



Munich Personal RePEc Archive

## **Atheists Score Higher on Cognitive Reflection Tests**

Da Silva, Sergio and Matsushita, Raul and Seifert, Guilherme and De Carvalho, Mateus

2015

Online at <https://mpra.ub.uni-muenchen.de/68451/>  
MPRA Paper No. 68451, posted 20 Dec 2015 23:58 UTC

# Atheists score higher on cognitive reflection tests

Sergio Da Silva<sup>a\*</sup>, Raul Matsushita<sup>b</sup>, Guilherme Seifert<sup>a</sup>, Mateus De Carvalho<sup>c</sup>

<sup>a</sup> *Department of Economics, Federal University of Santa Catarina, Florianopolis, S.C., 88049-970, Brazil*

<sup>b</sup> *Department of Statistics, University of Brasilia, Brasilia, D.F., 70910-900, Brazil*

<sup>c</sup> *Department of Economics, University of Birmingham, Birmingham, B15 2TT, United Kingdom*

\* Corresponding author. Tel.: +55 48 3721 9901; fax: +55 48 3721 9901. Email address: professorsergiodasilva@gmail.com (S. Da Silva).

## Abstract

We administrate the cognitive reflection test devised by Frederick to a sample of 483 undergraduates and discriminate the sample to consider selected demographic characteristics. For the sake of robustness, we take two extra versions that present cues for removing the automatic (but wrong) answers suggested by the test. We find a participant's gender and religious attitude to matter for the test performance on the three versions. Males score significantly higher than females, and so do atheists of either gender. While the former result replicates a previous finding that is now reasonably well established, the latter is new. The fact that atheists score higher agrees with the literature showing that belief is an automatic manifestation of the mind and its default mode. Disbelieving seems to require deliberative cognitive ability. Such results are verified by an extra sample of 81 participants using Google Docs questionnaires via the Internet.

**Subject areas:** cognitive psychology, behavioral economics

**Keywords:** cognitive reflection; religiosity; atheism; cognitive psychology

## 1. Introduction

These days, most cognitive psychologists favor a dual-system approach to higher cognition processes [1-2]. Behavioral economics, in turn, can be organized around such a two-system model of the mind [3]. The two systems compete for control of our inferences and actions. Intuitive decisions require little reflection and use “System 1.” By scanning associative memory, we usually base our actions in experiences that worked well in the past. However, whenever we make decisions by constructing mental models or simulations of future possibilities, we engage “System 2.” System 1 is evolutionarily older and we share it with other animals. Actually, it is made up of a set of autonomous subsystems. These subsystems include input modules related to specific-domain knowledge. In turn, System 2 is evolutionarily more recent and distinctively human. It allows abstract reasoning and way of thinking using hypotheses. System 2 correlates with general intelligence measures, such as IQ. However, it is bounded in terms of working memory capability. Decisions based on System 1 are fast and automatic, while those based on System 2 are slow and deliberative. Automatic decisions work well most of the time, but they also lead to predictable biases and heuristics (which are simple procedures that help find adequate, though often imperfect, answers to difficult questions) [3]. The early evolution of System 1 suggests its logic is related to an “evolutionary rationality,” while the logic of the lately evolved System 2 refers to the individual rationality [4]. Some decisions based on System 1 that seem irrational from an individual perspective may have an evolutionary logic from the genome perspective. Here, specific-domain reasoning makes more evolutionary sense than general-domain reasoning (System 2), and the System 1 mind is modular even when engaged in general-domain reasoning. The late emergence of System 2 occurred with little direct genetic control, and this allowed individuals to alternatively pursue their own objectives and not exclusively those of their genes [5]. This posits a potential conflict between individuals and genes that can possibly be the basis of the human psychology of self-deception [6].

The evidence for two minds come from experimental psychology, neuroscience, and possibly archaeology, too. In experimental tasks, participants who are asked to logically answer according to previous protocol mingle their prior beliefs with the logical deduction of the conclusion. When experimenters choose syllogisms where the premises conflict with the conclusion, a belief-bias effect emerges because participants tend to pick a conclusion if it is believable by itself. System 1 is biased to believe and confirm [3], while System 2 struggles to overcome (or otherwise leniently endorses) the beliefs suggested by System 1.

