STUDIA UNIVERSITATIS BABEȘ-BOLYAI OECONOMICA
VOLUME 63, ISSUE 3, 2018, pp. 56-67
DOI: 10.24193/subboec.2018.3.04

GENDER DIFFERENCES AND OTHER FINDINGS ON THE COGNITIVE REFLECTION TEST

Laura-Augustina AVRAM[*]
Babeș-Bolyai University, Romania

Abstract. Intelligence is the traditional element of interest when measuring the human cognitive abilities. However, intelligence is complex and researchers are constantly finding new angles of looking at it. One such angle is reflective reasoning. Sometimes individuals choose to override the intuitive answer and by engaging in further reflection they reach the correct answer. The cognitive reflection test (CRT) measures a person's ability to suppress their incorrect intuitive answer in favor of reflection that should then lead to the correct response. The test contains three short mathematically based problems, which measure, among others, cognitive ability, mathematical abilities and cognitive reflection. Using a sample of 195 students from a state university, one of the largest universities in Romania, we explore the extent to which a variety of phenomena and trends identified by previous findings on CRT show similar results on our sample.

JEL Classification: J16, D89, D91;

Keywords: cognitive reflection test, decision making, reflection.

1. Introduction

The Cognitive Reflection Test is a three question test that has been developed by Shane Frederick (2005). It is a mean to differentiate people with a higher cognitive ability from those with a lower one, the former ones differentiating from the later one's through their "ability to inhibit intuitive responses in favor of reflective and deliberative reasoning" (Travers et al., 2016). According to Juanchich et al., 2016, the CRT "is a powerful predictor of normative decision-making", but the existing literature is yet not sure about what the test actually measures. Traditionally this differentiation has been made with the help of tests like the IQ test, but such a

[*] Corresponding author: laura.avram@econ.ubbcluj.ro. Address: Faculty of Economics and Business Administration, Babeș-Bolyai University, Teodor Mihali, 58-60, Romania,
Tel. +40 264 418 652

test is usually rather lengthy and taking it can be quite time-consuming. The CRT is useful exactly for that reason: doing it takes a couple of minutes (unlike other tests that measure cognitive abilities that can take up to 3,5 hours) and its correlation to a person's IQ is very high (Frederick, 2005).

The purpose of the current study was to examine (a) the gender differences with regard the CRT in a representative adult sample, (b) the correlation between the results at the Romanian final high school exams *baccalaureate exam* and the CRT, (c) a few other previous findings from the literature.

What is such a differentiation between individuals with a higher cognitive ability and individuals with a lower cognitive ability useful for? It can be useful for many things, for example it gives researchers the possibility to separate individuals in different groups in accordance to their cognitive abilities (and without using a time-consuming test in the process). This can help them then see, for example, how people with a higher cognitive ability take decisions differently from people with a lower cognitive ability. According to research (Eysenck, 1979; Herrnstein and Murray, 1994; Jensen, 1980; Simonton, 1996; etc.), general intelligence assessed during childhood has consistently predicted behaviors that are maladaptive as well as adaptive. To this behaviors we can count, among others, "delinquent behavior, rate of learning, high-risk health behaviors", a better general health (Auld and Sidhu, 2005). A current issue when dealing with social problems has been that social sciences have neglected the influence of general intelligence when developing public policy and creating (effective) interventions, because they have concentrated mostly on specific groups instead of specific behaviors (Lubinski and Humphreys, 1997). Therefore, one of the uses of the CRT would be to easily help differentiate between people with higher and lower cognitive abilities as a means for creating better policies.

More recent studies on the CRT focused on, among others, expanding the CRT (see Toplak et al. 2014), explaining gender differences on the test (see Zhang et al. 2016), the prediction power of cognitive reflection in real-life decision situations while testing for personality and decision-making styles (see Juanchich et al. 2016), testing the dual process theory of reasoning of participants solving the CRT with the help of their mouse cursor movements (see Travers et al. 2016), intuitiveness – not something measured by the CRT (see Pennycook et el. 2015).

