Cognitive reflection and the disposition effect: A lab experiment

Newton da Costa Jr.1,2*, André Grezelle2, Marco Goulart1, H. K. Pradhan3

1Federal University of Santa Catarina, Brazil
2Pontifical Catholic University of Parana, Brazil
3Xavier School of Management, India
*Corresponding Author: ncacjr@gmail.com

Received: 20 June 2023  Accepted: 9 July 2023  DOI: https://doi.org/10.55942/jebl.v3i4.204

ABSTRACT

This research investigates the connection between cognitive reflection and the disposition effect, a well-known bias in the behavioral finance literature. Utilizing the Cognitive Reflection Test (CRT) developed by Frederick (2005), we measured cognitive abilities in a laboratory-based experiment comprising 55 students. The main goal was to investigate the extent to which these cognitive resources might modulate the disposition effect. The study was conceptualized within the framework of the dichotomy between a deliberative long-term self and an impulsive short-term self, as detailed in Kahneman’s 2011 two-systems theory. The findings indicate a significant negative correlation between cognitive abilities and the disposition effect, providing extra empirical support for Kahneman’s theory. This study presents new empirical evidence of the association between cognitive reflection and behavioral biases associated with decision-making under conditions of risky, thus providing a basis for possible interventions to mitigate these biases. The implications of this study are not limited to academia, but may provide information for the development of future financial education programs to enhance the decision-making process of individuals, whether in public or private companies.

Keywords: disposition effect, cognitive reflection test (CRT), behavioral biases, lab experiment

1. INTRODUCTION

The disposition effect (Shefrin and Statman, 1985) refers to the tendency of investors to sell winning stocks more often than losing ones. This behavioral bias has been observed across various trading environments, including options exercise (Heath, Huddart, and Lang, 1999), futures market operators (Locke and Mann, 2005; Coval and Shumway, 2005), equity investment fund shareholders (Chiu et al., 2004), and real estate sales (Genovese and Myer, 2001).

One of the main reasons why this effect is studied so much in finance is that individuals who suffer from the disposition effect, in general, underperform in the management of their investments compared with individuals who do not present this bias (Odean, 1998; Dhar and Zhu, 2006; Prates et al., 2019).

In general, it has been found that the more sophisticated and savvy an investor is, the less prone to the disposition effect and other cognitive anomalies he will be (Da Costa Jr. et al., 2013; Oechssler et al., 2009; Noori, 2016; Janssen et al., 2020). Frederick (2005) introduced the Cognitive Reflection Test (CRT), which comprises math problems with intuitive yet deceptive answers. Highly cognitively-abled individuals tend to exhibit less risk aversion in winning bets and display increased patience.

The results obtained by Frederick (2005) suggest that the human mind processes information in a dual way. Currently, according to Kahneman (2011), several authors agree that these processes can be unconscious, fast, automatic, and high capacity (System 1), or conscious, slow and deliberative (System 2).

In this study, we conducted a lab experiment to investigate the relationship between the disposition effect and individuals' cognitive abilities. The hypothesis is that high (low) CRT scores are associated with low (high) levels of the disposition effect. We simulated an exogenous stock market to capture the effect and assessed cognitive abilities using a questionnaire and CRT scores.

To be useful, a test of cognitive skills must be short and simple. The cognitive reflection test first proposed by
Frederick (2005) is a 3-item test that can be performed in less than five minutes and is a good predictor of cognitive skills, particularly with regard to mathematical skills.

Recent studies have already verified the relationship between behavioral biases and cognitive abilities (Oechssler et al., 2009; Benjamin et al., 2013; Deck and Jahedi, 2015; Noori, 2016; Blaywais and Rosenboim, 2019; Maloney and Retanal, 2020).

The only study, so far, that has analyzed the relationship between the disposition effect, emotions, and Systems 1 and 2 was that of Richards et al. (2018), which focused on real investors in the United Kingdom, but used a different approach from ours. As the present research was conducted in an experimental setting it has the potential to reduce the presence of noise that could directly interfere with investor behavior and thus can add a new dimension to the study of this effect.