Another experimental evidence for dual-process reasoning is given by the “matching bias,” which occurs when people select cards in “Wason selection tasks.” Wason [7] wondered whether ordinary people could be Popperian. Are they equipped to test hypotheses by searching for evidence that potentially falsifies them? Participants in Wason selection tasks commonly suggest the answer is “no.” People tend to pick matching cards according to their lexical content, regardless of their logical content. For this reason, they cannot find violations of descriptive or causal rules. The matching bias is a System 1 heuristic. Not surprisingly from an evolutionary standpoint, performance in the Wason selection task improves when social contracts are involved [8].

There is neuroscientific evidence that syllogistic reasoning occurs in two different brain regions. Using fMRI (functional magnetic resonance imaging) techniques in tasks that elicit the belief-bias effect [9], whenever participants make the logically correct decision there is activation of the right inferior prefrontal cortex. Incorrect answers biased by belief activate the ventral medial prefrontal cortex, which is part of the mammalian brain and implicated in both intuitive answers

and System 1 heuristics. Moreover, a PET (positron emission tomography) study finds the access to deductive logic involves the right ventral medial prefrontal cortex, an area linked to emotions [10].

After training participants to overcome the matching bias, those who succeed activated a distinct region of the brain. Under the matching bias, all the participants activated a posterior area, but those who overcome the bias activated the left prefrontal [11]. Moreover, fMRI shows neural differentiation for reasoning with abstract material and reasoning using semantic-rich problems [12]. Abstract problems activate the parietal lobe, while content-based reasoning recruits the left hemisphere temporal lobe.

The archaeological record also suggests humans developed System 2 for the general-purpose reasoning after the autonomous subsystems (System 1) had already evolved. About 50 thousand years ago, there was an abrupt emergence of representational art, religious imagery and fast adaptations in the design of tools and artifacts [13].

A simple test devised by behavioral economist Shane Frederick [14] assesses how individuals differ in their cognitive ability. His cognitive reflection test (CRT) measures the relative powers of Systems 1 and 2, that is, the individual ability to think quickly with little conscious deliberation (using System 1) and to think in a slower and more reflective way, using System 2. High scores on the CRT track how strong System 2 is for one individual. Not surprisingly, the CRT is correlated with other measures of cognitive ability, including the Wonderlic Personnel Test, the Need For Cognition scale, and self-reported SAT and ACT scores. However, the CRT is not just another IQ test. Because it measures the ability or disposition to resist reporting the response that first comes to mind (cognitive reflection), it is arguably better than the other measures in terms of predicting decision making [14, 15]. It may predict, for example, risk and intertemporal preferences [14]. Moreover, high intelligence does not make people immune to biases [15]. Intelligent people may still show superficial or lazy thinking, and this is related to a flaw in their reflective mind. This fact explains that university students usually perform poorly on the CRT.

Frederick's first use of the CRT [14] found that males score significantly higher than females. Males are more likely to reflect on their answers and less inclined to go with their intuitive responses. Can Frederick's result be used in waging a war between the sexes? Definitely not. Being automatic brings advantages in most daily tasks, although it also exposes us to predictable bias and heuristics in some situations. Interestingly, Frederick also found that "being smart makes women patient and makes men take more risks." Although Frederick himself found this result "unanticipated and suggesting no obvious explanation," the result makes sense if one thinks in evolutionary terms because picky females and risk-seeking males may have an evolutionary advantage. In this study, we replicate Frederick's finding. Indeed, in our sample males score significantly higher on the test than the sample females do.

Anything that makes it easier for associative memory to run smoothly will bias beliefs [3]. For System 1, understanding is first believing [3]. It is then up to System 2 to adopt or not the suggestions of System 1. Believing is our default mode of thinking. Moreover, there is an inborn readiness for us to separate physical and intentional causality [16, 17]. System 1 infers goals and desires where none exists, making us animists and creationists by default [18]. These two concepts of causality were shaped separately by evolutionary forces, and they may explain the ubiquity of religious beliefs [18]. They are rooted in System 1. System 1 sees the world of objects as separate from the world of minds, thus making it possible for us to envision soulless bodies and bodiless souls [18]. One consequence is to believe that an immaterial divinity is the ultimate cause of the

physical world; another is to believe in the existence of an afterlife. These two characteristics are common to all religions [18]. Because being a nonbeliever requires more cognitive effort, it is likely that atheists may be endowed with a more vigilant System 2. This inference is confirmed in our study, where individuals who declare themselves atheists outperform believers on the CRT.

