Mindware Instantiation as a Predictor of Logical Intuitions in Cognitive Reflection Test

Roman Burič [a] , Ľubica Konrádová [b]

[a] Institute of Experimental Psychology, Centre of Social and Psychological Sciences, Slovak Academy of Sciences, Dúbravská cesta 9, 841 04 Bratislava, Slovak Republic
[b] Department of Psychology, Faculty of Arts, Comenius University in Bratislava, Gondova 2, 811 02 Bratislava, Slovak Republic

Following the growing body of evidence suggesting that substantial individual differences in reasoning exist already at the early stages of the reasoning process and that reasoners might be able to produce logical intuitions, the model of mindware automatization posits that the mindware acquired to the extent that it is fully automatized can cue the logically correct type 1 response. We asked 908 participants to solve the Cognitive Reflection Test presented under the two-response paradigm, to obtain both intuitive and analytical response, while measuring mindware instantiation and conflict detection efficiency. These variables explained approximately 10% of the variance in the accuracy of intuitive answers. We also observed that in more than half of the cases when the response was correct in the final response stage, it was already correct at the initial response stage. These results are in line with the theoretical model of mindware automatization to a large extent and raise a question about the main attribute of the Cognitive Reflection Test – the ability to measure the tendency to override a misleading intuitive answer.

Key words: cognitive reflection, mindware, conflict detection, logical intuition, two-response paradigm

Introduction

The last five decades of reasoning and decision-making research have provided ample evidence of the misleading nature of intuitive judgments, especially in situations where the correct solution requires the involvement of logical-mathematical and probabilistic principles (e.g., Kahneman, 2011). A good example is the well-known bat & ball problem from the Fredericks (2005) Cognitive Reflection Test (CRT): 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? As the author famously showed, the majority of participants facing this problem tend to rely on their intuitive judgment which usually cues an incorrect answer – $ 0.10.

These fast heuristic-based type 1 processes have evolved through the evolution of humankind as processing mechanisms of low computational expense, which provided us with an evolutionary advantage in the form of the ability to make quick, energy-efficient and, in the majority of cases, the right decisions (Kahneman, 2011; Stanovich, 2018), given favorable circumstances (Kahneman & Klein, 2009). However, if the situation is rather hostile, it can give us misleading intuitive cues, which results in a wide spectrum of cognitive biases (Kahneman, 2011). From a perspective of the default-interventionist (DI) model of dual-processes, in such situations, one needs to detect the conflict between the heuristic intuition and logical structure of the situation and afterward engage in more demanding type 2 processes to come up with the normatively correct answer (Evans & Stanovich, 2013; Evans, 2007).

Although the DI model admits, that the type 1 processing can cue correct responses and, on the other hand, type 2 processing does not guarantee it (Evans, 2007), the general assumption of the model is that the correct responding requires correction of the intuitive, error-prone type 1 processes by slower, more demanding, analytic type 2 processes (Kahneman, 2011; Kahneman & Frederick, 2005). The nature of type 2 processes is therefore corrective. Another core assumption of the DI model is that reasoners fail to engage in the type 2 processing because they lack detection of the conflict between their heuristic intuition and the logical structure of the task, which is considered later, after the engagement of type 2 processes (De Neys, 2018). Both of these premises were recently questioned by a growing body of evidence (see De Neys, 2012, 2018 for reviews) showing that participants are in fact able to detect the conflict, even when providing biased response. Results of these studies show that they are less confident and need more time to give a response when facing a conflict (i.e., the intuition cues an answer conflicting with the logical structure of the task), compared to no-conflict problems (i.e., the intuitive response is in line with the logical structure). What is more, the evidence suggests that the conflict detection ability is unaffected even after the possible type 2 engagement is restrained by multiple interventions. The corrective nature of the type 2 processes is argued against by the studies using a two-response paradigm (Bago & De Neys, 2017, 2019b; Burič & Šrol, 2020; Thompson et al., 2011; Thompson & Johnson, 2014). In this paradigm, participants solve every task twice. They provide their initial response first, while their analytical thinking is knocked out by restrains such as a secondary cognitive task to burden their working memory or time limit, which assures that the response is quick and there is too little time to engage in type 2 processing. Then they provide their final answer, with no time limit or burdened working memory. One of the main findings of these studies is that if participants provided the correct final response, in the majority of cases the response was correct already at the initial response stage. The correct final answer given after the correction of the initial response is in these studies rather rare, which runs directly against the "corrective view" of the DI model.

These findings led researchers to study the possibility of normatively correct intuitions and even to formulate alternative models of dual-process, the so-called hybrid models (Bago & De Neys, 2017; Handley & Trippas, 2015; Pennycook et al., 2015) allowing for such a possibility. Even though this line

of research is relatively young, it provides us with some insights contributing to a better understanding of the logical intuitions. First, the above-mentioned conflict detection studies show that the detection also takes place largely intuitively (De Neys, 2012, 2018; Pennycook, 2018; Pennycook et al., 2015). In a recent study by Šrol and De Neys (2020), it was shown, that individual differences in form of mindware instantiation (i.e., the logical knowledge necessary to solve the heuristics and biases tasks), cognitive reflection and thinking dispositions emerged as significant predictors of the conflict detection. As stated by Burič and Šrol (2020), for the conflict to be detected, the participant needs to intuitively process the logical structure of the problem. The conflict detection studies therefore might be taken as the first piece of evidence, that there are individual differences in intuitive logical processing. A more direct approach was taken by Burič and Šrol (2020), who examined predictors of the normatively correct type 1 responses. Their results showed a predictive ability of both the mindware instantiation and cognitive reflection, which predicted the conflict detection in the study of Šrol and De Neys (2020).

