

Disfluent fonts do not help people to solve math and non-math problems regardless of their numeracy

Miroslav Sirota, Andriana Theodoropoulou & Marie Juanchich

Department of Psychology, University of Essex

Author Note

This manuscript was accepted for publication in *Thinking & Reasoning*. This postprint might differ from the published version.

Correspondence concerning this paper should be addressed to Miroslav Sirota, Department of Psychology, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom.

Email: msirota@essex.ac.uk, Phone: (+44) 1206 874 229.

All data sets, R code and materials used in this research are available at <https://osf.io/g65yj/>.

The preregistration protocol is available at <https://aspredicted.org/rd5zj.pdf>

Thanks to Andrew Meyer for sharing the data published in Meyer et al. (2015).

Abstract

Prior research has suggested that perceptual disfluency activates analytical processing and increases the solution rate of mathematical problems with appealing but incorrect answers (i.e., the Cognitive Reflection Test, hereafter CRT). However, a recent meta-analysis does not support such a conclusion. We tested here whether insufficient numerical ability can account for this discrepancy. We found strong evidence against the disfluency effect on the problem-solving rate for the Numerical CRT problems regardless of participants' numeracy and for the Verbal CRT non-math problems ($n = 310$, Exp. 1) even though simple instructions to pay attention to and reflect upon the Verbal CRT problems substantially increased their solution rate ($n = 311$, Exp. 2). The updated meta-analysis ($k = 18$) yielded close-to-zero effect, Hedge's $g = -0.01$, 95% CI[-0.05, 0.03] and decisive evidence against the disfluency effect on math problems, $BF_{0+} = 151.6$. Thus, perceptual disfluency does not activate analytical processing.

Keywords: Cognitive Reflection Test, Verbal Cognitive Reflection Test, disfluent font, numeracy, fluency

Dual-process theories of cognition postulate two types of cognitive processes: intuitive processes typically described as fast, automatic and frugal, and analytical processes typically described as slow, effortful and reflective (e.g., Darlow & Sloman, 2010; Evans, 2008; Evans & Stanovich, 2013; Morewedge & Kahneman, 2010; Reyna, 2004; Sloman, 1996). Dual-process theories have successfully accounted for various psychological effects in judgment and decision-making, deductive and inductive reasoning, moral reasoning and beliefs (e.g., Frederick, 2005; Liberali, Reyna, Furlan, Stein, & Pardo, 2012; Pennycook, Fugelsang, & Koehler, 2015; Sirota & Juanchich, 2018; Sirota, Juanchich, & Hagemayer, 2014; Toplak, West, & Stanovich, 2011, 2014). The recent theoretical discussion has focused on the mechanisms of activation of the analytical process (e.g., a serial or parallel activation model) and which cues can trigger them (e.g., Bhatia, 2017; De Neys, 2012 for a review).

In respect to this research problem, prior research has suggested that meta-cognitive difficulty—the subjective experience of ease or difficulty associated with processing information—might be one of the mechanisms triggering analytical processing when solving problems (Alter, Oppenheimer, Epley, & Eyre, 2007; Oppenheimer, 2008). In the absence of other cues, a feeling of ease associated with processing information will trigger more intuitive processing, whereas a feeling of meta-cognitive difficulty will trigger more systematic, analytical thinking. Some previous studies tested this idea by manipulating perceptual (dis-)fluency: presenting problems in a fluent (i.e., easy-to-read) font and a disfluent (i.e., difficult-to-read) font. This manipulation relied on the idea that the difficulty associated with reading a problem printed in the disfluent font would be substituted for the meta-cognitive difficulty of the problem itself: as in “if it is hard to read, it must be hard to solve”. For instance, Alter et al. (2007), presented participants with three mathematical problems that had appealing intuitive but incorrect answers, known as the Cognitive Reflection Test, (Frederick, 2005) either in a fluent or disfluent font. The participants

correctly solved more problems in the disfluent font condition than in the fluent font condition. According to the authors, the participants reading the problems in the disfluent font adopted a more systematic processing strategy, which resulted in improved performance. Thus, meta-cognitive experiences of disfluency served as a signal that the initially generated judgment was insufficient and that more elaborate systematic thinking was necessary (Alter et al., 2007).

However, several studies did not replicate the effect of disfluency on the solution rate of the three mathematical problems (Meyer et al., 2015; Thompson et al., 2013). A meta-analysis of the original study from Alter et al. (2007) and 16 subsequent replication studies found no overall disfluency effect on performance in the Cognitive Reflection Test (Meyer et al., 2015). None of the replication studies found a statistically significant disfluency effect. In theory, two main explanations could be considered for a possible discrepancy between the original study and the lack of disfluency effect reported in the replication studies.

The first explanation could be that either the original study or the set of replication studies do not adequately represent the actual population-level disfluency effect (or lack of it). It might be that the original study reported a false-positive finding (Type I error) or it might be that the subsequent replications reported false-negative findings (Type II error). The false-negative findings explanation can hardly be considered a plausible explanation: 7,327 participants amassed in the 16 studies would be sufficient to detect any reasonably small effect (i.e., $d = 0.1$ or bigger) with extraordinary statistical power (.99 or more with the increasing effect size). However, the false-positive finding of the original study remains quite possible. Supporting this possibility, Meyer et al. (2015) noted that, in the original study, the difference in performance between the disfluent and fluent font groups (each

having 20 people) was driven solely by the unusually low performance of the widget problem in the fluent font condition (4 out of 20 responses vs 16 out of 20 responses in the disfluent font condition). No difference was observed between the disfluent and fluent fonts in the bat and ball problem and the lily pad problem (16/20 vs 15/20, 18/20 vs 18/20, respectively). This means that sampling variation solely could explain the discrepant findings.

The second explanation for the discrepant findings could be the existence of hidden moderators. For instance, the authors of the original study suggested that cognitive ability might play a role: disfluency might benefit only those with high cognitive ability (Alter, Oppenheimer, & Epley, 2013). Indeed, some limited evidence suggested that this might be the case: only the participants with higher intelligence benefited from the disfluent manipulation in terms of enhanced performance (Thompson et al., 2013). However, the recent meta-analysis did not lend support to this possibility (Meyer et al., 2015). The authors provided several arguments against it: most notably, they showed that intelligence—measured either by selected items from Raven’s Progressive Matrices or by self-reported SAT math scores—did not interact with the font manipulation to predict enhanced performance in the Cognitive Reflection Test (Meyer et al., 2015). Also, the disfluency effect was not moderated by the experimental setting (in public, in a lab or online), presentation format (paper and pencil vs computer screen) and previous exposure to the problems (Meyer et al., 2015). Finally, it can be argued that the negligible amount of heterogeneity between the studies does not warrant an exhaustive search for hidden moderators (Meyer et al., 2015).