Other demographic characteristics we consider in our study are: age, handedness, parenthood, and current emotional state – but these proved to not be related to CRT performance in our sample. For a discussion of why these characteristics might matter, see reference [19].

The rest of this report is organized as follows: Section 2 presents the data collected through questionnaires and the methodology employed. Section 3 shows the results, and Section 4 concludes the study.

## 2. Materials and methods

Initially, a total of 483 undergraduates from the Federal University of Santa Catarina, Brazil took part in the study. The students came from five courses in social sciences and social work, two courses in mathematics and physics, and one course in education. The experimenters (GS and MDC) administered the tests to the participants in classrooms with the consent of teachers. After this round of tests was completed, the experimenters sent Google Docs questionnaires via the Internet (through email messages) to an additional sample of 81 distinct undergraduates, also from the Federal University of Santa Catarina. This aimed at verifying the results found with the previous sample. Therefore, in the end, 564 people participated. The questionnaire collected the aforementioned demographic variables (gender, age, handedness, parenthood, religiousness, and current emotional state) and then applied the CRT. Prior to the test, participants were asked to respond its three questions (below) in less than 30 seconds. At the end of the test, one respondent was asked whether he or she had really spent less than 30 seconds to answer the three questions. There was also a question inquiring whether one respondent already knew one or all of the three questions. Although we stated we would remove from the sample those who declared to have taken more than 30 seconds or to have known at least one of the questions, no participants reported either.

The questionnaire asked for each participant's gender (male or female), age, handedness (right-handed or left-handed), whether he or she had children and lived with them, whether he or she believed in God, and his or her current emotional state. As for the latter, the questionnaire presented a continuous affect scale, ranging from “very anxious” and “moderately anxious” to “emotionless,” “moderately excited” and “very excited” [19].

The three simple questions of the cognitive reflection test were conceived to elicit automatic responses that are wrong [14]. We call Frederick's CRT as Version A.

CRT: Version A

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

The correct answers are 5, 5 and 47. As for question 1, the intuitive answer that springs quickly to mind is “10 cents.” However, this is wrong. The difference between \$1.00 and 10 cents is only 90 cents, not \$1.00 as the problem stipulates. Let  $x$  be the ball cost; then  $(1.00 + x) + x = 1.10 \rightarrow x = 0.05$ . The automatic answer to question 2 is 100. This is wrong because taking the first sentence and multiplying the number of machines (5) by 20, it would take the same 5 minutes to make 100 widgets. In question 3, the intuitive answer is 24. This is also wrong. Let the exponential  $d^n$  represent the function where  $n$  is the number of days it takes the patch to cover the entire lake. Thus, half is always  $n - 1$ , and then the patch will cover half the lake in  $48 - 1 = 47$  days. When we asked the participants to respond the three questions above in less than 30 seconds we aimed to make sure an automatic choice was given.

Frederick observes that the three questions on the CRT are easy because their solutions are easily understood when explained [14]. However, reaching one correct answer requires the suppression of an erroneous answer that springs impulsively to mind. In his own study, Frederick reports that those who answered “10 cents” to the first question above also estimated that 92 percent of people would correctly solve it, whereas those who answered “5 cents” estimated that 62 percent would. In other words, those who performed poorly were also those overconfident.

Some participants received two alternative versions of the cognitive reflection test that minimally depart from Frederick’s version by making answers explicit. We call such variations as Version B and Version C. Our aim in introducing the two new versions of the CRT was to investigate whether System 2 is capable of suppressing the automatic drive to provide the wrong answers suggested by System 1.

CRT: Version B

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?  
 1 cent    5 cents    10 cents    15 cents    20 cents
2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?  
 1 minute    5 minutes    10 minutes    50 minutes    100 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?  
 10 days    12 days    24 days    40 days    47 days

CRT: Version C

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?  
 1 cent    5 cents    15 cents    20 cents    none of the above
2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?  
 1 minute    5 minutes    10 minutes    50 minutes    none of the above
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?  
 10 days    12 days    40 days    47 days    none of the above

Versions B and C aim at making System 2 more vigilant so that it can inhibit, if not suppress, the automatic and wrong answers made by System 1. By making the alternatives explicit in version B, System 2 is forced to make a comparison between them and thus to possibly become more engaged and capable of overriding the wrong answer System 1 is prone to suggest. In version C, the intuitive, wrong answers (10, 100, 24) are replaced with “none of the above.” We first hypothesized that this would push for further improvement of cognitive reflection, although the results shown in Section 3 will find Version C not to convey extra information. The idea of considering Versions B and C was inspired by reference [20].

