An Experimental Analysis of the Disposition Effect: Who and When?*

Carlos Cueva*  
Universidad de Alicante

Iñigo Iturbe-Ormaetxe  
Universidad de Alicante

Giovanni Ponti  
Universidad de Alicante  
The University of Chicago  
LUISS Guido Carli Roma

Josefa Tomás  
Universidad de Alicante

July 2019

Abstract

The disposition effect (DE) is a common bias by which investors tend to sell winning assets too soon and hold losing assets too long. We complement the existing evidence in three directions. First, we check whether the DE is robust to realistic features such as transaction costs and competitive payment schemes. Second, by using a gender-balanced design, we check for gender differences. Third, we search for psychological correlates of the DE. We find that the DE is positive and significant in all our treatments. We do not find significant differences across treatments, although transaction costs significantly reduce the propensity to sell both winners and losers. We find somewhat larger DE in women, but this effect is only significant in the second half of the experiment. On the other hand, women are more reluctant to sell losing assets throughout the experiment. Finally, we find that the most significant psychological predictors of the DE are difficulty recognizing one’s mistakes and optimism. Subjects scoring high in these traits are less likely to sell at a loss and therefore exhibit a larger DE. Our results provide further suggestive evidence of cognitive dissonance as an important determinant of the DE.

KEYWORDS: Behavioral Finance, Psychological characteristics, Gender, Cognitive Dissonance.

JEL CLASSIFICATION: C91, D70, D81, D91

* We thank Holger Rau for useful comments and for sharing his own experimental software and Esther Mata-Pérez, Haihan Yu and Vita Zhukova for valuable research assistance. The usual disclaimers apply. Financial support from the Spanish Ministries of Education and Science and Economics and Competitiveness (ECO2015-65820-P) and Instituto Valenciano de Investigaciones Económicas (IVIE) is gratefully acknowledged.

* Corresponding author. Departamento de Fundamentos del Análisis Económico, Universidad de Alicante, E-03071 Alicante (SPAIN). carlos.cueva@ua.es
1 Introduction

One of the most well-known decision-making biases in behavioral finance is the tendency of individual investors to sell winning stocks much more frequently (or earlier) than losing stocks. Following the seminal work by Shefrin and Statman (1985), this bias has come to be labeled as the *disposition effect* (DE). Such a tendency has been persistently observed in empirical studies of investor behavior. Odean (1998) finds that individual investors have a strong preference for selling winners and holding losers. Grinblatt and Keloharju (2000, 2001) use data of stock market investors in Finland, and find that investors are about twice as likely to sell a stock at a moderate gain as at a moderate loss. Shapira and Venezia (2001) confirm the prevalence of the DE among independent investors and professional brokers in Israel. Dhar and Zhu (2006), Brown *et al.* (2006) and Barber *et al.* (2007) document similar findings among investors in the United States, Australia and Taiwan, respectively. The DE has also been documented in real estate markets (Genesove and Mayer, 2001; Einiö *et al.*, 2008), sports betting (Hartzmark and Solomon, 2012) and in laboratory experiments (Weber and Camerer, 1998; Weber and Welfens, 2007; Da Costa *et al.*, 2008, 2013; Rau, 2014, 2015; Frydman *et al.*, 2014; Magnani, 2015; Jiao, 2017).

Various alternative behavioral theories have been put forward to rationalize the DE, which, for the sake of clarity, can be grouped into *preference-based*, *beliefs-based* or *mixed*. The most prominent preference-based theories of the DE rely on *prospect theory* (Kahneman and Tversky, 1979; Shefrin and Statman, 1985) or on *realization utility* (Barberis and Xiong, 2012; Ingersoll and Jin, 2013). According to both theories, investors employ a form of *mental accounting* by which gains and losses are evaluated separately for each stock. According to prospect theory, investors become risk averse when a stock has appreciated and risk loving when it has depreciated, resulting in a positive DE. Alternatively, investors who experience realization (dis)utility from selling winners (losers), may also exhibit a positive DE.\(^1\) The most prominent beliefs-based explanation of the DE is investors’ (mistaken) beliefs in mean reversion (Andreassen, 1988; Odean, 1998).\(^2\) Finally, an increasingly popular theory of the DE, which combines preferences and beliefs, is that of cognitive dissonance (Festinger, 1957; Antoniou *et al.*, 2013; Chang *et al.*, 2016).\(^3\) According to this theory, individuals tend to hold consistent beliefs, attitudes and actions, whereas events that upset this consistency are psychologically costly and often avoided, particularly when they can affect self-image negatively. Applied to financial decisions, selling a losing asset carries an additional psychological cost, since it involves the acceptance that prior beliefs and actions were wrong, while selling a winning asset has the additional value of confirming previous beliefs and actions.

---

\(^1\) For empirical evidence in support of prospect theory, see Barberis and Xiong (2009) and Kaustia (2010a). For experimental evidence of realization utility, see Frydman *et al.* (2014).

