before taking logs also yields similar results (see tables O1 and O2 in the Online Appendix). We choose the former specification because effect sizes are easier to interpret and very few observations are lost (1 in B, 1 in C, 4 in T and 4 in CT). We estimate four random effects regression models with incrementally more controls: Model (1) only includes a dummy for males, treatment dummies, all interactions between the male dummy and the treatment dummies, and group dummies; Model (2) adds 3 confidence dummies; Model (3) adds confidence×male interactions; Model (4) adds confidence×treatment and confidence×male×treatment interactions. Table 2 reports marginal effects of the male dummy according to these four regressions (full regression results can be found in Table O3 in the Online Appendix). Comparing Model 1 with Models 2–4 in Table 2 shows that the estimated gender gap remains highly significant and of very similar magnitude in every treatment, whether we control for confidence or not. In order to rule out learning, we also re-estimate the regressions using only the first measure of confidence collected in the session, which yields similar results (see Table O4 in the Online Appendix).

We redo the analysis focusing on overconfidence instead, and find similar results (see Table A1 in the Appendix). Controlling for overconfidence has no impact on the estimated gender gap in trade volume, which remains very large. Since overconfidence may be endogenous to trade volume, we also repeat the analysis using only overconfidence in the first treatment as a predictor of trade volume in the other three treatments. Results are similar also in this case (see Table O6 in the Online Appendix). Finally, since our measure of overconfidence is correlated with performance by construction (for instance, the best subject in the group can never be classified as overconfident), we estimate two alternative models. In the first one, we include the variable rank as an additional regressor; in the second one, we only include subjects with intermediate ranks (ranks 2 and 3). The results we get are, again, very similar (see tables O7 and O8 in the Online Appendix).

Table 2. Marginal effects of "male" on (log) average number of transactions per period. Model 1 includes a male dummy, treatment dummies, their interactions, and group dummies. Model 2 adds 3 confidence dummies. Model 3 adds confidence×male interactions. Model 4 adds confidence×treatment and confidence×male×treatment interactions.

	Model 1	Model 2	Model 3	Model 4
В	0.482***	0.491***	0.492***	0.443***
	(0.077)	(0.079)	(0.078)	(0.078)
С	0.504***	0.514***	0.514***	0.509***
	(0.094)	(0.096)	(0.097)	(0.099)
Т	0.303**	0.313**	0.313**	0.321**
	(0.133)	(0.130)	(0.131)	(0.131)
СТ	0.341***	0.349***	0.349***	0.363***
	(0.108)	(0.105)	(0.105)	(0.108)
	NO	VEG	VEG	VEG
Confidence	NO	YES	YES	YES
+ Gender interactions	NO	NO	YES	YES
+ Gender & Treatment				
interactions	NO	NO	NO	YES
N	730	730	730	730

Note: Estimates from random effects GLS regressions with group dummies and standard errors (in parentheses) clustered by group. See Table O3 in the Online Appendix for full regression results. *** p<0.01, ** p<0.05, * p<0.1

4.1 Trading Strategies

It is possible that gender differences in trade volume might be explained by differences in trading styles. In particular, if men are more overconfident they might take on larger portfolios than women, which could explain why they trade more. On the other hand, risk-averse investors might make smaller and more diversified investments, which is likely to result in fewer transactions.

Another possibility is that there are gender differences in the degree of trading rationality. To get a sense of how much a rational risk-neutral agent would trade, we need to make some auxiliary assumptions. Note that we do not tell subjects the underlying probabilities that each asset might have. Assuming that priors are not too pessimistic, so that the agent expects a positive expected payoff from investing on the asset with the highest price, the optimal policy would consist of investing the entire portfolio on the most expensive asset in each period. On average, the most expensive asset changes in our markets in 1 out of every 4 periods. A rational risk-neutral trader would therefore make in the order of 5000/100*1/4=12.5 transactions per period. Note that this policy applies to the case where there are no transaction costs (treatments B and C). An optimal policy with transaction costs (treatments T and CT) is harder to obtain, but it would clearly involve less trade. Interestingly, men average 10.6 trades in B and 12 trades in C, which is quite close to the rational benchmark. Women, on the other hand, average 6.6 trades in B and 7.4 trades in C.

We construct three measures of trading strategies. The first one is simply the total value of a trader's asset portfolio (V). With this measure, we intend to capture the size of a trader's asset positions. Our second measure is a "rationality index" (R). With this index, we measure how close a trader's strategy is to the risk-neutral rational benchmark. Our third measure is the Herfindahl index (H), which captures the degree of diversification.

$$V = \sum_{i=1}^{6} n_i p_i; \ R = \sum_{i=1}^{6} \frac{1}{6} s_i r_i; \ H = \sum_{i=1}^{6} s_i^2,$$

where n_i is the number of units of asset *i* in the portfolio, p_i is the price of asset *i*, s_i is the share of asset *i* in the portfolio, and r_i is the rank of asset *i*. Ranks are computed in each period according to an asset's price, assigning 1 to the cheapest asset, 2 to the second cheapest, and so on. As an example, let prices be 120 (A), 115 (B), 90 (C), 95 (D), 110 (E) and 135 (F). A subject holds a portfolio with 3 units of A, 5 of B, and 2 of F. The value of her portfolio is $V = 3 \cdot 120 + 5 \cdot 115 + 2 \cdot 135 = 1205$. If we rank assets starting from the cheapest one, we have CDEBAF. The rationality index for this subject is $R = \frac{1}{6} \left(5 \cdot \frac{3}{10} + 4 \cdot \frac{5}{10} + 6 \cdot \frac{2}{10} \right) \approx 0.78$. R ranges from 1/6 when the portfolio only contains the cheapest asset, to 1 when the portfolio only contains the most expensive asset. Similarly, $H = \left(\frac{3}{10}\right)^2 + \left(\frac{5}{10}\right)^2 + \left(\frac{2}{10}\right)^2 \approx 0.38$. *H* ranges from 1/6 when the portfolio has an equal share of every asset, to 1 when it only contains one asset.

[FIGURE 3 HERE]

Figure 3. Portfolio value (panel A), Rationality index (panel B) and Herfindahl index (panel C) for males and females averaged across treatments. Error bars represent ± 1 standard error.

