These Problems Sound Familiar to Me:
Previous Exposure, Cognitive Reflection Test, and the Moderating Role of Analytic Thinking

Jakub Šrol

Institute of Experimental Psychology, Center of Social and Psychological Sciences
Slovak Academy of Sciences

One of the current topics in research on the Cognitive Reflection Test (CRT) is its growing familiarity among the general public. Surprisingly, Bialek and Pennycook (2017) showed that previous exposure does not diminish the CRT's predictive power in Heuristics and Biases (H&B) tasks, but proposed that the relationship is moderated by analytic thinking, a conjecture tested in the present study. Participants (N = 365) filled in the CRT, Need for Cognition (NFC) scale, and a battery of H&B problems. While the CRT did retain its predictive power in the H&B performance, regardless of participants' self-reported thinking dispositions and exposure, both of these factors moderated the relationship, such that exposure increased CRT's predictive power in H&B tasks, albeit only among high-NFC individuals. Present results converge with studies showing that prior exposure does not invalidate the use of CRT, while offering some novel evidence for the metacognitive disadvantage account proposed by Bialek and Pennycook (2017).

Key words: Cognitive Reflection Test, familiarity issue, cognitive bias, analytic thinking, metacognitive disadvantage

Introduction

Over the past decade, Frederick's (2005) Cognitive Reflection Test (CRT), which consists of three tricky problems that give rise to com-pelling but wrong intuitive responses, became one of the most widely used methods in the research on individual differences in rational reasoning and decision making. As such, items of the CRT gradually came to be familiar to general public through research, popular books, or college psychology courses (e.g., Thomson & Oppenheimer, 2016). The familiarity issue is one of the current topics of the CRT research, with studies showing that people previously exposed to the test achieve better scores on it (Haigh, 2016; Pennycook, Cheyne, Koehler, & Fugelsang, 2015b), and that they might share different demographic characteristics from people unfamiliar with it (Stieger & Reips, 2016). Based on such results, researchers quickly started to believe that prior exposure invalidates the use of CRT as a predictor of various outcomes and several came up with alternative versions of the test to tackle the familiarity issue (Thomson &

Oppenheimer, 2016; Toplak, West, & Stanovich, 2014).

However, noting that the conjecture that previous exposure invalidates the CRT was not empirically tested, Bialek and Pennycook (2017) recently reanalyzed the data from six of their studies, where participants completed CRT as well as several outcome measures potentially related to analytic thinking. They have found that substantial proportion of participants in their research (22% – 60% of the samples) reported being familiar with the CRT and they scored higher on the test than the rest of the samples. More importantly, based on comparisons of correlations between CRT and outcomes among participants familiar and unfamiliar with the test, the authors found that the predictive validity of CRT never diminished as a result of prior exposure. Rather, it stayed similar or even became stronger among exposed participants, the latter option mostly occurring in the case of correlations between CRT and the composite of various heuristics and biases (H&B) tasks (Toplak, West, & Stanovich, 2011), such as ratio bias, conjunction fallacy, and base-rate neglect problems.

Further indication for the surprising non-effects of prior exposure to the CRT comes from Stagnaro, Pennycook, and Rand (2018), who have provided evidence that while scores on the CRT increased somewhat with the number of times participants were exposed to the test, the relationships between CRT and two measures related to analytical thinking – religious belief and political affiliation – remained surprisingly stable from the first time participants encountered CRT in their studies to the subsequent times they took the test. Moreover, in a comprehensive study by Meyer, Zhou, and Frederick (2018) it was shown not only that the CRT does not loose its predictive validity in one of the most notorious H&B problems (Linda task) and Raven's test of progressive matrices, but also that participants' real exposure to the CRT, as opposed to the self-re-

ported familiarity with the test, has only a very small effect on the CRT performance. Finally, the authors also conclude that people who improve their performance on the CRT with exposure are only those who continue to reflect upon the test, even with multiple exposures, and those who have performed well on the CRT the first time they took it.

Bialek and Pennycook (2017) offered several accounts to explain why CRT might not be negatively affected by participants' previous exposure to it. Firstly, *participants familiar with the CRT may also be familiar with the H&B composite,* as this exact battery of tasks has, for several years now, been frequently employed in other studies (e.g., Pennycook, Cheyne, Barr, Koehler, & Fugelsang, 2015a; Toplak, West, & Stanovich, 2011). Secondly, exposed participants may have scored higher on the CRT because of the *self selection effect*, i.e. highly reflective individuals may simply complete more studies and are therefore more likely to be familiar with CRT and have higher scores on the test, even if the prior exposure does not necessarily help them actually solve the problems. However, the self selection effect cannot really explain the differences in correlations between the CRT and other outcomes.

