

University of Heidelberg

Department of Economics



Discussion Paper Series | No. 465

Cognitive Abilities and Behavioral Biases

Joerg Oechssler,
Andreas Roider, and
Patrick W. Schmitz

May 2008

Cognitive Abilities and Behavioral Biases*

Jörg Oechssler
Department of Economics
University of Heidelberg

Andreas Roider[†]
Department of Economics
University of Heidelberg

Patrick W. Schmitz
Department of Economics
University of Cologne

May 19, 2008

Abstract

We use a simple, three-item test for cognitive abilities to investigate whether established behavioral biases that play a prominent role in behavioral economics and finance are related to cognitive abilities. We find that higher test scores on the Cognitive Reflection Test of Frederick (2005) indeed are correlated with lower incidences of the conjunction fallacy, conservatism in updating probabilities, and overconfidence. Test scores are also significantly related to subjects' time and risk preferences. We find no influence on anchoring. However, even if biases are lower for people with higher cognitive abilities, they still remain substantial.

Keywords: behavioral finance, biases, cognitive abilities, cognitive reflection test.

JEL-Classification: C91, D80, D90, J24.

*We are grateful to seminar participants in Bonn, Cologne, and Munich for helpful comments and discussions. We thank the experimental laboratories at the Universities of Bonn, Mannheim, and Cologne for allowing us to use their mailing lists. We are also very grateful to Alexander Voss who did an excellent job programming the experiment. Financial support by the DFG through SFB 504 and SFB TR15 is gratefully acknowledged.

[†]Corresponding author: Department of Economics, University of Heidelberg, Grabengasse 14, 69117 Heidelberg, Germany, email: roider@uni-hd.de, phone: +49-(0)6221-54-2952.

1 Introduction

Why should economists be interested in behavioral biases and cognitive abilities? The traditional view in economics and finance is that only outcomes matter. There is a strong presumption that behavioral biases do not play a role in the aggregate because either they are averaged out or they are corrected for by rational arbitrage. However, Barberis and Thaler (2003) argue forcefully that arbitrage cannot eliminate all effects of behavioral biases in financial markets.¹ But even if it could, on the individual level it would certainly still be valuable information if one could somehow discriminate between individuals that are less or more afflicted by behavioral phenomena. For example, this information would be useful when applied to potential employees or customers. In particular, if a concise test were available that could be administered in a few minutes but would reveal a lot about the probability that those individuals exhibit a number of well-known biases in decision-making, such a test should be quite instructive.

In this paper we provide an experimental test for the hypothesis that the incidence of behavioral biases is related to cognitive abilities. For it to be useful, a test for cognitive abilities must be short and simple. One such test is suggested by Frederick (2005) who shows that his Cognitive Reflection Test (CRT), which is a brief 3-item test that can be conducted in less than five minutes, is a good predictor of cognitive abilities, in particular with respect to mathematical abilities.

A number of recent studies have already demonstrated two stylized facts about the relationship between cognitive ability and economic behavior (see Frederick, 2005; Benjamin et al., 2006; Slonim et al., 2007, and Dohmen et al., 2007). First, individuals with high cognitive abilities seem to be less risk averse when gambling in the positive domain.² Second, they seem to be more patient.³

In this study, we replicate those earlier studies (and their main findings) with respect to risk and time preferences. But we also extend the hypothesis to a number of biases that play a crucial role, for example, in the behavioral finance literature

¹Such limits to arbitrage might, for example, arise from fundamental risk (e.g., due to the lack of short-selling opportunities), noise trader risk (see e.g., De Long et al., 1990; Shleifer and Vishny, 1997), or implementation costs (see e.g., Summers, 1986; Abreu and Brunnermeier, 2002). See also Shleifer (2000).

²See, however, Branas-Garza et al. (2008) who do not find evidence for a relationship between a GRE-like math test and risk attitudes.

³Moreover, for a representative sample of the German population, Dohmen et al. (2008) find that participants' education levels are systematically related to being prone to the gambler's fallacy (respectively, the hot hand fallacy).

(see Barberis and Thaler, 2003, for a comprehensive survey). In particular, we study the conjunction fallacy, anchoring, overconfidence, and conservatism with respect to probability updating. We find that individuals with low CRT scores are significantly more likely to be subject to the conjunction fallacy, to conservatism with respect to probability updating, and to overconfidence. We find no significant difference with respect to anchoring.

