Discussion Papers
Department of Economics
University of Copenhagen

No. 17-24

Thinking fast, thinking badly

Natalia Jimenez, Ismael Rodriguez-Lara, Jean-Robert Tyran and Erik Wengström

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
http://www.econ.ku.dk

ISSN: 1601-2461 (E)

Thinking fast, thinking badly

Natalia Jimenez • Ismael Rodriguez-Lara • Jean-Robert Tyran • Erik Wengström

October 30, 2017

Abstract

We test for the construct validity of the cognitive reflection test (CRT) by eliciting response times. We find that incorrect answers to the CRT are quicker than correct answers. At the individual level, we classify subjects into impulsive and reflective, depending on whether they choose the incorrect intuitive answer or the correct answer the majority of the time. We show that impulsive subjects complete the test quicker than reflective subjects.

Keywords: cognitive ability, cognitive reflection, response time, intuitive behavior, reflective behavior

JEL code: C91

Acknowledgements. We gratefully acknowledge financial support from the Ministerio de Economia y Competitividad under projects ECO2016-76789-P (Jimenez) and ECO2014-58297-R (Rodriguez-Lara), the Junta de Andalucia under project SEJ1436 (Jimenez), the Carlsberg Foundation (Tyran), and the Riksbankens Jubileumsfond (Wengström).

Contacts: Jimenez: natjj@ugr.es, Universidad de Granada (Spain), Rodriguez-Lara (corresponding author): I.Rodriguez-Lara@mdx.ac.uk, Middlesex University London (United Kingdom), Tyran: jean-robert.tyran@univie.ac.at, University of Vienna (Austria), Erik Wengström, erik.wengstrom@nek.lu.se, University of Lund (Sweden).

1 Introduction

Dual-system models of human thinking differentiate two cognitive processes: a type 1-system that is *fast*, automatic and non-conscious, and a type 2-system that is *slow*, controlled and conscious (Kahneman 2011, Stanovich and West 2000). Economists have recently become interested in the relation between these two cognitive processes and decision-making. The cognitive reflection test (CRT) introduced by Frederick (2005) has emerged as a popular tool to identify which way of thinking subjects use. The test consists of three questions that have "*an intuitive answer [that] does spring quickly to mind (…) but this "impulsive" answer is wrong. Anyone who reflects upon it for even a moment would recognize [the correct answer]*" (Frederick, 2005, pages 26-27).

While scores in the CRT have been related to risk preferences or behavioral biases (Frederick 2005, Oechssler et al. 2009, Bergman et al. 2010, Hoppe and Kusterer 2011, Cheung et al. 2014, Brañas-Garza et al. 2012, Andersson et al. 2016), we are not aware of any paper that directly tests the implicit assumption that the CRT measures the tendency to override an intuitive and spontaneous response that is incorrect and to engage in further reflection that leads to giving the correct response. More precisely, we lack evidence about the construct validity of the CRT showing that quick responses to the CRT are likely to be incorrect, while correct answers take longer. Our paper is an attempt to fill out this gap.

2 Data

Hard-copy invitation letters were sent out to a random sample of the Danish population aged between 18 and 80. A total of 2,347 subjects logged on to our webpage and participated (average age = 46.7, SD = 14.3; 1,209 males and 1,138 females). The experiment consisted of two incentivized parts, a public good game (see Thöni et. al 2012, Fosgaard et al. 2014) and a risk elicitation task (see Andersson et al. 2016). The incentivized part was followed by a questionnaire, which included the CRT (Frederick, 2005), as well as basic socio-economic questions, the Big Five personality test and a 20-

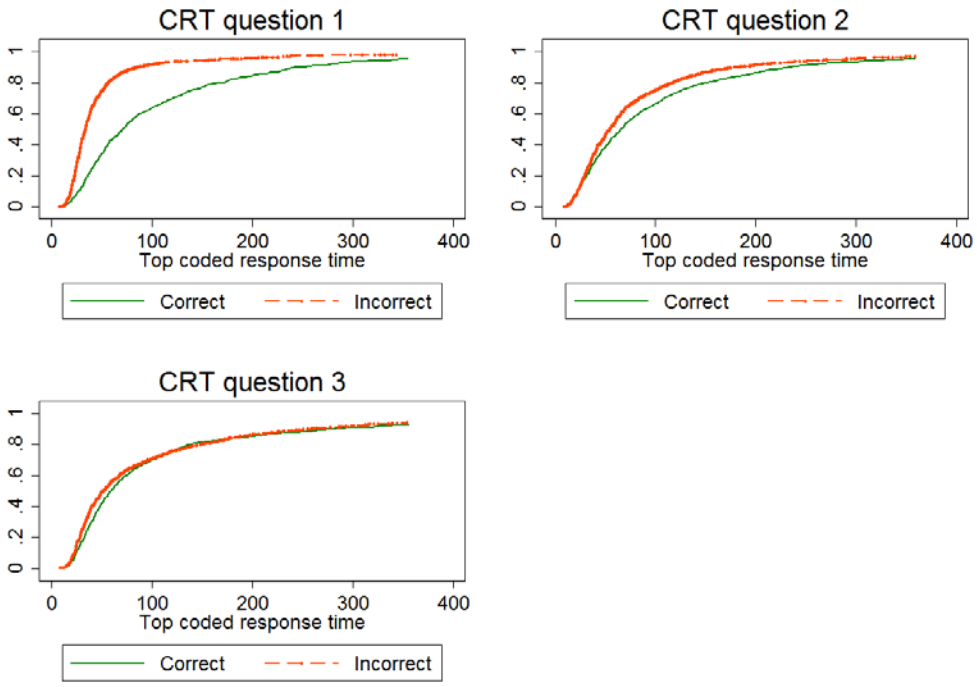
item cognitive ability test similar to a Raven's progressive matrices test (henceforth referred to as the cognitive ability test).[1]

3 Results

3.1 Correct answers and response times

Figure 1 depicts the cumulative response times for subjects that gave correct and incorrect answers to each question (see Section A3 in the Appendix for more detailed descriptive statistics).[2] We find that subjects who provided the correct answer devoted more time to each question ($p < 0.001$).[3]

Figure 1. Cumulative response times to each question.



[1] More information about the details of the questionnaire, the recruitment procedures and the sample composition is presented in Sections A1 and A2 of the Appendix.
[2] Response times of more than 360 seconds have been excluded since data contains outliers due to people taking a break or being interrupted. The choice of cut off is not important for any of our results.
[3] Unless otherwise noted, we use the Mann-Whitney and Kolmogorov-Smirnov tests.

Figure 1 also reveals that the difference in speed between correct and incorrect answers differs across questions. The difference is particularly striking in question 1, and much less pronounced in question 3. It could be that the first question has a more salient intuitive answer, or perhaps subjects figure out after the first question that they need to think longer since these are tricky if not trick questions.[4] Both explanations are consistent with our data since mean response times are increasing with questions (see Section A3 of the Appendix). As a robustness check, we collected additional data using an alternative measure of cognitive reflection (Toplak et al., 2014) with randomized and non-randomized questions to test for possible order effects. Overall, we do not find evidence of order effects, suggesting that the different patterns observed across questions is likely not due to the order of presentation, but rather due to characteristics of the questions.[5]

3.2 Intuitive but Incorrect Answers and Response time

Our previous findings support the hypothesis that fast responses are associated with incorrect answers, and vice versa for slow responses. While "impulsive" subjects are frequently defined as those who perform poorly in the CRT, subjects who provide the intuitive (wrong) answer might be treated differently than those who simply provided any incorrect answer (Noussair et al. 2014, Cueva et al. 2015, Ponti and Rodriguez-Lara 2015). We follow Cueva et al. (2015) and use the iCRT index which adds up the number of intuitive answers, $iCRT \in \{0,1,2,3\}$. We then define *Impulsive subjects* as those who scored two or more in the iCRT (39 % of the sample) and *Reflective subjects* as those who provided two or more correct answers in the CRT (49 % of the sample). The remaining 12% are classified as *Other*.