Finally, to explain why the CRT's predictive power does not diminish as a result of prior exposure, Bialek and Pennycook (2017) argue that people with low scores on the CRT may have a *metacognitive disadvantage*. That is, intuitive people may not realize the problems are tricky and therefore continue to do poorly even upon repeated exposure to the test. This conjecture stems from studies showing that intuitive individuals are worse at detecting the conflict in tasks which require suppressing misleading intuitions (Pennycook, Fugelsang, & Koehler, 2015b). It is also supported by recent study of Pennycook, Ross, Koehler, and Fugelsang (2017), who have shown that people who do poorly on the CRT strongly overesti-

mate their performance, suggesting they do not realize that the intuitive response they provided is incorrect. Note that this does not mean intuitive reasoners cannot improve their performance at all under repeated exposure to the test. If they encounter the problems in a context where they are presented along with the correct solution, their performance might subsequently increase even if they do not understand why the solution is correct. On the other hand, relatively analytic individuals will realize that the compelling intuitive response is incorrect and the problems require additional reflection, and therefore are more likely to increase their performance upon repeated exposure to the test. Although being already analytic, they might score high on the CRT as it is, and therefore would also not benefit much from their prior exposure, they are still expected to gain more from it than intuitive reasoners. Based on these propositions, Bialek and Pennycook (2017) concluded that familiarity may not be such a devastating problem for CRT's predictive validity as previously thought, as the test remains to distinguish well between intuitive and analytical reasoners even after multiple exposures. Of the accounts offered by Bialek and Pennycook (2017), the metacognitive disadvantage seems to fit the best with the results of Meyer et al. (2018), who have observed that the effect of familiarity is driven mostly by people, who continue to reflect upon the CRT even with multiple exposures.

The aim of the present study was to examine whether CRT retains its predictive validity in H&B problems among participants who are already familiar with the test, and to conduct a test of the metacognitive disadvantage account proposed by Bialek and Pennycook (2017). Participants were asked to answer the original CRT, indicate whether they were familiar with the test prior to taking a part in the study, solve a battery of H&B tasks, and fill in a separate measure of analytic thinking disposition, the Need for Cognition scale (NFC; Cacioppo, Petty, Feinstein, & Jarvis, 1996). The NFC represents a widely used self-report instrument designed to study predisposition toward effortful and analytic thought (e.g., Pennycook, Cheyne et al., 2015b; Toplak et al., 2014). However, as was shown by Pennycook et al. (2017), both genuinely reflective individuals and some intuitive reasoners, who lack insight into their true reflectivity, score highly on it. Therefore, the NFC should be distinguished from direct measures of analytic thinking disposition, such as the Cognitive Reflection Test, and rather thought of as an index of how strongly participants believe they are reflective. Concerning the reasoning and decision-making problems used in the present study, rather than using a composite H&B battery from previous research (Pennycook, Cheyne, Barr et al., 2015a; Toplak et al., 2011), new problems pertaining to several cognitive biases were created in order to reduce the possibility that participants were previously exposed to them.

Following the procedure of Bialek and Pennycook (2017), I have compared the correlations between CRT and outcome measures employed in the present research among unexposed and exposed participants, in order to determine whether the predictive power of CRT changes with previous exposure. Next, per the metacognitive disadvantage account, it was hypothesized that in analytic reasoners, as identified by their NFC scores, previous exposure would lead to a higher increase in the CRT performance than among relatively more intuitive participants. Finally, as Bialek and Pennycook (2017) observed that the relationships between CRT and H&B tasks were in some cases stronger among exposed participants, it was tested whether the predictive power of CRT in the performance on the H&B tasks would increase as a function of prior familiarity with the test, as well as participants' self-reported analytic thinking disposition.

Method

Participants

The study was presented in the form of an online survey and participants were students and alumni recruited through websites of several major Slovak universities and colleges. In total, 395 people participated in the study. However, based on attention check questions[1], 16 (4%) participants failed to follow instructions or their answers indicated random responding and their data were removed from subsequent analyses. Additionally, 14 (4%) participants failed to provide answers to one or more of the CRT items and were also dropped from further analyses. The final sample consisted of 365 participants of whom 92 (25%) were male and 273 (75%) female with the mean age of 23.39 ($SD = 4.103$). Most of the participants were university students, who reported having a high school diploma (45% of the sample), some had already finished their bachelor's (30%) or master's (23%) degree. Concerning the study fields, the participants were students of various universities and colleges with different majors, mostly economics and management (16%), pedagogy (14%), humanities (12%) and engineering (10%).

Sensitivity analysis was carried out in G*Power 3.1.9 software in order to determine effect sizes under the current sample size with 5% error probability and at least 80% statistical power. The results showed that the study should be powered enough to detect correlation coefficients of $r = .146$, differences between two independent correlations of $q = .296$, and differences between two independent means of $d = 0.294$ and higher[2].

Participants always first filled out the demographic information and the NFC, answered the original CRT and indicated their familiarity with the task, and then completed several blocks of H&B tasks and other measures not reported here. The order of the items within each block of problems was randomized. The materials and data for the present study are publicly accessible at OSF: https://osf.io/xfnsw.