Figure 3 displays V, R and H measures for men and women in every period, averaged across treatments. Men hold much larger portfolios, diversify less, and their trading strategies are closer to the risk-neutral rational benchmark than women. This confirms our conjecture that men tend to diversify

less and hold larger asset positions. Also, the fact that men's strategies are closer to the rational benchmark suggests that, to some extent, their greater confidence in this task may be justified. Finally, it explains why their profits have a slightly higher mean and a higher variance than women's, as shown in Table 1 and Figure A1 in the Appendix.

We next check whether these trading styles in fact correlate with confidence as predicted. We run three OLS regression with either mean V, R or H as dependent variables, and gender and mean ex-ante confidence as explanatory variables (see Table A2 in the Appendix). We find a positive and significant effect of confidence on H (p=0.015), a marginally significant effect of confidence on V (p=0.10), and no effect of confidence on R (p=0.58). The effect of *male* is very large and significant for V (p<0.0001), it is significant at the 5% level for H (p=0.032) and at the 10% level for R (p=0.063). These results suggest that, indeed, confident traders diversify less in our experiment. However, we find no support for the conjecture that more confident traders are also more likely to conform to the rational benchmark and we only find marginal support for the conjecture that they hold larger positions. Also, note that men still hold much larger and less diversified portfolios than women after controlling for confidence, which suggests that confidence cannot explain gender differences in trading styles along these dimensions.

Can the observed gender differences in trading styles explain the gap in trade volume? Figure 4 displays scatter plots of trade volume against portfolio value (panel A), R index (panel B) and H index (panel C). As expected, trade volume is highly correlated with portfolio value. Subjects that hold larger portfolios also trade more. Of course, part of this correlation is mechanical, since, in order to increase the size of the portfolio, one has to trade cash for assets. Note also that the gender gap in trade volume persists after controlling for mean portfolio value, although it is not as large as before. Averaging over all treatments and periods, men make 53% more trades than women. After controlling for portfolio value, this difference goes down to 24%. Therefore, it appears that a large portion of the gender gap in trade volume, but not all of it, can be explained by the fact that men take on bigger asset positions. Panels B and C of Figure 4 make two things clear: firstly, trade volume is uncorrelated with either index; secondly, a large gender gap in trade volume persists for any level of H and R.

Table A3 in the Appendix reports regression results of mean trade volume on portfolio value, R and H, which confirm these findings. In Section 4.3 we reexamine the gender gap in trade volume whilst controlling for several potential explanatory factors and again find similar results.

[FIGURE 4 HERE]

Figure 4. Scatter plot of mean number of trades against portfolio value (panel A), Rationality index (panel B) and Herfindahl index (panel C). Data is averaged at the subject level, over all periods and treatments. Solid lines represent the linear fit for males and dashed lines for females. Pearson's correlation between mean portfolio value and mean number of trades is high and significant (women:

 $\rho = 0.711$, p-value < 0.0001; men: $\rho = 0.543$, p-value < 0.0001). On the other hand, neither *R* or *H* indices are significantly correlated with trades for either group ($\rho < 0.1$, p-value > 0.3).

To conclude this section, we examine gender differences in beliefs. Recall that at the beginning of periods 1, 6 and 10, subjects are asked to rank each asset in terms of their underlying probability of a price increase. Under rational expectations, one should always select the most expensive asset as that with the highest probability of a price increase. This gives us an additional opportunity to examine the level of understanding displayed by subjects, and to test whether it correlates with confidence or trade volume. We construct an indicator variable which takes value 1 when a subject has rated the most expensive asset as that with the highest probability of a price increase, and 0 otherwise. In total, we have 12 observations per subject (since there are 4 treatments and 3 guesses per treatment). We then compute the average value of this indicator for each subject. Males guess correctly the best asset 80% of the time, while females do so 77% of the time, with no significant difference between these values (Mann-Whitney U test, p=0.295).

As one would expect, guessing accuracy is highly correlated with our rationality index for both men and women (men: Spearman's ρ =0.410, p<0.0001; women: Spearman's ρ =0.388, p=0.0001). Intuitively, subjects who are able to form expectations rationally also trade more rationally. However, there is no correlation between guessing accuracy and any other measure: Herfindahl index, confidence or trade volume.⁹

To sum up, we find that men take on larger and less diversified asset portfolios than women, and that their trading strategies are closer to that of a rational risk-neutral trader. However, only differences in portfolio size have any explanatory power on the gender gap in trade volume.

4.2 Treatment Effects

In this section, we check for gender differences in the response to treatments. The introduction of competition allows us to explore an alternative channel through which gender differences may create a gap in trading activity. To a large extent, speculative trading in real financial markets is a competitive activity, since a successful trader must predict price trends and therefore "outsmart the market". This competitive feature of trading is captured explicitly in treatments C and CT.

Both laboratory and field studies typically find that men are more willing to compete than women. For instance, Niederle and Vesterlund (2007) find that 73% of men in their experiment select a tournament payment scheme in a summation task compared to only 35% of women, despite no significant gender differences in performance. Furthermore, gender differences in confidence only

⁹ Herfindahl index, men: ρ =0.158, p-value=0.127, women: ρ =-0.003, p-value =0.976; confidence, men: ρ =0.167, p-value =0.108, women: ρ =0.013, p-value =0.901; number of trades, men: ρ =0.122, p-value =0.242, women: ρ =0.109, p-value =0.287.

explain part of the gap in willingness to compete in their experiment. Saccardo *et al.* (2017) replicate these results using a ball-tossing task. In one additional treatment, subjects can choose what percentage of their earnings will be determined by a tournament scheme. Men on average allocate 66% of their earnings to the tournament scheme, while women only allocate 35%. Again, a large gender gap remains after controlling for confidence.

This evidence suggests that, perhaps, men are more prone to trading in financial markets than women because they are generally keener to participate in competitive activities. In our experiment, subjects cannot opt out from the competitive treatments. However, as in the real world, they may modulate their degree of participation in these tasks, by deciding how much they trade, if at all. The introduction of a competitive payment scheme should then encourage male participation and discourage female participation, which would translate into a larger gender gap in trade volume.