\(^2\) However, with the exception of Jiao (2017) who finds partial experimental evidence for mean reversion, this theory has received little empirical support (Weber and Camerer, 1998; Odean, 1999; Kaustia, 2010a).

\(^3\) See also e.g. Kaustia (2010b), who refers to this theory also as *self-justification*. 
In this paper we follow Odean (1998) in defining the DE as the difference between the proportion of realized gains (PGR, a proxy for the tendency of selling a winner too soon) and the proportion of realized losses (PLR, a proxy for the tendency of holding a winner too long). Although both an increase in PGR and a reduction in PLR cause the DE to increase, many scholars agree that these are two independent biases (e.g. Weber and Welfens, 2008). For this reason, in our statistical analysis, we investigate not only the DE, but also its two building blocks PGR and PLR.

This paper contributes to the literature on the DE in several ways. First, we experimentally check for the robustness of the DE by designing novel alternative treatments that introduce two important institutional factors in real-world markets so far neglected in the experimental literature: transaction costs and competitive payment schemes. Transaction costs are prevalent in financial markets, so they should be incorporated into any experiment that tries to resemble real markets. The results that we might obtain from an experiment without transaction costs may not be transferable to the real world in which these costs are widespread. However, it is not clear a priori whether transaction costs can cause the DE to increase or decrease. On the one hand, transaction costs may potentially induce subjects to pay more attention and make fewer mistakes, since reverting unintended trades becomes more costly. This effect (paying more attention) could perhaps lead to a reduction of the DE. On the other hand, transaction costs may increase the relative weight of overconfident investors who are active in the market, which, in turn, may exhibit a larger DE (Kadous, et al. 2014). As it turns out, we do not find that adding transaction costs has any significant effect on the DE. We do find that subjects trade less, but they reduce both the sales of winning and of losing assets in approximately equal proportion.

We also introduce competition in another treatment. Competitive payment schemes are very common among professional traders and have been shown to affect portfolio choices, improving stock selection ability (Elton et al. 2003), but also affecting portfolio volatility (Brown et al., 1996; Dijk et al., 2014). Goulart et al. (2015) conduct an experiment in which profits are made public, finding that this increases the DE by inducing a higher rate of realized winners. However, competition here is only for prestige and does not affect subjects’ actual earnings in the experiment. Another reason for incorporating competition in our experiment has to do with our interest in identifying gender effects, since several authors have found that males and females react differently to competition. In particular, Gneezy et al. (2003) and Niederle and Vesterlund (2007) find that women tend to perform worse in competitive scenarios. Based on this evidence, we conjecture that competition might differentially increase the DE of women compared to men.

---

4 A prominent explanation of why individual investors actively trade in financial markets with significant transaction costs is that they are overconfident (Barber and Odean, 2000; 2001).

5 The analysis of (gender effects in) trade volume for this experiment is discussed in detail in Cueva et al. (2019).
Related to this is the more general issue of gender differences in the DE, which constitutes the second contribution of our paper. To our knowledge, gender differences in the DE have only been investigated in the field by Feng and Seasholes (2005), who find that women are more reluctant to sell losing stocks than men in a sample of approximately 1,500 (49% female) Chinese investors. On the other hand, Weber and Welfens (2007) find no evidence of gender differences in their field data, although this is not the focus of their study and could be simply driven by the fact that they have very few females in their sample. Experimental evidence is also far from conclusive. Da Costa et al. (2008) find that males exhibit a larger DE than females, although their results depend on a particular specification of the DE. In sharp contrast, Rau (2014) finds a significant DE for females and not for males. Our results are in line with Feng and Seasholes (2005), in that we find that women are significantly more reluctant to sell losers than men. As in Rau (2014), we find that the DE is larger for females, although the difference only becomes statistically significant towards the end of the experiment. Finally, we find that the introduction of competition does not affect the difference in DE between men and women. Our findings therefore complement the broader literature on gender differences in competition and trading.

The third contribution of our paper is to explore the relationship between the DE and individual characteristics, with particular emphasis on psychological factors which may modulate the bias according to prominent theories of the DE. In our view, this is an important underexplored aspect of the DE that could help inform the theoretical debate. To our knowledge, the only studies that examine psychological determinants of the DE are Chui (2001) and Kadous, et al. (2014). Chui (2001) finds suggestive evidence that subjects with an internal locus of control (a tendency to feel that events are under their own control) have a higher DE. Kadous, et al. (2014) find that subjects with lower self-regard and higher investment confidence hold losing assets longer. Both of these results are consistent with cognitive dissonance. Subjects with an internal locus of control feel more responsibility for their successes or failures, and therefore experience more cognitive dissonance when selling at a loss. Similarly, subjects with low self-regard or high investment confidence will be discouraged from selling losers in order to protect their self-image.