The three problems of the CRT are presented in Figure 1. When trying to answer the 3rd problem, if one was to respond intuitively, without deliberately reflecting on the answer, then they would likely answer 24 days. This choice of an answer can be explained by the dual-process theories, that have been mentioned in the literature quite often (Baron et al., 2015; Evans and Curtis-Holmes, 2005; Evans and Frankisch, 2009; Kahneman, 2011; Kahneman and Frederick, 2002; Sherman et al., 2014; Sloman, 1996). Different names have been used to differentiate between these two types processes; however, for simplicity, I will use Stanovich and West's terminology: "System 1" and "System 2" (Stanovich and West, 2000), even though in the latest literature many researchers, including Stanovich, prefer the terminology "type 1 processing" and "type 2 processing" (Sherman et al., 2014). The dual-process theory can be explained the following way: there are two types of cognitive operations, some are quick, associative, intuitive, heuristic, automatic,

unreflective, driven by affect, they are undemanding on our limited cognitive resources, and others that are slow, rule-based, reflective, and they require effortful thinking. Results at the CRT are explained in the following way: when the correct answer is given, then the System 2, the reflective one, manages to override the functioning of System 1, the intuitive one. However, when the wrong answer is given (to a problem that is not particularly difficult), then it is because System 2 didn't (most likely) *help* System 1 in the decision-making process (Baron et al., 2015; Campitelli and Gerrans, 2014; Kahneman and Frederick, 2002).

(1) A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? _____ cents

(2) If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? _____ minutes

(3) 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

Fig. 1: The Cognitive Reflection Test
Source: Frederick (2005)

Among Shane's most interesting results regarding the CRT was that the predictions of the prospect theory (Kahneman and Tversky, 1979) that "people will be more willing to take risks to avoid losses than to achieve gains; that respondents will switch from risk aversion to risk seeking when the valence of the gamble changes from positive to negative" worked very well when it came to the low CRT groups, but it was wrong when it came to the high CRT groups (Frederick, 2005).

2. Methodology

2.1. Participants and procedure

As far as we are aware, the CRT hasn't been applied to students at this faculty in a scholar environment and for scientific purposes. We developed a questionnaire that included a few additional questions next to the CRT and more groups of students filled in the questionnaire.

Our sample consisted of a total of 195 participants (140 females and 54 males), all students at one of the biggest universities in Romania. Since our population could clearly be divided into groups based on the characteristic study year, we used the stratified random sampling method based on the year of study of the students. Our sample was made of undergraduate students in the 1st year (44

students), 2nd year (64 students), 3rd year (77 students) and graduate students in the 2nd year (10 students). The participants were not paid for their participation. The students needed around 10 minutes to fill in the questions.

2.2. Results and Interpretation

The questions we tried to answer in this study, as well as our results can be found in the next sections.

Cognitive ability measured with the CRT

Firstly, we compared the results at the CRT of the students from our faculty to the ones of students from other universities and groups in which the test was applied, more precisely, the results mentioned in the original article by Frederick (Frederick, 2005). It can be seen in the table that almost all the other samples came from the US.

Table 1: Scores at the CRT, by Location

		Percentage scoring 0, 1, 2 or 3				
		"Low"			*"High"*	
Locations at which data were collected	*Mean CRT score*	*0*	*1*	*2*	*3*	*N=*
MIT	2.18	7%	16%	30%	48%	61
Princeton University	1.63	18%	27%	28%	26%	121
Boston fireworks display	1.53	24%	24%	26%	26%	195
Carnegie Mellon University	1.51	25%	25%	25%	25%	746
Harvard University	1.43	20%	37%	24%	20%	51
Univ. of Michigan: Ann Arbor	1.18	31%	33%	23%	14%	1267
Web-based studies	1.10	39%	25%	22%	13%	525
Bowling Green University	0.87	50%	25%	13%	12%	52
Univ. of Michigan: Dearborn	0.83	51%	22%	21%	6%	154
Michigan State University	0.79	49%	29%	16%	6%	118
Sample from one university in Romania	**0.77**	**51%**	**29%**	**12%**	**8%**	**195**
University of Toledo	0.57	64%	21%	10%	5%	138
Overall	1.2	**36%**	**24%**	**21%**	**17%**	**3623**

Source: compare to Frederick (2005)

As it can be seen in Table 1, the students from our sample who filled in the questionnaire had a mean CRT score of less than 1, which means that on average a person responded correctly to less than 1/3 of the questions. This result is very similar to the one of the sample from the Michigan State University (which had a mean CRT score of 0,78 vs. 0,79) (Frederick, 2005). However, the mean of our sample was lower than the overall mean of all the samples included in Table 1 and a lot lower than the sample with the highest results, a sample of students from MIT.