Our motivation for introducing transaction costs is threefold. Firstly, it is hard to interpret the economic relevance of large gender differences in trading activity unless trading is costly. Indeed, transaction costs are a key reason why excessive trading is irrational (Barber and Odean 2000). Secondly, when trading costs are large, only highly confident individuals may be willing to trade. If this is the case, we might observe a smaller reduction in trading activity in men than in women after the introduction of trading costs. Thirdly, Barber and Odean (2001) find that male traders experience a larger reduction in profits than females due to their excessive trading, which suggest that transaction costs may be particularly harmful to men. Therefore, we test whether the reduction in profits caused by transaction costs is larger for men than for women.

Table 1 shows that the gender gap in trade volume is persistent throughout the four treatments. We now take advantage of our within-subjects design to test for the statistical significance of treatment effects using paired non-parametric tests. Panel (a) of Figure 5 displays mean number of trades disaggregated by gender and treatment. First, we reject the hypothesis of no treatment differences within groups (Skillings-Mack tests, p<0.001 for both men and women). Second, competition increases trading activity and, surprisingly, the effect is significant at the 5% for women but not for men (B vs C – men: p=0.094, – women: p=0.019; T vs CT – men: p=0.061, – women: p=0.003). However, pairwise difference-in-differences tests fail to reject the hypothesis of no gender difference (Mann-Whitney *U*-tests, C – B: p=0.953; CT – T: p=0.676). In other words, we do not find any support for the hypothesis that willingness to compete contributes to the gender differences in trade volume.

[FIGURE 5 HERE]

Figure 5. Panel (a) Mean number of trades (\pm SE bars) per period disaggregated by gender and treatment. Panel (b) Mean earnings (portfolio value plus remaining cash \pm SE bars) at the end of each

treatment. Note that in C and CT these are not realized earnings, since only the winner in each group earns a profit in these treatments.

Transaction costs are clearly associated with a decrease in trading activity for both men and women (Wilcoxon signed-ranks tests, B vs T and C vs CT, p<0.001).¹⁰ However, pairwise difference-indifferences tests show that the negative impact is larger in men than in women (Mann-Whitney U-tests, T – B: p=0.006; CT – C: p=0.01). In other words, we find no support for the conjecture that men's willingness to trade is less sensitive to transaction costs than women's.¹¹

Panel (b) of Figure 5 displays mean earnings in each treatment for men and women. Pairwise comparisons between men and women are already reported in Table 1. We now test within-subjects treatment effects. Firstly, we reject the hypothesis of no treatment effects separately for both men and women (Skillings-Mack test, p<0.001). We find a significant negative effect of transaction costs for both men and women (Wilcoxon signed-ranks tests, B vs T and C vs CT, p<0.001), and no effect of competition in either group (all pairwise Wilcoxon signed-ranks tests, p>0.12). Lastly, we do not find any significant differences in differences in treatment effects between men and women (all pairwise Mann-Whitney U-tests, p>0.17). In other words, we do not find any support for the hypothesis that transaction costs have a larger negative impact in men's profits than in women's profits.

4.3 Alternative Explanations

Our evidence suggests that differences in confidence or overconfidence between men and women are not responsible for the gender gap in trading activity. Additionally, our analysis of treatment effects does not support the hypothesis that gender differences in willingness to compete explains it either. To rule out other potential mechanisms, we incorporate into the analysis a number of individual measures gathered from a questionnaire at the end of the experiment, as well as a measure of risk aversion estimated from the MPL lottery task, which subjects completed at the beginning of the experiment.

Among the various measures gathered from the questionnaire, we first focus on financial literacy, which, together with risk aversion, is *a priori* most likely to play a role. The rest of these measures are included later as additional controls (see Table A4 in the Appendix for summary statistics).

Table 3 displays mean values of risk aversion estimated from subjects' choices in the MPL lottery task. It also displays the proportion of correct answers to each of the financial literacy questions administered at the end of the experiment. In line with most results in the literature, our results show that men are significantly more risk-taking than women (Charness and Gneezy 2012, Croson and Gneezy 2009, Eckel and Grossman 2008). In particular, while women are on average risk averse

¹⁰ Additionally, the gender gap remains roughly constant for all values of transaction costs. See Figure O4 in the Online Appendix.

¹¹ These results continue to hold if we control for order and group effects using regression analysis. See Tables O9 and O10 in the Online Appendix.

(Wilcoxon signed-rank test: p=0.001), we cannot reject the null hypothesis of risk neutrality for men (p = 0.699). Consistently with earlier studies (e.g. Almenberg and Dreber 2015, Chen and Volpe 2002, Fonseca *et al.* 2012, Lusardi and Mitchell 2014), men tend to achieve a higher financial literacy score than women. As Table 3 shows, a higher proportion of men answer the three financial literacy questions correctly.

Table 3. Means and standard deviations (in parentheses) of risk aversion and financial literacy, disaggregated by gender.

	Risk Aversion	Fin. Lit. 1	Fin. Lit. 2	Fin. Lit. 3
Male	0.025 (0.351)	0.936 (0.246)	0.766 (0.426)	0.670 (0.473)
Female	0.196 (0.429)	0.762 (0.428)	0.454 (0.500)	0.474 (0.502)
Ν	149	191	191	191
p-value	0.009	0.001	< 0.001	0.008

Note: For risk aversion, p-values come from a Mann-Whitney U-test; for each financial literacy question, p-values come from two-tailed Fisher exact tests. The lower number of observations for risk aversion is because 42 subjects made inconsistent choices in the MPL task.

In Figure 6 we partition men and women into four quartiles of risk aversion and four categories of financial literacy. As Figure 6 shows, the gap in trading activity persists in every subgroup.

[FIGURE 6 HERE]

Figure 6. Mean number of trades disaggregated by gender and risk aversion quartile (left), and by gender and financial literacy (right). Male data for a zero score in financial literacy is omitted due to having only a single observation. Mann-Whitney U-tests of gender differences in average number of trades give the following p-values: Risk Q1, p=0.003; Risk Q2, p=0.007; Risk Q3, p=0.039; Risk Q4, p=0.003. Fin. Lit. 1, p=0.001; Fin. Lit. 2, p=0.021; Fin. Lit. 3, p=0.002.