We find that the DE is highly correlated with subjects’ self-assessed difficulty recognizing mistakes and with their general optimism. As expected, both effects operate via a reduced tendency to sell losing assets. We interpret these findings as further suggestive evidence of the cognitive dissonance explanation of investors’ reluctance to realize losses.

The remainder of this paper is arranged as follows. Section 2 describes our experimental design. Section 3 reports our experimental findings. Section 4 concludes. Additional results and the experimental instructions are available in an online appendix.
2 Experimental Design

A total of 192 subjects, 97 female and 95 male, participated in our study. We ran 8 sessions of 24 subjects each at the Laboratory of Theoretical and Experimental Economics (LaTEx) at the University of Alicante. We recruited subjects using ORSEE (Greiner, 2004), mainly from the undergraduate population. The experiment was programmed in z-Tree (Fischbacher, 2007) and it lasted for about two hours, with average earnings of 19.10 €.

We borrow the baseline experimental design from Weber and Camerer (1998) and Weber and Welfens (2007). In the baseline, subjects can buy and sell six risky assets (A, B, C, D, E, and F) over 9 periods using experimental currency. Each asset follows a different stochastic price path, independently of subjects’ actions. We introduce two variations of the baseline treatment: i) transaction costs (T) and ii) competitive payments (C). All subjects go through the baseline treatment, the T treatment, the C treatment, and the C+T treatment. The treatment order is counterbalanced. Subjects play a short trial before each treatment.

2.1 Prices

We randomly assign subjects into gender-balanced groups of four (two males and two females). They know that they are in a group of four but they are not told the identity, nor the gender, of their group mates. To ensure that subjects perceive a similar gender-balanced environment in every session, the positions of men and women are always interspersed in the laboratory. We tell subjects that each group of four faces the same price path of the six assets.

We randomized price paths between matching groups in order to generate exogenous variation across independent observations. On the other hand, we kept price paths equal within matching groups in order to test for gender differences in behavior within the same market environment and to facilitate the comparison of subjects’ performance within groups (with specific reference to the competitive treatments).

The history of each market begins in period −3, with the same initial price of 100 pesetas for all assets. From period −2 onwards, prices can go up by 6% or down by 5%. We tell subjects that each asset has a different probability of a price increase, but we do not reveal these probabilities. We tell subjects that the probability of a price increase is constant over the whole treatment and that price changes are independent of previous prices and subjects’ trades. In particular, we assign randomly to each asset a probability of a price increase that is chosen without

---

6 Our baseline treatment is essentially the same as the neutral treatment in Weber and Welfens (2007) except that we generate asset prices randomly for each group rather than using a unique predetermined price path. In our experiment, the experimental currency is called “pesetas”, and has exchange rate 1000 pesetas = 1 €.

7 Sessions 1 and 2: B, C, T, C+T; sessions 3 and 4: T, C+T, B, C; sessions 5 and 6: C, B, C+T, T; sessions 7 and 8: C+T, T, C, B.

8 This is done by asking male and female subjects to draw a number from different envelopes: one contains odd numbers and the other even numbers. Subjects are then seated on 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 sex are asked to draw a number from the envelope assigned to the other sex.
replacement from the set \( \{0.4, 0.45, 0.5, 0.5, 0.55, 0.6\} \). This implements a market in which there are always two “good” assets (those with a probability of a price increase above 0.5), two “neutral” assets (0.5), and two “bad” assets (< 0.5). Since the probability of a price increase for each asset is unknown and constant, a rational subject should infer that appreciating assets are more likely to continue appreciating than depreciating assets. This implies that rational subjects should be more likely to sell losing assets than winning assets. Therefore, a positive DE clearly constitutes a costly bias in this environment.

Subjects observe prices from period −3 to 10 but can only trade from period 1 to 9. They have an initial endowment of 5,000 pesetas to trade in the market. They cannot buy on credit or short-sell. Prices at period 10 are only used to liquidate portfolios.

To elicit beliefs about the asset types, in periods 1, 6, and 10 we ask subjects to guess which asset has the highest, second highest, lowest and second lowest probability of a price increase, respectively. The task is incentivized by paying 100 pesetas when all guesses in a given period turn out to be correct.

2.2 Treatments

In order to better control for individual heterogeneity in the analysis of treatment effects and to maximize power given our sample size, we employ a within-subjects design. In other words, all subjects participate in the four treatments. In the baseline treatment (B), they earn the value of the portfolio upon liquidation (period 10) plus their remaining cash. In treatment C, only the winner in each group of four earns the value of her portfolio plus remaining cash, multiplied by two. The remaining three group members do not receive the value of their portfolios, nor their remaining cash. In treatment T, subjects have to pay a fee for each transaction. The fee is a fixed rate, randomized across groups over the set \( \{1\%, 4\%, 7\%\} \) and constant throughout the session. The fee is the same for purchases and sales. Finally, in treatment C+T, we combine the features of treatments C and T. To elicit subjects’ self-confidence, we ask them to guess, at the beginning and at the end of each treatment, their ranking—in terms of earnings—within their group. They receive an additional 100 pesetas if they guess their position correctly.