Nevertheless, we believe that one additional moderator should be considered. One obvious candidate—a distinct ability required to solve the Cognitive Reflection Test, which is conceptually and operationally different from cognitive ability—is numeracy. Indeed, prior research found that solving the CRT requires numerical skills over and above cognitive

abilities (Liberali et al., 2012; Pennycook & Ross, 2016; Sinayev & Peters, 2015; Sirota, Kostovičová, Juanchich, Dewberry, & Marshall, 2018). The Cognitive Reflection Test has been used widely to measure the extent to which people rely on analytical processing (e.g., Pennycook & Rand, 2018, 2019), but it has also been widely criticised for measuring numeracy (e.g., Sinayev & Peters, 2015). This implies that the disfluent font (i) increases performance in the numerical version of the Cognitive Reflection Test only for those who have sufficient mathematical skills to benefit from the triggered analytical processing and (ii) increases performance in cognitive reflection problems that do not require numerical skills at all (e.g., Sirota et al., 2018).

We found mixed evidence to support this moderating role of numeracy. On one hand, perceptual disfluency enhanced the solution rate of the problems that did not require any numerical skills, but which were similar to those featured in the CRT in terms of offering an intuitive but incorrect first answer; for example, oversight problems such as the Moses illusion (Song & Schwarz, 2008) and the belief-bias syllogism (Alter et al., 2007). On the other hand, however, three pieces of evidence warrant some pessimism. First, some subsequent studies failed to replicate the effect of disfluency on non-numeric problems (Meyer et al., 2015; Morsanyi & Handley, 2012; Thompson et al., 2013), which might be because the original studies have their own methodological and analytical limitations (e.g., not correcting for multiple comparisons; excluding some items due to the floor/ceiling effect). Second, Meyer et al. (2015) described an unpublished study in which the font manipulation did not affect the solution rate of those participants who successfully solved the problem in the second round once they received a hint about the incorrect intuitive response. This means that the font manipulation did not affect the performance of those people who had sufficient mathematical skills to solve the problem since they managed to solve it in the second round. Finally, self-reported SAT math scores were not found to

moderate the disfluent font effect of the numerical problems across five different studies. This is strong evidence against numeracy as a moderator, but still needs to be replicated with an objective measure of numeracy, which would circumvent possible issues with inflated self-reported SAT scores (Cassady, 2001; Shepperd, 1993). Hence, this mixed evidence warrants further investigation of numeracy as a possible moderator of the disfluency effect on solving CRT problems, while measuring numeracy objectively, and of the disfluency effect on cognitive reflection problems that do not require any numerical skills—the verbal CRT (Sirota et al., 2018).

The Present Research

In the manuscript presented here, we aimed to test the effect of perceptual disfluency on cognitive reflection, while testing possible boundary conditions of the disfluency effect. Specifically, we made three advancements. First, we measured the numerical skills using the objective performance rather than the self-reported performance in a standard numeracy test (Lipkus, Samsa, & Rimer, 2001). Second, we measured cognitive reflection using an extended version of the Numerical Cognitive Reflection Test, which contains the three critical items of the original version of the CRT, but also four additional items (Frederick, 2005; Toplak et al., 2014). The extended version of the CRT has better statistical and psychometric properties, which makes it more likely to detect an effect since it has better score variability and internal consistency (Toplak et al., 2014). Finally, we also measured cognitive reflection using the Verbal Cognitive Reflection Test, which features problems that involve cognitive reflection but do not require any mathematical calculations (Sirota et al., 2018).

We derived three main hypotheses from the assumption that the disfluency effect exists. First, according to a strong version of the assumption that the disfluency effect exists

and operates independently from numeracy, we can expect that the disfluent font will increase the performance in the Numerical Cognitive Reflection Test compared with the fluent font (Hypothesis 1, Experiment 1). We expected, given the overwhelming evidence demonstrating the lack of the disfluency effect, that this hypothesis would not be confirmed. According to a weaker version of the assumption that the disfluency effect exists, the disfluent font triggers analytical processing but will manifest only when participants have sufficient numerical skills to solve the problems. We derived two hypotheses from this version of the assumption. First, disfluency will increase performance in the Numerical Cognitive Reflection Test, but only for those who have enough numeracy skills to solve the mathematical problems (Hypothesis 2, Experiment 1). Second, disfluency will increase performance in the Verbal Cognitive Reflection Test, since it does not require any mathematical skills (Hypothesis 3, Experiment 1). This last hypothesis hinges on the assumption that increased analytical processing translates into increased performance in the Verbal CRT. To test this assumption, we designed a new experiment in which we asked people explicitly to pay more attention to and reflect upon the tricky verbal problems. We expected that this explicit warning about the problems requiring reflection would elicit a higher solution rate in the Verbal CRT compared with the standard instructions (Hypothesis 4, Experiment 2).

Experiment 1

Method

Participants and design. We recruited 312 psychology undergraduate students from the University of Essex in exchange for course credits. The sample size was based on an a-priori stopping rule to reach a minimal sample size of 300 participants—comprising a sample of 296 participants required to detect a medium effect size, Cohen's $d = 0.50$ for an

independent samples t -test with high power ($\alpha = .05$, $1 - \beta = .99$ for a two-sided test) and an additional 4 students to account for attrition rate (Cohen, 1988). We collected data from slightly more students because we let all the students who had signed up for the study participate. We excluded one participant due to an error in materials (i.e., having two Verbal CRTs instead of one Verbal and one Numerical CRT) and one participant due to a failure to fulfil the eligibility criteria (i.e., reported age of 17 years old). The resulting sample size of 310 allowed us to detect a small effect size (Cohen's $d = 0.32$), while using an independent samples t -test and assuming $\alpha = .05$, $1 - \beta = .80$, and a two-sided test (Cohen, 1988). The achieved statistical power of the test was approximating 1.0 for the originally reported effect, Cohen's $d = 0.71$ (Alter et al., 2007). Hence, this was a sufficient sample size to assess the effect. Participants were mostly women (79.9%). The participants' ages ranged from 18 to 43 years ($M = 19.4$, $SD = 2.2$ years).

Participants solved the word problems from the Numerical Cognitive Reflection Test (Frederick, 2005; Toplak et al., 2014) and the Verbal Cognitive Reflection Test (counterbalanced for order) either in a fluent, easy-to-read font (i.e., black Myriad Web 12-point font) or in a disfluent, difficult-to-read font (i.e., 10% grey italicised Myriad Web 10-point font)—the same manipulation that was used in the original study (Alter et al., 2007). Thus, we had a 2(font: fluent font vs disfluent font) \times 2(CRT: Numerical vs Verbal) mixed-subjects design, with the font as a between-subjects factor and the CRT as a within-subjects factor with a counterbalanced order of presentation. We conducted a manipulation check on a separate sample of 20 participants from the same pool who rated the extent to which the same text (i.e., CRT questions) printed in the two fonts was easy/hard to read using a 5-point Likert scale (anchored as *1: extremely easy to read*, *5: extremely hard to read*). The counterbalanced presentation order of the two texts did not significantly alter the ratings for the fluent or disfluent font, $t(18) = 0$, $p = 1.000$, Cohen's $d = 0$; $t(18) = 0.75$, $p = .461$,

Cohen's $d = 0.34$, respectively; hence, it was not considered further. Similar to the prior reported research, the fluent font ($M = 1.2$, $SD = 0.4$) compared with the disfluent font ($M = 3.7$, $SD = 1.2$) was judged to be substantially and statistically significantly easier to read, $t(19) = -8.75$, $p < .001$, Cohen's $d = -1.96$. The manipulation was thus very effective.