Materials

Cognitive Reflection Test. The original three-item test developed by Frederick (2005) was used. After answering the three problems, participants were asked if they had ever encountered any of them before taking part in the present study.

Heuristics and Biases tasks. Four types of heuristics and biases tasks were used in the present study: 12 syllogistic reasoning tasks, 8 ratio bias items, 8 conjunction fallacy problems, and 6 base-rate neglect tasks. All of the items were constructed to evoke compelling but misleading intuitive response, which had to be suppressed in order to solve the problem in line with formal logic. Brief descriptions and example items for every type of H&B task can be found in the supplementary material. These four sorts of problems were selected because of their frequent use in the research on cognitive biases, and because they all have been shown in previous research to correlate with the CRT (e.g.,

---

[1] Two attention check questions were created. These items were randomly intermixed with the H&B problems and were constructed to resemble these tasks, but unlike the actual items they did not involve any catch and were actually very simple math problems. Participants who got any of the two attention check items wrong were automatically excluded from further analyses. Both attention check questions are available in the Supplementary materials for the study.

[2] The sensitivity analyses were calculated under the assumption that the two independent correlations or means come from groups of equal sample sizes, i.e. each group consists of 182 observations.

Oechssler, Roider, & Schmitz, 2009; Pennycook, Cheyne, Barr, Koehler, & Fugelsang, 2014; Toplak et al., 2014). For the purpose of all analyses, correct answers on the 34 problems were summed to form a single H&B composite score.

Need for Cognition scale (Cacioppo et al., 1996). Participants rated their agreement with 18 items such as *"I would prefer complex to simple problems"* on a 7-point scale ranging from 1 (not at all like me) to 7 (completely like me). After the data collection it was found that two of the NFC items did show unsatisfactory psychometric properties, they had negative correlations with some of the remaining items and their inclusion led to the decrease in the reliability of the scale. Therefore, these items were excluded from the analysis and NFC score was calculated as an average score reported on the remaining 16 items.

Results

Correlations between Measures

Descriptive statistics and correlation coefficients between all measures reported in the study are presented in Table 1. The results pertaining to the predictors and their relationships with H&B tasks are consistent with previous research on individual differences in cognitive biases. CRT and the NFC significantly corre-

lated with the H&B composite, however, the correlations tended to be somewhat stronger in the case of the former than the latter.

Differences between Exposed and Unexposed Participants

Table 1 also shows the correlations between participants' previous exposure to the original CRT and other measures in the study. Of the whole sample, 157 participants (43%) reported being familiar with one or more of the CRT tasks. As in other studies that asked for participants' prior exposure (Bialek & Pennycook, 2017; Haigh, 2016; Stieger & Reips, 2016), people already familiar with the CRT had higher scores on the test ($r = .408$). To facilitate the comparison with the results of previous studies, mean scores on the CRT were also compared, and this resulted in a large difference between exposed ($M = 2.36$, $SD = 0.99$) and unexposed ($M = 1.35$, $SD = 1.21$) participants, $t(360.4) = -8.75$, $p < .001$, $d = 0.91$. Furthermore, exposed participants also showed higher scores on the composite of H&B tasks presented in the study, although the relationships with exposure were lower than in the case of the CRT. The only non-significant correlation was observed between exposure and scores on the NFC, which shows that participants familiar and unfamiliar with the CRT did not differ significantly in their

Table 1 *Descriptive statistics and correlations between all methods in the present study*

	M	*SD*	*α*	1	2	3
1. Exposure	43%	–	–	1		
2. CRT	1.78	1.23	.78	**.408**	1	
3. NFC	4.17	0.82	.87	.090	**.193**	1
4. H&B composite	20.12	6.87	.89	**.240**	**.522**	**.303**

Note. $N = 365$. Table presents mean scores, standard deviations, and internal consistency of the measures employed in the study. Previous exposure to the CRT was coded as 0 = unexposed, 1 = exposed. Correlations that appear in bold are significant. Correlations of $r > .103$ are significant at $p = .05$, $r > .135$ are significant at $p = .01$, and $r > .172$ are significant at $p = .001$.

self-reported disposition toward analytic thinking. This seems to run counter to the notion of Bialek and Pennycook (2017) that increased scores on CRT among exposed participants might be the result of the self-selection effect. If more reflective people completed more studies, and therefore were more familiar with CRT, it could be expected that prior exposure would be correlated with the scores on the NFC scale. However, the lack of correlation could also reflect the fact that NFC captures only self-perception of participants as being analytic or not, and therefore, the potential relationship between number of studies completed and genuine analytic thinking disposition may have been attenuated by employing this measure.