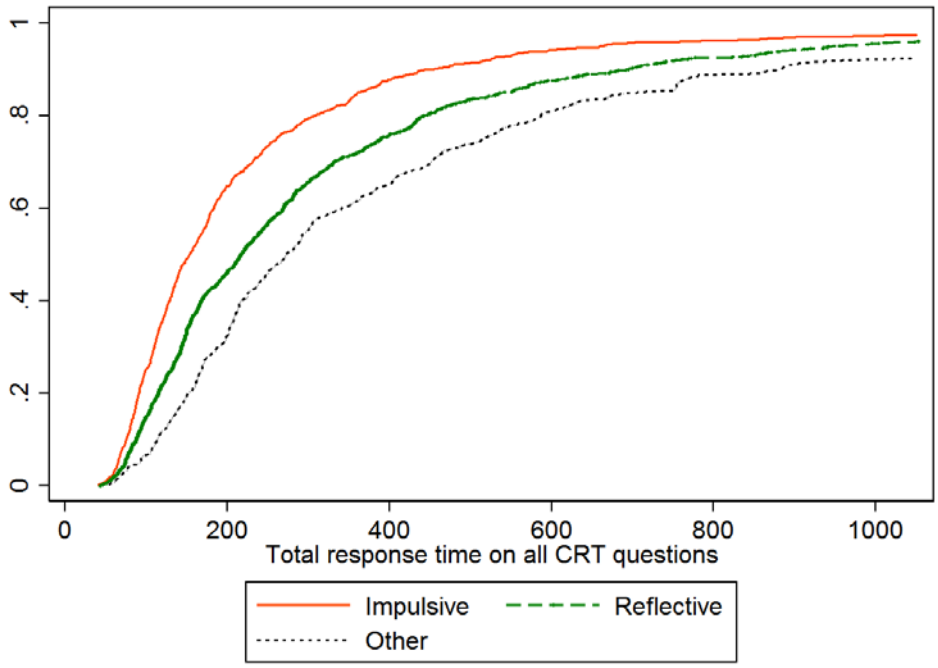
Figure 2 displays the cumulative response time distributions for the three types of subjects. We find that Impulsive subjects are faster (in total response times) than

[4] These arguments also relate to the "sequence effect" in Brañas et al. (2015). They report that subjects score better when questions are presented in the standard order, and the smallest (largest) proportion of correct answers is usually observed in question 1 (question 3).

[5] For further details, see Section A6 on the Appendix.

Reflective ones ($p < 0.001$), while Other subjects are slower than both the Impulsive and Reflective ones ($p < 0.001$).

Figure 2. Cumulative response times of Impulsive, Reflective and Other subjects



3.3 Regression Analysis

To learn more about the relationship between CRT scores and response times, we present a series of regressions in which we control for other factors that are likely to be correlated with both CRT scores and response times.

Table 1 displays the result from a series of OLS regressions using CRT score as the dependent variable. In column 1, we have included response time as the single explanatory variable. The response time has been top-coded at 1080 seconds, i.e. the response times of subjects that take more than 1080 seconds are recoded as 1080. Our results are robust to the choice of different cutoffs and also to replacing the OLS with an ordered logit or probit (see Table A6 on the Appendix). In columns 2-5, we include additional controls for gender, age, education, cognitive ability and big five personality traits.

Table 1. CRT Score, OLS regression

	(1)	(2)	(3)	(4)
Response time	0.00047***	0.00054***	0.00055***	0.00067***
	[9.10e-05]	[9.07e-05]	[8.91e-05]	[8.52e-05]
Female		-0.422***	-0.406***	-0.366***
		[0.044]	[0.045]	[0.046]
Age 30-39		-0.018	-0.139*	-0.011
		[0.081]	[0.081]	[0.078]
Age 40-49		-0.005	-0.107	0.0422
		[0.072]	[0.073]	[0.071]
Age 50-59		-0.008	-0.086	0.164**
		[0.074]	[0.075]	[0.075]
Age 60-80		-0.241***	-0.333***	0.040
		[0.077]	[0.078]	[0.081]
Basic Education			-0.095	-0.048
			[0.081]	[0.077]
Short Secondary Education			0.164***	0.115**
			[0.056]	[0.054]
Short Tertiary Education			0.580***	0.461***
			[0.069]	[0.067]
Cognitive ability				0.112***
				[0.007]
Big5 Personality Scores	No	No	No	Yes
Constant	1.328***	1.567***	1.473***	0.947***
	[0.034]	[0.065]	[0.065]	[0.065]
Observations	2,347	2,347	2,347	2,333
R-squared	0.011	0.053	0.088	0.185

Note. Standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

The main message of Table 1 is that longer response times are significantly associated with higher CRT scores. A one standard deviation increase in response time corresponds up to a 0.15 standard deviation change in the CRT score. The effect is relatively consistent across specifications and it becomes stronger as we include more control variables. Moreover, CRT scores are related to gender, education and cognitive ability. Notably, there is a negative effect of age effect in columns 2 and 3, but this effect vanishes once we include the cognitive ability test score (column 4). Hence, the decline in CRT with age effect appears to be driven by a decline in cognitive ability.

Table 1 only considers the total score on the CRT and does not distinguish between intuitive and other incorrect answers. Table 2 looks at this issue by means of OLS regressions, where the association to response times is different for Impulsive and Other subjects (the Reflective subjects constitute the left-out category.) [6]

Our estimates show that the relationship reported in Table 1 is driven by the Impulsive subjects, which are on average much faster than the Reflective ones. Impulsive subjects have response times that are up to 0.43 of a standard deviation shorter than the Reflective subjects. To the contrary, subjects who perform poorly on the CRT but are not classified as Impulsive (Other) take more time than the Reflective subjects did. Thus, to measure impulsive thinking, it is important to distinguish between different types of wrong answers and not only count the overall number of incorrect answers. As we include more covariates, the coefficient for the Impulsive subjects increases, whereas the coefficient for the Other subjects decreases. Response times are correlated with gender, age and cognitive ability.

[6] The dependent variable is total response time (in all three questions). Again the results are not sensitive to the choice of cutoff for the top-coding and hold if we instead use a Tobit model or median regressions without top-coding (see Tables A7-A8 in the Appendix).

Table 2. Total response time, OLS regression

	(1)	(2)	(3)	(4)
Impulsive	-80.89***	-85.82***	-87.78***	-107.1***
	[10.74]	[10.71]	[10.84]	[11.23]
Other	75.00***	63.12***	60.83***	43.54***
	[16.28]	[16.09]	[16.21]	[16.33]
Female		21.17**	21.24**	18.16*
		[10.04]	[10.27]	[10.98]
Age 30-39		8.955	12.49	6.900
		[17.99]	[18.43]	[18.48]
Age 40-49		22.85	26.20	14.75
		[16.16]	[16.57]	[16.91]
Age 50-59		46.47***	50.48***	26.92
		[16.48]	[16.94]	[17.90]
Age 60-80		128.1***	132.9***	94.68***
		[17.10]	[17.52]	[19.08]
Basic Education			-18.16	-19.48
			[18.31]	[18.24]
Short Secondary Education			-8.822	-7.907
			[12.78]	[12.75]
Short Tertiary Education			-24.93	-25.53
			[15.93]	[16.11]
Cognitive ability				-11.46***
				[1.73]
Big5 Personality Scores	No	No	No	Yes
Constant	306.6***	255.8***	263.8***	309.4***
	[7.167]	[14.39]	[15.47]	[62.95]
Observations	2,347	2,347	2,347	2,333
R-squared	0.044	0.077	0.078	0.097

Notes. Standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

4 Robustness check

We use post-experimental questionnaires in three other experiments to investigate the robustness of our findings. First, we elicited the response time of 311 students (M Age = 21.2 years, SD = 3; 132 males and 179 females) who participated in a laboratory experiment at the Universidad Pablo de Olavide in Seville (Spain). We utilize the alternative measure of cognitive reflection in Toplak et al. (2014) in experiments run at the Universidad de Valencia and the Universidad de Alicante, with a total of 312 participants (M Age = 22.7 years, SD = 5.8; 119 males and 193 females). In addition, Amazon's Mechanical Turk (MTurk) was used to collect data from 195 participants (M Age = 37.8 years, SD = 13.2; 81 males and 114 females) using the Toplak et al.'s version of the test, but with the questions presented in a randomized order.