Materials and procedure. After providing informed consent, participants were instructed to turn their phones off and were allocated to an individual booth without access to a computer or the internet. This was done to prevent the participants from searching online for the answers to the problems. Participants then solved the problems of the Numerical CRT or the Verbal CRT (depending on the random order) in a paper-and-pencil format. We used the extended 7-item version of the Numerical CRT (Toplak et al., 2014) for its good psychometric and statistical properties. It comprised the original three CRT items (Frederick, 2005) and four additional items (Toplak et al., 2014). A summation score (ranging from 0 to 7, transformed into a 0–100% scale for ease of comparison with the Verbal CRT performance) had an acceptable internal consistency (Cronbach's $\alpha = 0.74$); higher scores indicated more cognitive reflection. We used the 10-item version of the Verbal CRT (Sirota et al., 2018) requiring cognitive reflection while solving word problems that do not involve mathematical calculations (e.g., “Mary’s father has 5 daughters but no sons – Nana, Nene, Nini, Nono. What is the fifth daughter’s name probably?”). A summation score (ranging from 0 to 10, transformed into a 0–100% scale) had a good internal consistency (Cronbach's $\alpha = 0.83$); higher scores indicated more cognitive reflection.

Participants then answered numerical ability items using Lipkus et al.'s Numeracy Scale, a very common measure of numerical ability (Lipkus et al., 2001). It consists of 11 simple mathematical word problems that require an understanding of basic probability concepts, the ability to convert percentages to proportions, and the ability to compare

different risk magnitudes (e.g., “If Person A’s chance of getting a disease is 1 in 100 in ten years, and person B’s risk is double that of A’s, what is B’s risk?”) (e.g., Liberali et al., 2012; Sirota et al., 2018). A summation score (ranging from 0 to 11) had a questionable internal consistency (Cronbach’s $\alpha = 0.61$) but similar to the one reported by prior research (e.g., Liberali et al., 2012; Sirota et al., 2018); a higher score indicated better numeracy.

Finally, the participants answered socio-demographic questions about their age and gender and were debriefed. We conducted the study in accordance with the ethical standards of the University of Essex Ethics Committee and the APA ethical guidelines. We have reported all the experiments, measures, manipulations and exclusions. The data set, R code used for analysis and the materials used in this research are available on the Open Science Framework: <https://osf.io/g65yj/>.

Results and Discussion

Participants solved correctly, on average, 27.8% of the Numerical CRT problems¹. They solved slightly more word problems in the fluent font compared with the disfluent font condition (Figure 1). Thus, the observed effect was in the opposite direction to the prediction of the disfluency effect hypothesis even though it was not statistically significant, $t(308) = -0.39$, $p = .700$, $d = -0.04$. Since not rejecting the null hypothesis does not logically entail accepting the null hypothesis, we quantified the evidence for the disfluency effect hypothesis relative to the null hypothesis that included the opposite direction effect using a Bayes factor analysis (Lee & Wagenmakers, 2014; Morey & Rouder, 2015). We found strong evidence against the disfluency effect hypothesis, $BF_{0+} = 10.5$ (both using a default Cauchy

¹We found no evidence for CRT presentation order effect on cognitive reflection, $F(1, 306) = 2.76$, $p = .097$, its interaction with the font, $F(1, 306) = 1.10$, $p = .296$ or with the type of CRT, $F(1, 306) = 1.75$, $p = .186$. Therefore, we did not consider the presentation order factor in further analyses.

prior of 0.707) while robustness analysis found similar results using different Cauchy prior widths ($BF_{0+} = 14.8$ for wide and $BF_{0+} = 20.8$ for ultra-wide priors). In addition, we checked the disfluency effect for individual CRT items (given that the effect was mostly driven by differences for one item in the original study), but we did not find any significant differences (Table 1). Thus, we found evidence against the disfluency effect hypothesis applied to mathematical CRT problems (Hypothesis 1).

