What does the cognitive reflection test really measure: A process dissociation investigation.

Jonathon McPhetres

University of Regina

Massachusetts Institute of Technology

Word count: 10,523

Author Note

Jonathon McPhetres is a Postdoctoral Fellow at the University of Regina and at The Massachusetts Institute of Technology. Correspondence concerning this article can be addressed to Jonathon McPhetres. Email: jmcphetr@mit.edu or Jonathon.mcphetres@uregina.ca

Data and materials for all studies reported herein, along with preregistrations for studies 3 and 4 are available on the open science framework:

https://osf.io/4wv89/?view_only=51830f4661ed4ff9ab24fce069155fbe.

Abstract

The cognitive reflection test (CRT) is a series of brain-teaser questions believed to measure intuitive versus reflective thinking. However, those measures are confounded by the scoring method whereby a decrease in reflective thinking also results in an increase in intuitive thinking. This confound also precludes testing of the relation between the two processes—for example, whether intuition is independent from reflectiveness. The present studies directly address these limitations using process dissociation (PD) to quantify and manipulate each process independently. If the CRT measures both intuition and reflection then using PD to isolate each score should provide unique information about each process and allow for testing of models describing the relation between the two processes. However, four studies call in to question whether the CRT actually measures intuition (studies 1-3) and provide some limited evidence that reflection appears to be dependent on the processes occurring prior to it (consistent with time-series models of cognition). Moving forward, it is recommended that researchers 1) consider alternative measures of cognitive reflection, 2) are cognizant of the phrasing used when describing intuitions as inferred from the CRT, and 3) rethink measures of intuition.

Keywords: *Cognitive reflection test; intuition; dual-process; process dissociation; open data; open materials; preregistered.*

Introduction

Few would argue with the claim that some people are more “intuitive” and others are more “reflective.” Indeed, it seems intuitive to us that some like to feel and some like to think; some are emotional and some are rational. Exactly why and how these differences in reasoning occur, however, is a tantalizing problem for psychologists. How does one reason? What is the relation between intuition and reflection?

There are many names for intuitive and reflective processes which reflect varying assumptions about how these processes work (e.g., Type 1 and Type 2, system 1 and system 2, etc; for reviews see De Neys, 2006; Evans, 2008; Evans & Stanovich, 2013; Pennycook, 2017). The cognitive reflection test (CRT; Frederick, 2005; Toplak, Stanovich & West, 2014) is a series of brain-teaser type questions believed to measure reflective over intuitive thinking. The CRT is scored and interpreted such that a low score means more intuitive and a high score means more reflective. However, this is a problem because the poles are relative: an increase in reflection indicates a simultaneous decrease in intuition and vice-versa. This confound precludes accurate measurement of the processes and any testing of the relation between the two processes. Further, the use of the such scoring methods assumes that people will evince considerable variation on both ends of the continuum. Yet, such variation is not testable using traditional scoring methods, highlighting the need for alternative scoring methods for the CRT.

The Process Dissociation (PD) procedure (Jacoby, 1991; Payne & Bishara, 2009) provides a solution to these problems by allowing one to isolate the two confounding scores and measure or manipulate them independently. If the CRT measures both intuitive and reflective processes, then isolating the processes should provide information about each process as well as about the relation between the two processes. In the present research, I present a novel form of

the cognitive reflection test which is scored using PD procedures to provide independent intuitive and reflective parameters. In doing so, I am also able to examine some evidence for the different models of cognition. Below, I briefly outline the assumptions of the dual-process models and then test them experimentally.

Dual-Process Models of Cognition

Dual-process models describe two types of processing, referred to here as intuitive and reflective processes. Intuitive processes are quick, affective, and do not require cognitive resources. The term “reflection” often refers to one of two things (Evans, 2019). Cognitive decoupling (Stanovich & Toplak, 2012) is when one is able to reason based only on logic, *decoupled* from one’s beliefs about the world. Rationalization is when one rationalizes, rather than corrects, their intuitive responses; that is, sometimes reasoning can lead to incorrect conclusions. In both cases, reflective processes are slow, deliberative, and require resources such as time and energy to resolve.

A few main models have dominated the literature—here they are referred to as serial and parallel account for simplicity. Each has been written about extensively (Evans, 2007, 2008; Evans & Stanovich, 2013; Pennycook, 2017) and many variations exist, but they will be briefly summed up here.