2 Experimental Design

In total, 1250 subjects participated in our online, web-based experiment. After logging in on our website and providing some personal background information, all participants played a one-shot mini-ultimatum game (see our companion paper, Oechssler et al., 2008 for results). A subset of 564 subjects also answered a questionnaire with several decision-problems related to the well-known behavioral biases discussed above, and our results relate to this subset of subjects. Mixed in among those questions were the three questions that make up the Cognitive Reflection Test. A translation of the instructions can be found in the Appendix.

2.1 Cognitive Reflection Test

The Cognitive Reflection Test (CRT) introduced by Frederick (2005) is a quick and simple three-item test, and Frederick (2005) documents that the CRT compares very favorably (in terms of the relationship between test scores and observed behavior) to substantially more complex personality tests. The CRT differentiates between more impulsive and more reflective decision-makers. To achieve this, each of the three questions of the CRT has a seemingly intuitive (but incorrect) answer that springs quickly to mind, and the overwhelming majority of subjects indeed provides either the impulsive or the correct response (more than 90% of subjects in all three questions). The questions of the CRT are not difficult in the sense that the correct solution is easily understood when explained to subjects. Moreover, if a solution springs to mind it is easy for subjects to verify whether their response is indeed correct. However, arriving at the correct answer may require overcoming the initial, impulsive response. The CRT consists of the following three questions.

- 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 answer: 10 cents; correct answer: 5 cents).

- If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? (impulsive answer: 100 minutes; correct answer: 5 minutes).
- In a lake, there is a patch of lily pads. Every 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 answer: 24 days; correct answer: 47 days).

The average CRT score (i.e., the average number of correct responses) in our sample was 2.05, which places our subjects well between students of MIT and Princeton in Frederick’s (2005) sample. Of our subjects, 41.5% answered all three questions correctly, 30.7% answered two questions correctly, 17.7% answered one question correctly, and the remaining 10.1% answered none of the questions correctly. As in Frederick (2005), male subjects received a higher average test score (2.2) than female subjects (1.7), which is a significant difference ($p < 0.001$, MWU-test).⁴

2.2 Recruiting and Payment

Subjects were recruited via emails. Email addresses were obtained from the economic experimental laboratories in Bonn, Cologne, and Mannheim. All those contacted had indicated their interest in participating in economic experiments. Of the participants, 90% were university students, 25% studied economics or business, and 46% were female. Average age of participants was 24 years.

The website was linked to the Laboratory for Experimental Research in Economics at the University of Bonn to demonstrate that the experiment had a proper scientific background and that the promised financial rewards were credible. To prevent subjects from playing multiple times, each name-postal code combination and each email address was only allowed to play once.⁵

In addition to the payoff from the ultimatum game, which ranged from 0 to 8 euros for each participant, subjects were told that six of them would be drawn at random and paid according to their answers in the questionnaire (see the Appendix for details). All questions about risk preferences involved lotteries that were actually played out according to subjects’ chosen alternatives. Answers to the question

⁴However, the CRT does not merely proxy for gender as our below results continue to hold qualitatively when the analysis is done separately for men and women.

⁵For a more detailed discussion of methodological issues related to internet experiments, see e.g., Drehmann, Oechssler, and Roeder (2005, 2007).

about time preferences decided whether subjects would receive their accumulated experimental payments through a bank transfer immediately or with 10% interest in one month’s time (and this was, of course, known to subjects). Average payoff for the subjects who were paid for their answers was 151.13 euros.

3 Results

Table 1 lists the main results of our questionnaire. For each item, we present data for a “Low” CRT group, which are those subjects that correctly answered 0 or 1 question on the CRT, and a “High” CRT group, which are subjects with 2 or 3 correct answers.⁶ We consider the Low group as the subset of more *impulsive* decision makers, whereas the High group consists of the more *reflective* decision makers. The main question of our study is whether more reflective decision makers are less likely to display biases in decision making, which, for example, are commonly studied in behavioral finance. Before turning to these biases, we report our results on risk and time preferences, which have already been the subject of earlier studies.