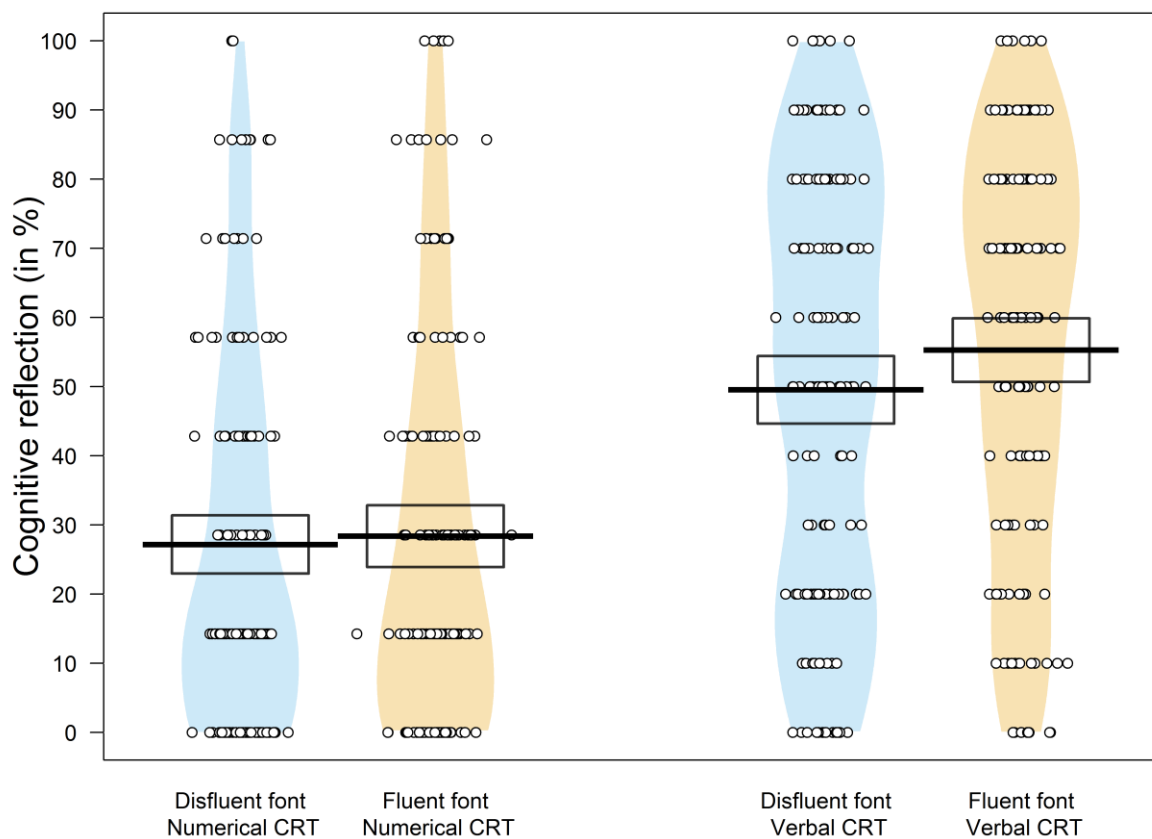


Figure 1: Effect of font and type of Cognitive Reflection Test (CRT) on performance in cognitive reflection (in % of correctly solved problems—cognitive reflection score).

Note. Horizontal lines represent means, boxes represent 95% confidence intervals, beans represent smoothed densities and circles represent individual responses.

Table 1 The effect of font manipulation on the number of correct responses (in %) for each problem in the Numerical Cognitive Reflection Test and the Verbal Cognitive Reflection Test.

	Correct responses (in %)		Differences between the disfluent and fluent font		
	Disfluent font	Fluent font	χ^2	p	ϕ
<i>Numerical CRT</i>					
NCRT 1	29.7	27.1	0.14	.706	-0.03
NCRT 2	27.7	27.7	<0.01	1.000	0.00
NCRT 3	29.7	31.6	0.06	.805	0.02
NCRT 4	14.8	22.6	2.57	.109	0.10
NCRT 5	14.2	14.2	<0.01	1.000	0.00
NCRT 6	33.5	27.1	1.24	.267	-0.07
NCRT 7	40.6	48.4	1.58	.209	0.08
<i>Verbal CRT</i>					
VCRT 1	19.4	21.3	0.08	.778	0.02
VCRT 2	65.8	72.3	1.22	.269	0.07
VCRT 3	65.2	78.7	6.39	.011	0.15
VCRT 4	60.6	68.4	1.71	.192	0.08
VCRT 5	68.4	75.5	1.60	.206	0.08
VCRT 6	42.6	48.4	0.83	.362	0.06
VCRT 7	41.3	46.5	0.61	.423	0.05
VCRT 8	45.2	45.2	<0.01	1.000	<0.01
VCRT 9	39.4	41.9	0.12	.729	0.03
VCRT 10	47.7	54.8	1.29	.256	0.07

Note. Given the number of comparisons (n), the critical p -value (p) of the chi-squared test (χ^2) was adjusted using the Bonferroni adjustment ($\alpha = .05/n$)—for the Numerical CRT the critical value is $p = .007$ and for the Verbal CRT it is $p = .005$; ϕ indicates the phi-coefficient.

Participants with higher numeracy performed substantially better with the CRT math problems than those with lower numeracy ($r = 0.44$), $t(308) = 8.62$, $p < .001$; for illustration purposes only: the lower numeracy group (median split)—with a numeracy score smaller than 9 out of 11—solved on average 18.5% of the problems, $SD = 21.7\%$ and the higher numeracy group—those with a numeracy score equal to 9 or more—solved on average 37.9% of the problems, $SD = 29.1\%$. This finding enabled us to persuasively test the second hypothesis that the disfluency effect is more pronounced in those with higher numeracy.

This was not the case. In a multiple linear regression model, we found no significant interaction between font and numeracy, $b = 0.2$, $t = 0.11$, $p = .912$. Using the Bayes factor framework, the model featuring only the main effects was preferred relative to the full model featuring the main effects and the interaction term, $BF_{12} = 8.4$ and the model featuring only numeracy was preferred relative to the full model, $BF_{12} = 62.6$. In other terms, we found substantive relative evidence against the interaction and very strong relative evidence against the effect of the font and its interaction with numeracy. Thus, we found evidence against the possibility that numeracy moderates the effect of disfluency on the numerical CRT (Hypothesis 2).

Participants solved correctly, on average, 52.4% of the Verbal CRT problems. They solved slightly more word problems in the fluent font compared with the disfluent font (Figure 1). Thus, the observed effect was again in the opposite direction to the one predicted by the disfluency effect hypothesis. However, it was not statistically significant, $t(308) = -1.69$, $p = .091$, $d = -0.19$. Using a Bayes factor analysis, we found strong evidence against the disfluency effect hypothesis, $BF_{0+} = 20.7$ (both using a default Cauchy prior of 0.707) while robustness analysis found similar results using different Cauchy prior widths ($BF_{0+} = 29.1$ for wide and $BF_{0+} = 41.1$ for ultra-wide priors). In addition, we tested the disfluency effect on individual items, but we did not find any significant differences (Table 1). Thus, even for the Verbal CRT, we found evidence against the disfluency hypothesis applied to non-mathematical problems tapping into cognitive reflection ability (Hypothesis 3).

Finally, to probe the robustness of the analyses reported above, we used a two-way analysis of variance with the font as a between-subjects factor and the Cognitive Reflection Test as a within-subjects factor and the cognitive reflection performance as the dependent variable. We reached the same conclusion. Consistent with the previous observations and

tests, the main effect of the font was not statistically significant, $F(1, 308) = 1.57, p = .210$; the effect of the type of CRT was significant (with the problems of the Verbal CRT being easier), $F(1, 308) = 207.49, p < .001$ and the interaction with the font was not, $F(1, 308) = 1.77, p = .185$. (Adding numeracy as a covariate into the model did not change the conclusions about the results.)

Data Synthesis: Estimating the disfluency effect on cognitive reflection with math problems

To update the meta-analytical disfluency effect on solving math problems reported by Meyer et al. (2015), we computed a meta-analysis across all published studies reported in Meyer et al. (2015) and the findings concerning the Numerical CRT reported here. We used a random-effects model as implemented in the R package ‘metafor’ (Viechtbauer, 2010). The overall effect was virtually zero, Hedge’s $g = -0.01$, 95% CI [-0.05, 0.03], $z = -0.44, p = .661$. We observed only minimal between-study heterogeneity in the effect sizes, $\tau^2 < 0.01$ (Cochran’s Q test was not significant, $Q(17) = 15.60, p = .553$) that occurred due to sampling error variability alone, $I^2 = 0.01\%$. We also computed meta-analytical Bayes factors on the same datasets using the R package ‘BayesFactor’ (Morey & Rouder, 2015). We found decisive evidence against the disfluency effect, $BF_{0+} = 151.6$ (using JSZ Bayes priors with an rscale of 1) relative to the null effect and the opposite effect, while robustness analysis found similar results using different Cauchy prior widths ($BF_{0+} = 107.2$ for rscale = 0.707 and $BF_{0+} = 214.4$ for rscale 1.414).