The serial model, sometimes called the “default-interventionist” model (Evans, 2007), involves a time-course assumption such that intuitive process occurs first, followed by reflective process. That is, one first has an intuitive response followed by reflection, which can then agree with, override, or ignore the intuitive response. Evidence for the serial models comes from studies where participants are asked to give two responses—the first quickly and the second after some deliberation time (Pennycook & Thompson, 2012; Thompson, Prowse, Turner, &

Pennycook, 2011). These studies found that participants often experienced the compelling intuitive response and then corrected it later after some deliberation. Further evidence comes from studies showing that responses to logic problems differ at various stages of the reasoning process (Pennycook, Fugelsang, & Koehler, 2015).

Parallel models assume that both intuitive and reflective processes occur simultaneously. Originally, this account sought to reconcile psychodynamic and cognitive accounts of the human consciousness by accounting for emotional and rational tendencies (e.g., Epstein, 1994). Later, more ‘strictly’ cognitive accounts (Sloman, 1996; Smith & Decoster, 2000) claimed that the two processes are different ‘systems’ in conflict. Thus, one process could be stronger than the other, or they could both be operating with equal strength, but they are in opposition to each other.

More recently, a “hybrid” account incorporates elements from both serial and parallel models (Bago & De Neys, 2019a). Specifically, evidence suggests that some elements of deliberative responding appear to be present from the beginning of the reasoning process: whether or not the conclusion is believable often influences intuitive responses from the start (Handley, Newstead, & Trippas, 2011; Newman, Gibb & Thompson, 2017; Pennycook et al., 2015). Additionally, some people can obtain the correct (i.e. reflective) answer quickly, intuitively, and under cognitive constraints (Bago & De Neys, 2017), and some “intuitions” are correct (Bago & De Neys, 2019b).

These models result in several different predictions. First, under the parallel models, one would expect people to show considerable variation in measures of both intuition and reflection because they are independent processes. That is, one could be both highly intuitive and highly reflective or one could be highly intuitive and have poor reflective abilities. However, including intuition and reflection on the same scoring continuum creates a problem of inference: a high

score means more reflective *and* less intuitive, confounding the result. Further, this claim is contradictory to serial processing models.

Under serial models, everyone would be expected to demonstrate a similar level of intuitiveness because everyone should have these gut-level responses—some people are just better at overriding them. Instead, variation should occur in measures of reflection, or at least at some stage *after* intuition—some process which affects one’s ability to *engage* in cognitive reflection. See, for example, Pennycook et al.’s (2015) “three-stage” model in which conflict detection leads to reflective reasoning.

Other predictions concern how the processes influence each other. For example, under serial models, a change in intuition would influence reflection because reflection comes afterwards and is thus dependent on the processes occurring prior to it (or, perhaps reflection is simply the opposite of intuition). Alternatively, a manipulation designed to influence reflection should only influence reflection under both models—either because the two processes are independent or because reflection comes later in time. However, the ability to independently measure and manipulate intuition and reflection is a necessary prerequisite to addressing any of these predictions.

The Cognitive Reflection Test

The cognitive reflection test (CRT; Frederick, 2005; Toplak, Stanovich & West, 2014) is a widely used measure of intuitive versus reflective thinking. There is a 3-question variant (Frederick, 2005) and a 7-question variant (Toplak et al., 2014) which includes the original three questions plus four additional questions. In both cases, the CRT is scored by summing up the correct responses which results in a score ranging from 0, indicating less reflection and/or greater intuition, to 3 or 7 indicating greater reflection and/or less intuition. Surprisingly, little attention

has been paid to meaning and content of these scoring procedures. One study reports that summing up the intuitive responses results in less reliable measures (Pennycook, Cheyne, Koehler, & Fugelsang, 2016), suggesting the CRT is a better measure of reflection than intuition. After all, it is called the cognitive *reflection* test, not the cognitive *intuition* test.

Such information should be both surprising and confusing because the test itself is predicated on the assumption that the questions contain intuitively attractive answers which lead people astray. There is some evidence for this claim. For example, most people are aware of the incorrect response, either because they chose it (Travers, Rolison, & Feeney, 2016) or because they deliberated against it (Mata, Ferreira, & Sherman, 2013) and the test was originally designed to bait people towards these responses (Frederick, 2005). In fact, much of psychological research has focused on humans' failures of rationality (e.g. Kahneman, 2011; Kahneman, Slovic, & Tversky, 1982; Pennycook, Fugelsang, & Koehler, 2015) and the CRT takes advantage of this irrationality. Thus, the test assumes that 1) the majority of people first come up with the incorrect response and 2) that people vary in their ability to override this response. As a result, the questions should measure one's likelihood of being duped by the incorrect response as much as they measure one's ability to avoid being duped. However, if most (or all) people experience the incorrect intuitive response (e.g., Travers, et al., 2016), then considerably less variability should be observed on levels of intuition—instead, variability should be observed on reflection.