In this way, we contribute to the literature by analyzing the relationship between the disposition effect and cognitive abilities via CRT scores in a lab environment. Real market tests of the disposition effect are inconclusive due to uncontrollable investor decisions and the interference of confounding variables. Laboratory experiments can be an important tool in finance research because they allow more control over the variables, allowing researchers to isolate specific factors and observe their impact on financial decision-making.

The remainder of this article is organized as follows. Section 2 describes the design of the experiment, Section 3 presents the results, and Section 4 concludes the work.

2. EXPERIMENTAL DESIGN

The procedure adopted here is characterized by a study in a laboratory environment with a hypothetical simulation design of investments through a specific software, with monetary incentives for participants throughout the experiment. The total number of participants was n=55.

The experiment design was similar to that of Weber and Camerer (1998) and Fischbacher et al. (2017). The simulation begins with all assets having the same price and participants receiving an initial endowment of 10,000 monetary units (=10 BRL) and can trade six different stocks labeled 1 to 6. As in Weber and Camerer (1998) and Fischbacher et al. (2017), the game consists of 34 rounds, beginning in Round –3 and ending in Round 30. In Rounds –3 to –1, the participants cannot trade but observe the price changes for the six stocks. In Rounds 0 to 30 participants can trade the six stocks.

In each period, the price can go up by 6% or down by 5%. The probability of the increase varies between stocks but is constant throughout the periods for each one. Participants know the size of the possible price increase or decrease, and they also know that the probabilities of price change vary from one asset to another, but are constant in all rounds and independent of their trading activities. Furthermore, in accordance with Fischbacher et al. (2017), we implemented the following six price increase probabilities for the six stocks: 40%, 45%, 50%, 55%, and 60%.

However, these probabilities were unknown to the participants. During the simulation, the participant can follow his results at any time. They are updated at the end of each of the 30 periods. The average time it took the 55 participants to complete the experiment was 56 minutes.

The cognitive reflection test (CRT), introduced by Frederick (2005), has proved to be one of the most useful measures in the study of individual differences in thinking, judgments, and decisions (Baron et al., 2015).

CRT differentiates between more impulsive and more reflective decision makers. To do so, each of the three CRT questions has an apparently intuitive (but incorrect) answer that quickly comes to mind. CRT questions are not difficult in the sense that the correct solutions are easily understood when explained to participants.

Frederick (2005) applied the CRT to measure the cognitive reflection capability (reflexivity x impulsivity) of undergraduate students at renowned universities. He found that Harvard students averaged only 1.43, while Princeton University students averaged 1.63 (on a score ranging from 0 to 3).

These intriguing results can be explained based on what researchers call the dual cognitive process (Epstein, 1994; Evans, 2003). The literature suggests that people use two types of cognitive processes: System 1 and System 2 (Stanovich and West 2000; Kahneman and Frederick, 2002; Kahneman, 2011). Thus, System 1 operates automatically and quickly, with little or no effort and no perception of voluntary control. In System 2, cognitive operations are performed from a logical judgment with predefined rules that require a high cognitive effort. While System 1 is related to an impulsive way of thinking, System 2 is a reflective and slower style of decision making. Thus, even when people know how to answer specific questions and how to make good decisions, they can misjudge if they resort to impulsive thinking.

The test consists of answering three questions (FREDERICK, 2005):

1) A bat and a ball together cost 110 cents. The bat costs 100 cents more than the ball. How much does the ball cost? (Impulsive response: 10 cents; correct answer: 5 cents).
2) In a lake, there's a patch of lily pads. Each day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? (Impulsive response: 24 days; correct response: 47 days).

3) If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets (in minutes)? (Impulsive response: 100 minutes; correct answer: 5 minutes).