Consistency with Kahneman's Linda-question

As seen in Appendix A, we added the reasonably famous Linda-question (Kahneman, 2011) - see Figure 2 - as a continuation to the CRT. The Linda-question is one of a series of similar questions developed by Tversky and Kahneman during their long collaboration which led to the later one receiving the Nobel prize in 2002.

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. Rank the following statements according to their probability: using 1 for the most probable and 2 for the least probable:

(a) Linda is a bank teller.

(b) Linda is a bank teller and is active in the feminist movement.

Fig. 2: Kahneman and Tverky's Linda-question
Source: Tversky and Kahneman (1983)

The particular phenomenon that Tversky and Kahneman discovered in 1983 regarding the previously mentioned question is the following: “people tend to believe that a conjunction of events (e.g., Linda is a bank teller *and* is active in the feminist movement) is more likely to occur than one of the conjuncts (e.g., Linda is a bank teller)”. Researchers have been trying to come up with reasons to explain this particular phenomena and the literature mentions usually two reasons: the *misunderstanding* of the problem or the presence of a *reasoning bias* (Moro, 2009). However, those particular things don't concern us right now. The reason why we tested whether there is a correlation between the two is related to the following: if achieving a good result at the CRT implies having mathematical abilities, and properly solving questions like the Linda-question implies following the conjunction rule of the probability theory, which has something to do with mathematical abilities again, then common sense would dictate a correlation between the two.

Table 2. Correlations between the different questions of the CRT and the Linda question

		CRT question 1	*CRT question 2*	*CRT question 3*	*Linda question*
CRT question 1	*Pearson Correlation*	1.00	.43	.38	-.02
	Sig. (2 tailed)		.00	.00	.81
	N	152	152	152	145
CRT question 2	*Pearson Correlation*	.43	1.00	.21	.01
	Sig. (2 tailed)	.00		.01	.88
	N	152	152	152	145
CRT question 3	*Pearson Correlation*	.38	.21	1.00	-.08
	Sig. (2 tailed)	.00	.01		.35
	N	152	152	152	145
Linda question	*Pearson Correlation*	-.02	.01	-.08	1.00
	Sig. (2 tailed)	.81	.88	.35	
	N	145	145	145	145

Source: Designed by the authors based on own calculations

Looking at Table 2, which presents the correlations between the different questions of the CRT and the Linda-question, we can see the following: there is no significant correlation between the Linda-question and any of the three CRT questions. Even though none of the three CRT questions points out to the conjunction fallacy like the Linda-question, result were a little bit surprising and they differed from the usual findings of the literature (see, e.g., Brañas-Garza et al., 2015).

Gender trend

More researches concluded that on average men have a better result at the CRT in comparison to women (Campitelli and Gerrans, 2014; Frederick, 2005; Pennycook et al., 2015; Primi et al., 2016) and we wanted to test next whether this applies to our participants as well. However, it's worth mentioning that this trend can be seen not only in relation to the CRT; when it comes to math tests, in general, men get better results than women (Benbow and Stanley, 1980; Halpern, 2004; Hedges and Nowell, 1995; Hyde, Fennema and Lamon, 1990). Frederick noticed that in the case of the CRT the types of mistakes that women made were different from the type of mistakes than men made: while men made a larger variety of mistakes, women often made the intuitive type of mistakes, e.g., at the 2nd question of the CRT they answered 100 (Frederick, 2005).