### 3. Results

Table 1 shows the number of scores 0, 1, 2 or 3 on the CRT for the 483 participants considering the three versions of the test A, B, and C. In Version A (Frederick’s) most participants scored 0 out of 3 (53 percent) and only a minority scored 3 out of 3 (8.6 percent). This replicates his results [14] and confirms most participants act automatically and make mistakes using System 1.

Table 1. CRT scores, by version

	<i>Percentage scoring 0, 1, 2 or 3</i>			
<i>Version</i>	0 (low)	1	2	3 (high)
A	53.37	23.93	14.11	8.59
B	39.52	35.33	13.17	11.98
C	33.99	28.10	20.26	17.65

The results for Versions B and C show the cues improved cognitive reflection. The high CRT group increases across the three versions from nearly 9 to 12 to 18 percent. The low group reduces from 53 to 40 to 34 percent. However, the middle groups show no consistency. This fact prompted us to investigate whether Version C really conveys useful information in addition to that already existent in Version B.

Thus, to assess the level of statistical significance of version differences, we compute the  $p$ -values from chi-square and likelihood-ratio chi-square tests, which are appropriate for discrete responses. Considering the three versions A, B, and C, the responses do depend on the CRT version (chi-square  $p$ -value = 0.0035; likelihood-ratio chi-square  $p$ -value = 0.0040). However, if we compute the  $p$ -values considering only Versions B and C, we find no statistically significant differences (chi-square  $p$ -value = 0.1050; likelihood-ratio chi-square  $p$ -value = 0.1043). For this reason, we merged the sample responses coming from Versions B and C and recalculated the percentage scoring. The updated results are displayed in Table 2. Now, Version A and the merged Version BC present statistically significant differences (chi-square  $p$ -value = 0.0050; likelihood-ratio chi-square  $p$ -value = 0.0048).

Table 2. CRT scores: Version A and the merged Version BC

	<i>Percentage scoring 0, 1, 2 or 3</i>			
<i>Version</i>	0 (low)	1	2	3 (high)
A	53.37	23.93	14.11	8.59
BC	36.88	31.88	16.56	14.69

Table 3 shows males scored significantly higher than females on the CRT (considering Version A). The gender differences are statistically significant (chi-square  $p$ -value  $< 0.0001$ ; likelihood-ratio chi-square  $p$ -value  $< 0.0001$ ). Thus, the responses given to the CRT do depend on a participant's gender. This replicates Frederick's version. The CRT questions have mathematical content, and males generally score higher than females on math. However, Frederick [14] shows females' mistakes tend to be of the intuitive variety because males score higher than females on the CRT, even controlling for SAT math scores. Moreover, the fact that CRT scores are related to gender does not depend in turn on the CRT version considered, either A or BC: chi-square  $p$ -value = 0.7185; likelihood-ratio chi-square  $p$ -value = 0.7185.

Table 3. CRT scores, by gender: Version A

	<i>Percentage scoring 0, 1, 2 or 3</i>			
<i>Gender</i>	0 (low)	1	2	3 (high)
Male	32.04	30.10	19.90	17.96
Female	50.18	28.52	12.64	8.66

Table 4 shows the main novelty of this work, namely that atheists score significantly higher than believers on the CRT. The differences are statistically significant (chi-square  $p$ -value = 0.0227; likelihood-ratio chi-square  $p$ -value = 0.0240). Thus, the responses given to the CRT do depend on a participant's religious preferences. The fact that atheists score higher suggests believers are the group with stronger System 1.

Table 4. CRT scores, by religiosity: Version A

	<i>Percentage scoring 0, 1, 2 or 3</i>			
<i>Group</i>	0 (low)	1	2	3 (high)
Believers	46.31	28.69	14.49	10.51
Atheists	32.52	30.08	19.51	17.89

As for the rest of the demographic variables, they show no statistical significance (age: chi-square  $p$ -value = 0.6432; handedness: chi-square  $p$ -value = 0.8234; parenthood: chi-square  $p$ -value = 0.9787; current emotional state: chi-square  $p$ -value = 0.1075).