The findings of both Šrol and De Neys (2020) and Burič and Šrol (2020) are in line with the recent theoretical model presented by Stanovich (2018). In line with the earlier model of De Neys and Bonnefon (2013), the author claims that the mindware instantiation is the key element of normatively correct responses. However, Stanovich (2018) goes even further by postulating a relationship between mindware, conflict detection, and logical intuitions. According to the model, the conflict cannot be detected and the response will be incorrect if the mindware is absenting. If the mindware is present, the better it is adopted the easier it can interfere with incorrect intuition, thus the conflict will be detected, which results in a higher probability of correction of the incorrect intuition. Finally, if the mindware is adopted to such an extent that it is automatized, no conflict detection or inhibition is needed, as the reasoner does not need to engage in analytical thinking and can provide normatively correct type 1 response.

The main goal of the present study was to examine the role of mindware instantiation in CRT. We chose the CRT for one main reason. As mentioned above, the studies using the two-response paradigm applied to conflict problems have found that it is rather rare for participants to correct their erroneous intuitive answer and override it with the correct one by tapping into the type 2 processing. In the majority of cases, if the answer was correct, it was already correct at the initial response stage (e.g., Bago & De Neys, 2017, 2019b; Burič & Šrol, 2020; Thompson et al., 2011; Thompson & Johnson, 2014). The CRT works in a similar way to the conflict problems used in the studies. However, the CRT tasks were designed to measure participants' tendency to override incorrect intuitions (Frederick, 2005), therefore, we find it interesting to examine the possibility of logically correct intuitive answers and the role of mindware instantiation in this context. We used the two-response paradigm, which allowed us to take a closer look at its role in both intuitive and analytical responses. In line with the theoretical model of De Neys and Bonnefon (2013), we also employed conflict detection measures – confidence in the response and the time needed for the response. We hypothesize that the mindware automatization (i.e., number of correctly solved neutral problems and the time needed to solve them) along with the conflict detection measures will significantly predict the num-

ber of logically correct intuitive responses. We believe that examining the role of mindware automatization can help to better understand logical intuitions and advance the debate of the dual-process models of higher cognition.

Method

Participants

The study was run online and participants were students and alumni recruited via websites of major Slovak universities and social networks. The data were collected from 908 participants aged between 15 and 58 years (M = 21.10; SD = 5.22; 594 women, 314 men). 53.2% of participants stated that they had completed high school education, 36.9% had completed at least the first degree of university education.

Reading Pretest

To determine the time limit for the intuitive response, we ran a reading pretest in which the participants were instructed only to read the given problems, not to solve them. This pilot was run via the same server as the main study. We determined the time limit based on the average time participants needed to read the problems. The reading time was measured as the latency from the time the problem was presented to the time participants clicked on the submit button. There was no other button on the screen, to avoid participants trying to solve the problems. Responses from ten participants (6 females, 4 males, age: M = 24.3; SD = 4.14) were analyzed and the resulting average time was 6.72 seconds (SD = 2.55). Like Bago and De Neys (2017), we rounded this time up to 7 seconds and then used it as the time limit for the first answer in the main research for all of the CRT items.

Materials

Cognitive Reflection Test

To measure participants' cognitive reflection, we used five items, including three items from the original CRT (Frederick, 2005) with altered content and numbers (e.g., "If it takes 3 printers 3 minutes to print 3 magazines, how long would it take 100 printers to print 100 magazines?"). Two other items were taken from previous research (Burič & Šrol, 2020; Šrol, 2019). As the problems differed in length, we shortened the items to be as similar in length as possible and at the same time to preserve its nature – to trigger intuitive responses (we provide all of the problems in the Online supplement – section A). Each participant solved 5 CRT items and 5 no-conflict versions of these items, which were designed to be as similar to the CRT as possible, except it cues an intuitive answer that is in line with the normative correctness of the problem. Participants always had to choose from three options – the correct one, the intuitive one, and the third option – either calculated as the half value of the correct answer or a random value still relevant for the given context. In no-conflict problems, the intuitive response is also the correct one, therefore we chose the other two options as random values still relevant for the given context. Internal consistency of the conflict CRT items was quite low (α = .56) at the initial response, which could have been caused by a small number of items (Streiner, 2003; Vaske et al., 2017). An alternative explanation, given that the reliability was relatively high at the final response stage (α = .80), is the restriction of the type 2 processing. Participants were rushed to respond with burdened working memory, which are not typical conditions when solving the CRT. As a result, participants could simply answer

randomly on some items. We address the possibility of random responses in the discussion. In the no-conflict versions, we observed even higher reliability, both at the initial (α = .73) and the final response stage (α = .97).