2.3 Individual measures

Before going through the four treatments, we elicit risk preferences using an incentivized Multiple Price List (MPL, Holt and Laury, 2002). Subjects go through a sequence of 21 binary decisions between a lottery and a safe option. The lottery is always the same: 5,000 or 0 pesetas with equal probability. Safe options range from zero to 5,000 pesetas, increasing by a constant step of 250 pesetas along the sequence. At the end of the experiment, one of the 21 choices is randomly selected and the relevant lottery is played out for payment.
At the end of the session we ask subjects to complete a questionnaire. The questionnaire includes some basic socio-demographics and a number of questions that aim to capture different psychological traits that have been associated to the DE or to closely related phenomena. We include measures of **indecisiveness** (Germeijs and De Boeck, 2002), **self-control** (Tangney et al., 2004), **difficulty recognizing one’s mistakes**, **optimism**, **willingness to take risks**, **confidence**, **competitiveness**, **enjoyment of winning** and financial literacy (Lusardi and Mitchell, 2008, see the online appendix for details). We investigate the impact of these variables on the DE for the following reasons:

1. Indecisive individuals often report regretting their decisions. In turn, regret aversion has been proposed as a possible source of the DE (Shefrin and Statman, 1985; Frydman and Camerer, 2016).

2. Self-control has also been related to the DE (Shefrin and Statman, 1985). Under the view that the DE is caused by “realization utility” (Shefrin and Statman, 1985; Barberis and Xiong, 2012; Ingersoll and Jin, 2013), one might also expect self-control, willingness to take risks, competitiveness and enjoyment of winning to be correlated with the DE.

3. Optimism and confidence feature prominently as explanations of overtrading, momentum, and the so-called “January effect” (Odean, 1999; Ciccone, 2011; Antoniou et al., 2013). Cognitive dissonance theory suggests that these factors could reduce traders’ willingness to realize losses.

4. Difficulty recognizing one’s mistakes has been proposed as a contributing factor to the reluctance to sell losers (Gross, 1982; Shefrin and Statman, 1985) and is a key reason why investors struggle to sell losers according to cognitive dissonance theory.

5. Financial literacy is a proxy for expertise/sophistication in finance, which has been identified as a key individual correlate of the DE (Shapira and Venezia, 2001; Feng and Seasholes, 2005; Brown et al., 2006).

The questionnaire also includes a reduced version of the Big Five test.\(^9\) We do not hypothesize an effect of any of these measures on the DE, which are simply included as additional controls.

---

\(^9\) The Big Five test is a widely used psychological questionnaire originally developed by Costa and MacCrae (1992) which classifies human personality into five factors: openness, conscientiousness, extraversion, agreeableness and neuroticism. We use a Spanish translation of the reduced form of the Big Five test (Benet-Martinez and John, 1998; John and Srivastava, 1999). Big Five personality traits have been used as predictors of risk aversion (Borghans et al., 2009; Dohmen et al., 2011) and of household debt and financial market participation (Brown and Taylor, 2014), but to our knowledge, not yet as predictors of the DE.
3 Results

To calculate the disposition effect, we follow Odean (1998) by computing the number of units sold at a price above the reference price (“Realized Gains”), the number of units sold at a price below the reference price (“Realized Losses”), the number of units not sold and whose price exceeds the reference price (“Paper Gains”), and the number of units not sold and whose price is below the reference price (“Paper Losses”).

The DE is defined as the difference between the Proportion of Gains Realized (PGR) minus the Proportion of Losses Realized (PLR), where $PGR = \frac{\text{Realized Gains}}{\text{Realized Gains} + \text{Paper Gains}}$, and $PLR = \frac{\text{Realized Losses}}{\text{Realized Losses} + \text{Paper Losses}}$.

Since $PGR$ and $PLR$ are between 0 and 1, the DE lies between $-1$ and 1. An alternative proxy for the DE, proposed by Weber and Camerer (1998), is the so-called “alpha” coefficient, defined as $\alpha = \frac{S^+ - S^-}{S^+ + S^-}$, where $S^+$ ($S^-$) is the number of sales of stocks whose price has gone up (down) in the last period.

The first key issue when calculating $PGR$ and $PLR$, is related to the reference price from which gains and losses are computed. As Odean (1998, p. 1782) puts it: “Any test of the DE is a joint test of the hypothesis that people sell gains more readily than losses and of the specification of the reference point from which gains and losses are determined.”

A number of different reference prices have been used in the literature, such as the weighted average price, the first period price, the last period price, the first-in-first-out rule (FIFO) or the last-in-first-out rule (LIFO). Most empirical studies follow Odean (1998) and use the weighted average price (which is also how capital gains are usually computed for tax purposes). However, other experimental studies use alternative reference prices. For example, Weber and Camerer (1998) use FIFO and LIFO, Da Costa et al. (2008) use LIFO and last price, while Rau (2014) uses only LIFO.