The test is used to separate people into those who tend to use System 1 and those who tend to use System 2. Studies indicate that individuals with a higher score for CRT questions tend to be less risk-averse in gains (FREDERICK, 2005). Noori (2016) found in his sample that individuals with lower scores in CRT are significantly more likely to exhibit conjunction fallacy, illusion of control, overconfidence, base rate fallacy and conservatism biases.

As already pointed out, in this study the interest lies in identifying whether the score in the CRT (S1 and S2) is associated with the level of the disposition effect.

The average number of correct CRT responses in our sample was 1.75 questions. Of our subjects, 25.5% answered all three questions correctly, 41.8% answered two questions correctly, 16.4% answered one question correctly, and the remaining 16.4% answered none of the questions correctly. As in Frederick (2005), male subjects received a higher average score (2.03 questions) than female subjects (1.25 questions), which is a significant difference (p < 0.004, Mann-Whitney U test).

2.1 The Disposition Coefficient (DC)

As defined previously, the disposition effect is the tendency to realize gains before losses. The term was coined by Shefrin and Statman (1985).

Detecting the disposition effect on market data is a difficult task because once detected there is always an additional plausible hypothesis to explain it, such as the tendency of investors to believe in the mean reversion phenomenon, portfolio rebalancing, taxes, etc. Thus, the use of data collected in controlled laboratory experiments can help to explain this and many other phenomena.

The disposition effect of each participant in the experiment can be estimated, based on Odean (1998), by the disposition coefficient (DC):

\[
\begin{align*}
PRG &= \frac{RG}{RG+PG} \\
PRL &= \frac{RL}{RL+PL}
\end{align*}
\]

where PRG (PRL) is the proportion of realized gains (losses); RG (RL) is the number of trades of an investor/participant with a realized gain (loss); PG (PL) is the amount of the paper gain (loss) of the investor/participant. Each stock that is in the investor's/participant's portfolio at the beginning of each simulation period, but is not sold, is considered a paper gain or a paper loss (not realized).

The disposition coefficient (DC) of the investor/participant is then

\[
DC = PRG - PRL
\]

2.2 Recruiting and Payment

Fifty-nine students were randomly selected from several classes in the Business and Accounting undergraduate courses at a private Brazilian university. Four participants were excluded due to prior CRT knowledge, resulting in a sample of 20 females and 35 males. The majority (81%) were below 25 years old. Data was collected during two experimental sessions in the computer lab on March 15 and 18, 2019.

As an incentive for the participants, an award system was included. Thus, each participant was given a number that was used to identify him/her. Twenty participants from the 55 who carried out the experiment were randomly selected. There were three levels of awards: the participant with a final balance up to 10,000.00 monetary units received R$5.00; final balance from 10,001.00 to 20,000.00 received R$10.00; and those who achieved a final balance above 20,001.00 received R$15.00. This information was available at the beginning of the experiment. The exchange rate on March 29, 2019, was 1 USD=3.89 BRL.

3. FINDINGS AND DISCUSSIONS

In Table 1 we observe that 80% (44 individuals) of the sample presented DC greater than zero, a result like other experiments (Weber and Camerer, 1998; Da Costa Jr. et al., 2013). It is also observed that the average assets held in portfolio per period and per participant was 3.54 (maximum would be 6).

<table>
<thead>
<tr>
<th>Sample (n)</th>
<th>Students</th>
<th>Random trades</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>55</td>
<td>55</td>
</tr>
<tr>
<td>Average Portfolio Returns (%)</td>
<td>16%</td>
<td>11.4%</td>
</tr>
<tr>
<td>Average transactions per period</td>
<td>33</td>
<td>37</td>
</tr>
<tr>
<td>Average assets in portfolio per period</td>
<td>3.54</td>
<td>4.33</td>
</tr>
<tr>
<td>Subjects with DC &gt; 0 (%)</td>
<td>80%</td>
<td>56%</td>
</tr>
</tbody>
</table>