Experiment 2

In Experiment 1, we found evidence against the disfluency effect on the Numerical Cognitive Reflection Test in people with low and high numeracy. However, a critical part of this evidence relied on a new cognitive reflection measure that does not require any numerical

abilities—the Verbal Cognitive Reflection Test. The null effect on the Verbal CRT might indicate either the authentic absence of the disfluency effect or a lack of sensitivity in our measure to detect it—perceptual disfluency might have triggered analytical processing but the measure was not sensitive enough to capture its effect. To disentangle these two interpretations, in a pre-registered experiment, we conducted a positive control check of this new instrument following the recommendation of Alter et al. (2013) to test whether explicit instructions to be analytical improves accuracy in the task. In the experimental condition, we instructed participants to pay attention and direct their mental effort to the problems to solve them accurately. In the control condition, we used standard instructions. As a manipulation check, we expected that the participants in the experimental condition would spend more time on the problems than the participants in the control condition. We hypothesised that the participants in the experimental condition would be more likely to solve the verbal problems correctly than the participants in the control condition (Hypothesis 4).

Method

Participants and design. We recruited 311 participants from an online panel (Prolific Academic) who were invited to complete a short study in exchange for financial compensation (the average reward per hour was £7.16). The sample size was based on an a-priori stopping rule to reach a minimal sample size of 310 participants with an identical power consideration to Experiment 1. Participants were eligible to take part only if they had a minimal 90% approval rate in previous studies, did not participate in the previous studies using the Verbal CRT run by our lab, their first language was English, they were UK nationals and were at least 18 years old. Three participants were automatically timed out by the recruiting system, and we did not exclude any participants from the rest of the sample.

Participants were women (65.0%), men (34.7%) or of other gender identity (0.3%). Their ages ranged from 18 to 74 years ($M = 38.3$, $SD = 13.2$ years).

In a between-subjects design, participants solved five word problems selected from the Verbal Cognitive Reflection Test following standard instructions (control condition) or instructions to be analytical (experimental condition).

Materials and procedure. After providing informed consent, participants were randomly allocated to one of the two conditions. In the control condition, they read a standard set of instructions (equivalent to those in Experiment 1: “On the next pages you will see five problems that vary in difficulty. Answer as many as you can.”) In the experimental condition, they read a set of instructions designed to trigger analytical processing: “On the next pages you will see five problems that vary in difficulty. Answer as many as you can. Please note that all five of them are ‘tricky’ problems. So, if you want to solve them correctly you should pay close attention to the wording of the questions and reflect on how the problem is trying to trick you.” They were then asked to confirm that they had read the instructions by answering a multiple-choice question (see Supplementary Materials). If they failed to answer correctly, they were asked to read the instructions and the associated multiple-choice question again. Only those who selected the correct response (“Based on the instructions I have just read, the problems I am going to solve are: a/ tricky problems that require my full attention”) were allowed to proceed. Participants in both conditions then solved five problems from the Verbal Cognitive Reflection Test (Sirota et al., 2018): the Mary’s father problem, the Moses illusion, the electric train problem, the match problem and the widow problem. These were presented in a random order for each participant. A summation score (ranging from 0 to 5) had an acceptable internal consistency (Cronbach’s $\alpha = 0.74$); higher scores indicated more cognitive reflection. Finally, the participants answered socio-

demographic questions about their age and gender and were debriefed. We conducted the study following the ethical standards of the University of Essex Ethics Committee and the APA ethical guidelines. We have reported all the experiments, measures, manipulations and exclusions. The pre-registration, data set, R code used for analysis and the materials used in this research are available on the Open Science Framework: <https://osf.io/g65yj/>.

Results and Discussion

Participants spent more time trying to solve the problems in the experimental condition ($Mdn = 120.3$, $IQR = 85.3$ seconds) compared with those in the control condition ($Mdn = 97.0$, $IQR = 64.8$ seconds). The effect was statistically significant, $W = 9308$, $p < .001$ (as pre-registered, a non-parametric Mann–Whitney U test was used due to a high positive skewness of the time variable, i.e., $+2.3$). Thus, the participants who were instructed to be analytical spent more time trying to solve the problems, which indicates that the manipulation was effective.

Crucially, participants solved substantially more of the CRT items correctly in the experimental condition compared with the control condition (Figure 1). This medium-to-large effect was statistically significant, $t(309) = -5.65$, $p < .001$, $d = -0.64$. The data provided decisive evidence supporting our warning effect hypothesis, $BF_{.0} = 575000$ (using a default Cauchy prior of 0.707), while robustness analysis found similar results using different Cauchy prior widths ($BF_{.0} = 519702$ for wide and $BF_{.0} = 428737$ for ultra-wide priors). In addition, we checked the effect of the explicit instructions to be analytical for each item and found significant differences in the predicted direction for every single one (Table 2). Thus, as predicted, we confirmed that explicit instructions to be analytical translated into improved accuracy (Hypothesis 4).