Overall, we find similar results when we look at these data (a detailed analysis can be found in the Appendix, sections A5-A6). We confirm that Impulsive subjects are always faster (in terms of total response time) than Reflective ones ($p < 0.045$). We do not find that the order of presentation affects the difference in response times between correct and incorrect answers in the MTurk experiment.

5 Conclusion

To the best of our knowledge, this paper is the first providing evidence that fast answers to the CRT tend to be incorrect, while subjects who take longer tend to provide the correct answer. Our findings lend support to the assumption underlying the test that responses to the CRT can be used to measure the tendency to override intuitive responses. However, our findings also show that merely summing the number of incorrect answers will provide a poor measure of intuitive thinking. Instead, it is important to distinguish between intuitive answers and other types of incorrect answers.

References

Andersson, O., Holm, H. J., Tyran, J.-R., Wengström, E. (2016). Risk aversion relates to cognitive ability: preferences or noise? *Journal of the European Economic Association*, 14(5): 1129-1154.

Bergman, O., Ellingsen, T., Johannesson, M., Svensson, C. (2010). Anchoring and cognitive ability. *Economics Letters*, 107(1), 66-68.

Brañas-Garza, P., Garcia-Muñoz, T., González, R. H. (2012). Cognitive effort in the beauty contest game. *Journal of Economic Behavior and Organization*, 83(2), 254-260.

Brañas-Garza, P., Kujal, P., Lenkei, B. (2015). Cognitive Reflection Test: Whom, how, when. MPRA Paper No. 68049.

Cueva, C., Iturbe-Ormaetxe, I., Mata-Perez, E., Ponti, G., Yu, H., Zhukova, V. (2015). Cognitive (ir)reflection: New experimental evidence. *Journal of Behavioral and Experimental Economics*, 64(5): 81-93.

Cheung, S. L., Hedegaard, M., Palan, S. (2014). To see is to believe. Common expectations in experimental asset markets. *European Economic Review*, 66: 84–96.

Fosgaard, T. R., Hansen, L. G., Wengström, E. (2014). Understanding the nature of cooperation variability. *Journal of Public Economics*, 120: 134-143.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4): 25-42.

Hoppe, E. I., Kusterer, D. J. (2011). Behavioral biases and cognitive reflection. *Economics Letters*, 110(2): 97-100.

Kahneman, D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux.

Noussair, C. N., Tucker, S. J., Xu, Y. (2014). A futures market reduces bubbles but allows greater profit for more sophisticated traders. University of Waikato, Working Paper in Economics 12/14.

Oechssler, J., Roider, A., Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior and Organization*, 72(1): 147-152.

Ponti, G., Rodriguez-Lara, I. (2015) Social preferences and cognitive reflection: Evidence from a dictator game experiment. *Frontiers in Behavioral Neuroscience*, 9, 146.

Stanovich, K. E., West, R. F. (2000). Individual difference in reasoning: implications for the rationality debate? *Behavioural and Brain Sciences*, 23: 645-726.

Thöni, C., Tyran, J.-R., Wengström, E. (2012). Microfoundations of social capital. *Journal of Public Economics*, 96(7): 635-643.

Toplak, M. E., West, R. F., Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, 20(2): 147-168.

Appendix

This appendix presents the details of the questionnaires used to elicit cognitive reflection (Frederick 2005, Toplak et al. 2014). The appendix also contains information about the recruitment procedures, sample composition and additional tests for the Internet experiment. We present details and data analysis of our laboratory experiments in Spain and the experiment we ran using Mechanical Turk, in which the order of the CRT questions was randomized.

Contents

A1. Cognitive reflection tests in our studies

A2. Internet data

A3. Descriptive statistics

A4. Internet data: additional robustness checks

A5. Results from the Lab and MTurk experiments – detailed analysis

A6. Order effects

A1. Cognitive reflection tests in our studies

A1.1 Frederick test

The Cognitive Reflection Test (CRT) in Frederick (2005) consists of the following three questions, each of which has an intuitive but incorrect answer:

(1) **Bat and ball.** A bat and a ball cost $1.10 in total. The bat costs a dollar more than the ball. How much does the ball cost? ____ cents [Correct answer = 5 cents; intuitive answer = 10 cents]

(2) **Machines.** It takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? _____ minutes [correct answer = 5 minutes; intuitive answer = 100 minutes]

(3) **Lily pads.** 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 [Correct answer = 47 days; intuitive answer = 24 days]

We use this questionnaire in our Danish internet sample and our laboratory experiment in Seville, Spain (henceforth, Lab experiment 1). We believe that the exposure to the CRT test was relatively low among the general population in Denmark at the time of the experiment in 2008. To our knowledge, subjects participating in the Lab experiment 1 were inexperienced and never exposure to the test.

A1.2 Toplak et al. test

We decided to check the robustness of our findings using questions in Toplak et al. (2014):

(1) **Barrels.** If John can drink one barrel of water in 6 days, and Mary can drink one barrel of water in 12 days, how long would it take them to drink one barrel of water together? _____ days [correct answer = 4 days; intuitive answer = 9]

(2) **Marks.** Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are in the class? _____ students [correct answer = 29 students; intuitive answer = 30]

(3) **Pig.** A man buys a pig for $60, sells it for $70, buys it back for $80, and sells it finally for $90. How much has he made? _____ dollars [correct answer = $20; intuitive answer = $10]

(4) **Stocks.** Simon decided to invest $8,000 in the stock market one day early in 2008. Six months after he invested, on July 17, the stocks he had purchased were down 50%. Fortunately for Simon, from July 17 to October 17, the stocks he had purchased went up 75%. At this point, Simon has: a. broken even in the stock market, b. is ahead of where he began, c. has lost money [correct answer = c, because the value at this point is $7,000; intuitive response = b].

This version of the CRT was used in our second laboratory experiment in Spain (henceforth, Lab experiment 2).[7] Sessions of this experiment were run in Alicante and Valencia, where there is an increasing exposure to the three classic items; however, participants in these locations were never exposed to the Toplak et al. (2014) test. This test was also used in our experiment in Amazon's Mechanical Turk (MTurk), where the order of the questions was randomized.

[7] According to the authors, the four-item measure displays a 0.58 correlation with the original version.

A2. Internet data

This section presents information about the recruitment procedures and the sample characteristics of our internet data. In addition, it gives an overview of the cognitive and psychometric tests that are used to construct the control variables in the regression analyses.

A2.1 Recruitment of subjects

The participants were recruited as follows:

- Statistics Denmark, the official statistics office in Denmark, randomly selected 40,000 individuals from the Danish population.
- In total, 18,027 individuals were randomly selected out of the 40,000 and hard-copy letters were sent out to the respondents in two waves on May 15 and May 30, 2008.
- The letters invited subjects to log on to our webpage at the University of Copenhagen, using a personal identification number printed in the letter. Subjects had one week to complete the experiment.
- In total, 3,107 subjects logged on to our web page and out of these, 2,037 completed the cognitive reflection test (CRT).