Moreover, in order to examine whether prior exposure to CRT influenced the predictive validity of the test, Fisher's z test was used to compare the correlations between CRT and other measures among exposed and unexposed participants. There were no significant differences in the correlation between CRT and NFC among unexposed ($r = .189$) and exposed ($r = .143$) participants, $z = 0.444, p = .660$. Similarly, there was no significant difference in correlations observed among unexposed ($r = .509$) and exposed ($r = .452$) participants between CRT

and the H&B battery, $z = 0.696, p = .484$. The observed differences were both very small in size and not significant, however, the correlations were never higher in exposed participants than unexposed participants, contrary to what was observed in the study of Bialek and Pennycook (2017).

Does Exposure to the CRT Predict Responses on the CRT?

Next, to test the hypothesis that prior exposure would benefit only analytic but not intuitive individuals when solving the CRT, a simple moderation analysis was conducted, where exposure was used as a predictor (X) of the scores on CRT (Y) with self-reported analytic thinking style as a moderator (M). All moderation analyses in the present paper were conducted with Hayes' (2013) macro implemented in the IBM SPSS v.20 software. The variables were always mean centered prior to the analyses. The results of this moderation (Table 2) show that both exposure and NFC are predictors of the responses on the CRT, but there is no moderation present. Thus, people with high and low self-reported analytic cognitive styles did not benefit to different extent from prior exposure, when

Table 2 *Simple moderation analysis of the CRT exposure as a predictor of scores on the CRT and the NFC as a moderator*

	b [95% CI]	SE	t	p
Scores on the CRT				
Constant	1.784 [1.670; 1.899]	0.058	30.639	$p < .001$
NFC	0.232 [0.092; 0.371]	0.071	3.268	$p = .001$
CRT exposure	0.978 [0.746; 1.209]	0.118	8.308	$p < .001$
Interaction	-0.094 [-0.378; 0.190]	0.144	-0.649	$p = .517$

Model: $R^2 = .19, F(3,361) = 28.605, p < .001$; Change: $\Delta R^2 = .0009, F(1,361) = 1.130, p = .517$

Note. $N = 365$. Table contains unstandardized regression coefficients (b's) with their corresponding 95% confidence intervals, standard errors, t-values and levels of significance. ΔR^2 denotes R-squared change due to interaction (adding the moderator to the regression). Variables were mean centered before the analysis.

solving the problems of the original CRT. Such possibility seems to run counter to the meta-cognitive disadvantage account of Bialek and Pennycook (2017), however, it should be noted that the result is based on participants' self-reported reflectivity, rather than the genuine one. Therefore, the failure of high-NFC scorers to benefit more from their previous exposure when solving the CRT than their counterparts with lower self-report reflectiveness may be due to the fact that the former group consists of both genuinely analytic individuals and intuitive participants, who only believe themselves to be analytic. This point is further explicated in the discussion.

Does Exposure and NFC Moderate the Role of CRT as a Predictor of H&B Tasks?

To address the possibility that previous exposure and self-reported analytic thinking disposition will amplify the relationship between

Table 3 *Moderated moderation analysis of the CRT as a predictor of scores on the H&B tasks with the CRT exposure and NFC as moderators*

	b [95% CI]	SE	t	p
Scores on the H&B composite				
Constant	19.925 [19.276; 20.574]	0.330	60.398	$p < .001$
CRT exposure	0.023 [-1.304; 1.345]	0.674	0.033	$p = .973$
CRT	2.702 [2.164; 3.240]	0.274	9.881	$p < .001$
CRT*CRT exposure	0.492 [-0.623; 1.607]	0.567	0.868	$p = .386$
NFC	1.375 [0.541; 2.209]	0.424	3.242	$p = .001$
CRT*NFC	0.214 [-0.469; 0.896]	0.347	0.616	$p = .538$
CRT exposure*NFC	1.449 [-0.283; 3.181]	0.881	1.645	$p = .101$
Interaction	1.466 [0.024; 2.907]	0.733	2.000	$p = .046$

Model: $R^2 = .34$, $F(7,357) = 26.145$, $p < .001$; Change: $\Delta R^2 = .0074$, $F(1,357) = 3.999$, $p = .046$

Conditional effects of CRT on H&B composite at values of moderators:				
Low NFC – unexp	2.834 [1.966; 3.702]	0.441	6.424	$p < .001$
Low NFC – exp	2.119 [0.813; 3.425]	0.664	3.190	$p = .002$
Mean NFC – unexp	2.491 [1.840; 3.142]	0.331	7.525	$p < .001$
Mean NFC – exp	2.983 [2.077; 3.888]	0.460	6.478	$p < .001$
High NFC – unexp	2.147 [1.209; 3.085]	0.477	4.502	$p < .001$
High NFC – exp	3.847 [2.444; 5.249]	0.713	5.393	$p < .001$

Conditional effects of CRT*CRT exposure interaction at values of NFC:				
Low NFC	-0.715 [-2.283; 0.853]	0.797	-0.897	$p = .370$
Mean NFC	0.492 [-0.623; 1.607]	0.567	0.868	$p = .386$
High NFC	1.699 [0.012; 3.387]	0.858	1.980	$p = .048$

Note. $N = 365$. Table contains unstandardized regression coefficients (b's) with their corresponding 95% confidence intervals, standard errors, *t*-values and levels of significance. ΔR^2 denotes R-squared change due to three-way interaction (adding the moderators to the regression). Low NFC and High NFC reflect one standard deviation below and above the mean of NFC scores in the present sample. Variables were mean centered before the analysis.