In a post-experimental questionnaire, we elicit subjects’ most likely reference price and find that it is the average price (see the online appendix). This approach is also used in Kohsaka et al. (2013), with similar results. Therefore, we shall use the weighted average reference price to calculate both $PGR$ and $PLR$ throughout the paper. In the online appendix we replicate our key results using the various other reference prices and the alpha measure.\(^{10}\)

The average DE in the baseline treatment is 0.069 and is significantly different from zero ($p = 0.002$, Wilcoxon signed-rank test). In this respect, our findings are comparable to those obtained in related works. For instance, Odean (1998, Table 1) also uses the average reference price and finds a DE of 0.05.

\(^{10}\) In the online appendix we also show that all these different measures of the DE are strongly correlated in the baseline.
3.1 Treatment effects

Figure 1 displays average DE, PGR and PLR in each treatment. Considering the four treatments combined, the DE is 0.078 (± 0.011 SE). The DE is 0.069 (± 0.023 SE) in B, 0.070 (± 0.024 SE) in C, 0.105 (± 0.019 SE) in T, and 0.068 (± 0.020 SE) in C+T. Wilcoxon tests reject the null hypothesis that DE is equal to zero in all treatments (p < 0.01). In other words, our subjects display a positive and significant DE in all treatments.

![Figure 1](image)

**Figure 1.** Mean DE, PGR, PLR (±SE bars) by treatment.

Although the DE is higher in treatment T than in B, this difference is not significant (Wilcoxon test, p = 0.391). Using the Skillings-Mack (SM) test (a generalization of the Friedman test that allows for missing observations), we also find no significant differences in DE across treatments (p = 0.363). By contrast, we find that PGR and PLR are significantly affected by treatment conditions (SM tests, p < 0.01). Both of them decrease significantly in the treatments involving transaction costs compared to the baseline (Wilcoxon tests, p < 0.01), with this effect being somewhat stronger for PLR. Finally, we find no differences in PGR or PLR between treatments B and C.

In the online appendix we show that an individual’s DE is highly correlated across treatments. We also report the results corresponding to all our treatments and reference prices and show that the DE is significantly different from zero for most treatments and reference prices. The DE tends to be higher in T although, as before, the difference is not significant.

3.2 Gender differences

Earlier experimental studies report mixed results regarding gender differences in the DE. Da Costa *et al.* (2008) find that only men have a positive DE, while Rau (2014) finds a negative average DE for males and a positive one for females. To facilitate the comparison, we first
examine gender differences in the baseline treatment. The average DE is lower for males (0.049) than for females (0.088), although this difference is not statistically significant (p = 0.358, Wilcoxon signed-rank test). To run this test, we proceed as follows. Recall that our subjects are divided into a total of 48 groups of four, each group facing a different market. First, we compute the average DE for the two women and for the two men in each group, which yields 96 data points in 48 pairs. Next, we compute the difference between males and females in the 48 groups. The null (alternative) hypothesis is that the difference between pairs follows a symmetric (non-symmetric) distribution around zero. As mentioned above, we cannot reject the null. A different approach is to disregard the group structure of our data and run a simple Mann-Whitney U test comparing the average DE of men and women. Again, we cannot reject the null of no gender difference with this test (p = 0.473). Regarding the other three treatments, results are very similar. Males have a lower DE than females, but these differences are never statistically significant.

There is a number of differences between our experimental set-up and that of Da Costa et al. (2008) and Rau (2014), which could explain the discrepancy of our results. First, in these studies all participants face the same market, which corresponds exactly to the price path in Weber and Camerer (1998). In our experiment, instead, each group of four subjects (2 males and 2 females) faces a different market, with randomly generated price paths for each group. Second, our sample size is larger. We have 192 subjects, roughly half of each gender. Rau’s sample is composed of 55 subjects, of which 28 are female. Da Costa et al. (2008) have 96 subjects, of which 52 are female.† Finally, our experiment is incentivized, whereas Da Costa et al. (2008) use hypothetical payoffs.

In Figure 2, we plot paired gender differences within groups of DE, PGR and PLR, disaggregated by treatment. As the figure shows, we get very similar values for the PGR of males (0.255) and females (0.235) in the baseline treatment, with no significant differences (p = 0.217). However, we do find clear gender differences in the baseline treatment for PLR, with females having significantly lower PLR than males (0.145 vs. 0.209, p = 0.005).†† Regarding the other treatments, we find that PGR is higher for women in treatments with transaction costs, but these differences are statistically significant only in treatment C+T (p = 0.048). We also find a marginally significant difference in PLR in treatment C (p = 0.093). In sum, we only find strong gender differences in the propensity to realize losers in the baseline treatment. This finding is consistent with Rau (2014) and with field evidence by Feng and Seasholes (2005).