Risk preferences. Items 1 and 2 in Table 1 present results for two lottery choice questions in which subjects have to choose between a sure payment of x and a lottery where they have a 75% probability of getting $2x$ and a 25% probability of getting 0. Item 1 is in the domain of gains ($x = 10\text{€}$), Item 2 is in the domain of losses ($x = -10\text{€}$). For both items, the High CRT group is more likely to choose the alternative that is compatible with risk neutrality. However, although we use lotteries similar to those of Frederick (2005), we find much less evidence for loss aversion than he finds. Only about 7.7% of Low types and 2.9% of High types chose the sure payment in the domain of gains and the lottery in the domain of losses.⁷

Time preferences. Again supporting earlier evidence,⁸ we find that Low CRT types are less patient than High CRT types, although the difference is significant at the 10% level only. The test question asked subjects whether they would choose to receive their payment for the experiment now or rather in one month with 10% interest. Compared to Frederick (2005), who elicited time preferences through a similar question, our subjects are, however, generally much more patient (the monthly

⁶Splitting subjects into extreme groups of those that answered 0 or 3 questions correctly would not alter results qualitatively. However, the group with 0 correct answers is relatively small.

⁷Although Frederick (2005) does not report individual data on how many subjects chose a gamble in losses while choosing the sure payoff in gains, his aggregate data show that 54% of the Low CRT subjects gamble in losses while only 19% do so in gains.

⁸See e.g., Frederick (2005), Benjamin et al. (2006), Slonim et al. (2007) and Dohmen et al. (2007).

interest rate in Frederick is at 11.8% comparable to our 10%).⁹

Table 1: Behavioral biases and CRT scores

category	item	CRT group		sign. level
		Low	High	
risk	% preferring €10 for sure to 75% chance of winning € 20	66.0%	47.7%	$p < 0.001$
	% preferring – €10 for sure to 75% chance of losing – €20	88.5%	95.1%	$p < 0.005$
time	% preferring payment now rather than 10% more in 1 month	12.9%	8.1%	$p < 0.09$
conjunction	% thinking Linda bank teller less likely than bank teller & feminist	62.6%	38.3%	$p < 0.001$
conservatism	avg. assessed prob. for urn A (correct prob.: 0.97)	0.59	0.69	$p < 0.001$
anchoring	avg. guess when anchor 60 – avg. guess when anchor 10	7.8	10.6	n.s.
overconf.	% assessing their own # of correct answers correctly	9.7%	67.8%	$p < 0.001$

Note: Number of obs. for CRT group Low (CRT scores of 0 or 1): 157, for CRT group High (CRT scores of 2 or 3): 407. Significance levels refer to two-sided χ^2 -tests (first 4 categories) or MWU-tests (last 3 categories), respectively, for differences between CRT groups.

Conjunction Fallacy. Subjects were presented the following story due to Kahneman and Tversky (1974): “Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.” Then, they were asked to indicate which of the following two statements is more likely to be true: 1) Linda is a bank teller. 2) Linda is a bank teller and is active in the feminist movement. Of our subjects in the Low CRT group, 62.6% thought that 1) is less likely than 2), which, of course, is impossible. This percentage is much lower for the High CRT group at 38.3% ($p < 0.001$, χ^2 -test).

Conservatism. In an experiment first run by Edwards (1968), subjects were asked the following question. “Imagine there are two urns - urn A and urn B. Urn A contains 3 blue balls and 7 red balls. Urn B contains 7 blue balls and 3 red balls. Balls are now randomly drawn from one of these urns where the drawn ball is always

⁹Note that most subjects in Frederick’s (2005) study received a show-up fee only, i.e., his time-preference question was hypothetical.

placed back into the same urn. Twelve such random draws yielded 8 red balls and 4 blue balls. What do you estimate the probability that the balls were drawn from urn A?” The correct answer is 0.97, but most subjects will say something much closer to the base rate of 0.5. Again, subjects in the High CRT group are less affected by this bias in our experiment. The average probability for the High CRT group is 0.69, which is significantly higher than the 0.59 of the Low CRT group ($p < 0.001$, MWU-test).

Anchoring. When forming estimates, people often start with some initial, possibly arbitrary value, and then (insufficiently) adjust away from it. This effect is called anchoring (Kahneman and Tversky, 1974). Half of our subjects received the following question: “Do you believe that more or less than 10% of all participants answered all three [CRT] questions correctly?” The second half received the same question with the anchor 10% replaced by the anchor 60%. Thereafter, all subjects were asked: “What do you think, what is the percentage of participants who answered all three [CRT] questions correctly?” We find a significant effect of anchoring in both CRT groups. When the anchor is 10%, average guesses are about 43%. When the anchor is 60%, average guesses are about 52%. This difference is significant at $p < 0.01$ according to a MWU-test. Furthermore, the effect of anchoring is not diminished by cognitive ability. In fact, if anything, the High CRT group seems to be more susceptible to anchoring, although this effect is not significant (see Table 1).