CRT and the H&B performance, a moderated moderation analysis[3] was employed, where the CRT (X) was entered as a predictor of H&B tasks (Y) with both previous exposure to the CRT (M) and NFC (W) as moderators. The results are presented in Table 3.

Both CRT and NFC were significant predictors of the performance on the composite of H&B tasks. Importantly, there was a three-way interaction present, indicating the moderated moderation effect, although it was just below the conventional threshold for significance. The middle part of the table shows that CRT was a significant predictor of H&B tasks among unexposed and exposed participants of both intuitive and analytic self-reported cognitive style. However, as the results in the bottom part suggest, the presence of interaction between prior exposure and CRT in predicting H&B tasks was only significant in participants who believed themselves to be analytic, specifically, as the Johnson-Neyman technique (Hayes, 2013) shows, among people who scored above 4.964 on the NFC scale (16% of the present sample). The interaction is depicted in Figure 1. When looking at the high-NFC individuals, CRT predicted H&B scores both among unexposed and exposed participants, but the relationship was stronger among the latter. The same was not observed among people with moderate scores on NFC, who think of themselves as not especially analytic, nor intuitive. In them, as can be seen from the middle panel of Figure 1, CRT predicted H&B tasks to similar extent, regardless of the previous exposure. Interestingly, when looking at the participants with intuitive self-reported cognitive style, the interaction observed among high-NFC individuals seems to reverse (upper panel in Figure 1), and the

conditional effect indeed is in the opposite direction, although it does not reach significance among participants, who scored one standard deviation below the mean NFC.

Does Exposure to the CRT Predict Responses on the H&B Tasks?: An Exploratory Analysis

As shown in the correlation analysis, not only did participants exposed to the CRT achieve higher accuracy on the test itself, but they also scored significantly higher on the H&B composite. To examine this surprising result, I have decided to carry out one additional exploratory moderation analysis, where composite of H&B tasks (Y) was predicted by the CRT exposure (X) and this relationship was moderated by the NFC (M). The results of this analysis are presented in Table 4.
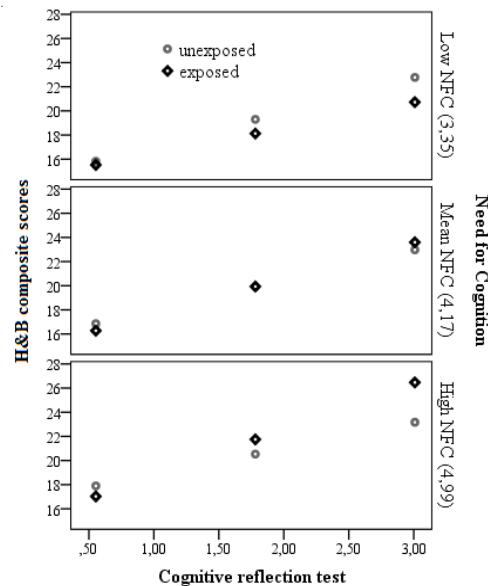

Figure 1 Cognitive reflection test as a predictor of scores on the H&B composite at low, mean, and high levels of NFC among participants exposed and unexposed to the CRT

---

[3] As a template for this analysis, model 3 in Hayes (2013) was used. A visual representation of the analysis is included in the Supplementary material (see Figure S1).

Table 4 *Simple moderation analysis of the CRT exposure as a predictor of scores on the H&B composite and the NFC as a moderator*

	b [95% *CI*]	*SE*	*t*	*p*
Scores on the H&B composite				
Constant	20.054 [19.396; 20.712]	0.335	59.922	*p* < .001
NFC	2.416 [1.615; 3.218]	0.408	5.929	*p* < .001
CRT exposure	2.919 [1.589; 4.249]	0.676	4.317	*p* < .001
Interaction	1.815 [0.183; 3.447]	0.830	2.187	*p* = .029

Model: R^2 = .15, $F(3,361)$ = 20.970, p < .001; Change: ΔR^2 = .0113, $F(1,361)$ = 4.783, p = .029

Conditional effects of CRT exposure on H&B composite at values of NFC:

Low NFC	1.424 [-0.496; 3.344]	0.976	1.459	*p* = .146
Mean NFC	2.919 [1.589; 4.249]	0.676	4.317	*p* < .001
High NFC	4.414 [2.553; 6.275]	0.947	4.664	*p* < .001

Note. N = 365. Table contains unstandardized regression coefficients (*b'*s) with their corresponding 95% confidence intervals, standard errors, *t*-values and levels of significance. ΔR^2 denotes R-squared change due to interaction (adding the moderator to the regression). Low NFC and High NFC reflect one standard deviation below and above the mean of NFC scores in the present sample. Variables were mean centered before the analysis.