Regression analysis shows that the inclusion of risk aversion and financial literacy to the set of covariates makes a negligible difference to the estimated gender gap in trading activity. Table 4 presents the estimated gender effects using incrementally more controls (see Table O11 in the Online Appendix for full regression results). Moving from Model 1 to Model 4, we can observe that the addition of confidence (Model 2), risk aversion (Model 3) and financial literacy (Model 4) makes little difference to the estimated gender difference in the number of trades. Finally, in Model 5, we incorporate the full set of questionnaire variables as controls in the regression, which again makes no difference to the estimated gender effect.

Table 4. Marginal effects of "male" on (log) average number of transactions per period. Model (1) only includes gender, treatment dummies, their interactions, and group dummies; Model (2) adds 3 confidence dummies; Model (3) adds risk aversion; Model (4) adds financial literacy. Model 5 adds the full set of controls from the questionnaire.

	Model 1	Model 2	Model 3	Model 4	Model 5
В	0.581***	0.598***	0.577***	0.566***	0.572***
	(0.106)	(0.108)	(0.111)	(0.121)	(0.124)
С	0.600***	0.615***	0.593***	0.583***	0.587***
	(0.117)	(0.122)	(0.124)	(0.129)	(0.121)
Т	0.362**	0.379**	0.358**	0.348**	0.312**
	(0.161)	(0.158)	(0.158)	(0.157)	(0.132)
СТ	0.384***	0.396***	0.373***	0.363***	0.327***
	(0.132)	(0.131)	(0.133)	(0.126)	(0.099)
Confidence	NO	YES	YES	YES	YES
Risk Aversion	NO	NO	YES	YES	YES
Financial literacy	NO	NO	NO	YES	YES
Full set of controls	NO	NO	NO	NO	YES
Ν	566	566	566	566	566

Note: Estimates from random effects GLS regressions with group dummies and standard errors clustered by group. See Appendix for full regression results. To keep the four models comparable, 43 subjects were excluded from the regressions due to making inconsistent choices in the MPL task, which made it impossible to determine their risk aversion. *** p<0.01, ** p<0.05, * p<0.1

5 Discussion

Our experimental evidence replicates well-documented findings in the empirical literature that men are (*i*) more confident and (*ii*) trade more than women. However, we do not find any support for the widespread conjecture by which (*i*) explains (*ii*). The gender gap in trading activity persists when we compare men and women with the same confidence level, or after controlling for confidence in our regression analyses. Overconfidence does not help to explain any portion of the gender gap in trade volume either. In this respect, our results do not lend support to the conclusions of Barber and Odean (2001) that overconfidence is responsible for the gender gap in trading activity.

Three previous experimental studies have attempted to assess jointly the relationship between gender, overconfidence and trading activity. However, for reasons that will become apparent below, it is difficult to draw clear conclusions from their results.

Biais *et al.* (2005) study miscalibration and self-monitoring (a self-assessed measure of attention to social cues and adaptation to the social environment) in an experimental market with asymmetric information. They find that subjects with high miscalibration and low self-monitoring are more likely to suffer from the winner's curse and, consequently, earn lower profits. However, even though they find

significant gender differences in trading activity, neither miscalibration nor self-monitoring differ significantly between men and women.

Deaves *et al.* (2008) measure three forms of overconfidence: miscalibration, better-than-average ability and illusion of control. Better-than-average ability is measured after trading by asking subjects to guess how many people in the experiment had earned more money than themselves. Illusion of control is measured before trading by asking subjects for their level of agreement with two statements concerning their ability to detect or buy securities which will perform well in the future. In their experiment, subjects trade in several markets with asymmetric information. However, the informativeness of private signals depends on subjects' performance in the general knowledge miscalibration task. The relationship between miscalibration and subsequent trading behavior is, therefore, built into the experimental design. The authors find significant gender differences in trading activity only in sessions with a balanced gender mix. In contrast, they do not find significant gender differences in any of their overconfidence measures.¹²

Finally, Fellner-Röhling and Krügel (2014) conduct experimental markets with asymmetric information after measuring participants' miscalibration (both in general knowledge and in time series forecasting) and their degree of overweighting of private signals in a prediction task. Although trading activity is significantly higher in men, no gender differences are found in any of their overconfidence measures except in the general knowledge miscalibration task, in which women are in fact more overconfident than men.

The first and most important limitation of the three experimental studies above is that neither of them observe clear gender differences in the chosen measures of confidence. In contrast with the better-than-average effect, which has been shown to vary consistently between men and women, miscalibration has not been documented to differ systematically in the same way. This raises doubts as to whether miscalibration is the right measure to explain the gender gap in trading activity. A second limitation of these studies is that experimental asset markets are typically volatile and very sensitive to the actions of a single individual. The behavior of each trader in a given market does not constitute an independent observation, which makes it difficult to make reliable *ceteris paribus* comparisons between men and women. This does not happen in our experiment, where prices are determined independently of the actions of traders. A third limitation is that neither of these experiments involve trading costs. In contrast, the empirical finance literature emphasizes trading costs as a key reason why trading too much is costly (Barber et al., 2009; Barber and Odean, 2000; Odean, 1999). Lastly, with the exception of the

¹² Half of the sessions are conducted in Canada with groups of subjects with similar miscalibration and the same gender. The other half of the sessions are conducted in Germany with a balanced gender mix and either the most overconfident subjects or the least overconfident subjects in terms of the miscalibration task. Firstly, it is likely that the homogenous gender composition of some of these markets had a large effect on trading activity, as found in Cueva and Rustichini (2015). Secondly, the study is conducted with only 34 men and 30 women in total, which, combined with the heterogeneous market conditions, severely limits the power of their analysis.

signal-based prediction task in Fellner-Röhling and Krügel (2014), none of the confidence measures in these studies are incentivized.