Overconfidence. We asked each subject to estimate the number of questions on the CRT that he had solved correctly. The fraction of subjects who judge their performance in the CRT correctly is dramatically higher for the High CRT group at 67.8% than for the Low group (9.7%). Of course, this has to do with the property of the CRT questions that induce impulsive decision makers to judge their intuitive but wrong answers as correct. In fact, very few subjects underestimated their performance (1% of subjects in the Low CRT group, less than 5% in the high CRT group).

4 Conclusion

In a large-scale internet-based experiment, we studied the relationship between cognitive abilities and biases that play a prominent role in behavioral economics and finance. To measure cognitive abilities we used the Cognitive Reflection Test

(CRT) first introduced by Frederick (2005), which has several desirable properties. The test is very short and easy to conduct. Moreover, it has been shown to be able to discriminate well between impulsive and reflective decision makers, and our hypothesis was that this distinction matters for the extent to which individuals would display some of the well-known behavioral biases.

This hypothesis was largely confirmed in the sense that biases are significantly more pronounced for individuals with low cognitive abilities. In particular, we found that, as in earlier studies, risk aversion in the positive domain and impatience were significantly more pronounced among individuals with low test scores on the CRT. But we also found that individuals with low CRT scores were significantly more likely to be subject to the conjunction fallacy, to conservatism with respect to probability updating, and to overconfidence. We found no significant difference with respect to anchoring.

We should be careful to point out, though, that our findings do not imply that behavioral biases are unimportant in financial markets on the grounds that market participants are likely to have high cognitive abilities.¹⁰ Although we found significant influences of cognitive abilities on behavioral biases, even the group with high cognitive abilities showed substantial incidences of all the biases we studied.

¹⁰For example, Christelis et al. (2006) document a positive relationship between mathematical skills, verbal fluency, and recall skills and the level of stock market participation.

References

- ABREU, D., AND M. BRUNNERMEIER (2002): “Synchronization Risk and Delayed Arbitrage,” *Journal of Financial Economics*, 66, 341–360.
- BARBERIS, N., AND R. THALER (2003): “A Survey of Behavioral Finance,” in *Handbook of the Economics of Finance*, ed. by H. M. Constantinides, G., and R. Stulz, pp. 1051–1121. Elsevier Science B.V.
- BENJAMIN, D., S. BROWN, AND J. SHAPIRO (2006): “Cognitive Ability and Anomalous Preferences,” *mimeo*, *Harvard University*.
- BRANAS-GARZA, P., P. GUILLEN, AND R. LOPEZ DEL PASO (2008): “Math Skills and Risk Attitudes,” *Economics Letters*, forthcoming.
- CHRISTELIS, D., T. JAPPELLI, AND M. PADULA (2006): “Cognitive Abilities and Portfolio Choice,” *CEPR Discussion Paper No. 5735*.
- DE LONG, J., A. SHLEIFER, L. SUMMERS, AND R. WALDMAN (1990): “Noise Trader Risk in Financial Markets,” *Journal of Political Economy*, 98, 703–738.
- DOHMEN, T., A. FALK, D. HUFFMAN, F. MERKLEIN, AND U. SUNDE (2008): “The Economic Consequences of Biased Probability Judgement: Representative Evidence,” *mimeo*, *University of Maastricht*.
- DOHMEN, T., A. FALK, D. HUFFMAN, AND U. SUNDE (2007): “Are Risk Aversion and Impatience Related to Cognitive Ability?,” *IZA Discussion Paper No. 2735*.
- DREHMANN, M., J. OECHSSLER, AND A. ROIDER (2005): “Herding and Contrarian Behavior in Financial Markets: An Internet Experiment,” *American Economic Review*, 95(5), 1403–1426.
- (2007): “Herding With and Without Payoff Externalities: An Internet Experiment,” *International Journal of Industrial Organization*, 25, 391–415.
- EDWARDS, W. (1968): “Conservatism in Human Information Processing,” in *Formal Representation of Human Judgement*, ed. by B. Kleinmütz, pp. 17–52. Wiley, New York.
- FREDERICK, S. (2005): “Cognitive Reflection and Decision Making,” *Journal of Economic Perspectives*, 19(4), 25–42.