A2.2 Representativeness of sample

The sample of participants considered in the current project is generally representative of the Danish population. Table A1 reports the gender, age and educational characteristics of our sample and the Danish population, respectively. As can be seen from the table, the gender and age distributions of the participants in our sample match the corresponding distributions of the Danish population quite closely, although there are exceptions. For example, females in the age range 41-50 are overrepresented in our sample. The educational distribution of the sample does not follow the general population as closely as the gender and age distributions. People with a vocational educational background are under-represented, whereas people with tertiary education are overrepresented.

Table A1. Representativeness of sample

	Our Sample			Danish population*		
	Male	Female	Total	Male	Female	Total
No. observations	1,209	1,138	2,347			
Gender Women			48%			50%
Age						
18-30	15%	14%	14%	21%	20%	20%
31-40	14%	17%	16%	19%	19%	19%
41-50	24%	29%	26%	20%	19%	20%
51-60	24%	23%	24%	18%	18%	18%
61-70	16%	14%	15%	15%	15%	15%
71-80	7%	3%	5%	8%	9%	8%
Education						
Basic education (up to 10 years)	11%	10%	11%	26%	27%	26%
High school (up to 12 years)	13%	12%	12%	6%	7%	6%
Vocational education (up to 12 years)	18%	7%	13%	42%	36%	39%
Short tertiary education (less than 3 years)	11%	19%	15%	6%	4%	5%
Medium tertiary education (between 3 and 4 years)	26%	39%	32%	11%	20%	16%
Long tertiary education (more than 4 years)	21%	13%	17%	8%	6%	7%

**Source*: Statistics Denmark (http://www.dst.dk/HomeUK.aspx). For gender and age the population is restricted to individuals between 18-80 years of age. For education the population is restricted to individuals between 18 and 69. The education variables for the subjects of the experiment include ongoing education whereas the figures for the Danish population only refer to completed education.

A2.3 Overview of the experiment

In short, the participants were invited to log on to our web page twice, once during the period in which the experiment was open and once during a feedback period after the experiment was closed. The first time they logged on they participated in two public goods games and completed a series of other questionnaires and tests. After the experiment closed, participants were matched together in groups for the public good game and payments were calculated. Participants logged on to our web page again to see the results of their group and provided us with their bank details necessary for distributing the payments. Below we describe the different parts of the experiment.

A2.4 Login and information screens

The first screen of the experiment was a simple login screen where subjects had to enter the personal identification code printed in the invitation letter. Upon login, subjects saw a welcome screen providing information about the experiment. They were informed that their participation in the experiment would be valuable to research in economics and reminded of the importance that the person participating was the person named in the invitation letter. Moreover, they were informed that they could earn money in the experiment (within the range of 8 to 510 DKr, corresponding to approximately 1.6 to 102 USD) and that this is standard procedure in economic experiments. They were also cautioned that they had to complete the experiment to get their money by electronic transfer. All subjects were then informed that the experiment would last approximately 50 minutes. Finally, they were reassured that they would remain anonymous.

After answering some questions about their socioeconomic background (age, gender and highest completed education), subjects proceeded to a public goods experiment which was the main incentivized part of the experiment.

A2.5 The public good games

Subjects played two variants of the public good game. First they played a standard linear one-shot public good game involving one unconditional contribution choice (referred to as the Standard game). Afterwards they played a public goods game using the strategy method which involves an unconditional choice as well as a series of conditional choices (referred to as the Strategy game).

A2.6 Other measures

After the public good game, a series of other tests including the CRT follows. In addition to the CRT test, the subjects completed the visual IST 2000R[8] **Cognitive ability test** (Beauducel et al. 2010). This test asks the subjects to solve 20 different logic puzzles.

[8] Used with permission from the Danish Psychology Publisher, www.dpf.dk.

The task in each puzzle is to identify one of five candidate symbols, which would finalize a sequence of pictures constituting a logical graphical string (for a snapshot example, see the appendix). For instance, subjects see three solid square boxes in a row as the logical string. Subjects are asked which of five suggested symbols would logically prolong the presented string. If subjects, for instance, can choose between a triangle, a line, a circle and a squared solid box, the correct answer is to choose the solid box, which is the only logical continuation of the sequence of symbols. The subjects were given 10 minutes to solve as many of the puzzles as possible, and were allowed to jump back and forth between the puzzles as they wished. The assumption is that the higher the number of puzzles solved, the higher the cognitive ability of the participant.

We also applied a Danish version of the **Big 5 personality test**.[9] The test consists of 60 statements covering personality traits in five dimensions: agreeableness, conscientiousness, extraversion, neuroticism, and openness.[10] Based on the answers to these statements, each subject is assigned a score for each of the big 5 dimensions. A high score for a given trait indicates that the trait is an important part of the subject's personality.

[9] We used the Danish NEO-PI-R Short Version test by permission of Danish Psychology Publishing (www.dpf.dk).

[10] The Danish NEO-PI-R Short Version consists of five 12-item scales which measure each of the 5 domains. The 12 items for each domain are chosen from the original 48 items (of the full NEO-PI-R test) as follows: for each facet, the two items (out of eight) with the highest correlation with the total factor score are chosen (this is different from the American 60-item version of NEO-PI-R, called NEO-FFI, where the 12 items with the highest correlation with the total factor score are picked, regardless of which facet the single items belong to). In the Danish short version, all facets are therefore equally represented within each domain.

A3. Descriptive statistics

Our study comprises data from four experiments. Besides the Internet experiment that is described in detail above (and analyzed extensively in the main text) we also collected data from two laboratory experiments in Spain and one experiment carried out using MTurk. Tables A2-A5 provide descriptive statistics for the four experiments. Panel a) summarizes the frequency of correct and incorrect answers and the mean/median response time to each question. Panel b) reports the frequency of reflective, impulsive and other subjects and their mean/median response time (see section 3.2 in the manuscript for the criteria we used to classify subjects).

Table A2. Internet experiment: Descriptive statistics (N = 2,347)

a) Responses to the CRT (Frederick, 2005)

	Correct	*Incorrect Intuitive*	*Incorrect Other*	*All*
CRT1	845 (36%)	1430 (61%)	72 (3%)	
Mean [Median] time	106.0 [71]	48.0 [33]	106.1 [63]	70.7 [40]
CRT2	1,197 (51%)	781 (33%)	369 (16%)	
Mean [Median] time	98.9 [65]	70.6 [46]	106.7 [72]	90.7 [60]
CRT3	1,388 (59%)	741 (32%)	218 (9%)	
Mean [Median] time	100.6 [59]	70.5 [42]	178.9 [158]	98.4 [56]

b) Classification of subjects

	Reflective	*Impulsive*	*Other*	*All*
Number of subjects (%)	1,148 (49%)	923 (39%)	276 (12%)	
Mean [median] total time	306.6 [217.5]	225.7 [154]	381.6 [274]	283.6 [192]

Note. All response times are top coded at 360 seconds (or 1080 for total response time).