Another empirical study closely related to ours is that of Glaser and Weber (2007). In their survey of online broker investors, they find a positive correlation between trade volume and better-thanaverage confidence, but not with measures of miscalibration. Their study, which focuses on the role of overconfidence rather than on gender, is conducted with a sample of approximately 95% male traders. Such a small sample of women (fewer than ten) makes it impossible to obtain clear results regarding gender differences.

In our experiment, confidence does not explain why men trade so much more than women, despite large gender differences in confidence. Of course, we cannot rule out the possibility that this result is the consequence of our specific experimental design. For instance, confidence is measured in one particular way, so we cannot be sure that other forms of confidence or overconfidence could have explained the gender gap in trade volume. In addition, the trading task does not involve multi-person trading nor real stocks, which may have dampened a potential effect of (over)confidence on trading activity.

As we argued at the end of Section 3, there are important advantages to a design in which subjects do not trade amongst themselves; namely, greater experimenter control and statistical power. Furthermore, this task has been very popular in the study of the disposition effect, with a sizeable number of studies employing it following Weber and Camerer (1998). The disadvantage of our design is that it departs significantly from the standard setup in behavioral finance models of overconfidence. However, as we formally show in the Appendix, overconfidence also leads to more trade in our environment. Notwithstanding the simplicity of our task, our findings raise the following question:

What factors, other than confidence, might explain gender differences in trading activity?

Our first approximation has been to examine gender differences in competitiveness, risk aversion and financial literacy. However, none of these measures are able to explain a significant portion of the gender gap in trading activity. Furthermore, we control for a number of psychological and demographic variables (such as Big Five personality traits and field of studies) in our regression analysis, but do not find any change in the estimated gender difference in trade volume.

One possibility that we have not examined is that men might enjoy trading more than women do.¹³ In their survey, Dorn and Sengmueller (2009) find that traders who report that they enjoy investing and gambling trade much more frequently than the rest. They also find that men are more likely to agree with the statements "I enjoy investing" and "I enjoy taking risky positions". Controlling for these survey responses somewhat reduces the gender gap in trading activity, although it still remains large and

¹³ Financial literacy, which tends to be higher in men than in women, may be correlated with enjoyment of trading. However, the inclusion of FLT scores in our analysis does not help to reduce the estimated gender gap in trade volume, which suggests that the part of enjoyment of trading that FLT may capture cannot help to explain the gap.

significant. Along similar lines, Grinblatt and Keloharju (2009) argue that, since speculative trading is a risky activity, it may be particularly attractive to sensation seekers. In support of this hypothesis, the authors find a significant positive correlation between speeding tickets (their proxy for sensation seeking) and trading activity. However, they continue to find a large gender effect and do not report how it changes by the inclusion of their sensation seeking measure.

Sensation-seeking and gambling attitudes have been found to systematically vary by gender. For instance, Cross *et al.* (2011) find in a meta-analysis that men are more sensation seeking than women on both questionnaire and behavioral measures, while Salonen *et al.* (2017) observe that men have more positive attitudes towards gambling than women.

In a recent experimental study, Corgnet *et al.* (2018) find that fluid intelligence, cognitive reflection and theory of mind are significant positive predictors of trading performance. In line with previous results in the literature, the authors also find that men outperform women on cognitive reflection (Bosch-Domènech et al., 2014; Cueva et al., 2016; Frederick, 2005; Oechssler et al., 2009), while women outperform men in theory of mind (Baron-Cohen et al., 1997). In their experiment, Corgnet *et al.* (2018) also find that men trade significantly more than women. However, they do not explore whether gender differences in cognitive reflection or theory of mind have any explanatory power.

How much of the gender gap in trading activity can be accounted for by a combination of these measures remains an open question. However, our results suggest that future research should look beyond overconfidence.

Appendix A

A model of overconfidence and portfolio choice

We present a simple model of portfolio choice using the standard mean-variance model, following the notation of Eeckhoudt *et al.* (2005), to which we add overconfidence.

There are *n* risky assets, indexed by i = 1, ..., n, and a risk-free asset. The return of asset *i* is x_i , with mean μ_i and variance σ_i^2 . In our experimental setup, the risk-free asset (cash) pays no returns, and the returns of the different risky assets are independently distributed, so all covariances are zero. Investors choose the quantity to invest in each risky asset, a_i . Normalizing initial wealth to 1, final wealth equals:

$$z=1+\sum_{i=1}^n a_i x_i \, .$$

Individuals maximize a utility function defined on the mean-variance space as:

$$u(z) = E[z] - \frac{1}{2}A \operatorname{var}[z],$$

where A is the coefficient of absolute risk aversion.

Solving this maximization problem yields optimal amounts invested in each asset *i*:

$$a_i^* = \frac{\mu_i}{\sigma_i^2 A},$$

and the amount held in the safe asset (cash):

$$a_0^* = 1 - \sum_{i=1}^n a_i^*$$

We introduce overconfidence in this model in a very simple way. We assume that, for any given asset, an overconfident investor underestimates the variance of the asset's returns by a factor $\lambda \in [0,1]$. Therefore, an unbiased investor has $\lambda = 1$, while an overconfident investor has $\lambda < 1$. This captures the idea that overconfident investors have the same information as unbiased investors (namely, assets' prices), but believe that their information is more precise than it really is.

The optimal choice of an investor with overconfidence λ will then be:

$$a_i^{OC} = \frac{\mu_i}{\lambda \sigma_i^2 A}$$
,

which means that the optimal amount invested in each risky asset is an increasing function of overconfidence. Because of the simple way in which we model overconfidence, the relative share of each asset is independent of overconfidence. Overconfidence affects only the total fraction of wealth invested in risky assets, not how individuals choose among the different risky assets. In particular, compared to an unbiased trader, an overconfident trader's portfolio is larger by a factor of $1/\lambda$.

Since overconfidence entails bigger portfolios, it also implies more trading. Let asset prices change from period *t* to period *t*+1 such that an unbiased trader would need to make *m* transactions to update her optimal portfolio from *t* to *t*+1. An overconfident trader would then need to make m/λ transactions.

[FIGURE A1 HERE]

Figure A1. Cumulative distribution of earnings of men and women.