- KAHNEMAN, D., AND A. TVERSKY (1974): “Judgement under Uncertainty: Heuristics and Biases,” *Science*, 185, 1124–1131.
- OECHSSLER, J., A. ROIDER, AND P. SCHMITZ (2008): “Cooling-Off in Negotiations - Does It Work?,” *CEPR Discussion Paper No. 6807*.
- SHLEIFER, A. (2000): *Inefficient Markets: An Introduction to Behavioral Finance*. Oxford University Press, Oxford.
- SHLEIFER, A., AND R. VISHNY (1997): “The Limits of Arbitrage,” *Journal of Finance*, 52, 35–55.
- SLONIM, R., J. CARLSON, AND E. BETTINGER (2007): “Possession and Discounting Behavior,” *Economics Letters*, 97(3), 215 – 221.
- SUMMERS, L. (1986): “Does the Stock Market Rationally Reflect Fundamental Values?,” *Journal of Finance*, 41, 591–601.

5 Appendix: Instructions

In the following we present an English translation of the German instructions. Subsequent to entering their personal data (such as address, age, gender, and field of study), subjects played an (unrelated) mini-ultimatum game. Afterwards, they proceeded to a questionnaire containing various decision problems, which are the topic of the present paper. The post-experimental questionnaire also contained additional, unrelated questions pertaining to behavior in the ultimatum game.

Questionnaire

Some questions

We would be delighted if you could answer a few additional questions. The answers will take a few minutes only.

In addition, you can win a prize of up to 265 euros: At the end of the experiment we will randomly draw 6 out of the 600 participants that will be paid for their answers in the following way.¹¹ Please note that prizes will be positive. A winner will obtain:

- As a thank you for answering all of the questions a lump sum of 60 euros.
- For questions 1, 4, 6, 8, and 10: 5 euros for each correct answer.
- For questions 3 and 5: The amount in euros that is realized in the respective lottery (we will actually carry out the lotteries).
- For questions 9 and 11: For each of the questions 10 euros minus an amount that is increasing in the deviation of your answer from the correct answer. Formally, we use the following formula: $10 * (1 - ((\text{correct answer} - \text{your answer})/100)^2)$. Here is an example: Your answer is 70%. Suppose the correct answer is 50%. The resulting payoff for this question is $10 * (1 - 0.2^2) = 9.6$ euros.
- For question 7: 10 euros minus an amount that is increasing in the deviation of your answer from the correct answer. Formally, we use the following formula: $10 - (\text{correct answer} - \text{your answer})^2$.

¹¹While this phase of the experiment was designed for 600 participants, only 564 subjects filled in the questionnaire raising the individual probability to win a prize.

Questions:

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? [_____ cents]
2. If you are one of the six winners who receive a prize for this questionnaire, you can pick between two alternatives. Alternative 1: We will transfer your winnings immediately after the end of the experiment. Alternative 2: We will transfer your winnings one month after the end of the experiment with a 10% premium. Which alternative do you choose? [1/2]
3. You have the choice between two alternatives. Alternative 1: You receive 10 euros. Alternative 2: You receive a lottery ticket that yields a 75% chance of winning 20 euros. With 25% probability it is worthless. Which alternative do you choose? [1/2]
4. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? [_____ minutes]
5. You have to pay 10 euros. Would you rather replace this payment through the following alternative: With a probability of 75% you must pay 20 euros. With 25% probability you don't have to pay anything. [yes/no]
6. In a lake, there is a patch of lily pads. Every 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? [_____ days]

You have just answered three questions about the prices of a bat and a ball (B), about the production time of certain machines (M) and about a patch of lily pads (P). In the following we shall refer to these three questions as “BMP-questions”.¹²

7. What do you think, how many of the BMP-questions did you answer correctly? [0,1,2,3]
8. Do you believe that more or less than 10% [60%] of all participants answered all three BMP-questions correctly? [less/more]

¹²At this stage, subjects were prevented from going back to alter their responses to these questions.

9. What do you think, what is the percentage of participants who answered all three BMP-questions correctly? [number between 0 and 100]
10. Consider the following description. “Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.” Which of the following two statements is more likely to be true? Statement 1: “Linda is a bank teller”. Statement 2: “Linda is a bank teller and is active in the feminist movement”. [1/2]
11. Imagine there are two urns – urn A and urn B. Urn A contains 3 blue balls and 7 red balls. Urn B contains 7 blue balls and 3 red balls. Balls are now randomly drawn from one of these urns where the drawn ball is always placed back into the same urn. Twelve such random draws yielded 8 red balls and 4 blue balls. What do you estimate the probability that the balls were drawn from urn A? [Please enter a number between 0 and 100, where 100 means “it was certainly urn A” and 0 means “it was certainly urn B”]