Considering the relationship between CRT scores and gender, we adjust the linear regression (Table 5):

$$\text{CRT scores} = 0.902 + 0.431g - 0.297q, \quad (1)$$

where  $g$  is a dummy for the gender (male = 1; female = 0), and  $q$  is a dummy for the type of questionnaire (Version A = 1; Version BC = 0). Thus, on average males score 43 percent more than females and Version BC makes participants score nearly 30 percent more. Therefore, the cues help improve performance on the CRT. A low  $R$ -square, however, indicates these results hold only for averages, and cannot be used to predict individual behavior.

Using Akaike information criterion for variable selection, religiosity appears as significant, but not the other demographic variables. The adjusted linear regression is (Table 6):

$$\text{CRT scores} = 1.125 + 0.401g - 0.295q - 0.290r, \quad (2)$$



where  $r$  is a dummy for religious attitude (believer = 1; atheist = 0). On average, believers score 29 percent less than atheists. Regarding the other significant variables ( $g$  and  $q$ ), these do not change a great deal in equation (2) relative to their estimation using equation (1).

Table 5. CRT scores, gender, and type of questionnaire

Analysis of variance					
Source	DF	Sum of squares	Mean square	$F$ -value	$Pr. > F$
Model	2	32.4194	16.2097	15.80	< 0.0001
Error	480	492.4791	1.0260		
Corrected total	482	524.8985			
	Root MSE	1.0129		$R$ -square	0.0618
	Dependent mean	0.9855		Adj. $R$ -sq.	0.0579
	Coeff. var.	102.7812			
Parameter estimates					
Variable	DF	Parameter estimate	Standard error	$t$ -value	$Pr. >  t $
Intercept	1	0.90199	0.06979	12.92	< 0.0001
$g = 1$	1	0.43117	0.09324	4.62	< 0.0001
$q = 1$	1	-0.29743	0.09752	-3.05	0.0024

Table 6. CRT scores, gender, type of questionnaire, and religious attitude

Analysis of variance					
Source	DF	Sum of squares	Mean square	$F$ -value	$Pr. > F$
Model	3	40.3352	13.4451	13.29	< 0.0001
Error	479	484.5633	1.0116		
Corrected total	482	524.8986			
	Root MSE	1.0058		$R$ -square	0.0768
	Dependent mean	0.9855		Adj. $R$ -sq.	0.0711
	Coeff. var.	102.0582			
Parameter estimates					
Variable	DF	Parameter estimate	Standard error	$t$ -value	$Pr. >  t $
Intercept	1	1.1251	0.1057	10.65	< 0.0001
$g = 1$	1	0.4015	0.0932	4.31	< 0.0001
$q = 1$	1	-0.2952	0.0968	-3.05	0.0024
$r = 1$	1	-0.2899	0.1036	-2.80	0.0054

The results using the CRT questionnaires via the Internet do confirm the role of gender and religiosity. Table 7 shows the scores using Version A for the smaller sample of 81 distinct participants. Here again, males score significantly higher than females (chi-square  $p$ -value = 0.0002; likelihood-ratio chi-square  $p$ -value = 0.0001).

Table 7. Version A CRT scores, by gender: confirmatory sample

	<i>Percentage scoring 0, 1, 2 or 3</i>			
<i>Gender</i>	0 (low)	1	2	3 (high)
Male	15.79	23.68	26.32	34.21
Female	53.49	27.91	13.95	4.65

Table 8 shows the scores for religious attitude using Version A. As with the larger sample, again atheists score higher using this smaller sample (chi-square  $p$ -value = 0.0184; likelihood-ratio chi-square  $p$ -value = 0.0275).

Table 8. Version A CRT scores, by religiosity: confirmatory sample

<i>Group</i>	<i>Percentage scoring 0, 1, 2 or 3</i>			
	0 (low)	1	2	3 (high)
Believers	40.32	25.81	22.58	11.29
Atheists	21.05	26.32	10.52	42.11

#### 4. Conclusion

Daniel Gilbert [21] once observed: “Findings from a multitude of research literatures converge on a single point: People are credulous creatures who find it very easy to believe and very difficult to doubt. In fact, believing is so easy, and perhaps so inevitable, that it may be more like involuntary comprehension than it is like rational assessment.” And Daniel Kahneman [3] added in a follow up: “As the psychologist Daniel Gilbert observed, disbelieving is hard work and System 2 is easily tired.” Our findings in this study makes experimental sense of this point of view.