As can be seen from the table, while CRT exposure was in itself a significant predictor of the H&B composite, a significant interaction between exposure and the NFC emerged, indicating the presence of the moderation effect. Conditional effects analysis showed that among participants with low NFC previous exposure to the CRT did not help increase the accuracy on various H&B tasks, but in the people with higher self-reported analytic cognitive style, it did. Johnson-Neyman technique showed that prior CRT exposure did not significantly predict scores on the H&B composite among people, who scored less than 3.516 on the NFC (22% of the present sample). If the participants who were familiar with the CRT also already knew some of the H&B tasks, then it might be the case that those who perceive themselves as reflective did gain some insight into the tasks and thus scored higher on them in the present study. On the other hand, participants who believed themselves to be of little reflectivity did not benefit from this potential familiarity with the H&B problems, which could be regarded as an evidence for their metacognitive disadvantage. However, due to the exploratory nature of this analysis, such interpretation should remain cautious.

Discussion

In this study I examined the relationships between prior exposure to CRT, self-reported analytic thinking disposition, and H&B performance. Similarly to the previous research (Haigh, 2016; Stieger & Reips, 2016), 43% of the participants in the present sample indicated they were familiar with the items of CRT before their participation, and these people achieved substantially higher scores on the test than the rest of the sample (*d* = 0.91). As was shown in the first moderation analysis (Table 2), previous

exposure predicted participants' CRT responses independently of their self-reported thinking dispositions. This result might be seen as running counter to the metacognitive disadvantage conjecture of Bialek and Pennycook (2017), who suggested that intuitive participants, unlike their more analytically disposed counterparts, would not realize the tricky nature of the CRT problems even upon repeated exposure and thus would not benefit much from prior exposure when solving the CRT. Yet, as was shown in the study by Pennycook et al. (2017), high self-reported NFC is not only representative of genuinely reflective participants, but also of some intuitive reasoners, who misestimate their true reflectivity. Then high-NFC scorers in the present study may not have benefitted more from their exposure when solving the CRT than participants with low NFC, precisely because the effect of presumed metacognitive advantage of the former may have been attenuated by the subset of genuinely intuitive individuals, who self-identified themselves as analytic. However, before reaching any conclusions on this matter, I shall review other findings of the present study that are also relevant to the metacognitive disadvantage account.

The account of Bialek and Pennycook (2017) was further explored in the model where CRT predicted scores on the H&B composite with both exposure to the CRT and NFC included as possible moderators of this relationship (Table 3). While both CRT and self-reported analytic thinking disposition predicted H&B scores, as in the previous research (Toplak et al., 2011), CRT exposure did not emerge as significant independent predictor of the susceptibility to cognitive biases. Yet, there was a three-way interaction between the predictors indicating the presence of moderated moderation. Conditional effects showed that while CRT was a predictor of H&B tasks regardless of participant's exposure and NFC score, prior familiarity with the CRT increased its predictive value, albeit only

individuals, who perceived themselves as analytic. Moreover, as was shown in an additional exploratory moderation analysis (Table 4), exposure to the CRT in itself predicted H&B performance, and this relationship was further amplified by the self-reported analytic thinking disposition.

Thus, even if the results of the simple moderation might be seen as not in line with the metacognitive disadvantage proposition (Bialek & Pennycook, 2017), subsequent analyses showed that self-reported analytic thinking disposition did play a role in the predictive power of CRT on the H&B tasks among exposed participants, and even moderated the link between exposure to the CRT and the ability to solve various H&B problems, which is quite consistent with the aforementioned account. Furthermore, there are two things that one has to consider when looking at the results of the first moderation reported in this paper. First, a possible explanation for why high-NFC scorers did not show predicted higher performance increase on CRT upon exposure, in comparison with participants who self-identified as intuitive, is that their scores might have been already almost at the ceiling and therefore could not improve much more. The average CRT performance in the present study was quite high even among participants unfamiliar with the test ($M = 1.35$), but among those who were familiar, it was not too far from perfect ($M = 2.36$). As a sidenote, these values are in line with some other studies on the CRT exposure (e.g., Haigh, 2016; Stieger & Reips, 2016). It is plausible then that participants who scored highly on NFC in the present research could not benefit much from previous exposure because the performance of a substantial amount of them was already at the ceiling. If this was the case, then since participants with low self-reported analytic thinking disposition had far more space for improvement on the CRT upon exposure, the fact that they only improved as much as the more analytic

reasoners, who could only perform a little better to begin with, could actually be taken as evidence for their presumed metacognitive disadvantage (Bialek & Pennycook, 2017; Mata, Ferreira, & Sherman, 2013).