Table A3. Lab experiment 1: Descriptive statistics (N = 311)

a) Responses to the CRT (Frederick, 2005)

	Correct	*Incorrect Intuitive*	*Incorrect Other*	*All*
CRT1	86 (28%)	212 (68%)	13 (4%)	
Mean [Median] time	72.4 [58.5]	43.1 [32]	95.5 [81]	53.3 [39]
CRT2	68 (22%)	191 (61%)	52 (17%)	
Mean [Median] time	75.8 [58]	58.9 [50]	80.8 [68]	66.2 [55]
CRT3	99 (32%)	149 (48%)	63 (20%)	
Mean [Median] time	57.8 [47]	63.5 [53]	99.5 [86]	69.0 [54]

b) Classification of subjects

	Reflective	*Impulsive*	*Other*	*All*
Number of subjects (%)	72 (23%)	188 (60%)	51 (17%)	
Mean [median] total time	167.4 [151.5]	196 [175.5]	255.7 [242]	189 [170]

Table A4. Lab experiment 2: Descriptive statistics (N = 312)

a) Responses to the CRT (Toplak et al., 2014)

	Correct	*Incorrect Intuitive*	*Incorrect Other*	
CRT1	88 (28%)	79 (25%)	145 (47%)	
Mean [Median] time	77.3 [55]	50.1 [47]	47.5 [45]	53.8 [47]
CRT2	57 (18%)	140 (45%)	115 (37%)	
Mean [Median] time	23.4 [19]	21.7 [18]	21.8 [20]	20.0 [19]
CRT3	130 (42%)	106 (34%)	76 (24%)	
Mean [Median] time	28.7 [26.5]	25.0 [23]	28.1 [24]	27.3 [25]
CRT4	154 (49%)	149 (45%)	18 (6%)	29.3
Mean [Median] time	31.4 [29]	27.1 [23]	27.7 [22.5]	[25]

b) Classification of subjects

	Reflective	*Impulsive*	*Other*	*All*
Number of subjects (%)	107 (34%)	128 (41%)	77 (25%)	
Mean [median] total time	144.7 [132]	125.4 [118]	126.8 [122]	132.4 [121]

Table A5. Mechanical Turk Experiment: Descriptive statistics (N = 195)

a) Responses to the CRT (Toplak et al., 2014)

	Correct	*Incorrect Intuitive*	*Incorrect Other*	*All*
CRT1	55 (28%)	27 (14%)	113 (58%)	
Mean [Median] time	49.2 [33.9]	37.1 [25.6]	31.5 [22.2]	37.3 [26.5]
CRT2	33 (17%)	103 (53%)	59 (30%)	
Mean [Median] time	19.4 [14.8]	16.8 [13.3]	15.2 [11.9]	16.7 [12.9]
CRT3	71 (36%)	73 (37%)	51 (27%)	
Mean [Median] time	35.2 [28.4]	34.7 [24.3]	33.7 [22.2]	34.6 [25.7]
CRT4	85 (44%)	91 (47%)	19 (9%)	
Mean [Median] time	35.2 [31.1]	28.0 [21.1]	25.4 [15.6]	30.9 [24.6]

b) Classification of subjects

	Reflective	*Impulsive*	*Other*	*All*
Number of subjects (%)	65 (33%)	86 (44%)	44 (23%)	
Mean [median] total time	161.3 [131.4]	105.1 [90.9]	111.1 [73.9]	125.3 [95.7]

A3.1 Experiments using Frederick (2005) (Internet and Lab experiment 1)

The mean and median response times of those giving the correct answer are higher than for those giving an incorrect answer, with the only exception of CRT3 in the Lab experiment 1, where the relationship is reversed. The internet data yield a higher percentage of correct answers than the Lab experiment 1 for all the three questions. One possible explanation is that Internet participants are not time constrained, thus they have time to ask for help or Google the correct answer. An alternative explanation is that the proportion of females in the lab data is significantly higher than in the Internet one (58% vs 48%, respectively). As it is shown in the regression analysis, females score less than males in the CRT score (see Table 1 in the main text or Table A6 below). Importantly, the difference in the percentage of correct answers translates into more reflective subjects in the Internet experiment (49% vs 23%) and more impulsive subjects in the Lab experiment 2 (39% vs 60%).[11] Despite the different distribution of types, our main

[11] Danish subjects might indeed be more reflective than Spanish ones, but we would need to run an experiment in both countries with the same characteristics to verify this last hypothesis.

conclusion still holds: reflective subjects take more than impulsive ones to response the test (and other subjects take more than both impulsive and reflective ones).

A3.2 Experiments using Toplak et al. (2014) (Lab experiment 2 and MTurk)

The proportion of correct answers and the distribution of types is remarkably consistent in the two experiments in which we used the version of Toplak et al. (2014). Across all 4 questions, the mean and median response times of those giving correct answer are higher than for those giving an incorrect answer. In our Lab experiment 2, we classify 41% of subjects as impulsive and 34% as reflective, which is very close to percentages observed in the MTurk experiment (44% and 33%) that use the same CRT measure. In both experiments, reflective subjects take more than impulsive subjects to response the test. Contrary to what is observed in experiments using Frederick (2005) the category for other subjects seems to be somewhere in the middle (if anything, other subjects behave as impulsive).

Our next section presents the robustness check for our Internet data. Section A5 extends the analysis to the three studies not discussed in the main text: Laboratory experiments 1 and 2 and the MTurk experiment. This section presents also the cumulative response times and the regression analysis for the three experiments.

A4. Internet data: additional robustness checks

In this section, we present results from a series of robustness checks. Table A6 presents Ordered Logit estimates comparable with the OLS estimates of Table 1 in the main text. Table A7 and A8 shows robustness checks of Table 2 in the main text with respect to the estimation technique and without top coding of the response time variable (Tobit regressions in Table A7 and Median regressions in Table A8). In all cases, we obtain results very similar to those presented in the main text.

Table A6. CRT Score, Ordered Logit regression

	(1)	(2)	(3)	(4)
Response time	0.0007***	0.0009***	0.0009***	0.00127***
	[0.0001]	[0.0001]	[0.0001]	[0.0002]
Female		-0.708***	-0.691***	-0.681***
		[0.0756]	[0.0777]	[0.0849]
Age 30-39		-0.0234	-0.227	-0.0101
		[0.138]	[0.142]	[0.145]
Age 40-49		0.000480	-0.183	0.108
		[0.123]	[0.127]	[0.132]
Age 50-59		0.00358	-0.135	0.331**
		[0.126]	[0.130]	[0.140]
Age 60-80		-0.389***	-0.555***	0.0995
		[0.132]	[0.136]	[0.150]
Basic Education			-0.169	-0.0872
			[0.140]	[0.143]
Short Secondary Education			0.275***	0.198**
			[0.0967]	[0.0985]
Short Tertiary Education			0.978***	0.809***
			[0.121]	[0.125]
Cognitive ability				0.206***
				[0.0138]
Big5 Personality Scores	No	No	No	Yes
cut1				
Constant	-0.867***	-1.281***	-1.158***	-0.349
	[0.0626]	[0.117]	[0.124]	[0.495]
cut2				
Constant	0.266***	-0.111	0.0411	0.953*
	[0.0602]	[0.114]	[0.122]	[0.496]
cut3				
Constant	1.459***	1.114***	1.300***	2.312***
	[0.0672]	[0.116]	[0.124]	[0.498]
Observations	2,347	2,347	2,347	2,333

Note. Standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A7. Total response time, Tobit regression