Indeed, in our study, individuals who declare themselves atheists outperform believers on cognitive reflection tests. Believers are more automatic. Our study also confirms the role of gender on CRT performance. Males score higher, as previously shown [14]. We also consider other demographic characteristics – age, handedness, parenthood, and current emotional state – only to find these unrelated to CRT performance.

Our results are robust to slight changes in the CRT. We consider two other versions of the cognitive reflection test that minimally depart from the original formulation by making alternative answers explicit. Our aim in introducing these is to investigate whether System 2 is capable of suppressing the automatic drive to provide the wrong answers suggested by System 1. We find that although responses become less automatic, System 2 is ultimately incapable of dissuading the wrong cues. Some tests were administered in classrooms to a sample of 483 undergraduates and via the Internet to a confirmatory, smaller sample of 81 distinct participants.

## References

- [1] Evans JSBT (2003) In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences* 7 (10), 454-459.
- [2] Evans JSBT (2008) Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology* 59, 255-278.
- [3] Kahneman D (2011) *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
- [4] Stanovich KE, West RF (2000) Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences* 23 (5), 645-665.
- [5] Stanovich KE (2004) *The Robot's Rebellion: Finding Meaning in the Age of Darwin*. Chicago: Chicago University Press.
- [6] Trivers R (2011) *The Folly of Fools: The Logic of Deceit and Self-Deception in Human Life*. New York: Basic Books.
- [7] Wason P (1966) Reasoning. In *New Horizons in Psychology*, ed. Foss BM. Harmondsworth: Penguin.
- [8] Cosmides L, Tooby J (1992) Cognitive adaptations for social exchange. In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, ed. Barkow JH, Cosmides L, Tooby J, pp. 163-228. Oxford: Oxford University Press.
- [9] Goel V, Dolan RJ (2003) Explaining modulation of reasoning by belief. *Cognition* 87, B11–B22.
- [10] Houde O, Zago L, Crivello F, Moutier S, Pineau A, Mazoyer B, Tzourio-Mazoyer N (2001) Access to deductive logic depends upon a right ventromedial prefrontal area devoted to emotion and feeling: Evidence from a training paradigm. *Neuroimage* 14, 1486-1492.
- [11] Houde O, Zago L, Mellet E, Moutier S, Pineau A, Mazoyer B, Tzourio-Mazoyer N (2000) Shifting from the perceptual brain to the logical brain: The neural impact of cognitive inhibition training. *Journal of Cognitive Neuroscience* 12, 721-728.
- [12] Goel V, Buchel C, Frith C, Dolan RJ (2000) Dissociation of mechanisms underlying syllogistic reasoning. *Neuroimage* 12, 504-514.
- [13] Mithen S (2002) Human evolution and the cognitive basis of science. In *The Cognitive Basis of Science*, ed. Carruthers P, Stich S, Siegel M, pp. 23-40. Cambridge: Cambridge University Press.

- [14] Frederick S (2005) Cognitive reflection and decision making, *Journal of Economic Perspectives* 19 (4), 25-42.
- [15] Toplak ME, West RF, Stanovich KE (2011) The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition* 39 (7), 1275-1289.
- [16] Heider F, Simmel M (1944) An experimental study of apparent behavior. *The American Journal of Psychology* 57 (2), 243-259.
- [17] Leslie AM, Keeble S (1987) Do six-month-old infants perceive causality? *Cognition* 25 (3), 265-288.
- [18] Bloom P (2005) Is God an accident? *The Atlantic* 12, 105-112.
- [19] Da Silva S, Baldo D, Matsushita R (2013) Biological correlates of the Allais paradox. *Applied Economics* 45 (5), 555-568.
- [20] Kirkpatrick LA, Epstein S (1992) Cognitive-experiential self-theory and subjective probability: Further evidence for two conceptual systems. *Journal of Personality and Social Psychology* 63 (4), 534-544.
- [21] Gilbert DT (1991) How mental systems believe. *American Psychologist* 46 (2), 107-119.