Two-Response Paradigm

All of the CRT problems were presented under a two-response paradigm. In this paradigm, participants are asked to solve each problem twice. The first answer should be intuitive – the first answer that comes to mind after reading the problem. In the second response, participants have an unlimited amount of time and can think well about the solution. To make the first response truly intuitive, we adopted multiple restrictions, previously shown to be an effective manipulation to knock out type 2 processing (Bago & De Neys, 2017; De Neys, Moyens, & Vansteenwegen, 2010; Thompson et al., 2011) – instruction, time limit resulting from the reading pretest, and secondary cognitive task. As for the secondary cognitive task, we used a task in which participants had to memorize a pattern of dots in a matrix while solving CRT problems, which should have burdened their working memory, and thus limited the type 2 processing. We provide an example of the task in Section B of the Online supplement. Cases in which participants did not provide the first answer within the time limit or incorrectly solved the secondary cognitive task were excluded from the final analysis (see Section G of the Online supplement).

Conflict Detection

For each response, two indicators of conflict detection between the heuristic and logical responses were recorded (Bago & De Neys, 2017, 2019a, 2019b; Burič & Šrol, 2020; De Neys et al., 2013; Thompson & Johnson, 2014). The first was response time – the time from the onset of the problem until participants selected a response, which should be lower for no-conflict problems when compared with the conflict ones, meaning the conflict was detected. After selecting the answer, participants were asked to indicate on a scale from 0 to 10 how confident they are that their response is correct (0 = "I'm not sure at all", 10 = "I'm completely sure"). This confidence was used as a second measure of conflict detection, as it should be higher in no-conflict versions of the CRT, meaning the conflict was detected. In cases where participants failed to respond within the time limit, neither the time nor the certainty of their response was analyzed. To be able to use these indicators in the analyses, it was necessary to determine indices of conflict detection capabilities. We calculated the indices for each participant (Frey, Johnson, & De Neys, 2018; Šrol & De Neys, 2020) as the sum of cases in which the participant successfully detected a conflict (i.e., needed more time to respond or entered lower confidence in incorrectly solved conflict problem compared to a correctly solved no-conflict problem). However, the number of items through which this index could be calculated varied across participants. Therefore, for each participant, we divided the sum of successfully detected conflicts by the total number of cases from which the index could be calculated (i.e., the sum of incorrectly solved conflict problems and correctly solved no-conflict problems) to obtain an index of conflict detection capabilities for participants across problems.

We calculated the indices based on time (M = 0.54; SD = 0.46; α = .01) and confidence (M = 0.75; SD = 0.38; α = .16) separately for the initial response, and also the index based on time (M = 0.55; SD = 0.36; α = .28) and confidence (M = 0.86; SD = 0.46; α = .57) for the final response. When a participant did not

provide any incorrect answers to the conflict problems (*n* = 25) or correct answers to the no-conflict problems (*n* = 22), it was not possible to calculate the detection capability index and these participants were excluded from further analysis.

Mindware Instantiation and Automatization

As in other studies (Frey, Johnson, & De Neys, 2018; Šrol & De Neys, 2020), we used the correctness of neutral problems to measure the mindware instantiation, which reflected basic knowledge needed to solve the CRT. They are called neutral because, unlike conflict and no-conflict problems, they do not elicit any heuristic answer. Since many studies have shown that items in CRT are based on mathematical operations (Baron et al., 2015; Låg et al., 2014; Liberali et al., 2012; Weller et al., 2013), we assume the mindware needed to solve CRT is represented in form of such operations. According to Bago and De Neys (2019b), the mathematical operation on which the famous bat and ball problem is based is $100 + 2x = 110$. Meaning, for a participant to solve the bat and ball problem correctly, he must possess the basic knowledge in the form of $100 + 2x$ operation first. We have also successfully determined the mathematical operation for the second problem of the original CRT test, and since we were unable to determine the operation for the third problem, this operation was not represented in the series of neutral problems measuring mindware instantiation. Overall, we used four tasks to measure the presence of mindware for each of the four CRT problems (except problem 3, see Section C of the Online supplement for examples of the neutral problems based on the CRT problems), meaning, if a participant solved a given neutral problem correctly, he/she possesses the mindware needed to solve the CRT item. Each participant solved a total of 16 problems (4 variations for each problem) measuring the mindware (*M* = 12.89; *SD* = 2.69; *α* = .78).

As we also wanted to tap the level of mindware automatization, we measured the participants' time they needed to solve the neutral problems (*M* = 21.63; *SD* = 13.51; *α* = .72). We hypothesized, that if the participants' answer on the neutral problem was correct, the shorter time of response should reflect a greater degree of the mindware automatization. We also tried to calculate an index that would capture the degree of mindware automatization, however, we decided to use the number of correctly solved problems and the time needed to solve them as the two separate measures. We provide an explanation and the calculations in Section D of the Online supplement.