Table 3. Gender differences at the CRT given to students from our sample consisting of students from one of the biggest universities in Romania

	Mean CRT score	*0*	*1*	*2*	*3*	*N=*	*Significance of group difference*
Female	0.65	54%	32%	9%	5%	140	P<0.001
Male	1.06	44%	22%	17%	17%	54	

Source: Designed by the authors based on own calculations

We compared the mean CRT score of males (=1.06) to the mean CRT of females (=0.65) – see Table 3 – to test to test whether in a bigger population males tend to have a better score at this test than women or if the difference in our sample was caused by chance. The results showed a significant difference between the male and female populations, which confirmed the results of previous studies that male perform better than women at this test. However, this test doesn't tell us why such a difference exists. According to Frederick (2005) the reason why men have better results at this test is their mathematical ability or interest. The author of the original article further explains that men are known to be better at solving math tests than women in general and the CRT is another test that can be included in this category.

Results at BAC and CRT – same strong correlation as between SATs and CRT?

Frederick wanted to see in his study how strong was the correlation between the CRT results of the participants in the study and their results at different cognitive measurement tests, among others: the WPT, NFC, ACT, SATs and SATs in the field of mathematics. There was a positive and significant correlation between these measurements (Frederick, 2005). Additionally, research has shown that people who perform well at the CRT tend to have good results at some other types of tests, namely numeracy tests, other general ability tests (Cokely and Kelley, 2009; Frederick, 2005; Liberali et al., 2011; Oechssler et al., 2009; Toplak et al., 2011). Additionally, the CRT also shows substantial correlation with, e.g., common biases in judgment and decisions (Campitelli and Labollita, 2010; Toplak and Stanovich, 2002), utilitarian moral judgments (Baron et al., 2015; Paxton et al, 2012), disbelief in God and the supernatural (Gervais and Norenzayan, 2012; Pennycook et al., 2012; Shenhav et al., 2011). We tested to find out whether there is a correlation between the Romanian version of the SATs, *examenul de Bacalaureat*, and the results at the CRT.

Table 4. Correlation between Cognitive Measures: Comparison between Frederick's CRT+SAT (compare to Frederick, 2005) and CRT+SATm and my sample's CRT+BAC and CRT+BACm

	CRT	*SAT*	*SATm*			*CRT*	*BAC*	*BACm*
CRT		.44	.46		*CRT*		.25 (p<0.05)	.16 (p>0.05)
SAT	.44		.77		*BAC*	.25 (p<0.05)		.67 (p<0.05)
SATm	.46	.77			*BACm*	.16 (p>0.05)	.67 (p<0.05)	

Source: Designed by the authors based on own calculations

The first part of Table 4 shows the correlations from the original article of Frederick (2005) between the CRT and the overall SATs results (=.44) and to the SATs in Mathematics (=.46), respectively. In both cases a moderate positive relationship was found between the two. The second part of the table shows the correlations between the results at the CRT and the overall results at the Romanian final high school exams (*examenul de bacalaureat*), respectively the results at mathematics. A weak positive relationship has been found between the CRT and the overall BAC (=.25). For the CRT and the BACm the correlation we found wasn't statistically significant. We can conclude that while there is a moderate positive relationship between the results of American students at their SATs and their cognitive abilities (which is what the CRT measures), the situation is different when it comes to Romanian students. There is less respectively no significant correlation between the results of the students at the CRT and their BAC results. This can lead further to the question of what does ultimately the *examenul de bacalaureat* intend to measure and what does it actually measure. We weren't able to find many studies about the Romanian baccalaureate. The ones that we found had interesting results, but their research questions were different from ours. For example, a study of Popa & Bochiş (2016) concluded after testing a sample (N=125) of Romanian students that their baccalaureate averages and their GPA tend to be consistent with their overall results during their studies. Marincas & David (2013) spoke about the reasons for the high rates of failure at the baccalaureate. What they mention and is of interest for the current study is the major changes that have been proposed and have taken place during the last years regarding the BAC. We consider that these constant yearly changes and unclear long-term plans of the education boards when it comes to the purpose and use of this exam in Romania are a reason for the lack of correlation that we found.

If you had a bad results at the CRT are you more likely to think it was easy?