Despite these concerns, the scores are often interpreted to provide information about *both* intuition *and* reflection. For example, religious people are *more intuitive* (Shenhav, Rand & Green, 2012; c.f. Farias et al., 2017); atheists are *more reflective* (Pennycook, Ross, Koehler, & Fugelsang, 2016; *see also* Gervais & Norenzayan, 2012 and failed replication attempts by

Camerer et al., 2018, and Sanchez, Sundermeier, Gray, & Calin-Jageman, 2017); liberals are *more reflective* (Deppe et al., 2015); participants are referred to as *intuitive* or *deliberative responders* (pg. 3, Mata et al., 2013). One way to resolve this confound is to isolate measurements of intuitive and reflective processes and then attempt to manipulate them independently.

Process Dissociation

The Process Dissociation (PD) procedure is a mathematical technique allowing the independent quantification of two processes which would normally be confounded (Jacoby, 1991; Payne & Bishara, 2009). The main idea underlying PD is that both automatic and deliberative processes underpin a wide range of judgments and decisions; separating these processes allows for more precise measurement and clearer interpretation. PD has its roots in various areas of psychological research based on dual-process accounts of reasoning where both automatic and deliberative responses contribute to a final judgment: biases in weapon identification (Payne, 2001) and implicit judgments (e.g. the quad model; Conrey, Sherman, Gawronski, Hugenberg & Froom, 2005), memory research (Hense, Penner & Nelson, 1995) and moral decision-making (Conway & Gawronski, 2013).

As an example, Conway and Gawronski (2013) applied PD to moral decision making. Across a series of 20 scenarios, utilitarian (outcome-based) decisions were summed to create a single-continuum (bipolar) score ranging from deontological to utilitarian, much like the intuitive-reflective scores on the CRT. However, experimental manipulations were ineffective on this bipolar score. Using PD, the authors report that cognitive load undermined the utilitarian (reflective processes) while leaving the intuitive process intact.

Thus, PD was formulated specifically to distinguish between automatic and deliberative processes, making it an ideal approach to dissect what the CRT is actually measuring and to test the relation between the two processes. Thus, the application of PD to the CRT may be a possible solution to the problem of confounding the intuitive and reflective processes.

The Present Studies

The present research had several goals. The first goal was to apply PD to the scoring of the CRT. It was expected that, if the CRT accurately measures both intuition and reflective, then isolating each process via PD should result in independent parameters which correlate with other indices thought to be related to individual differences in intuition and/or reflection.

Second, having separated the intuitive and reflective processes, the dual-process models can be tested. If intuitive and reflective processes operate in parallel, then separate experiments should demonstrate that these processes can be manipulated independently. For example, experimental procedures aimed at influencing the intuitive parameter should not affect the reflective parameter, and vice-versa. However, if the processes operate in serial then any manipulation intended to influence the intuitive process should also influence the reflective process. That is, if reflection is dependent on the processes occurring prior to it, then attempts at manipulating earlier processes (e.g., intuition) should instead result in effects on later processes (e.g., reflection).

The third goal was to create a new but substantially similar version of the CRT. The reason for this is two-fold. First, the nature of PD requires questions to have specific formats and some of the original CRT items did not follow this format, so items needed to be replaced. Second, the CRT is widely used and a recent popular media article provided members of the general public with the correct answers to the CRT (McCall, 2017), which may severely

undermine its validity. Further, several researchers expressed concern about the validity of measurements after multiple exposures and some efforts have been made to develop new versions (e.g., Toplak et al., 2014; Thomson & Oppenheimer, 2016). However, there is also some evidence that performance on the CRT is stable after multiple exposures (Bialek & Pennycook, 2016; Meyer, Zhou & Shane, 2018; Raelison, & De Neys, 2019).

Finally, it is possible that the relations evidenced in the literature are reliant upon the specific problems—that is, whether the CRT measures “performance on the CRT” or whether it measures actual cognitive reflection or intuitions (in which case, reliable results should be obtained with any similar measure that actually captures intuition and reflection). Thus, I sought to create a new “version” of the CRT that also included items from previous research investigating intuitive and reflective responding.