Procedure

The experiment was programmed in the Java Enterprise Edition and run on a private server. Participants were first presented with informed consent and several demographic questions,

Table 1 *Examples of conflict, no-conflict, and neutral versions of the CRT tasks*

Conflict versions	No-conflict versions	Neutral version
An apple and an orange weigh 160 grams in total. The apple weighs 100 grams more than the orange. How much does the orange weigh?	The handyman and the electrician work for 240 days in total. The electrician works for 200 days. How many days does the handyman work?	$100 + 2x = 160$

followed by the CRT problems along with their no-conflict versions presented under a two-response paradigm, and finally with a block consisting of neutral problems presented as open-ended questions. The order of problems within the blocks was randomized. Before participants begun solving the problems, instruction along with two practice tasks were presented, so that participants could familiarize themselves with the type of problems they were about to solve. The block with neutral problems was always presented after the one with the CRT, to prevent the influence on participants subsequent responses (Frey et al., 2018).

Results

The data for this study are publicly available at: https://osf.io/3gzxb/

Accuracy, Conflict Detection, and Direction of Change Analysis

As our sample was relatively large, we can assume the sampling distribution to be normal. Therefore, we opted for parametric tests in further analysis. However, we provide more information about the data distribution in the Online supplement, section E. As expected, the accuracy at the initial response stage was higher in no-conflict problems (M = 2.42; SD = 1.26) compared to the conflict items (M = 1.16; SD = 1.08; t(437) = -16.35; p < .001; d = 1.07). Similar results were observed at the final response stage, as the accuracy was again higher in no-conflict (M = 4.45; SD = 1.02), compared to conflict items (M = 2.30; SD = 1.47; t = -30.54; p < .001; d = 1.69). At the initial response stage, participants managed to solve 37.66% of the conflict problems correctly, in 51.15% of the cases they reached for the incorrect heuristic answer, and in 11.10% for the incorrect third option. At the final response stage, the number of correctly solved

problems increased to 49.25%. In 46.70% of cases, participants opted for an incorrect heuristic response and just 4% for an incorrect third option. These results show that participants solved more problems correctly when they could engage in type 2 processing, which is in line with the classic default-interventionist view (Evans & Stanovich, 2013). Also, participants solved on average 12.89 (SD = 2.69) of neutral problems correctly, with average time of correctly solved problems of 21 seconds (SD = 13.51).

Regarding conflict detection, we present an overview of the findings in Table 2. We calculated average times and confidences for conflict and no-conflict items and explored whether we could observe this trend in other conflict detection studies (e.g., Frey et al., 2018; Mevel et al., 2015) – higher time needed for a response and lower confidence of the response in incorrectly solved conflict problems in comparison with correctly solved no-conflict problems. As Table 2 shows, the effect was indeed observed. With an exception of the time of intuitive responses, in which the difference was not significant, participants were able to detect the conflict in all of the remaining measures.

In line with previous research using the two-response paradigm (Bago & De Neys, 2017, 2019b; Burič & Šrol, 2020), we also decided to conduct the direction of change analysis to explore the direction of answer change. Simply put, we wanted to see whether participants tend to change their initial answer when given enough time for rethinking and their working memory is not burdened with a secondary cognitive task. As a consequence, four possible scenarios might take place. We labeled these scenarios as: 11 – both answers correct, 00 – both answers incorrect, 01 – incorrect initial answer and correct final answer, and 10 – correct initial answer and incorrect final answer. Frequencies of these four scenarios are presented in Table 3.

Table 2 *Summary of the conflict detection analysis*

	Time *(ms)*		Confidence *(%)*	
	Initial	Final	Initial	Final
n	314	381	341	402
No-conflict (*SD*)	5473.25 (863.94)	13157.35 (8579)	57.28 (29.78)	82.67 (24.32)
Conflict (*SD*)	5542.50 (835.23)	18056.64 (20016.40)	33.51 (27.64)	41.57 (25.24)
Difference	*t* = 1.31	*t = 5.02*	*t = -13.13*	*t = -30.60*
p - value	.192	*< .001*	*< .001*	*< .001*
Cohen's *d*	0.08	*0.31*	*0.83*	*1.66*

Note. The values reflect participants' average times of answers, and confidence at the two response stages averaged across all conflict CRT problems. The reported confidence was converted into % for interpretative purposes (e.g., confident reported as 1 = 10%; 5 = 50%, 10 = 100%). The reported differences represent the pairwise comparison of the average (initial or final) time or response confidence between the conflict and the no-conflict version of the task. In the initial response stage, only those responses that were provided in the time limit along with the correctly solved secondary cognitive task were analyzed. Cohen's *d* is reported as a measure of effect size. Significant paired differences are presented in italics (*p* < .05).

Table 3 *Direction of change analysis for conflict and no-conflict CRT problems*

Task	Direction of change			
	11	00	01	10
No-conflict	74.79% (825)	3.63% (40)	20.39% (225)	1.18% (13)
Conflict	28.35% (319)	36.97% (416)	25.33% (285)	9.33% (105)

Note. The table contains frequencies and percentages of trials on which a specific direction of change for conflict and no-conflict CRT problems was observed.