Another interesting phenomena described by Frederick (2005) is the following: the persons who took the test and had a worse result were more likely to say that they found the test easy, compared to people who had better results at it, who claimed the test as harder. In order to test this hypotheses, our sample of students was asked to rate the difficulty of the test on a scale from 1 (very easy) to 5 (very difficult). The results of my sample were different than the one's of Frederick: the correlation that we found wasn't statistically significant, so people who did better in the test didn't necessarily find it harder and the other way around. This results were unexpected. A potential explanation for this difference is the different methodologies used. While Frederick (2005) asked the participants in the study how many percent of people they thought answered each of the questions correctly, I used the previously described question. What the author of the original article also noted was that leaving aside whether the participants of the study answered the question correctly or not, they all overestimated by a lot how many other people would give the correct answer to the questions.

3. Conclusions

Some of the results we found confirmed what previous studies concluded, while some of the results were different. The results at the CRT of our sample of students from a university in Romania were similar to the ones of other samples of students from other universities. Men did indeed have better results than women in our sample, which was what previous testing found as well. We did get to wonder and tried to come up with reasons, e.g., why the BAC results showed a much lower correlation compared to the SATs, but this was outside the scope of this study. Testing a bigger and more varied sample of Romanian pupils, adults, and students would be interested in order to see, e.g., how non-university students answer, and to then further study differences between populations, as well as compare the results to results of similar tests from Western Europe.

One of the limits of the paper is the small sample. In a further study we would like to include a bigger sample from Romanian universities and include a few additional questions in order to test whether they could be a possible replacement to the three original ones, which are starting to be known. Additionally, the sampling methodology could be improved in the sense that in the future we could include students from all years of study and in a higher number.

In the context of the discussions about the *baccalaureate exam* in Romania, not finding any significant correlation between the CRT results and the results at the *baccalaureate exam* was surprising. However, even though it was outside of the scope of the current article, an expansive analysis of what the exam actually measures is a possible further direction of study.

References

Auld, M. C., Sidhu, N. (2005) Schooling, cognitive ability and health, Health Economics, 14(10), 1019–1034

Baron, J., Scott, S., Fincher, K., Metz, E. (2015) Why does the Cognitive Reflection Test (sometimes) predict utilitarian moral judgment (and other things)?, Journal of Applied Research in Memory and Cognition, 4(3), 265–284

Benbow, C. P., Stanley, J. C. (1980) Sex Differences in Mathematical Ability: Fact or Artifact? Science, 210(12 December), 1978–1980

Brañas-Garza, P., Kujal, P., & Lenkei, B. (2015). Cognitive Reflection Test: Whom, how, when. *Working Paper*.

Campitelli, G., Gerrans, P. (2014) Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach, Memory & Cognition, 42(3), 434–47

Campitelli, G., Labollita, M. (2010) Correlations of cognitive reflection with judgments and choices, Judgment and Decision Making, Vol. 5, No. 3, June 2010, pp. 182–191

Cokely, E. T., Kelley, C. M. (2009) Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation, Judgment and Decision Making 4(1), 20–33.

Evans, J., Curtis-Holmes, J. (2005) Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning, Thinking & Reasoning, 11(4), 382–389

Evans, J., Frankisch, K. (2009) In two minds: Dual processes and beyond, Oxford University Press, Oxford

Eysenck, H. J. (1979) The structure and measurement of intelligence, Springer-Verlag, New York

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

Gervais, W. M., Norenzayan, A. (2012) Reminders of secular authority reduce believers' distrust of atheists, Psychological Science, 23(5), 483–91

Halpern, D. (2004) A cognitive-process taxonomy for sex differences in cognitive abilities, Current Directions in Psychological Science, 13(4), 135–139

Hedges, L., Nowell, A. (1995) Sex differences in mental test scores, variability, and numbers of high-scoring individuals, Science, 269(5220), 41–45

Herrnstein, R. J., Murray, C. (1994) The bell curve: intelligence and class structure an American life, Free Press.

Hyde, J. S., Fennema, E., Lamon, S. J. (1990) Gender Differences in Mathematics Performance : A Meta-Analysis, Psychological Bulletin, 107(2), 139–155

Jensen, A. (1980) Bias in Mental Testing, Free Press.