Study 1

The purpose of study 1 was to construct and test the feasibility of a process dissociation measure to capture intuitive and reflective processes independently of one another. Thus, following previous research using PD, a survey was conducted to assess the PD reflective and PD intuitive parameters as they relate to several scales indicative of reflective and intuitive cognitive styles.

Participants

I recruited 182 students (142 female) enrolled in an undergraduate psychology course in exchange for extra course credit. All subjects indicated they were native (or fluent in) English. Ten students were excluded because they demonstrated no variance on the congruent problems (Conway & Gawronski, 2013; Friesdorf, Conway & Gawronski, 2015), precluding the calculation of the PD parameters. This resulted in a sample of 172 students who were 19.81 years

old ($SD = 1.15$) on average, were 56% white, 24% Asian, 8% Latino, 5% Black, and 8% mixed/other. Five students (3%) reported having already obtained a four-year degree.

Materials and Procedure

Participants signed up for a study on problem solving and completed all materials on a computer using Qualtrics survey software in a private laboratory cubicle. Participants first responded to the 28-item CRT measure developed for this study (described below). Items were based on various versions of the CRT (Finucane & Gullion, 2010; Frederick, 2005; Thomson & Oppenheimer, 2016; Toplak, West, & Stanovich, 2011; Toplak et al., 2014). The final items were selected from a larger pool of items based on pilot testing with an independent sample in which items demonstrating the most variability were retained.

Process Dissociation Procedure. PD works by providing people with two version of each problem, an *incongruent* and a *congruent* trial. Take the ubiquitous bat/ball problem for example:

“A bat and a ball cost \$1.10 in total. If the bat costs \$1.00 more than the ball, how much does the ball cost?”

This problem represents an *incongruent* version because it pits the intuitive response against the correct/reflective response. In the case of the CRT, the questions each have two possible responses: the intuitive/incorrect response and the correct/reflective response. However, PD requires the addition of another set of problems, *congruent* versions of the same problem in which the intuitive and reflective processes would yield the same answer. For example, the bat/ball problem can be reworded as follows:

“A bat and a ball cost \$1.10 in total. If the ball costs \$.05, how much does the bat cost?”

In this case, using either reflective or intuitive processes, one would yield the answer of 5 cents, which is correct. Because both intuitive and reflective processes are presumed to contribute to both versions of the problem, two scores can be calculated from the combination of these two versions. First, the reflective parameter (R) can be calculated using formula 1:

$$R = p(\text{intuitive} \mid \text{congruent}) - p(\text{intuitive} \mid \text{incongruent}) \quad (1)$$

In this formula, the influence of the reflective process (R) is calculated by subtracting the probability of giving the intuitive response on incongruent trials from the probability of giving the intuitive response on congruent trials. Next, the intuitive parameter (I) can be calculated using formula (2):

$$I = p(\text{intuitive} \mid \text{incongruent}) / (1 - R) \quad (2)$$

In this formula, the influence the intuitive process (I) is calculated by dividing the probability of giving the intuitive response on incongruent trials by the inverse of R.

As an example, if a person gives the intuitive answer on 3 out of 10 incongruent problems (.3) and the reflective answer on 6 out of 10 congruent problems (.6), the two equations above would yield an R parameter score of .3 and an I parameter score of .75. These scores reveal that intuition has a stronger influence on this persons' decisions than does reflection. Importantly, these scores are interpreted as tendencies rather than proportion of responses, as in one could have stronger intuitive tendencies and weaker reflective tendencies. Likewise, one may score high on both parameters.

Process Dissociation Scoring. Participants respond to 28 items in total: 14 *incongruent* items in random order followed then by 14 *congruent* items in random order. From these 28 items, I calculated three scores. I calculated a bipolar intuitive-reflective score by summing the correct responses on the incongruent questions resulting in a score ranging from 0 (*relatively*

intuitive) to 14 (*relatively reflective*). On incongruent trials, 41% of the total number of answers were the intuitively incorrect response, 52% were correct, and 7% were incorrect but non-intuitive. The full distribution of response types is presented in the supplementary materials (Table S1).

I calculated a PD reflective parameter using formula (1), above, where higher scores indicate more reflective tendencies. I also calculated a PD intuitive parameter using formula (2), where higher scores indicate stronger intuitive tendencies. I also measured the amount of time taken to solve each problem and averaged this into two separate scores for incongruent and congruent problems, as well as an average time across all problems.