The second consideration regarding the first moderation analysis comes from the surprising link between previous exposure to the CRT and increased performance on H&B composite. It might be that people, who were familiar with the CRT also knew the H&B problems, as these methods are often employed together in psychological studies on cognitive biases (Bialek & Pennycook, 2017), and therefore achieved better performance on the latter. While this possibility cannot be ruled out, it would have been more plausible if the present sample came from a participant pool, which is known to be especially likely to be familiar with the CRT, such as Mechanical Turk service or undergraduate psychology students (Haigh, 2016; Thomson & Oppenheimer, 2016). In this study, no recruitment service was used and most of the participants were majoring in various subjects, with only small proportion coming from social sciences (4%). Also, instead of the H&B battery, which is most frequently employed in studies along with the CRT (Bialek & Pennycook, 2017; Toplak et al., 2011), in the present research different sets of problems were used, some of which were constructed specifically for this study and should not have been previously seen by the participants. Importantly, even if some of the participants exposed to the CRT may have seen several of the H&B problems before, although not exactly the ones used here, this would not explain why only those who identified themselves as analytic ended up benefiting from the exposure when solving the H&B battery. However, it could be that only high-NFC scorers benefited from exposure to the CRT when solving H&B problems because they actually understood the logic of the tasks, while more intuitive participants may have encoun-

tered H&B problems already, but they either did not recognize them, or were unable to gain insight into the logic of these tasks and therefore did not improve their performance. Such explanation would then again speak in favor of metacognitive disadvantage of intuitively disposed reasoners (Bialek & Pennycook, 2017; Mata et al., 2013).

Alternative interpretation of the link between exposure to the CRT and H&B performance comes from the suggestion of Meyer et al. (2018) that the ability to recall previous exposure may be related to general intelligence. Thus, people better at recalling their familiarity with the CRT may have had higher cognitive abilities, which have been linked to superior performance on various H&B problems (Pennycook, Cheyne et al., 2015a; Toplak et al., 2014). Exploratory moderation also showed that this relationship was further amplified by the NFC, which would likewise be consistent with the fact that avoidance of cognitive biases requires both intelligence and analytic thinking disposition (e.g., Toplak et al., 2011). Such account would also explain why exposure did not predict H&B performance after the CRT score itself was included in the analysis (Table 3). As the ability to solve the test has been known in part to reflect general cognitive abilities (Frederick, 2005), its inclusion in the regression model may have explained away any difference in intelligence among participants who were able to recall CRT exposure and those who were not.

Putting the differences between analytically disposed reasoners and their intuitive counterparts aside, one additional result of the moderated moderation (Table 3) deserving further notice is that the CRT does not lose its predictive validity in the H&B performance upon previous exposure. While several researchers in the past assumed that familiarity with the CRT automatically invalidates its use (e.g., Haigh, 2016; Stieger & Reips, 2016), the results of the present study converge with recent investiga-

tions, which show that while exposure may lead to an improvement in the CRT performance, it does not seem to affect the test's predictive power in various outcomes related to analytic thinking (Meyer et al., 2018; Stagnaro et al., 2018). Similarly to the observation by Bialek and Pennycook (2017), if anything, exposure to the CRT seemed to increase the predictive power of the test in H&B performance, although this was true only among individuals with high self-reported reflectivity, presumably because of their metacognitive advantage. As the CRT is known to be an important predictor of H&B performance over and above the measures of cognitive ability and thinking dispositions (Toplak et al., 2011), it is important to reiterate that as far as the results of this study go, the test retains its predictive power in the performance on H&B tasks among all participants who claim they have seen or taken it previously, regardless of their self-reported thinking dispositions.

Limitations

The metacognitive disadvantage account (Bialek & Pennycook, 2017) tested in the present study relies on the prediction that intuitive reasoners will not benefit from previous exposure when solving CRT as much as their analytic counterparts because they fail to gain insight into the tricky nature of the problems. However, the performance of intuitive participants might improve if they were exposed to the CRT in a context in which it was presented along with a correct response. As I did not ask participants where they have encountered the CRT before, I cannot rule out the possibility that they saw the test during some academic course or in an internet video, where they were able to learn correct responses without actually having to understand the logic behind the tricky problems. If this was the case with substantial number of intuitive participants, it might

explain why intuitive people also benefitted from their exposure when solving the CRT. To circumvent this possibility, researchers in the future might want to ask their participants not only whether they know the CRT but also whether they know the correct answers to the problems.

Other caveat of the present research stems from the use of self-report NFC scale to study participants' thinking dispositions. While it has been used for this purpose in a great number of studies, recently Pennycook et al. (2017) showed that people who score highly on NFC are actually a mix of genuinely analytical reasoners and intuitive participants, who overestimate their true reflectivity. Based on this, the authors recommend relying on performance measures of thinking style instead of the self-report ones. However, while the CRT or H&B composite score are often employed as performance measures of analytic thinking disposition (Pennycook, Fugelsang et al., 2015a), the ability to solve both of them is also dependent on other factors, such as numeracy or cognitive ability, which may confound the intended effect of the cognitive style. For this reason, a self-report NFC scale was used here to specifically reflect participants' propensity to engage in analytic thinking, without tapping into other related constructs. Still, based on the conclusions of Pennycook et al. (2017), it is important to realize that NFC may be an imperfect indicator of a participant's disposition for analytical thinking. Some of the participants, who scored highly on NFC in the present study, may actually have been intuitively disposed, and this might explain why the expected effect of analytic individuals benefitting more from exposure while solving the CRT did not emerge.