	(1)	(2)	(3)	(4)
Impulsive	-81.96***	-87.00***	-88.98***	-108.7***
	[11.13]	[11.10]	[11.22]	[11.61]
Other	78.08***	65.80***	63.46***	45.77***
	[16.91]	[16.70]	[16.81]	[16.91]
Female		21.62**	21.72**	18.67
		[10.40]	[10.63]	[11.35]
Age 30-39		8.482	12.07	6.425
		[18.63]	[19.07]	[19.09]
Age 40-49		23.28	26.69	15.15
		[16.73]	[17.15]	[17.47]
Age 50-59		47.75***	51.85***	27.92
		[17.07]	[17.54]	[18.50]
Age 60-80		132.4***	137.2***	98.12***
		[17.72]	[18.15]	[19.73]
Basic Education			-18.62	-20.09
			[18.96]	[18.86]
Short Secondary Education			-9.087	-8.119
			[13.23]	[13.18]
Short Tertiary Education			-25.26	-25.85
			[16.50]	[16.66]
Cognitive ability				-11.69***
				[1.793]
Big5 Personality Scores	No	No	No	Yes
Sigma	251.6***	247.2***	247.1***	244.3***
	[3.799]	[3.733]	[3.731]	[3.699]
Constant	309.3***	257.1***	265.3***	317.1***
	[7.436]	[14.90]	[16.01]	[65.08]
Observations	2,347	2,347	2,347	2,333

Notes. Standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A8. Total response time without top coding, Median regression

	(1)	(2)	(3)	(4)
Impulsive	-64***	-63***	-62.33***	-73.47***
	[9.140]	[9.027]	[9.088]	[9.395]
Other	56***	42***	41.67***	31.48**
	[13.86]	[13.56]	[13.59]	[13.67]
Female		15*	15*	11.52
		[8.458]	[8.610]	[9.191]
Age 30-39		-4	-6.333	-5.266
		[15.16]	[15.45]	[15.47]
Age 40-49		10	7.667	2.046
		[13.62]	[13.90]	[14.15]
Age 50-59		35**	36**	13.96
		[13.89]	[14.21]	[14.98]
Age 60-80		91***	93.33***	64.67***
		[14.41]	[14.70]	[15.96]
Basic Education			-13.67	-8.169
			[15.35]	[15.26]
Short Secondary Education			-2.333	-6.259
			[10.72]	[10.67]
Short Tertiary Education			-15.33	-15.63
			[13.36]	[13.48]
Cognitive ability				-7.725***
				[1.451]
Big5 Personality Scores	No	No	No	Yes
Constant	218***	183***	187.7***	184.9***
	[6.102]	[12.12]	[12.97]	[52.68]
Observations	2,347	2,347	2,347	2,333

Notes. Standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

A5. Results from the Lab and MTurk experiments – detailed analysis

A5.1 Correct answers and response times

Figures A1 and A2 present the cumulative response times in our laboratory experiments, while Figure A3 depicts the data for the MTurk experiment.

Figure A1. Lab experiment 1: Cumulative response times for correct and incorrect answers

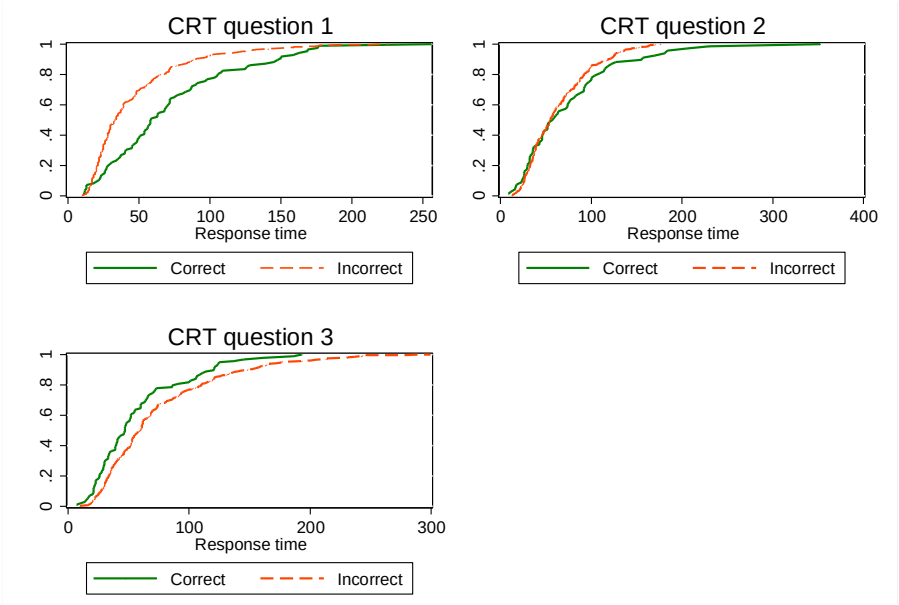


Figure A2. Lab experiment 2: Cumulative response times for correct and incorrect answers

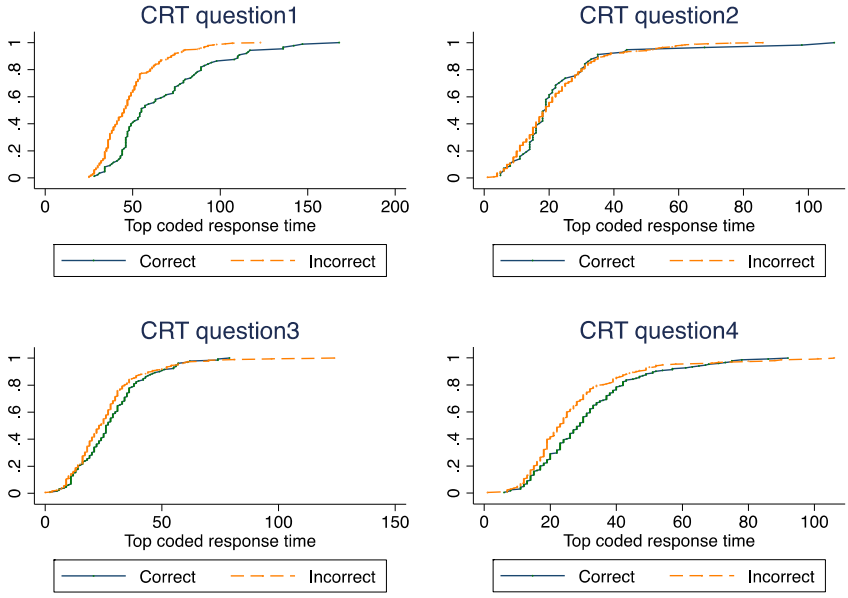
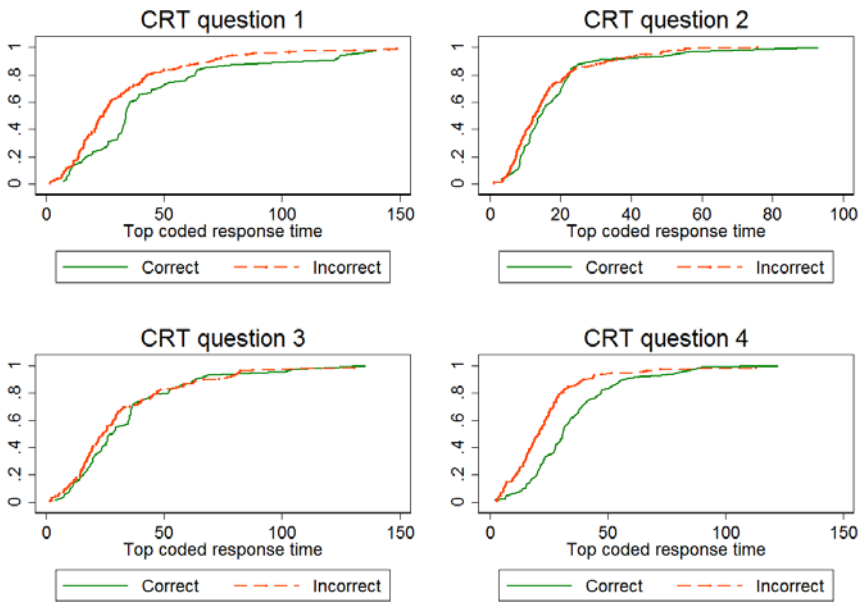


Figure A3. MTurk Experiment: Cumulative response times for correct and incorrect answers



In our Lab experiment 1, the distributions of response times for subjects with correct and incorrect answers are significantly different using the Mann-Whitney test for the first question ($p < 0.001$). The difference is not significant for question 2 ($p = 0.484$) and for question 3 the difference is significant but in the opposite direction ($p = 0.002$). In our Lab experiment 2, the distributions of response times for subjects with correct and incorrect answers are significantly different using the Mann-Whitney test for all questions, except for the third one (CRT_1: $p < 0.001$, CRT_2: $p = 0.051$, CRT_3: $p = 0.807$ and CRT_4: $p = 0.003$). A similar pattern is observed in the MTurk experiment. The distributions of response times for subjects with correct and incorrect answers are significantly different using the Mann-Whitney test for the first and fourth questions ($p < 0.01$). The difference is not significant for the second and third questions ($p > 0.18$).