As can be seen from Table 3, the most frequent situation in conflict problems was the 00 scenario, which is in line with the DI model (Evans & Stanovich, 2013), according to which reasoners tend to give a biased response, which they do not correct due to absence of conflict detection. However, in opposition to the models assumption, the second most prevalent situation was the 11 scenario, which means, that if the final response was correct, in more than half of these cases the response was correct already at the initial response stage. These findings are supported by previous studies (Bago & De Neys, 2017, 2019b; Burič & Šrol, 2020), which similarly showed, that people are often capable of normatively correct responses already at the initial response stage, which is in favor of the above-mentioned hybrid models of dual processing.

Correlations between Measured Variables

As the first step in testing our hypothesis, we have calculated correlations between reasoning accuracy at both initial and final response

stage, conflict detection indices, number of correctly solved neutral problems, and the time needed to correctly solve the neutral problems. Correlations are presented in Table 4.

The analysis shows significant correlations between both of the mindware measures and reasoning accuracy at the initial and the final response stage. Regarding conflict detection indices, only the one based on the time of initial response showed a significant relationship with final reasoning accuracy.

Correlations between the mindware measures and reasoning accuracy are in line with the model of mindware automatization (Stanovich, 2018), according to which the automatized mindware should enable participants to provide normatively correct type 1 response. The model also postulates the relationship between mindware automatization and conflict detection, which is only partially supported by our results. As Table 4 shows, a significant correlation was observed only between neutral problem accuracy and conflict detection at the final response stage based on confidence, and between the average time of correctly solved neutral problems and conflict detection at the initial response stage based on time. To get a better estimate of the fit with the Stanovich's model, we provide a more direct examination of the relationship in Section F of the Online supplement.

Predicting Initial and Final Reasoning Accuracy

Finally, to directly test our hypothesis, we have conducted two regression analyses, to examine mindware instantiation and automatization and conflict detection as independent predictors of intuitive and analytical reasoning accuracy. Based on recent theoretical models (De Neys & Bonnefon, 2013; Pennycook et al., 2015; Stanovich, 2018) we have entered the mindware measures in the first step, and conflict detec-

Table 4 *Correlations between initial and final reasoning accuracy, neutral problems accuracy, average time of correctly solved neutral problems and conflict detection indices*

	1.	2.	3.	4.	5.	6.	7.
1. Initial reasoning accuracy	1						
2. Final reasoning accuracy	.312**	1					
3. Mindware – accuracy	.192**	.367**	1				
4. Mindware – time	-.124*	-.143**	.103*	1			
5. Conflict detection at the initial response stage – confidence	.009	.026	-.003	-.058	1		
6. Conflict detection at the initial response stage – time	-.033	.154*	.126	.145*	.028	1	
7. Conflict detection at the final response stage – confidence	-.028	.003	-.103*	-.032	.250**	-.029	1
8. Conflict detection at the final response stage – time	-.032	-.052	-.025	.018	.037	.044	.077

Note. Mindware – accuracy represents the accuracy of the neutral problems, Mindware – time represents the average time needed to correctly solve the neutral problems. Correlations that appear in italics are significant.

* $p < .05$; ** $p < .01$.

Table 5 *Summary of the regression analyses predicting reasoning accuracy at the initial and final response stage*

	Initial reasoning accuracy			Final reasoning accuracy		
	b (SE)	β	t	b (SE)	β	t
Step 1						
Constant	2.30 (0.45)		5.8	4.82 (0.36)		14.36
Mindware – accuracy	0.15 (0.03)	.28	4.47***	0.21 (0.02)	.39	8.06***
Mindware – time of answer	-2.26 (0.72)	-.20	-3.15**	-1.72 (0.48)	-.17	-3.58***
R^2 = .104; $F(2, 227)$ = 14.23; p <0.001				R^2 = 0.164; $F(2, 360)$ = 35.36. p < 0.001		
Step 2						
Constant	2.35 (0.47)		5.2	4.89 (0.36)		5.2
Mindware – accuracy	0.15 (0.03)	.29	4.54***	0.21 (0.03)	.39	8.04***
Mindware – time	-2.16 (0.72)	-.19	-2.98**	-1.70 (0.48)	-.17	-3.54***
Conf. detection – confidence	-0.20 (0.23)	-.06	-0.86	-0.22 (0.18)	-.06	-1.18
Conf. detection – time	0.04 (0.22)	.012	0.19	0.12 (0.20)	.03	0.60
ΔR^2 = 0.003; $F(2, 225)$ = 0.385; p = 0.681				ΔR^2 = 0.004; $F(2, 358)$ = 0.845; p = 0.430		

Note. The table contains unstandardized (*b*) and standardized regression coefficients (*β*) with their respective *t*-ratio and significance. R^2 and ΔR^2 denote adjusted *r*-square for the initial model and change in *r*-square at the 2nd step of the regression with appropriate change statistics. Mindware – accuracy represents the accuracy of the neutral problems. Mindware – time represents the average time needed to correctly solve the neutral problems. Significant regression coefficients are presented in italics.
* p < .05; ** p < .01; *** p < .001

tion measures in the second step of the model. The results are summarized in Table 5.