One last limitation that I would like to mention is that while this study presents evidence that the exposure to the CRT increases its predictive power in the H&B tasks among high-NFC individuals, the three-way interaction this finding was based upon was just below the

conventional threshold for significance ($p = .046$). Therefore, it would be wise to wait for other researchers to independently replicate this finding before drawing any strong conclusions from it. Most likely this effect reached only marginal significance because the increase of predictive power of CRT in H&B tasks among exposed analytic participants was not particularly strong. This again might be seen as a consequence of the possibility that the high-NFC group may have been contaminated with some intuitive individuals, who lacked insight into their true reflectivity.

Conclusion

While the CRT remains a popular individual difference measure in the research on cognitive biases, as well as other areas related to analytic thinking (Pennycook, Fugelsang et al., 2015a), many researchers now realize that there are problems with this method stemming from unsatisfactory psychometric properties (Bialek & Pennycook, 2017), questionable nature of the construct that is being measured by it (Pennycook & Ross, 2016), and its increasing familiarity to the general public (Haigh, 2016; Stieger & Reips, 2016). Other issues notwithstanding, the present research converges with the results of several recent studies, which suggest that mere familiarity with the CRT may not actually present such a problem as was previously thought (Bialek & Pennycook, 2017; Meyer et al., 2018; Stagnaro et al., 2018). While participants familiar with the test do achieve higher scores on it, regardless of their self-reported thinking dispositions, the predictive power of CRT on the H&B performance is not lost among exposed participants. If anything, it grows stronger although only among high-NFC individuals, who due to their presumed metacognitive advantage (Mata et al., 2013; Pennycook et al., 2017) gain insight into the tricky nature of the problems and therefore subsequently improve their performance with multiple encounters of the CRT, unlike their intuitive counterparts. While this conjecture remains to be examined in more detail by future research, the results of the present study point out some discrepancies in reasoning processes among participants with analytic and intuitive self-reported thinking disposition, and thus highlight the need to change the focus of research in this area to the individual differences among subgroups of reasoners (Mata et al., 2013; Svedholm-Häkkinen, 2015).

References

Bialek, M., & Pennycook, G. (2017). The cognitive reflection test is robust to multiple exposures. *Behavior Research Methods*, 1–7.

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, *119*(2), 197–253.

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

Haigh, M. (2016). Has the standard Cognitive Reflection Test become a victim of its own success? *Advances in Cognitive Psychology*, *12*(3), 145–149.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis*. London: The Guilford Press.

Mata, A., Ferreira, M. B., & Sherman, S. J. (2013). The metacognitive advantage of deliberative thinkers: A dual-process perspective on overconfidence. *Journal of Personality and Social Psychology*, *105*(3), 353–373.

Meyer, A., Zhou, E., & Frederick, S. (2018). The non-effects of repeated exposure to the Cognitive Reflection Test. *Judgment and Decision Making*, *13*(3), 246–259.

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

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015a). Everyday consequences of analytic thinking. *Current Directions in Psychological Science*, *24*(6), 425–432.

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015b). What makes us think? A three-stage dual-

process model of analytic engagement. *Cognitive Psychology*, *80*, 34–72.

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2014). Cognitive style and religiosity: The role of conflict detection. *Memory & Cognition*, *42*(1), 1–10.

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2015a). On the reception and detection of pseudo-profound bullshit. *Judgment and Decision Making*, *10*(6), 549–563.

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

Pennycook, G., & Ross, R. M. (2016). Commentary: Cognitive reflection vs. calculation in decision making. *Frontiers in Psychology*, *7*(9).

Pennycook, G., Ross, R. M., Koehler, D. J., & Fugelsang, J. A. (2017). Dunning–Kruger effects in reasoning: Theoretical implications of the failure to recognize incompetence. *Psychonomic Bulletin & Review*, *24*(6), 1774–1784.

Stieger, S., & Reips, U.-D. (2016). A limitation of the Cognitive Reflection Test: familiarity. *PeerJ*, *4*(2014), e2395.

Svedholm-Häkkinen, A. M. (2015). Highly reflective reasoners show no signs of belief inhibition. *Acta Psychologica*, *154*, 69–76.

Stagnaro, M. N., Pennycook, G., & Rand, D. G. (2018). Performance on the Cognitive Reflection Test is stable across time. *Judgment and Decision Making*, *13*(3), 260–267.

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, *11*(1), 99–113.

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.