Additional measures. Next, participants completed the need for cognition ($\alpha = .80$) and faith in intuition ($\alpha = .77$) subscales of the rational-experiential inventory (Epstein, Pacini, Denes-Raj & Heier, 1996). The need for cognition scale measures the extent to which one enjoys engaging in complex thinking and problems solving (e.g., *I prefer to do something that challenges my thinking abilities rather than something that requires little thought.*). The faith in intuition scale measures the extent to which one relies on their intuitions when making decisions (e.g., *I believe in trusting my hunches*)¹. All items were rated on a scale ranging from 1(*strongly disagree*) to 7(*strongly agree*). Participants also completed the 7-item cognitive reflection test (Toplak et al., 2014) which has been used extensively to measure reflective versus intuitive thinking. I summed the total correct responses resulting in a score ranging from 0(*intuitive*) to 7(*reflective*). I also measured the time spent solving each problem and averaged this into a composite score. Participants also responded to a prompt measuring correspondence bias which was included for purposes not related to this study (correlating this score with religiosity) and it

is not discussed here. Finally, participants responded to demographics questions, which included a 6-item religiosity scale (Cohen, Shariff & Hill, 2008; $\alpha = .95$).

Results and Discussion

Zero-order correlations for all measures were examined first. As shown in Table 1, the reflective and intuitive PD parameters correlated negatively but weakly with each other. The reflective and intuitive PD parameters correlated positively and negatively, respectively, with the relative intuitive-reflective score, which is consistent with meta-analytic results on these patterns (Friesdorf et al., 2015).

As expected, the PD reflective parameter correlated positively with time spent on the incongruent and congruent problems (as well as the total average time), suggesting more reflective thought (Pennycook et al., 2015). The PD reflective parameter also correlated positively with the Need for Cognition scale, negatively with the Faith in Intuition scale, and positively with the 7-item CRT. The PD reflective parameter correlated negatively with religiosity. The PD intuitive parameter correlated negatively but weakly with average time spent on the problems and with need for cognition. It correlated positively but weakly with the faith in intuition scale, with time spent on the 7-item CRT, and with religiosity. It correlated negatively with the 7-item CRT scale, indicating more intuitive/less reflective scores.

Table 1. Zero-order correlations for all measures in Study 1.

	1	2	3	4	5	6	7	8	9	10
1. PD reflective	-									
2. PD Intuitive	-.15*	-								
3. Bipolar I/R	.88**	-.47**	-							
4. Incongruent time	.22**	-.07	.19*	-						
5. Congruent time	.24**	-.05	.18*	.73	-					
6. Average total time	.24**	-.07	.20*	.97**	.88**	-				
7. Need for Cognition	.25**	-.06	.22**	.11	.09	.11	-			
8. Faith in Intuition	-.17*	.08	-.15†	-.15†	-.19*	-.17*	-.19*	-		
9. CRT-7	.63**	-.36**	.73**	.11	.14†	.13†	.30	-.14†	-	
10. CRT-7 response time	.28**	.01	.19*	.56**	.56**	.60**	.12	-.31**	.18*	-
11. Religiosity	-.30**	.05	-.30**	-.03	.07	.01	-.06	.14†	-.23**	-.01

Note: ** $p < .01$, * $p < .05$, † $p < .09$.

Thus, these patterns show that the PD reflective and intuitive parameters are divergently related to a number of measures indicating reflective and intuitive tendencies. This provides preliminary evidence that the intuitive and reflective processes can be dissociated from one another and may indeed be quantified independently, though this study does not provide any evidence for or against different models of processing.

Further, this study provides only weak evidence for the accuracy of the PD intuitive parameter because it correlated strongly with only the original CRT scores. That is, if the measure were accurate, it should have correlated strongly with other measures (e.g., FII). In further support of this interpretation, participants showed considerably less variance on PD intuitive scores (see Table 2). Scores are remarkably consistent across each study for those in the control conditions. Raw means are displayed because the standardized means are difficult to compare directly due to differing manipulations across studies and only data within the control condition is shown. Study 2 extends these findings with an attempt to manipulate the reflective parameter independently.

Table 2. Raw means, standard deviations, and average variance for control conditions across all studies.

	Study 1			Study 2			Study 3			Study 4		
	<i>M</i>	<i>(SD)</i>	Variance	<i>M</i>	<i>(SD)</i>	Variance	<i>M</i>	<i>(SD)</i>	Variance	<i>M</i>	<i>(SD)</i>	Variance
PD Reflective	.4157	(.250)	.063	.4080	(.300)	.09	.4222	(.268)	.072	.4085	(.257)	.07
PD Intuitive	.6753	(.183)	.033	.6743	(.170)	.03	.6588	(.204)	.042	.6470	(.200)	.04

Note: Study 1 includes all participants, studies 2-4 include only the control condition.