Juanchich, M., Dewberry, C., Sirota, M., & Narendran, S. (2016). Cognitive Reflection Predicts Real-Life Decision Outcomes, but Not Over and Above Personality and Decision-Making Styles. Journal of Behavioral Decision Making, 29, (52-59).

Kahneman, D. (2011) Thinking, Fast and Slow, Farrar, Straus and Giroux, New York

Kahneman, D., Frederick, S. (2002) Representativeness revisited: Attribute substitution in intuitive judgmen, Heuristics of Intuitive Judgment: Extensions and Applications

Kahneman, D., & Tversky, A. (1979) Prospect Theory: An Analysis of Decision under Risk, Econometrica: Journal of the Econometric Society, 47(3), 263–291

Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., Pardo, S. T. (2011) Individual Differences in Numeracy and Cognitive Reflection, with Implications for Biases and Fallacies in Probability Judgment, Journal of Behavioral Decision Making, 25(4):361-381

Lubinski, D., Humphreys, L. G. (1997) Incorporating general intelligence into epidemiology and the social sciences, Intelligence, 24(1), 159–201

Marincas, D., David, O. A. (2013) Psycho-social Predictors of Baccalaureate Performance of Romanian Students: what Makes the Difference between Students who Fail and those who Succeed?, 14(2), 209

Moro, R. (2009) On the nature of the conjunction fallacy, Synthese, 171(1), 1–24

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

Paxton, J. M., Ungar, L., Greene, J. D. (2012) Reflection and reasoning in moral judgment, Cognitive Science, 36(1), 163–177

Pennycook, G., Cheyne, J. A., Koehler, D. J., Fugelsang, J. (2015) Is the cognitive reflection test a measure of both reflection and intuition?, Behavior Research Methods, 48(1), 341–348

Pennycook, G., Cheyne, J. A., Seli, P., Koehler, D. J., Fugelsang, J. A. (2012) Analytic cognitive style predicts religious and paranormal belief, Cognition, 123(3), 335–346

Popa, C. A., Bochiş, L. (2016) The influence of baccalaureate average score on academic achievement of undergraduates, Studia Universitatis Babes-Bolyai, Psychologia-Paedagogia, 61 (1), 95-106

Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., Hamilton, J. (2016) The Development and Testing of a New Version of the Cognitive Reflection Test Applying Item Response Theory (IRT), Journal of Behavioral Decision Making, 29(5), 453–469

Shenhav, A., Rand, D. G., Greene, J. D. (2011) Divine intuition: Cognitive style influences belief in God, Journal of Experimental Psychology, 141(3), 423–8

Sherman, J., Gawronski, B., Trope, Y. (2014) In two minds, Dual-Process Theories of the Social Mind, ebook: The Guilford Press.

Simonton, D. K. (1996) Genius: The natural history of creativity, Behaviour Research and Therapy, 34(4), 395

Sloman, S. A. (1996) The empirical case for two systems of reasoning, Psychological Bulletin, 119(1), 3–22

Stanovich, K. E., West, R. F. (2000) Individual differences in reasoning: implications for the rationality debate? The Behavioral and Brain Sciences, 23(5), 645-665-726

Toplak, M. E., Stanovich, K. E. (2002) The domain specificity and generality of disjunctive reasoning: Searching for a generalizable critical thinking skill, Journal of Educational Psychology, 94(3), 197–209

Toplak, M. E., West, R. F., Stanovich, K. E. (2011) The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks, Memory & Cognition, 39(7), 1275–1289

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

Travers, E., Rolison, J. J., Feeney, A. (2016) The time course of conflict on the Cognitive Reflection Test, Cognition, 150, 109–118

Tversky, A., Kahneman, D. (1983), Extension versus intuitive reasoning: The conjunction fallacy in probability judgment, Psychological Review. 90 (4): 293–315

Zhang, D. C., Highhouse, S., & Rada, T. B. (2016). Explaining sex differences on the Cognitive Reflection Test. *Personality and Individual Differences*, *101*, 425–427.