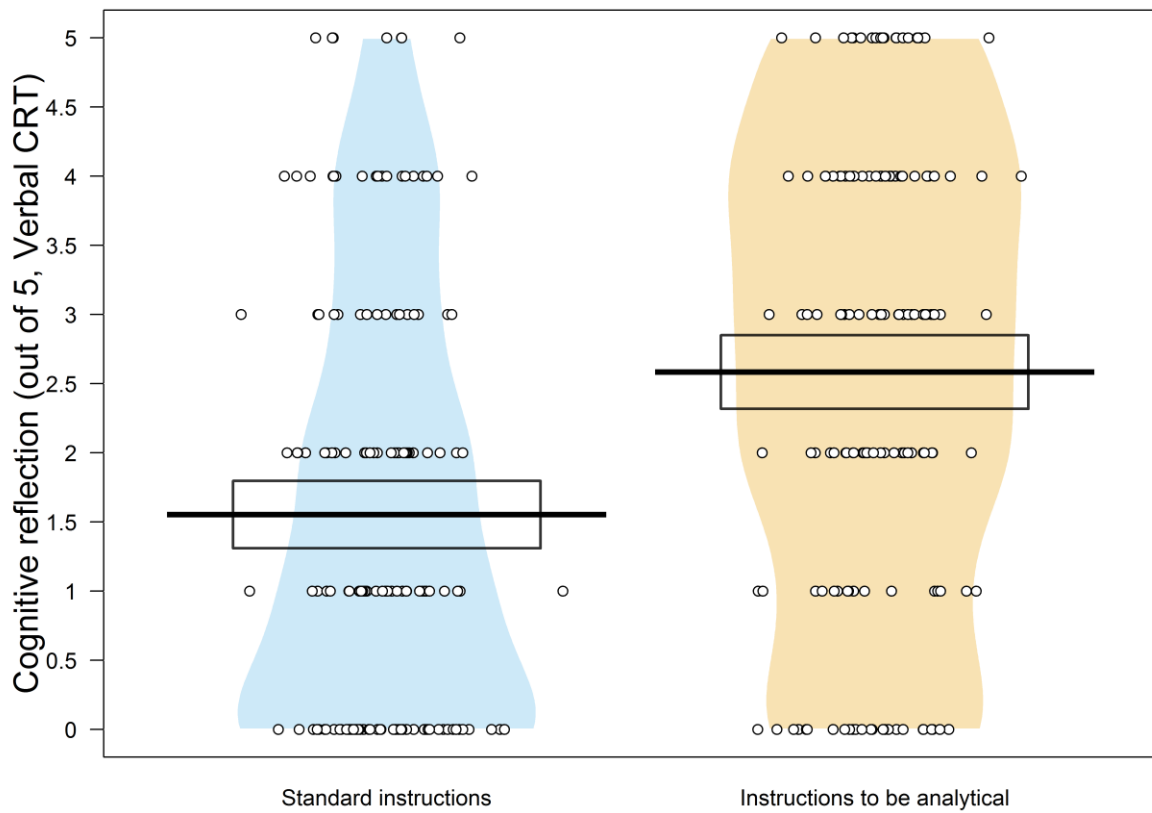


Figure 2: Effect of explicit instructions to be analytical (vs. standard instructions) on performance in cognitive reflection (the number of correctly solved Verbal CRT problems out of five).

Note. Horizontal lines represent means, boxes represent 95% confidence intervals, beans represent smoothed densities and circles represent individual responses.

Table 2 The effect of explicit instructions to be analytical (vs. standard instructions) on the number of correct responses (in %) for each selected problem of the Verbal Cognitive Reflection Test.

	Correct responses (in %)		Differences between the two conditions		
	Standard instructions	Be analytical instructions	χ^2	p	ϕ
<i>Verbal CRT</i>					
VCRT 1	46.5	68.8	14.97	<.001	0.23
VCRT 2	21.7	39.0	10.23	.001	0.19
VCRT 3	33.1	50.0	8.44	.004	0.17
VCRT 4	33.8	61.7	23.21	<.001	0.28
VCRT 5	20.4	39.0	12.01	<.001	0.20

Note. Given the number of comparisons (n), the critical p -value (p) of the chi-squared test

(χ^2) was adjusted using the Bonferroni adjustment ($\alpha = .05/n$)—the critical value is $p = .01$;

ϕ indicates the phi-coefficient.

General Discussion

In the two experiments reported here, we found additional evidence against the hypothesis that a disfluent font increases analytical reasoning that would manifest into improved performance in two versions of cognitive reflection tests: the numerical CRT (Frederick, 2005; Toplak et al., 2014) and the verbal CRT (Sirota et al., 2018). The null effect of the disfluent font on performance in the numerical Cognitive Reflection Test contrasts with the original findings (Alter et al., 2007), but it is aligned with the reported results from other authors (Meyer et al., 2015; Thompson et al., 2013). Furthermore, the updated meta-analysis, which included the original study (Alter et al., 2007), 16 additional studies (Meyer et al., 2015) and our study (Experiment 1), yielded a close-to-zero estimate of the disfluency effect on accuracy and a Bayes factor meta-analysis of these studies allowed us to quantify the strength of this evidence as decisive evidence against the effect of disfluency on cognitive reflection.

We also did not find support for the version of the disfluency hypothesis that considers numeracy as a boundary condition of the disfluency effect. According to this hypothesis, the effect of disfluency might not translate into higher cognitive reflection performance because people also needed to have sufficient numerical skills to benefit from increased analytical reasoning. Two pieces of evidence presented here go against this hypothesis. First, one would expect the disfluency effect to occur for those with higher numerical skills. However, we found no significant interaction of the font manipulation with numeracy on cognitive reflection. This is aligned with the evidence that self-reported SAT math scores did not moderate the disfluency effect across five different studies (Meyer et al., 2015). Our results extend these findings as we measured numeracy objectively. Second, one would expect the disfluency effect to occur for the problems that do not require numerical skills, as reaching the correct solution here would be easier once people inhibit their intuition and adopt a more analytical approach. However, we only found evidence against the disfluency effect on performance in the Verbal Cognitive Reflection Test (Sirota et al., 2018). This is despite the fact that enhanced analytical processing induced by an explicit warning about the tricky nature of the problem increased accuracy on selected Verbal CRT problems in Experiment 2. This further undermines the idea that a lack of numeracy explains the absence of the disfluency effect, since one does not require numerical skills to answer these items. Finally, our results with the Verbal CRT also add more weight to the disconfirming evidence that a disfluent font activates analytical processing in the non-numerical problems requiring deliberation such as belief bias syllogism or oversight problems (Meyer et al., 2015; Morsanyi & Handley, 2012; Thompson et al., 2013).

Taken together, our results indicate that a disfluent font does not trigger analytical processing when solving cognitive reflection problems (Alter et al., 2007). A statistical variation can account for the fact that the original study reported a significant effect on CRT

problems. As noted by Meyer et al. (2015), the disfluency effect was driven by unusually low performance in the control, fluent condition for the “widgets problem”; the rest of the problems had very similar performances across the conditions. This, in combination with very small samples used in the original research (20 participants per condition), can account for the discrepancy between the original study and the remaining 17 studies trying to replicate the finding. Although such an explanation seems the most parsimonious, we should also consider alternative explanations. The most obvious one is that solving problems featured in either of the cognitive reflection tests does not require extensive analytical processing. Supporting this possibility, recent evidence coming from thinking aloud protocols and dual-response paradigms indicates that most of the participants who answered the CRT problems correctly provided the answers relatively quickly and without extensive corrective deliberation (Bago & De Neys, 2019; Szaszi, Szollosi, Palfi, & Aczel, 2017). This could mean that a disfluent font perhaps activated analytical processing, but that this was not required to solve the problems. However, we observed that instructing participants to be analytical substantially increased the accuracy of the Verbal CRT problems. Besides, if analytical processing was not required, then disfluency should not have affected cognitive reflection performance in the original study either. Hence, regardless of whether the CRT requires extensive analytical reasoning or not, one needs to assume that the original study reported a false-positive finding (Alter et al., 2007).

Several limitations of our research deserve more attention here. First, our study was not an exact replication of the original study, because it included an additional CRT problem. While we used the same font manipulation as in the original research with an effective manipulation check, we cannot exclude the possibility that presenting more items is eroding the disfluency effect. One could counter-argue, however, that longer exposure to a disfluent font (via an increased number of written materials) should activate a stronger—not

a weaker—effect. One would also expect at least some disfluency effect on the first set of items since the order of the items within each instrument was fixed. This was not the case here. Second, our sample consisted of undergraduate psychology students who performed substantially lower in the Cognitive Reflection Test than the students of Princeton University participating in the original research. Finally, although we showed that the fluency manipulation was effective and the Verbal CRT measure was sensitive to explicit analytical instructions, further positive controls for the Numerical CRT will be needed.