As evident from the table, the number of correctly solved neutral problems ($β$ = .28; p < .001) and the time needed to solve them ($β$ = -.20; p = .002) were both significant independent predictors of initial reasoning accuracy already in the first step of the regression and together these variables accounted for 10% of explained variance. As we also measured the time needed for correctly solving the neutral problems, the two mindware measures were stronger predictors of initial accuracy than a single mindware measure in form of the number of correctly solved neutral problems in the study of Burič and Šrol (2020). In general, these results support predictions of theoretical models (De Neys & Bonnefon, 2013; Pen-nycook et al., 2015; Stanovich, 2018) and also recent empirical findings (Burič & Šrol, 2020; Šrol & De Neys, 2020), according to which the mindware instantiation could be the key determinant of correctly solved conflict problems. However, after adding conflict detection indices in the second step of the regression we did not observe a significant increase in its predictive power.

Regression predicting final accuracy shows very similar results. Both the number of correctly solved neutral problems and the time needed to solve them have again emerged as significant predictors in the first step of the regression and have explained 16% of the variance. Once again, after adding conflict detection indices as predictors in the second step of the regression, the predicting power

of the model did not increase. The results regarding the role of mindware in the second step of the regression are again in line with previous studies (Burič & Šrol, 2020; Šrol & De Neys, 2020), however, we cannot say the same about the conflict detection, which has support in previous research and theoretical models (Burič & Šrol, 2020; De Neys & Bonnefon, 2013; Stanovich, 2018; Šrol & De Neys, 2020) but which was not a significant predictor of either initial or final accuracy.

Discussion

The main goal of this study was to examine the role of mindware and its possible automatization along with the conflict detection in CRT, in both intuitive and analytical responses. The results support the findings of the conflict detection studies (Bago & De Neys, 2017, 2019a, 2019b; Burič & Šrol, 2020; De Neys, 2012, 2018; Pennycook, 2018; Pennycook et al., 2015; Šrol & De Neys, 2020; also see De Neys, 2018), as at both initial and final response stage participants were able to detect the conflict between their heuristic intuition and logical structure of the CRT, which was manifested by longer time needed for an answer and lower confidence in conflict items when compared with no-conflict versions. The only exception was the time of intuitive answer. This could be due to the strict time limit which was set to restrict possible type 2 engagement. Participants needed to provide the answer immediately after they read the problem, which could have resulted in low variability in the times of the answers. These findings once again run against the DIs assumption of biased response due to a lack in conflict detection.

One could also argue that the strict time limit along with burdening the working memory could make the tasks too demanding and participants could end up simply guess-

ing the answer. If this was the case, participants would opt for all of the three possible responses in around 33% of cases. It can be alarming that in the case of correctly solved problems, the frequency is indeed similar to this number. However, if participants just guessed, they would end up with a heuristic answer and an incorrect third answer in just as many cases, which did not happen.

By using the two-response paradigm, we were able to conduct the direction of change analysis. The results are in line with the previous research (e.g., Bago & De Neys, 2017, 2019b; Burič & Šrol, 2020), as the 11 scenario was much more frequent, as predicted by the DI model. However, the DI model postulates, that after an incorrect intuitive response people engage in type 2 process intervention and correct their initial answer. In this study, we labeled such a scenario as 01. Unlike in the studies mentioned above, it was almost as frequent as the 11 scenario. There are a couple of possible explanations. First, not only did we observe a higher frequency of the 01 scenario when compared with the previous studies, but also a lower frequency of the 00 scenario, meaning that it was more often the case that our participants corrected their initial response. 25.3% of our participants stated that they are studying mathematical science, physics, chemistry, computer science, or other fields of study with a technical focus, while this was the case in only 14.7% of participants in the study of Burič and Šrol (2020). We assume that these participants are better at solving mathematical problems similar to the ones we used for measuring the mindware, as it is a part of their full-time study, which might result in a higher correction rate. Nevertheless, this assumption needs to be further analyzed in future research, as it is not supported by relevant data. Second, there is not much research combining the CRT with the two-response paradigm, as most of the studies have

been using syllogisms, conditional reasoning tasks or base-rate neglect tasks (Bago & De Neys, 2017; Burič & Šrol, 2019; Thompson et al., 2011; Thompson & Johnson, 2014), with an exception of Bago and De Neys (2019b), who used the bat-and-ball problem, but not the rest of the CRT. Thus, we are not provided with a relevant benchmark.

To the main point of the study, we tried to build on the previous research (Burič & Šrol, 2020; Frey et al., 2018; Šrol & De Neys, 2020) and theoretical models (De Neys & Bonnefon, 2013; Stanovich, 2018) linking the mindware and conflict detection with logical intuitions. Also, we attempted to add the time of a response to the neutral problems as an approximate estimation of the mindware automatization. Results showed multiple findings that deserve attention, starting with the mindware measure itself. Neutral problems showed much higher reliability than the items in previous studies. This can be explained by several factors – for example, we used only one type of task – the CRT, which also showed good reliability. Other studies (Burič & Šrol, 2020; Šrol & De Neys, 2020) used several types of problems and measured the mindware instantiation using a single index calculated from all of the neutral items combined. However, according to Stanovich (2018), mindware is task-specific, as it represents a different kind of knowledge needed to solve each type of problem. In addition, these studies used only two items from each task while our set of neutral tasks consisted of 16 items.