Table A1. Marginal effects of "male" on (log) average number of transactions per period. Model 1 includes a male dummy, treatment dummies, their interactions, and group dummies. Model 2 overconfidence. Model 3 adds overconfidence×male interactions. Model 4 adds overconfidence×treatment and overconfidence×male×treatment interactions.

	Model 1	Model 2	Model 3	Model 4
В	0.482***	0.479***	0.479***	0.487***
	(0.077)	(0.077)	(0.077)	(0.076)
С	0.504***	0.504***	0.504***	0.503***
	(0.094)	(0.094)	(0.094)	(0.093)
Т	0.303**	0.299**	0.299**	0.274**
	(0.133)	(0.134)	(0.134)	(0.137)
СТ	0.341***	0.338***	0.339***	0.333***
	(0.108)	(0.109)	(0.109)	(0.111)
Overconfidence	NO	YES	YES	YES
+ Gender interactions	NO	NO	YES	YES
+ Gender & Treatment interactions	NO	NO	NO	YES
N	730	730	730	730

Note: Estimates from random effects GLS regressions with group dummies and standard errors clustered by group. See Table O5 in the Appendix for full regression results. *** p<0.01, ** p<0.05, * p<0.1

Table A2. Linear regressions. Dependent variable: (1) Portfolio value, (2) Rationality index, (3) Herfindahl index. Data is averaged at the subject level.

	(1)	(2)	(3)
	Portfolio value	Rationality index	Herfindahl index
Male	790.6***	0.0358*	0.0430**
	(183.2)	(0.0188)	(0.0194)
Confidence	227.1	0.0100	0.0401**
	(135.8)	(0.0180)	(0.0159)
Observations	191	191	191

Robust standard errors clustered by group in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table A3. Linear regressions. Dependent variable: log of number of trades. All variables are averaged at the subject level, over all periods and treatments.

	(1)	(2)	(3)	(4)	(5)
Male	0.391***	0.144*	0.374***	0.404***	0.161**
	(0.0795)	(0.0729)	(0.0824)	(0.0796)	(0.0662)
Portfolio value		0.000280***			0.000308***
		(2.29e-05)			(2.62e-05)
Rationality index			0.437		-0.418
			(0.270)		(0.272)
Herfindahl index				-0.226	-0.420*
				(0.282)	(0.234)
Observations	191	191	191	191	191

Robust standard errors clustered by group in parentheses

*** p<0.01, ** p<0.05, * p<0.1

	Male	Female	Difference test
Trades	9.079 (4.076)	5.95 (3.136)	p<0.0001***
Confidence	3.229 (0.614)	2.829 (0.593)	p<0.0001***
Earnings	5097 (332)	5020 (225)	p=0.1336
Overconfidence	0.663 (0.861)	0.395 (0.836)	p=0.0319**
Risk Aversion	$0.025~(0.351)^{\dagger}$	$0.196~(0.429)^{\dagger}$	p=0.0085***
Fin. Lit. 1	0.936 (0.246)	0.763 (0.428)	p=0.0009***
Fin. Lit. 2	0.766 (0.426)	0.454 (0.5)	p<0.0001***
Fin. Lit. 3	0.67 (0.473)	0.474 (0.502)	p=0.0064***
Optimism	0.745 (0.438)	0.526 (0.502)	p=0.0017***
Risk willingness	0.596 (0.493)	0.557 (0.499)	p=0.5861
High Confidence	0.713 (0.455)	0.526 (0.502)	p=0.0080***
Difficulty recognizing mistakes	0.606 (0.491)	0.557 (0.499)	p=0.4877
Competitiveness	0.713 (0.455)	0.577 (0.497)	p=0.0513*
Enjoys Winning	0.734 (0.444)	0.649 (0.48)	p=0.2073
Decisiveness	0.617 (0.489)	0.577 (0.497)	p=0.5770
Self-control	0.489 (0.503)	0.577 (0.497)	p=0.2243
Average grade	6.902 (0.916)	6.957 (0.896)	p=0.6804
Neuroticism	-0.271 (0.97)	0.261 (1.012)	p=0.0005***
Extraversion	-0.044 (1.047)	0.093 (0.982)	p=0.3154
Agreeableness	-0.001 (1.088)	0.037 (0.943)	p=0.7020
Openness	0.108 (0.997)	-0.04 (1.01)	p=0.3978
Conscientiousness	-0.119 (0.925)	0.18 (1.057)	p=0.0523*
Age	22.308 (3.520)	21.814 (2.848)	p=0.4070
Art	0.096 (0.296)	0.113 (0.319)	p=0.6911
Health	0.245 (0.432)	0.392 (0.491)	p=0.0297**
Science	0.096 (0.296)	0.103 (0.306)	p=0.8657
Engineering	0.223 (0.419)	0.134 (0.342)	p=0.1013
Social Sciences	0.34 (0.476)	0.258 (0.44)	p=0.2130
Observations	94 †76	97 †73	

Table A4. Summary statistics

Note: difference tests are Mann-Whitney U tests of the null hypothesis of no difference between male and female.

Appendix C

Questionnaire items

• Financial Literacy Test (Lusardi and Mitchell, 2008).

Question 1: Suppose you had \$100 in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow? a) More than \$102; b) Exactly \$102; c) Less than \$102; d) Do not know; e) Refuse to answer.

Question 2: Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, how much would you be able to buy with the money in this account? a) More than today; b) Exactly the same; c) Less than today; d) Do not know; e) Refuse to answer.

Question 3: Please specify whether this statement is true or false. "Buying a single company's stock usually provides a safer return than a stock mutual fund." a) True; b) False; c) Do not know; d) Refuse to answer.