A5.2 Response times for different types of subjects (Impulsive, Reflective and Others)

We classify subjects into *Impulsive*, *Reflective* or *Other* depending on their score in the CRT and the *i*CRT (see Tables A2-A5 for the distribution of types in each experiment). Figures A3 and A4 present the cumulative distribution of response times in our laboratory studies, while Figure A5 replicates the analysis for the MTurk experiment.

Figure A4. Laboratory experiment 1: Cumulative response times for different types of subjects

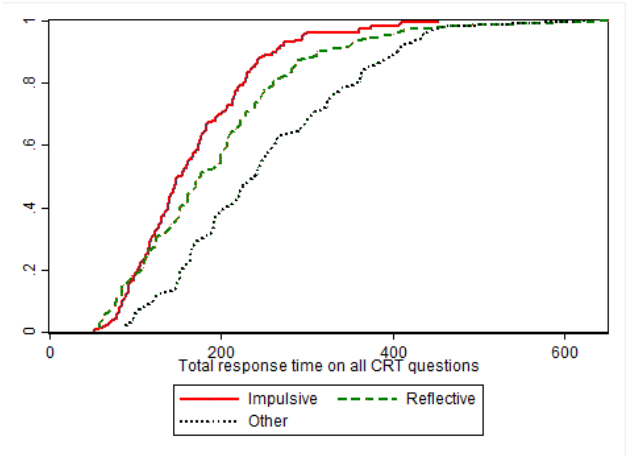


Figure A5. Laboratory experiment 2: Cumulative response times for different types of subjects

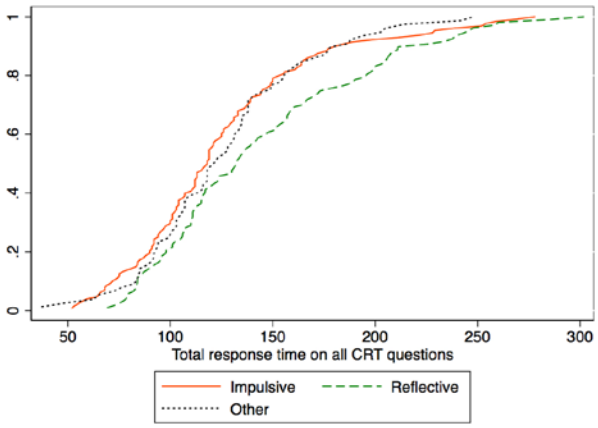
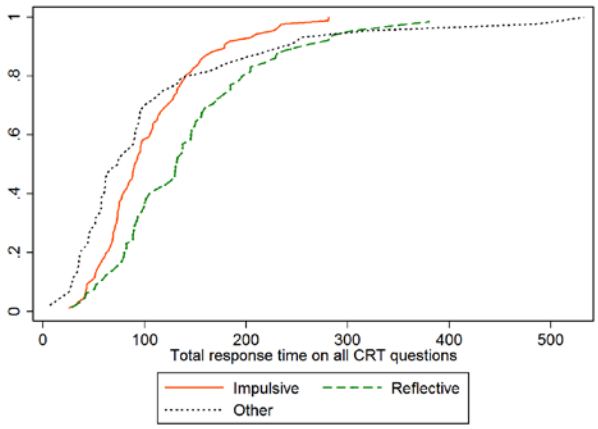


Figure A6. MTurk experiment: Cumulative response times for different types of subjects



The conclusion from Figures A4–A6 is in line with our working hypothesis that Impulsive subjects are faster than reflective subjects. The Mann-Whitney or Kolmogorov-Smirnov tests support this conjecture (Lab study 1: $p < 0.045$, Lab study 2: $p < 0.031$, MTurk: $p \leq 0.005$). As it occurs in our Internet study, Other subjects seem to be slower than both the Impulsive and Reflective ones in our Lab experiment 1. Recall that we use the CRT in Frederick (2005) in both experiments. The relationship between Other subjects and Impulsive or Reflective subjects does not seem to be clear-cut when using the CRT in Toplak et al. (2014); if anything, Other subjects seem to mimic the behavior of Impulsive subjects in our Lab experiment 2 and MTurk experiment (see also Section A3).

A5.3 Regression analysis

In this section, we present the econometric results. Following our previous discussion, Table A9 displays the result from a series of OLS regressions using CRT score as the dependent variable. We present the econometric results from the Lab experiment 1, 2 and MTurk in Panel A, B and C, respectively.. Our analysis in all three cases confirms that longer response times are associated with higher scores in the test. We also find that females perform worse than males in the test.[12]

Table A10 uses the classification of subjects (Impulsive, Reflective, Other) to estimate total response time. We confirm our previous findings that the Impulsive subjects respond quicker than the Reflective ones to the test. We also observe that Other subjects take longer than Reflective ones in our Lab experiment 1. They seem to mimic the behavior of Impulsive subjects and take less time than Reflective subjects to complete the CRT in Toplak (2014) (Lab experiment 2 & MTurk experiment). However, the difference between Reflective and Other subjects is only significant in the Lab experiment 2.

[12] In Lab experiment 1, subjects faced two 4x4 matrices and their task was to compute subtractions and then to pick the highest outcome from each matrix; we measure the response time in this task. Most of our subjects (60%) solved the 2 matrices correctly. We define the "Math Measure" as a dummy variable equal to 1 if the subject solved the 2 matrices correctly and her response time is in the first quartile and 0 otherwise. This dummy is robust to choosing other cut-offs for response times.

Table A9. CRT score, OLS regression

Panel A. Laboratory experiment 1

	(1)	(2)	(3)
Response time	0.0009*	0.0009*	0.001*
	[0.0004]	[0.0004]	[0.0006]
Female		-0.541***	-0.493***
		[0.109]	[0.113]
Age		-0.00656	-0.00690
		[0.0180]	[0.0280]
Math measure			0.313**
			[0.121]
Constant	0.650***	1.087***	0.939
	[0.125]	[0.413]	[0.629]
R-squared	0.007	0.080	0.100
Observations	311	311	311

Panel B. Laboratory experiment 2

	(1)	(2)	(3)
Response time	0.005***	0.005***	0.005***
	[0.001]	[0.001]	[0.001]
Female		-0.799***	-0.718***
		[0.126]	[0.133]
Age		0.001	-0.001
		[0.010]	[0.011]
Big5 Personality Scores	No	No	Yes
Constant	0.703***	1.200***	2.286***
	[0.188]	[0.300]	[0.744]
R-squared	0.045	0.156	0.169
Observations	312	312	312

Panel C. MTurk experiment

	(1)	(2)
Response time	0.002***	0.00252***
	[0.001]	[0.000682]
Female		-0.181
		[0.173]
Age		0.00196
		[0.00646]
Constant	0.945***	1.148***
	[0.118]	[0.343]
R-squared	0.064	0.064
Observations	195	195