---

† This refers to the deliberate selling treatment, which is the case comparable to ours. Da Costa et al. (2008) and Rau (2014) use the alpha measure and the LIFO rule, respectively. In the online appendix, we report values of the DE for men and women according to all reference prices and the alpha measure and still find no significant differences.

†† This result is robust to alternative reference prices, except for LIFO, where no gender differences are found (see online appendix).
3.3 Experience

Recall that treatment conditions were administered within subjects, in a balanced order. Each treatment was played first, second, third or fourth depending on the session. This makes it possible to measure how experience – acquired by playing in similar markets under different treatment conditions – can help subjects to correct their biases.

Figure 3 reports mean values of DE, PGR, and PLR comparing the two treatments played first (experienced = 0) with the two treatments played last (experienced = 1). We find that experience matters in reducing DE. Overall, we find a significant negative effect of experience on the DE (p = 0.048, Wilcoxon matched-pairs signed-rank test). However, as Figure 3 shows, the effect is mainly driven by men. In particular, men exhibit a positive DE in the first two treatments (DE = 0.085, p = 0.006, Wilcoxon signed-rank test), but not in the last two (DE = 0.018, p = 0.601). If we test for differences according to experience, we reject the null of no differences (p = 0.040). By contrast, women’s DE is greater than zero in both halves of the experiment (first two treatments: DE = 0.110, p < 0.001; last two treatments: DE = 0.097, p < 0.001). For women, we cannot reject the null of no differences by experience (p = 0.588), even though, on average, it also decreases over time. This differential effect of experience in men results in a significant gender difference in the DE during the latter half of the experiment (p = 0.043).

Interestingly, experience only appears to attenuate the propensity to sell winners. Pooling men and women together, we find a very clear effect of experience on PGR (p = 0.002) and no effect on PLR (p = 0.548). While both men and women reduce sales of winners with experience, this reduction is significant only for men (p = 0.007). On the other hand, experience has no significant effect on PLR for either group.
The result that the DE is attenuated by experience is in line with previous field and laboratory evidence (Feng and Seasholes, 2005; Dhar and Zhu, 2006; Weber and Welfens, 2007). However, the field studies by Feng and Seasholes (2005) and Dhar and Zhu (2006) find that this effect mainly operates via an increase in PLR, while Weber and Welfens (2007) find a more symmetric effect on both the PGR and PLR in the field and in the laboratory. Our results, therefore, add to the growing evidence that experience ameliorates the DE. Whether this effect is stronger for males, as our evidence suggests, will need further investigation in the field.

3.4. Individual characteristics

In this section we evaluate the potential effect of individual attributes on the DE, PGR and PLR. The original questionnaire measures (described in Section 2.3) typically take integer values from 1 (= disagree completely) to 7 (= agree completely). We divide our subjects into two groups according to whether they are below or above the median score in each of these variables and then calculate the mean values of DE, PGR, and PLR for these two groups, separately by gender.\textsuperscript{13}

Among all the variables we consider, the only two which exhibit a significant impact are difficulty recognizing mistakes and optimism. Figures 4 and 5 display the results for each of these two variables, respectively. The corresponding figures for the other variables can be found in the online appendix.

As Figure 4 shows, men with below-median scores in difficulty recognizing mistakes do not fall into the DE (DE = −0.011, p = 0.474, Wilcoxon signed-rank test), while those above do (DE = 0.095, p = 0.002). A difference test between men below and above the median rejects the

\textsuperscript{13} The online appendix contains descriptive statistics of individual characteristics separated by gender.
null of no differences ($p = 0.014$, Mann-Whitney $U$ test). By contrast, women in both groups exhibit a positive DE (below median: $DE = 0.077$, $p = 0.006$; above: $DE = 0.127$, $p < 0.001$), and, although the effect goes in the same direction, we cannot reject the null of no differences ($p = 0.616$). With respect to PGR and PLR, we see that things go in the expected direction, since both men and women below the median reduce sales of winners and increase sales of losers. However, these differences are only significant for men (PGR for men, $p = 0.034$, PGR for women, $p = 0.700$; PLR for men, $p = 0.040$, PLR for women, $p = 0.387$; Mann-Whitney $U$ test).

Figure 4. Mean DE, PGR and PLR depending on difficulty recognizing mistakes

Turning to optimism, Figure 5 shows a fairly similar pattern. Men with below-median optimism do not exhibit a DE ($DE = -0.007$, $p = 0.638$) while above-median optimists do ($DE = 0.075$, $p = 0.012$). However, optimism appears to only affect men’s PLR ($p = 0.019$), not their PGR. Women do not appear to be affected by this trait, exhibiting similar levels of the DE regardless of optimism ($DE \approx 0.1$, $p < 0.01$ in both groups).