- Big Five questionnaire: Spanish version of the brief form (Benet-Martinez and John, 1998)
- Self-control: Brief Self Control measure (13 items) by Tangney *et al.* (2004).
- **Other items:** (Answers from 1 = completely disagree to 7 = completely agree)
 - Are you an optimistic or a pessimistic person?
 - Are you a person willing to take risks or do you try to avoid risks?
 - Are you a confident person?
 - I find it hard to recognize my mistakes
 - I am a very competitive person
 - I enjoy very much winning in a game

References

- Agnew, J., Balduzzi, P., Sundén, A., 2003. Portfolio Choice and Trading in a Large 401 (k) Plan. Am. Econ. Rev. 401, 193–215.
- Almenberg, J., Dreber, A., 2015. Gender, stock market participation and financial literacy. Econ. Lett. 137, 140–142. doi:10.1016/j.econlet.2015.10.009
- Barber, B.M., Lee, Y.-T., Liu, Y.-J., Odean, T., 2009. Just How Much Do Individual Investors Lose by Trading? Rev. Financ. Stud. 22, 609–632. doi:10.1093/rfs/hhn046
- Barber, B.M., Odean, T., 2001. Boys Will be Boys: Gender, Overconfidence, and Common Stock Investment. Q. J. Econ. 116, 261–292.
- Barber, B.M., Odean, T., 2000. Trading is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors. J. Finance 55, 773–806. doi:10.2139/ssrn.219228
- Baron-Cohen, S., Jolliffe, T., Mortimore, C., Robertson, M., 1997. Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger Syndrome. J. Child Psychol. Psychiatry 38, 813–822.
- Benet-Martínez, V., John, O.P., 1998. Los Cinco Grandes Across Cultures and Ethnic Groups:
 Multitrait Multimethod Analyses of the Big Five in Spanish and English. J. Pers. Soc. Psychol. 75, 729–750. doi:10.1037/0022-3514.75.3.729
- Bengtsson, C., Persson, M., Willenhag, P., 2005. Gender and overconfidence. Econ. Lett. 86, 199– 203. doi:10.1016/j.econlet.2004.07.012
- Benos, A. V, 1998. Aggressiveness and survival of overconfident traders. J. Financ. Mark. 1, 353–383.
- Beyer, S., 1990. Gender Differences in the Accuracy of Self- Evaluations of Performance. J. Pers. Soc. Psychol. 59, 960–970. doi:10.1037/0022-3514.59.5.960
- Beyer, S., Bowden, E., 1997. Gender differences in seff-perceptions: Convergent evidence from three measures of accuracy and bias. Personal. Soc. Psychol. Bull. 23, 157–172.
- Biais, B., Hilton, D., Mazurier, K., Pouget, S., 2005. Judgmental Overconfidence, Self-Monitoring, and Trading Performance in an Experimental Financial Market. Rev. Econ. Stud. 72, 287–312.
- Bosch-Domènech, A., Brañas-Garza, P., Espín, A.M., 2014. Can exposure to prenatal sex hormones (2D:4D) predict cognitive reflection? Psychoneuroendocrinology 43, 1–10. doi:10.1016/j.psyneuen.2014.01.023
- Campbell, N.K., Hackett, G., 1986. The effects of mathematics task performance on math selfefficacy and task interest. J. Vocat. Behav. 28, 149–162. doi:10.1016/0001-8791(86)90048-5
- Charness, G., Gneezy, U., 2012. Strong Evidence for Gender Differences in Risk Taking. J. Econ. Behav. Organ. 83, 50–58. doi:10.1016/j.jebo.2011.06.007
- Chen, H., Volpe, R.P., 2002. Gender Difference in Personal Financial Literacy Among College

Students. Financ. Serv. Rev. doi:10.5897/AJBM10.1267

- Corgnet, B., Desantis, M., Porter, D., 2018. What Makes a Good Trader? On the Role of Intuition and Reflection on Trader Performance. J. Finance LXXIII. doi:10.1111/jofi.12619
- Croson, R., Gneezy, U., 2009. Gender Differences in Preferences. J. Econ. Lit. 47, 1–27. doi:10.1257/jel.47.2.448
- Cross, C.P., Copping, L.T., Campbell, A., 2011. Sex differences in impulsivity: A meta-analysis. Psychol. Bull. 137, 97–130. doi:10.1037/a0021591
- Cueva, C., Iturbe-Ormaetxe, I., Mata-Pérez, E., Ponti, G., Sartarelli, M., Yu, H., Zhukova, V., 2016. Cognitive (ir)reflection: New experimental evidence. J. Behav. Exp. Econ. 64. doi:10.1016/j.socec.2015.09.002
- Cueva, C., Iturbe-Ormaetxe, I., Ponti, G., Tomás, J., 2018. An experimental analysis of the Disposition Effect: who and when? Work. Pap.
- Cueva, C., Rustichini, A., 2015. Is financial instability male-driven? Gender and cognitive skills in experimental asset markets. J. Econ. Behav. Organ. 119, 330–344. doi:10.1016/j.jebo.2015.08.014
- Daniel, K., Hirshleifer, D., 2015. Overconfident Investors, Predictable Returns, and Excessive Trading 29, 61–88.
- Daniel, K., Hirshleifer, D., Subrahmanyam, A., 1998. Investor Psychology and Security Market Under- and Overreactions. J. Finance 53, 1839–1885. doi:10.1111/0022-1082.00077
- Daniel, K.D., Hirshleifer, D., Subrahmanyam, A., 2001. Overconfidence, Arbitrage, and Equilibrium Asset Pricing. Finance LVI, 921–965. doi:10.1111/0022-1082.00350
- Deaves, R., Luders, E., Luo, G.Y., 2008. An Experimental Test of the Impact of Overconfidence and Gender on Trading Activity. Rev. Financ. 13, 555–575. doi:10.1093/rof/rfn023
- Dorn, D., Sengmueller, P., 2009. Trading as Entertainment? Manage. Sci. 55, 591–603. doi:10.1287/mnsc.1080.0962
- Duxbury, D., 2015. Behavioral finance: insights from experiments II: biases, moods and emotions. Rev. Behav. Financ. 7, 151–175.
- Eckel, C.C., Füllbrunn, S.C., 2015. Thar she blows? gender, competition, and bubbles in experimental asset markets. Am. Econ. Rev. 105, 906–920. doi:10.1257/aer.20130683
- Eckel, C.C., Grossman, P.J., 2008. Men, Women and Risk Aversion: Experimental Evidence, in: Handbook of Experimental Economics Results. pp. 1061–1073. doi:10.1016/S1574-0722(07)00113-8
- Eeckhoudt, L., Gollier, C., Schlesinger, H., 2005. Economic and Financial Decisions under Risk. Princeton University Press.
- Estes, R., Hosseini, J., 1988. The Gender Gap on Wall Street: An Empirical Analysis of Confidence in Investment Decision Making. J. Psychol. 122, 577–590. doi:10.1080/00223980.1988.9915532

Fellner-Röhling, G., Krügel, S., 2014. Judgmental overconfidence and trading activity. J. Econ.