The first mindware measure – the number of correctly solved neutral problems along with the second one – time needed to solve them both emerged as significant predictors of accuracy at both response stages. These findings support theoretical models (De Neys & Bonnefon, 2013; Stanovich, 2018), according to which the mindware is the key element in bias susceptibility. However, the relation-

ship between the accuracy in CRT and the time needed to correctly solve the neutral items is not clear, as we did not manage to compute a single index representing the mindware automatization. Therefore, we cannot be certain that participants who solved more neutral problems in a shorter time also solved more of the CRT problems. The relationship between the number of correctly solved neutral problems and mindware instantiation is clear – the more neutral problems participant solved, the better instantiated mindware he/she possessed. It does not have to be the same case with the time of the answers, as the relationship might not be linear.

Unlike other conflict detection studies, we did not observe the predictive ability of conflict detection in CRT accuracy. One of the possible explanations might lay in the measure of conflict detection and its low reliability, which certainly is not a novel issue, since it was reported by other authors as well (Burič & Šrol, 2020; Šrol & De Neys, 2019). It could be caused, for example, by the fact that the number of detections from which indices were calculated differed between participants and indices could not be calculated for participants who responded to all CRT items correctly, as the correct answers require not only conflict detection but also inhibition of the initial response. According to Stanovich (2018), conflict detection is domain-specific, so it is possible that it plays a greater role in certain tasks than in others. By this we do not want to question its importance, we rather point to the lack of evidence reflecting the weight of mindware, conflict detection, and inhibition – the individual components of the theoretical model (De Neys & Bonnefon, 2013) in different types of tasks and biases.

The last point we feel the urge to mention, is the direction of change analysis and its implication for the nature of the CRT. In the last fifteen years of research, cognitive reflection

has been seen as the ability of participants to engage in type 2 processing, which should lead to inhibition of the type 1 response and its replacement by a correct response (Frederick, 2005; Kahneman, 2011; Toplak et al., 2011). In such a description, it is easy to see why it is considered a reflection, since an intuitive response is being reflected upon. Our results indicate that this statement is correct in only about half of the cases. In the second half, the participants answered correctly already at the initial response stage, and thus there was no need for the reflection. Moreover, the results of Bago and De Neys (2019b) suggest, that reflection is needed in far fewer cases. We are aware that extraordinary claims require strong evidence and we do not want to draw any conclusions that would contradict the current understanding of cognitive reflection. However, we consider it necessary to subject type 1 responses in the CRT to further, much more detailed examination.

We are also aware of the limits of this study. First, we did not control for the prior exposure of our sample to the CRT, which can be of concern, given the popularity of the CRT and findings of previous studies showing that familiarity with the test can affect the performance (Chandler et al., 2014; Stieger & Reips, 2016). However, Chandler et al. (2014) also found that the correlation between prior exposure and the performance disappeared when authors used structurally identical items with modified content. As we did not use the CRT with the original content, but only the modified one, we assume the impact of prior exposure on the performance is rather unlikely. Second, as already mentioned, is the low reliability of conflict detection indices, which might be a great obstacle to future research. The third limitation is the main focus of this study itself – mindware automatization. Even though both the number of correctly solved neutral problems and the time needed to solve them emerged as predictors of accuracy in CRT,

we cannot say that by measuring these variables we successfully captured the extent of mindware automatization. To do so, it is necessary to calculate a uniform index expressing the degree of automatization for each participant. However, the nature of the relationship between the time needed to correctly solve the problem with its accuracy is not clear, as both too short and too long response time needed for response may lead to the incorrect answer. Also, we used 5 CRT problems, while we measured the mindware via mathematical operations derived just from the four of them. However, as mentioned above, the reliability of the items was much better compared to the previous studies (Burič & Šrol, 2020; Šrol & De Neys, 2020). The length of the CRT problems could also be problematic, as the items consisted of a different word count and we used a uniform time limit to reduce the possibility of type 2 intervention, for all participants, across all of the problems. We tried to take this into account to some extent by adjusting the CRT items to be as short and as similar in word count as possible. Nevertheless, it could be a case of the performance being affected by individual differences in time needed to read and to process the problems. Perhaps it may be a good idea to set a time limit for each participant separately, to avoid such a possibility.

Recent years have brought a line of research indicating that substantial individual differences exist already at the intuitive stages of a reasoning process. This study adds to these findings, suggesting that not only do they exist, but they can also be closely related to the extent of mindware instantiation and automatization.

Acknowledgement

"Examining unfounded beliefs about controversial social issues".