The BNI software's calibration for calculating disposition coefficients was tested using random asset transactions in an Excel spreadsheet. The null hypothesis, that the coefficient is
Table 2: Descriptive statistics of the participants and random trades

<table>
<thead>
<tr>
<th>Disposition Coefficients</th>
<th>DC Random</th>
<th>PRG</th>
<th>PRL</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>55</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.058</td>
<td>0.169</td>
<td>0.110</td>
</tr>
<tr>
<td>Median</td>
<td>0.057</td>
<td>0.173</td>
<td>0.100</td>
</tr>
<tr>
<td>Max.</td>
<td>0.397</td>
<td>0.398</td>
<td>0.550</td>
</tr>
<tr>
<td>Min.</td>
<td>-0.288</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Standard Dev.</td>
<td>0.118</td>
<td>0.092</td>
<td>0.113</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov (D)</td>
<td>0.131</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t stat (mean=0)</td>
<td>3.63***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wilcoxon test</td>
<td>1214.5***</td>
<td>-0.084</td>
<td></td>
</tr>
</tbody>
</table>

Note:
* p < 0.10; ** p < 0.05; *** p < 0.01

Table 3 displays CRT score statistics for 55 students, with an average score of 1.75 correct questions. Students were split into "high" (2-3 correct answers; 36 students) and "low" (0-1 correct answers; 19 students) groups, reflecting their decision-making styles. Women scored significantly lower than men (p < 0.004), indicating greater reliance on intuition. Another fact to note is that participants with positive DCs achieved a lower score than those with negative CDs, but the difference was not significant (p < 0.139).

Table 3: Descriptive statistics for the CRT score

<table>
<thead>
<tr>
<th></th>
<th>CRT</th>
<th>CRT Low</th>
<th>CRT High</th>
<th>CRT Male</th>
<th>CRT Female</th>
<th>CRT (CD&lt;0)</th>
<th>CRT (CD&gt;0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>55</td>
<td>19</td>
<td>36</td>
<td>20</td>
<td>35</td>
<td>11</td>
<td>44</td>
</tr>
<tr>
<td>Mean</td>
<td>1.75</td>
<td>0.52</td>
<td>2.39</td>
<td>2.03</td>
<td>1.25</td>
<td>2.18</td>
<td>1.64</td>
</tr>
</tbody>
</table>

Note:
(1) "CRT low" is the group that answered 0 or 1 question correctly, and "CRT high" is the group that answered 2 or 3 questions correctly.

3.1 Regression Analysis

Next, in Table 4, we present a series of regressions to provide more details regarding the association between DC and CRT, along with five control variables, namely: sex, transactions, overconfidence, average number of assets in portfolio and profitability (portfolio’s final return).

A robust regression (via Stata) was used to check how much the variables interfere with or relate to one another. Robust regression is able to correct patterns of nonnormality in the residual regression. Its objective is to reduce the influence of discrepant points that affect the quality of the estimation of the regression model parameters.

It can be seen in Table 4 that the control variables, in general, showed no great explanatory power for DC variations. Only the CRT variable was (negatively) significant across all regressions, and the overconfidence variable at a lower significant level.