To summarise, we found several pieces of evidence *against* the hypothesis that a disfluent font activates analytical processing that would, in turn, increase the solution rate of the mathematical problems used in the Cognitive Reflection Test. Such a disfluency effect was not moderated by numeracy since it was absent from the performance of people with lower as well as higher numerical ability. Finally, the disfluent font did not increase the solution rate of the non-mathematical and easier problems used in the Verbal Cognitive Reflection Test. The evidence presented here, along with prior failed replications, corroborates the null effect of disfluent fonts on cognitive reflection.

References

- Alter, A. L., Oppenheimer, D. M., & Epley, N. (2013). Disfluency prompts analytic thinking-
But not always greater accuracy: Response to Thompson et al. (2013). *Cognition*, *128*,
252-255. doi: 10.1016/j.cognition.2013.01.006
- Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition:
Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology-
General*, *136*, 569-576. doi: 10.1037/0096-3445.136.4.569
- Bago, B., & De Neys, W. (2019). The smart System 1: Evidence for the intuitive nature of
correct responding on the bat-and-ball problem. *Thinking & Reasoning*, *25*, 257-299.
- Bhatia, S. (2017). Conflict and bias in heuristic judgment. *J Exp Psychol Learn Mem Cogn*, *43*,
319-325. doi: 10.1037/xlm0000307
- Cassady, J. C. (2001). Self-reported GPA and SAT: A methodological note. *Practical
assessment, research & evaluation*, *7*, 1-6.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. (2nd ed. ed.). Hillsdale,
NJ: Lawrence Erlbaum.
- Darlow, A. L., & Sloman, S. A. (2010). Two systems of reasoning: architecture and relation
to emotion. *Wiley Interdisciplinary Reviews-Cognitive Science*, *1*, 382-392. doi:
10.1002/wcs.34
- De Neys, W. (2012). Bias and Conflict: A Case for Logical Intuitions. *Perspectives on
Psychological Science*, *7*, 28-38. doi: 10.1177/1745691611429354
- Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social
cognition *Annual Review of Psychology* (Vol. 59, pp. 255-278).
- Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition.
Perspectives on Psychological Science, *8*, 223-241. doi: 10.1177/1745691612460685

- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*, 25-42. doi: 10.1257/089533005775196732
- Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*: Cambridge University Press.
- Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., & Pardo, S. T. (2012). Individual Differences in Numeracy and Cognitive Reflection, with Implications for Biases and Fallacies in Probability Judgment. *Journal of Behavioral Decision Making*, *25*, 361-381. doi: 10.1002/bdm.752
- Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, *21*, 37-44. doi: 10.1177/0272989x0102100105
- Meyer, A., Frederick, S., Burnham, T. C., Pinto, J. D. G., Boyer, T. W., Ball, L. J., et al. (2015). Disfluent fonts don't help people solve math problems. *Journal of Experimental Psychology-General*, *144*, E16-E30. doi: 10.1037/xge0000049
- Morewedge, C. K., & Kahneman, D. (2010). Associative processes in intuitive judgment. *Trends in Cognitive Sciences*, *14*, 435-440. doi: 10.1016/j.tics.2010.07.004
- Morey, R. D., & Rouder, J. N. (2015). BayesFactor: An R package for Bayesian data analysis. (Version 0.9.10-2).
- Morsanyi, K., & Handley, S. J. (2012). Logic feels so good—I like it! Evidence for intuitive detection of logicity in syllogistic reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 596-616. doi: 10.1037/a0026099
- Oppenheimer, D. M. (2008). The secret life of fluency. *Trends Cogn Sci*, *12*, 237-241. doi: 10.1016/j.tics.2008.02.014

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

10.1177/0963721415604610

Pennycook, G., & Rand, D. G. (2018). Cognitive Reflection and the 2016 U.S. Presidential Election. *Personality and Social Psychology Bulletin*, 45, 224-239. doi:

10.1177/0146167218783192

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188, 39-50. doi: <https://doi.org/10.1016/j.cognition.2018.06.011>

Pennycook, G., & Ross, R. M. (2016). Commentary: Cognitive reflection vs. calculation in decision making. *Frontiers in Psychology*, 7. doi: 10.3389/fpsyg.2016.00009

Reyna, V. F. (2004). How people make decisions that involve risk - A dual-processes approach. *Current Directions in Psychological Science*, 13, 60-66. doi: 10.1111/j.0963-7214.2004.00275.x

Shepperd, J. A. (1993). Student derogation of the Scholastic Aptitude Test: Biases in perceptions and presentations of College Board scores. *Basic and Applied Social Psychology*, 14, 455-473.

Sinayev, A., & Peters, E. (2015). Cognitive reflection vs. calculation in decision making. *Frontiers in Psychology*, 6. doi: 10.3389/fpsyg.2015.00532

Sirota, M., & Juanchich, M. (2018). Effect of response format on cognitive reflection: Validating a two- and four-option multiple choice question version of the Cognitive Reflection Test. *Behavior Research Methods*, 50, 2511-2522. doi: 10.3758/s13428-018-1029-4

Sirota, M., Juanchich, M., & Haggmayer, Y. (2014). Ecological rationality or nested sets?

Individual differences in cognitive processing predict Bayesian reasoning. *Psychonomic Bulletin & Review*, *21*, 198-204. doi: 10.3758/s13423-013-0464-6

Sirota, M., Kostovičová, L., Juanchich, M., Dewberry, C., & Marshall, A. C. (2018). Measuring

Cognitive Reflection without maths: developing and validating the verbal Cognitive Reflection Test. *PsyArXiv*. doi: <https://doi.org/10.31234/osf.io/pfe79>

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*,

119, 3-22. doi: 10.1037//0033-2909.119.1.3

Song, H. J., & Schwarz, N. (2008). Fluency and the detection of misleading questions: Low

processing fluency attenuates the Moses illusion. *Social Cognition*, *26*, 791-799. doi: 10.1521/soco.2008.26.6.791

Szaszi, B., Szollosi, A., Palfi, B., & Aczel, B. (2017). The cognitive reflection test revisited:

exploring the ways individuals solve the test. *Thinking & Reasoning*, *23*, 207-234.

Thompson, V. A., Turner, J. A. P., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., et al.

(2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition*, *128*, 237-251. doi:

10.1016/j.cognition.2012.09.012

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*, 1275-1289. doi: 10.3758/s13421-011-0104-1

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information

processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, *20*, 147-168. doi: 10.1080/13546783.2013.844729

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of*

Statistical Software, *36*, 1-48.