To sum up, we only find evidence for a significant impact of two individual traits on the DE: difficulty recognizing mistakes and optimism. In line with our intuition, both of these traits correlate with a reduction in the proportion of losses realized, although difficulty recognizing mistakes also appears to increase the proportion of gains realized, further exacerbating the DE. Interestingly, neither of these two traits appears to have a significant impact on women’s DE.

Our evidence is suggestive of cognitive dissonance theory. Investors are more prone to selling winners than selling losers because selling losers involves the painful act of accepting a past mistake (Shefrin and Statman, 1985) or because doing so conflicts with their optimistic attitudes and beliefs (Antoniou et al., 2013). Interestingly, our results suggest the possibility that
men might be particularly sensitive to the effects of cognitive dissonance in the financial domain, perhaps because this domain is more relevant for their self-image (Prince, 1993).

![Mean DE, PGR and PLR depending on optimism](image)

**Figure 5.** Mean DE, PGR and PLR depending on optimism

On the other hand, we do not find that the DE is significantly higher for subjects who would be expected to be most vulnerable to the DE according to realization utility theory, such as those with low self-control, or high enjoyment of winning.

### 3.5. Regression analysis

To conclude our results, we run three random–effects linear regressions, one for each variable of interest (DE, PGR and PLR), including the same set of explanatory variables in all of them. First, we include treatment dummies, the female dummy, and the experience dummy (= 1 in the last two treatments played, = 0 otherwise), together with all their possible interactions. These are all the variables explicitly randomized in our experimental design. Second, we include the set of individual characteristics of interest detailed in Section 2.3, which are not interacted. To ease the interpretation of the quantitative effects associated with each of these variables, we collapse them into dummy variables, depending on whether the original measure (an integer between 1 and 7) is below or above the median. As for financial literacy, our dummy is positive if the three questions of the corresponding test have been answered correctly. Finally, we include as additional controls the standardized scores of each of the Big Five personality traits, four field of studies dummies (Art, Health, Science, Engineering, with Social Sciences as the reference category), a “high grades” dummy (indicating whether the subject’s GPA is 7 or higher out of 10)
and an “erratic” dummy, indicating whether the subject switched more than once in the MPL lottery task.\textsuperscript{14}

To facilitate the exposition, we divide the results of the regressions into two tables. In Table 2 we report the coefficients corresponding to the randomized covariates, while in Table 3 we report the coefficients corresponding to the individual characteristics of interest. The online appendix contains summary statistics of all the variables we use, and the full regression results.

\textbf{Table 2. Regression results I (marginal effects of the randomized covariates)}

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>DE</th>
<th>PGR</th>
<th>PLR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment effects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0.008</td>
<td>0.012</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.017)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>T</td>
<td>0.031</td>
<td>-0.027</td>
<td>-0.064***</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.020)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>C+T</td>
<td>0.004</td>
<td>-0.050***</td>
<td>-0.056***</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.016)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>Female</td>
<td>0.025</td>
<td>-0.000</td>
<td>-0.025</td>
</tr>
<tr>
<td></td>
<td>(0.034)</td>
<td>(0.028)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>Experienced</td>
<td>-0.036*</td>
<td>-0.027**</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.013)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>Observations</td>
<td>706</td>
<td>720</td>
<td>710</td>
</tr>
<tr>
<td>Number of id</td>
<td>192</td>
<td>192</td>
<td>192</td>
</tr>
</tbody>
</table>

Note: In the first three rows (treatment effects), we calculate the marginal effects of the treatment dummies taking treatment B as the default. In the fourth row we calculate the effect of the female dummy pooling together the four treatments. Robust standard errors clustered by group and session in parentheses. *** p<0.01, ** p<0.05, * p<0.1

As Table 2 shows, we do not detect any treatment effect on the DE, although we find that transaction costs reduce both PGR and PLR, with competition enhancing this effect in the case of PGR in treatment C+T. Although not shown in Table 2, the reduction of PGR in treatments with transaction costs has a strong gender component. This effect is driven mostly by males (–0.056 and –0.088, p = 0.052 and p = 0.001 in treatments T and T+C, respectively). While we already know from Figure 2 that PLR is lower for females in treatments B and C, the overall marginal effect of gender is not significantly different from zero. However, taking into account the

\textsuperscript{14}To avoid reducing the sample size significantly, we do not include the measure of risk aversion gathered from the MPL lottery task, since we would have to exclude 42 subjects who switched from the safe to the risky option more than once. This variable was nevertheless insignificant when included in the regression. Instead, we included the self-assessed general “willingness to take risks” measure of risk and used this “erratic” measure as a control in the regression. For simplicity and consistency with the other psychological measures, we also only included the self-assessed general measure of confidence and not the group ranking forecasts, although results do not change if the latter measure is used instead (these forecasts are reported and used in Cueva et al. 2019).
female×treatment interaction confirms a significant negative effect of female on PLR in treatments B and C (see the online appendix). Finally, in line with Figure 3, we find a significant marginal effect of experience on PGR and (to a lesser extent) on the DE.