Table 4: Robust regression between DC and six independent variables

<table>
<thead>
<tr>
<th>DC</th>
<th>Intercept</th>
<th>CRT</th>
<th>Sex</th>
<th>Transactions</th>
<th>Overconf.</th>
<th>Assets</th>
<th>Return</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>.1080***</td>
<td>-.0283**</td>
<td>(.0262)</td>
<td>(.136)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.0595**</td>
<td>(.1157)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td>.0914***</td>
<td>-.0381**</td>
<td>(.0279)</td>
<td>(.162)</td>
<td>(.0378)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.1004*</td>
<td>(.1142)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 3</td>
<td>.0853**</td>
<td>-.0392**</td>
<td>(.0384)</td>
<td>(.171)</td>
<td>(.0389)</td>
<td>(.0009)</td>
<td></td>
</tr>
<tr>
<td>.1020*</td>
<td>(.1152)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 4</td>
<td>.0654</td>
<td>-.0337**</td>
<td>(.0412)</td>
<td>(.166)</td>
<td>(.0362)</td>
<td>-.19e-06</td>
<td>.0862*</td>
</tr>
<tr>
<td>.1462</td>
<td>(.1135)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 5</td>
<td>.0613</td>
<td>-.0339**</td>
<td>(.0416)</td>
<td>(.165)</td>
<td>(.0364)</td>
<td>.00001</td>
<td>.0868*</td>
</tr>
<tr>
<td>.1463</td>
<td>(.1146)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 6</td>
<td>.0641</td>
<td>-.0331*</td>
<td>(.0734)</td>
<td>(.0169)</td>
<td>(.0365)</td>
<td>(.0528)</td>
<td>(.0144)</td>
</tr>
<tr>
<td>.1157</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:
(1) This table presents the regression coefficients and, in parentheses, the robust standard error; in bold, the R-sq, and below the R-sq the RMSE, both estimated via Stata.
(2) DCi = a + bX1i + b2X2i + b3X3i + b4X4i + b5X5i + b6X6i + ui
Where DCi is the disposition coefficient; X1 is the CRT test used to measure the participants' attention level, whether they were in a more intuitive or rational mode; X2 is the sex of the participants; X3 is the average number of transactions carried out by each participant during the simulation; X4 is the level of confidence in participants' decision-making identified by means of a questionnaire; X5 is the average number of assets in portfolio in each period per participant; X6 is the profitability achieved by each participant (loss or profit).
(3) Sample size n=55 participants.
(4) To build a confidence index, which was used as a proxy to measure the participants' overconfidence, a questionnaire like that used by Busenitz and Barney (1997) was applied.
(5) * p < 0.10; ** p < 0.05; *** p < 0.01
4. CONCLUSIONS

Our study found participants' disposition effect significantly different from zero at a 1% level, consistent with past research on the subject in both lab experiments and archival data (Shefrin & Statman, 1985; Odean, 1998; Weber & Camerer, 1998; Genovese & Myer, 2001; Locke & Mann, 2005; Dhar & Zhu, 2006; da Costa Jr. et al., 2013).

However, new in terms of a lab experiment, we showed that the disposition effect is lower among students with higher measured cognitive ability. To the extent that the cognitive ability test (CRT) proposed by Frederick (2005) can be taken as a proxy for available cognitive resources, our results are consistent with the two systems theories (Kahneman, 2011), which postulate that when a subject decides (shows his preferences) he usually faces a conflict between a deliberative, long-term self and an impulsive, short-term self.

Despite the popularity of the CRT, one challenge hinders its interpretation: the numerical characteristic of the CRT may cause reflection skills to be confounded with mathematical skills (Sirotka et al. 2021). However, our findings align with studies showing a negative correlation between cognitive abilities and biases (Frederick, 2005; Benjamin et al., 2013; Oechssler et al., 2009; Noori, 2016; Richards et al., 2018; Blaywais & Rosenboim, 2019; Maloney & Retanal, 2020).

Last, the experiment was conducted in 2019 using the traditional CRT, however, currently the questions of this test are already well known among students, and it would be recommended to apply a more updated version of it, such as CRT-7, among others (Toplak et al., 2014; Primi et al., 2016, Sirotka et al., 2021). However, self-reported prior exposure does not affect the test’s predictive validity (Sirotka, 2021).

Funding
This study was financed by Grant 306385/2018-0, National Council for Scientific and Technological Development (CNPq), Brazil (Newton da Costa Jr.)

Data Sharing Statement
The data that support the findings of this study are openly available in figshare at https://doi.org/10.6084/m9.figshare.19727380.v1

REFERENCES


Maloney, E. A.; Retanal, F. (2020). “Higher math anxious people have a lower need for cognition and are less reflective in their thinking.” Acta Psychologica, 202, 102939.