Supplementary Materials

Materials used in 'Pretest'

Manipulation check questions

1. How hard/easy was to read the example text A? (circle one option)

1 Extremely easy-to-read	2 Easy-to-read	3 Neither easy nor difficult-to- read	4 Difficult-to-read	5 Extremely difficult-to-read
--------------------------------	-------------------	--	------------------------	-------------------------------------

2. How hard/easy was to read the example text B? (circle one option)

1 Extremely easy-to-read	2 Easy-to-read	3 Neither easy nor difficult-to- read	4 Difficult-to-read	5 Extremely difficult-to-read
--------------------------------	-------------------	--	------------------------	-------------------------------------

Note. Example text A and B were represented either in the fluent or the disfluent font of the materials described below. The order of the example texts was counterbalanced.

Materials used in 'Experiment 1'

Numerical Cognitive Reflection Test (Toplak, West, & Stanovich, 2014)

Instructions

You will see on the next pages several items that vary in difficulty. Answer as many as you can. Please solve the problems in the order they are presented.

Items

1. A bat and a ball cost £1.10 in total. The bat costs £1.00 more than the ball. How much does the ball cost?

_____ pences

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

_____ minutes

3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

_____ days

4. If John can drink one barrel of water in 6 days, and Mary can drink one barrel of water in 12 days, how long would it take them to drink one barrel of water together?

_____ days

5. Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are in the class?

_____ students

6. A man buys a pig for £60, sells it for £70, buys it back for £80, and sells it finally for £90. How much has he made?

_____pounds

7. Simon decided to invest £8,000 in the stock market one day early in 2008. Six months after he invested, on July 17, the stocks he had purchased were down 50%. Fortunately for Simon, from July 17 to October 17, the stocks he had purchased went up 75%. At this point, Simon has:

- a) broken even in the stock market
- b) is ahead of where he began
- c) has lost money

Verbal Cognitive Reflection Test (Sirota, Kostovičová, Juanchich, Dewberry, & Marshall, 2018)

Instructions

You will see on the next pages several items that vary in difficulty. Answer as many as you can. Please solve the problems in the order they are presented.

Items

1. Mary's father has 5 daughters but no sons – Nana, Nene, Nini, Nonno. What is the fifth daughter's name probably?

2. If you were running a race, and you passed the person in 2nd place, what place would you be in now?

3. It's a stormy night and a plane takes off from JFK airport in New York. The storm worsens, and the plane crashes - half lands in the United States, the other half lands in Canada. In which country do you bury the survivors?

4. A monkey, a squirrel, and a bird are racing to the top of a coconut tree. Who will get the banana first, the monkey, the squirrel, or the bird?

5. In a one-story pink house, there was a pink person, a pink cat, a pink fish, a pink

computer, a pink chair, a pink table, a pink telephone, a pink shower— everything was pink!
What color were the stairs probably?

6. How many of each animal did Moses put on the ark?

7. The wind blows west. An electric train runs east. In which cardinal direction does the smoke from the locomotive blow?

8. If you have only one match and you walk into a dark room where there is an oil lamp, a newspaper and wood - which thing would you light first?

9. Would it be ethical for a man to marry the sister of his widow?

10. Which sentence is correct: a) "the yolk of the egg are white" or b) "the yolk of the egg is white"?

Note. Participants solved the word problems from the Numerical Cognitive Reflection Test and the Verbal Cognitive Reflection Test (counterbalanced for presentation order) either in a fluent, easy-to-read font (i.e., black Myriad Web 12-point font) or in a disfluent, difficult-to-read font (i.e., 10% gray italicised Myriad Web 10-point font). The presentation order of the word problems themselves was fixed and presented as shown above.

Numeracy Scale (Lipkus, Samsa, & Rimer, 2001)

1. Imagine that we rolled a fair, six-sided die 1,000 times. Out of 1,000 rolls, how many times do you think the die would come up even (2, 4, or 6)?

.....

2. In the BIG BUCKS LOTTERY, the chances of winning a \$10.00 prize is 1%. What is your best guess about how many people would win a \$10.00 prize if 1,000 people each buy a single ticket to BIG BUCKS?

.....

3. In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1,000. What percent of tickets to ACME PUBLISHING SWEEPSTAKES win a car?

.....

4. Which of the following numbers represents the biggest risk of getting a disease?

- 1 in 100
- 1 in 1000
- 1 in 10

5. Which of the following numbers represents the biggest risk of getting a disease?

- 1%
- 10%
- 5%

6. If Person A's risk of getting a disease is 1% in ten years, and person B's risk is double that of A's, what is B's risk?

.....

7. If Person A's chance of getting a disease is 1 in 100 in ten years, and person B's risk is double that of A's, what is B's risk?

.....

8. If the chance of getting a disease is 10%, how many people out of 100 would be expected to get the disease?

.....

9. If the chance of getting a disease is 10%, how many people out of 1000 would be expected to get the disease?

.....

10. If the chance of getting a disease is 20 out of 100, this would be the same as having a _____% chance of getting the disease.

.....

11. The chance of getting a viral infection is .0005. Out of 10,000 people, about how many of them are expected to get infected?

.....

Sociodemographic questions

What is your age? _____ years

What is your gender? [circle one]

1 male

2 female

References

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making, 21*, 37-44. doi: 10.1177/0272989x0102100105

Sirota, M., Kostovičová, L., Juanchich, M., Dewberry, C., & Marshall, A. C. (2018). Measuring Cognitive Reflection without maths: developing and validating the verbal Cognitive Reflection Test. *PsyArXiv*. doi: <https://doi.org/10.31234/osf.io/pfe79>

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning, 20*, 147-168. doi: 10.1080/13546783.2013.844729

Materials used in 'Experiment 2'

Standard instructions

On the next pages you will see five problems that vary in difficulty. Answer as many as you can.

Warning instructions

On the next pages you will see five problems that vary in difficulty. Answer as many as you can.

Please note that all five of them are "tricky" problems. So, if you want to solve them correctly you should pay close attention to the wording of the questions and reflect on how the problem is trying to trick you.

Reading check question

Based on the instructions I have just read, the problems I am going to solve are:

- tricky problems that require my full attention
- not tricky problems that I can answer quickly
- problems about which it is impossible to say whether they are tricky or not based on the instructions

Problems

Mary's father has 5 daughters but no sons – Nana, Nene, Nini, Nonno. What is the fifth daughter's name probably?

How many of each animal did Moses put on the ark?

The wind blows west. An electric train runs east. In which cardinal direction does the smoke from the locomotive blow?

If you have only one match and you walk into a dark room where there is an oil lamp, a newspaper and wood - which thing would you light first?

Would it be ethical for a man to marry the sister of his widow?

Sociodemographic questions

What is your age?

What is your gender?

- Male
- Female
- Other