Table 3: Regression results (II). Individual characteristics.

<table>
<thead>
<tr>
<th></th>
<th>DE</th>
<th>PGR</th>
<th>PLR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diff. recog. mistakes</td>
<td>0.085***</td>
<td>0.032</td>
<td>-0.057***</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.029)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Optimism</td>
<td>0.071**</td>
<td>0.010</td>
<td>-0.061**</td>
</tr>
<tr>
<td></td>
<td>(0.034)</td>
<td>(0.035)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>Risk</td>
<td>-0.026</td>
<td>-0.024</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.031)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Confidence</td>
<td>-0.053</td>
<td>-0.006</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.034)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>Competitiveness</td>
<td>0.045</td>
<td>0.014</td>
<td>-0.028</td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
<td>(0.029)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>Enjoys Winning</td>
<td>-0.050*</td>
<td>-0.042</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.026)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>Decisiveness</td>
<td>-0.016</td>
<td>-0.027</td>
<td>-0.014</td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.036)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Self-control</td>
<td>0.057</td>
<td>0.058*</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>(0.040)</td>
<td>(0.031)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>High Fin. Literacy</td>
<td>-0.030</td>
<td>-0.016</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
<td>(0.035)</td>
<td>(0.024)</td>
</tr>
</tbody>
</table>

Observations 706 720 710
Number of id 192 192 192

Robust standard errors clustered by group and session in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 3 shows the estimated effects of individual characteristics. Confirming the evidence in Figure 4, the coefficient on difficulty recognizing mistakes is a highly significant predictor of the DE and PLR, but not of PGR. The size of the effect on the DE (0.085) is very large, compared to the average value of the DE (0.078). The regression results also confirm the evidence in Figure 5. Much like difficulty recognizing errors, optimism has a large and significant effect on the DE and a negative effect on PLR. These results are robust to the alternative reference prices used to compute the DE/PGR/PLR (see the online appendix). They are also robust to (i) bootstrapped standard errors and (ii) removing 10% of the subjects with the most extreme values of the DE.
4. Conclusion

We study several issues related to the DE. First, we check the robustness of the DE to the addition of transaction costs and competitive payment schemes. We find that these variations do not have a significant effect on the size of the DE. Second, we check for the existence of gender differences. We find that women have a larger DE than men, but the difference is only significant during the second half of the experiment, when subjects have gained some experience with the task. We also find that women are more reluctant to sell losing assets throughout the experiment, which confirms previous field and laboratory evidence (Feng and Seasholes, 2005; Rau, 2014). Third, we study individual determinants of the DE, including various psychological measures such as enjoyment of winning, self-control, decisiveness, optimism, and difficulty recognizing mistakes. Interestingly, only the latter two variables turn out to be significant predictors of the DE. In particular, we find that both difficulty recognizing mistakes and optimism increase the DE by reducing the willingness to sell losing assets.

Our findings provide direct support for the claim, already discussed by Shefrin and Statman (1985) that “investors are also reluctant to accept and realize losses because the very act of doing so proves that their first judgment was wrong” (Gross, 1982, p. 150). As Summers and Duxbury (2012) show, personal responsibility in the investment decision induces feelings of rejoice or regret which may differentially affect subjects who have difficulty recognizing mistakes.

More generally, our evidence offers novel support for cognitive dissonance theory as an explanation of the DE and fits well with several recent findings in the literature. Specifically, it is consistent with the fact that investors tend to find it much easier to sell losers when doing so does not reflect badly on their own initial judgment, such as when the assets are initially purchased by the experimenter (Summers and Duxbury, 2012), when they are received as gifts (Lehenkari, 2012), when they are inherited from a previous fund manager (Jin and Scherbina, 2010), or when an external manager can be blamed (Calvet et al. 2009; Ivković and Weisbenner, 2009; Chang et al. 2016). In all these contexts, the decision maker can avoid feeling responsible for the original investment decision, so that selling a loser no longer involves the recognition of a past own mistake.

With regards to optimism, our evidence is again supportive of cognitive dissonance and in line with previous field evidence. Antoniou et al. (2013) find that small investors are slow to sell losers during optimistic periods and also interpret their evidence in line with cognitive dissonance theory. In their view, investors are slower to incorporate bad news because they are particularly dissonant when overall sentiment is optimistic. While Antoniou et al. (2013) provide evidence of this effect by exploiting temporal variation in optimistic sentiments, we obtain similar findings exploiting cross-sectional variation in self-assessed optimism.
References


Cueva, C., Iturbe-Ormaetxe, I., Ponti, G., and Tomás, J., 2019, Boys will still be boys: Gender differences in trading activity are not due to differences in (over)confidence, Journal of Economic Behavior & Organization, 160, 100-120.