Notes: Standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$ and * $p < 0.1$

Table A10. Total response time, OLS regression

Panel A. Laboratory Experiment 1

	(1)	(2)	(3)
Impulsive	-28.58**	-30.35**	-35.04**
	[13.75]	[13.99]	[13.85]
Other	59.73***	57.11***	54.05***
	[20.16]	[20.23]	[20.28]
Female		7.519	3.531
		[10.45]	[10.47]
Age		-1.292	-1.192
		[1.614]	[1.576]
Math measure			-31.03***
			[11.47]
Constant	196***	220.5***	233.2***
	[12.65]	[38.51]	[39.03]
R-squared	0.114	0.117	0.138
Observations	311	311	311

Panel B. Laboratory Experiment 2

	(1)	(2)	(3)
Impulsive	-19.26***	-20.51***	-20.63***
	[6.274]	[6.591]	[6.580]
Other	-17.86**	-18.95**	-18.87**
	[7.158]	[7.346]	[7.376]
Female		4.593	3.395
		[5.880]	[6.114]
Age		0.501	0.552
		[0.466]	[0.491]
Big5 Personality Scores	No	No	Yes
Constant	144.7***	131.3***	72.09**
	[4.631]	[11.93]	[32.84]
R-squared	0.034	0.039	0.067
Observations	312	312	312

Panel C. MTurk experiment

	(1)	(2)
Impulsive	-35.33*	-37.48*
	[18.93]	[19.03]
Other	-21.35	-21.43
	[21.04]	[21.21]
Female		31.05*
		[18.19]
Age		-0.0271
		[0.688]
Constant	148.8***	101.8***
	[15.58]	[38.66]
R-squared	0.018	0.034
Observations	195	195

Notes: Standard errors in brackets. *** p < 0.01, ** p < 0.05 and * p < 0.1

A6. Order effects

Thereafter, we investigate potential order effects. We used MTurk to collect additional data using the Toplak et al. (2014) version of the CRT used in Lab experiment 2. The main design difference compared to Lab experiment 2 was that the order of the questions of the CRT was randomized. Next, we show that response time differences between correct and incorrect answers are not affected by the order of the questions.

We start by visually comparing Figure A3 from the MTurk experiment (randomized order) with Figure A2 from Lab experiment 2 (fixed order). The patterns are strikingly similar. In both cases, there is pronounced difference in cumulative distributions for questions 1 and 4, with much smaller differences for questions 3 and 4. This indicates that the order does not explain why certain questions are associated with larger differences in response times between correct and incorrect answers. Similar conclusions can be drawn from Figure A7, which displays the cumulative distributions of correct and incorrect answers in the MTurk experiment, based on the order in which the questions were displayed (i.e. each panel contains aggregate data from all four questions). There is no clear order effect. If anything, the difference in distributions seems to be smallest for the question that was presented first. When comparing differences in response times between correct and incorrect answers using the Mann-Whitney test, we observe significant differences in all four cases (p-values <0.035).

Figure A7. MTurk experiment: Cumulative response times by order of presentation.

In Table A11 we present some key descriptive statistics broken down by question and order of presentation. The fractions of correct answers (Fraction Correct) show no clear pattern over time, indicating that the people did not perform better on a given question if it was presented later in the sequence. Median response times (Median RT) indicate that in general response time decreased over time, but the effect is quite modest and does not hold for CRT3. We also compare the difference in median response times between incorrect and correct answers, referred to as Diff RT Incorrect-Correct in the table. This variable shows no clear pattern over time, again indicating that there are no strong order effects.

Table A11. Response times and fraction of correct answers by order of presentation

Question	Measure	Order of presentation			
		1	2	3	4
CRT1	Fraction Correct	0.2	0.33	0.33	0.27
	Median RT	30.22	27.53	25.31	26.52
	Diff RT Incorrect-Correct	24.87	6.60	10.45	11.87
	Observations	41	52	43	59
CRT2	Fraction Correct	0.15	0.18	0.11	0.24
	Median RT	14.34	13.58	12.03	11.9
	Diff RT Incorrect-Correct	0.47	7.45	1.44	-3.75
	Observations	48	57	44	46
CRT3	Fraction Correct	0.36	0.4	0.35	0.35
	Median RT	25.55	25.9	25.23	29.38
	Diff RT Incorrect-Correct	4.60	14.12	-3.73	2.14
	Observations	59	50	46	40
CRT4	Fraction Correct	0.43	0.42	0.42	0.48
	Median RT	27.14	23.44	25.05	23.48
	Diff RT Incorrect-Correct	4.90	10.81	12.29	19.00
	Observations	47	36	62	50

Notes: Median RT is the median response time. Fraction correct is the fraction of correct answer. Diff RT Incorrect-Correct is the difference between the median response times between subjects providing incorrect and correct answers.

Finally, we report coefficient estimates from two random effects panel models in Table A12. The dependent variable is the top-coded response time for each question, and as independent variables we include a dummy variable whether the answers was correct (Correct), question dummies (CRT2-CRT4), interactions between the correct and question (Correct*CRT2- Correct*CRT4*)*, order of presentation (Order) and an interaction between correct answer and order (Correct*Order). In model 2 we also include variables capturing gender and age. Providing a correct answer is associated with a longer response time in CRT1, but as indicated by the interaction terms the effect is not present for CRT2 and CRT3 (which was visible already in Figure A3) and weaker for CRT4. Interestingly, the Order coefficient shows that response times decrease over time, this

suggests some learning effect of the task despite there is no apparent relationship between the questions in the CRT test. Nevertheless, the interaction term between correct answer and order is very small and insignificant. Again, this shows that the order of presentation appears to play no role for the difference in response times between correct and incorrect answers.

Table A12. Response time, correct answer, order, Random effects regression

	(1)	(2)
Correct	17.47**	18.32***
	[7.140]	[7.105]
CRT2	-16.27***	-16.13***
	[3.211]	[3.205]
CRT3	2.609	2.844
	[3.457]	[3.450]
CRT4	2.609	2.844
	[3.570]	[3.562]
Correct *CRT2	-15.75**	-15.98**
	[7.138]	[7.109]
Correct *CRT3	-19.99***	-20.86***
	[6.330]	[6.311]
Correct *CRT4	-12.85**	-13.87**
	[6.200]	[6.181]
Order	-2.319**	-2.339**
	[1.083]	[1.081]
Correct*Order	-0.0731	-0.0327
	[2.025]	[2.015]
Age		0.152
		[0.105]
Female		9.977***
		[2.798]
Constant	38.45***	16.68**
	[3.798]	[6.572]
Observations	780	780
Number of randomid	195	195

Notes: Standard errors in brackets. Correct indicate that the subject gave a correct answer. CRT2-CRT4 are question dummies. Correct*CRT2- Correct*CRT4 denote interactions between correct and question. Order takes values between 1 and 4 gives the position of the question. Correct*Order is an interaction term between Correct and Order. Age and Female are self-explanatory. *** p<0.01, ** p<0.05, * p<0.1

6 References

Beauducel, A., D. Leipmann, S. Horn, and B. Brocke (2010): Intelligence Structure Test, Hogrefe.

Brañas-Garza, P., Kujal, P., Lenkei, B. (2015). Cognitive Reflection Test: Whom, how, when. MPRA Paper No. 68049.

Cueva, C., Iturbe-Ormaetxe, I., Mata-Perez, E., Ponti, G., Yu, H., Zhukova, V. (2015) Cognitive (Ir)reflection: New Experimental Evidence. *Journal of Behavioral and Experimental Economics*, 64(5): 81-93.

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, 19(4): 25-42.

Toplak, M. E., West, R. F., Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, 20(2): 147-168.