Behav. Organ. 107, 827-842. doi:10.1016/j.jebo.2014.04.016

- Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Exp. Econ. 10, 171–178. doi:10.1007/s10683-006-9159-4
- Fonseca, R., Mullen, K.J., Zamarro, G., Zissimopoulos, J., 2012. What Explains the Gender Gap in Financial Literacy? The Role of Household Decision Making. J. Consum. Aff. 46, 90–106. doi:10.1111/j.1745-6606.2011.01221.x
- Frederick, S., 2005. Cognitive Reflection and Decision Making. J. Econ. Perspect. 19, 25–42. doi:10.1257/089533005775196732
- French, K.R., 2008. Presidential address: The cost of active investing. J. Finance 63, 1537–1573. doi:10.1111/j.1540-6261.2008.01368.x
- Glaser, M., Weber, M., 2007. Overconfidence and trading volume. Geneva Risk Insur. Rev. 32, 1–36. doi:10.1007/sl07
- Gneezy, U., Leonard, K.L., List, J.A., 2009. Gender Differences in Competition: Evidence From a Matrilineal and a Patriarchal Society. Econometrica 77, 1637–1664. doi:10.3982/ECTA6690
- Gneezy, U., Niederle, M., Rustichini, a., 2003. Performance in Competitive Environments: Gender Differences. Q. J. Econ. 118, 1049–1074. doi:10.1162/00335530360698496
- Gneezy, U., Rustichini, A., 2004. Gender and Competition at a Young Age. Am. Econ. Rev. 94, 377–381. doi:10.1257/0002828041301821
- Greiner, B., 2004. The Online Recruitment System ORSEE 2.0 A Guide for the Organization of Experiments in Economics, University of Cologne Working Paper Series in Economics. University of Cologne.
- Grinblatt, M., Keloharju, M., 2009. Sensation seeking, overconfidence, and trading activity. J. Finance 64, 549–578. doi:10.1111/j.1540-6261.2009.01443.x
- Holt, C.A., Laury, S.K., 2002. Risk Aversion and Incentive Effects. Am. Econ. Rev. 92, 1644–1655.
- Hyde, J.S., Fennema, E., Ryan, M., Frost, L.A., Hopp, C., 1990. Gender comparisons of mathematics attitudes and affect. Psychol. Women Q. 14, 299–324.
- John, O.P., Srivastava, S., 1999. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. Handb. Personal. Theory Res. 2, 102–138. doi:citeulike-article-id:3488537
- Lundeberg, M.A., Fox, P.W., Puncochar, J., 1994. Highly Confident but Wrong: Gender Differences and Similarities in Confidence Judgments. J. Educ. Psychol. 86, 114–121.
- Lusardi, A., Mitchell, O.S., 2014. The Economic Importance of Financial Literacy: Theory and Evidence. J. Econ. Lit. 52, 5–44. doi:http://dx.doi.org/10.1257/jel.52.1.1
- Lusardi, A., Mitchell, O.S., 2008. Planning and financial literacy: How do women fare? Am. Econ. Rev. 98, 413–417. doi:10.1257/aer.98.2.413
- McCrae, R.R., John, O.P., 1992. An Introduction to the Five-Factor Model and Its Applications. J. Pers. 60, 175–215. doi:10.1111/j.1467-6494.1992.tb00970.x
- Michailova, J., Schmidt, U., 2016. Overconfidence and Bubbles in Experimental Asset Markets. J.

Behav. Financ. 17, 280–292. doi:10.1080/15427560.2016.1203325

- Moore, D.A., Healy, P.J., 2008. The Trouble With Overconfidence. Psychol. Rev. 115, 502–517. doi:10.1037/0033-295X.115.2.502
- Niederle, M., Vesterlund, L., 2007. Do women shy away from competition? do men compete too much? Q. J. Econ. 122, 1067–1101.
- Odean, T., 1999. Do investors trade too much? Am. Econ. Rev. 89, 1279–1298.
- Odean, T., 1998. Volume, Volatility, Price, and Profit When All Traders Are Above Average. J. Finance 53, 1887–1934. doi:10.1111/0022-1082.00078
- Oechssler, J., Roider, A., Schmitz, P.W., 2009. Cognitive abilities and behavioral biases. J. Econ. Behav. Organ. 72, 147–152. doi:10.1016/j.jebo.2009.04.018
- Prince, M., 1993. Women, men and money styles. J. Econ. Psychol. 14, 175–182. doi:10.1016/0167-4870(93)90045-M
- Saccardo, S., Pietrasz, A., Gneezy, U., 2017. On the Size of the Gender Difference in Competitiveness. Manage. Sci. 64, 1541–1554. doi:10.1287/mnsc.2016.2673
- Salonen, A.H., Alho, H., Castrén, S., 2017. Attitudes towards gambling, gambling participation, and gambling-related harm: cross-sectional Finnish population studies in 2011 and 2015. BMC Public Health 17, 122. doi:10.1186/s12889-017-4056-7
- Smith, V., Suchanek, G., Williams, A., 1988. Bubbles , Crashes , and Endogenous Expectations in Experimental Spot Asset Markets. Econometrica 56, 1119–1151.
- Tangney, J.P., Baumeister, R.F., Boone, A.L., 2004. High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. J. Pers. 72, 271–324. doi:10.1111/j.0022-3506.2004.00263.x
- Weber, M., Camerer, C.F., 1998. The disposition effect in securities trading: an experimental analysis.J. Econ. Behav. Organ. 33, 167–184. doi:10.1016/S0167-2681(97)00089-9
- Weber, M., Welfens, F., 2007. An Individual Level Analysis of the Disposition Effect : Empirical and Experimental Evidence. Work. Pap.
- Yablonsky, L., 1991. The Emotional Meaning of Money, Gardner Pr. ed. New York.