Authors' ORCID
Roman Burič
https://orcid.org/0000-0002-3762-3354
Lubica Konrádová
https://orcid.org/0000-0002-7698-3993

References

Bago, B., & De Neys, W. (2017). Fast logic?: Examining the time course assumption of dual process theory. *Cognition*, *158*, 90–109. https://doi.org/10.1016/j.cognition.2016.10.014

Bago, B., & De Neys, W. (2019a). Advancing the specification of dual process models of higher cognition: A critical test of the hybrid model view. *Thinking & Reasoning*, 1–30. https://doi.org/10.1080/13546783.2018.1552194

Bago, B., & De Neys, W. (2019b). The Smart System 1: Evidence for the intuitive nature of correct responding on the bat-and-ball problem. *Thinking & Reasoning*, *25*(3), 257–299. https://doi.org/10.1080/13546783.2018.1507949

Bonner, C., & Newell, B. R. (2010). In conflict with ourselves? An investigation of heuristic and analytic processes in decision making. *Memory & Cognition*, *38*(2), 186–196. https://doi.org/10.3758/MC.38.2.186

Burič, R., & Šrol, J. (2019). *Individual differences in logical intuitions on reasoning problems presented under two-response paradigm*. https://doi.org/10.31234/osf.io/w6un2

Burič, R., & Šrol, J. (2020). Individual differences in logical intuitions on reasoning problems presented under two-response paradigm. *Journal of Cognitive Psychology*, 1–18. https://doi.org/10.1080/20445911.2020.1766472

De Neys, W. (2012). Bias and conflict: A case for logical intuitions. *Perspectives on Psychological Science*, *7*(1), 28–38. https://doi.org/10.1177/1745691611429354

De Neys, W. (Ed.). (2018). *Dual process theory 2.0* (1 Edition). Routledge.

De Neys, W., & Bonnefon, J.-F. (2013). The 'whys' and 'whens' of individual differences in thinking biases. *Trends in Cognitive Sciences*, *17*(4), 172– 178. https://doi.org/10.1016/j.tics.2013.02.001

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Evans, J. St. B. T. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning*, *13*(4), 321–339. https://doi.org/10.1080/13546780601008825

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Frey, D., Johnson, E. D., & De Neys, W. (2018). Individual differences in conflict detection during reasoning. *Quarterly Journal of Experimental Psychology*, *71*(5), 1188–1208. https://doi.org/10.1080/17470218.2017.1313283

Handley, S. J., & Trippas, D. (2015). Dual processes and the interplay between knowledge and structure: A new Parallel Processing Model. In *Psychology of Learning and Motivation* (Issue 62, pp. 33–58). Elsevier. https://doi.org/10.1016/bs.plm.2014.09.002

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods*, *46*(1), 112–130. https://doi.org/10.3758/s13428-013-0365-7

Kahneman, D. (2013). *Thinking, fast and slow* (1st edition). Farrar, Straus and Giroux.

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In *The Cambridge handbook of thinking and reasoning* (pp. 267–293). Cambridge University Press.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, *64*(6), 515–526. https://doi.org/10.1037/a0016755

Mevel, K., Poirel, N., Rossi, S., Cassotti, M., Simon, G., Houdé, O., & De Neys, W. (2015). Bias detection: Response confidence evidence for conflict sensitivity in the ratio bias task. *Journal of Cognitive Psychology*, *27*(2), 227–237. https://doi.org/10.1080/20445911.2014.986487

Pennycook, G. (2018). A perspective on the theoretical foundation of dual process models. In

Dual process theory 2.0. (pp. 5–27). Routledge/ Taylor & Francis Group.

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology*, *80*, 34–72. https://doi. org/10.1016/j.cogpsych.2015.05.001

Stanovich, K. E. (2018). Miserliness in human cognition: The interaction of detection, override and mindware. *Thinking & Reasoning*, *24*(4), 423–444. https://doi.org/10.1080/13546783.2018.1459314

Stieger, S., & Reips, U.-D. (2016). A limitation of the Cognitive Reflection Test: Familiarity. *PeerJ, 4, e2395.* doi: 10.7717/peerj.2395 « iscience. (n.d.). Cit 14. October 2020, from https://iscience.uni-konstanz.de/en/2016/01/10/stieger-s-reips-u-d-2016-a-limitation-of-the-cognitive-reflection-test-familiarity-peerj-4-e2395-doi10-7717peerj-2395/

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99–103. https://doi.org/10.1207/S15327752JPA8001_18

Šrol, J., & De Neys, W. (2019). *Predicting individual differences in conflict detection and bias susceptibility during reasoning* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/2uf6g

Šrol, J., & De Neys, W. (2020). Predicting individual differences in conflict detection and bias susceptibility during reasoning. *Thinking & Reasoning*, 1–31. https://doi.org/10.1080/13546783.2019.1708793

Thompson, V. A., & Johnson, S. C. (2014). Conflict, metacognition, and analytic thinking. *Thinking & Reasoning*, *20*(2), 215–244. https://doi.org/10.1080/13546783.2013.869763

Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, *63*(3), 107–140. https://doi.org/10.1016/j.cogpsych.2011.06.001

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. https://doi.org/10.3758/s13421-011-0104-1

Vaske, J. J., Beaman, J., & Sponarski, C. C. (2017). Rethinking internal consistency in Cronbach's alpha. *Leisure Sciences*, *39*(2), 163–173. https://doi.org/10.1080/01490400.2015.1127189

Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J., & Peters, E. (2013). Development and testing of an Abbreviated Numeracy Scale: A Rasch analysis approach. *Journal of Behavioral Decision Making*, *26*(2), 198–212. https://doi.org/10.1002/bdm.1751