Study 2

The purpose of Study 2 was to manipulate the PD reflective parameter independent of the PD intuitive parameter. To undermine the reflective processes, participants were put under

cognitive load. As has been shown in previous research (Conway & Gawronski, 2013), cognitive load can influence outcomes normally associated with reflective processes, presumably by limiting access to the cognitive resources needed to process information. Thus, it was hypothesized that cognitive load would diminish the PD reflective parameter, but would leave the PD intuitive parameter intact, observed in previous research (Conway & Gawronski, 2013; McPhetres, Conway, Hughes, & Zuckerman, 2018). No predictions were made regarding the relative intuitive-reflective score because any reduction in this parameter equates to a simultaneous increase in the intuitive parameter, confounding the result.

Participants

A power analysis using G*Power (Faul, Erdfelder, & Buchner, 2007) indicates that, for a small-medium sized ($d = .45$) difference between two groups, 158 participants are needed. I recruited 158 undergraduates (106 female) enrolled in an undergraduate psychology course in exchange for extra course credit. All subjects indicated they were native (or fluent in) English. Participants who completed Study 1 were not eligible to participate. As with past research (Friesdorf et al., 2015), 6 participants were excluded for whom PD scores could not be calculated: 2 demonstrated no variance on the congruent or incongruent problems and 4 did not respond to all problems. Thus, the final sample was 152 (104 female) students. Participants were 19.83 ($SD = 1.29$) years of age on average, were 52% white, 27% Asian, 7% Latino, 7% black, and 6% mixed/other. Four participants (3%) indicated having earned a four-year degree.

Materials and Procedure

Upon arriving in the lab, participants were randomly assigned to one of two conditions where they responded to the same 28 randomized logic problems as described in Study 1. In the cognitive load condition, participants were asked to memorize an 8-digit string (e.g., Lgwm?zx1)

before responding to each problem. Participants then responded to one of the logic problems and, on the following page, were asked to report the digit string they had memorized. This process was repeated for each of the 14 incongruent problems in random order, and then for each of the 14 congruent problems in random order. In the control condition, participants were shown the same digit strings but were asked to ignore them. As in Study 1, PD reflective, PD intuitive, and bipolar scores were calculated. On incongruent trials, 43% of responses were intuitive and incorrect, 50% were correct, and 7% were incorrect and non-intuitive. The full distribution of response types is presented in the supplementary materials (Table S1).

I also measured the time spent on each problem and calculated an average time for the incongruent problems, for the congruent problems, and for all 28 problems on average.

Participants then responded to demographics questions, including the 6-item religiosity scale (Shariff, Cohen & Norenzayan, 2008; $\alpha = .97$).

Results and Discussion

Zero-order correlations were analyzed first (See Table 3). The PD reflective and intuitive parameters correlated negatively; the bipolar score was significantly correlated with both the PD reflective and intuitive parameters. Again, the PD reflective parameter correlated negatively with religiosity, $r = -.29, p < .001$, but the PD intuitive parameter did not, $r = .06, p = .471$.

Table 3. Means, standard deviations, and correlations in Study 2.

Variable	<i>M</i>	<i>SD</i>	1	2	3	4
1. PD Reflective	0.33	0.98				
2. PD Intuitive	0.24	0.89	-.34**			
3. Bipolar score	7.88	3.66	.93**	-.50**		
4. Incongruent time (seconds)	26.72	10.53	.24*	-.01	.26*	
5. Congruent time (seconds)	16.02	5.56	.31**	.04	.25*	.64**

Note. *M* and *SD* are used to represent mean and standard deviation, respectively;

* indicates $p < .05$. ** indicates $p < .01$.

The main hypothesis was examined using a t-test to compare the standardized PD intuitive, PD reflective, and raw bipolar intuitive-reflective scores between the cognitive load and control conditions. As hypothesized, the PD reflective scores were reduced in the cognitive load condition compared to the control condition. The PD intuitive scores did not differ between the cognitive load and control. These results are depicted in Figure 1. The relative intuitive-reflective scores did not differ between the cognitive load and control. Standardized means and inferential statistics are displayed in Table 4.

Figure 1.
The effect of condition on PD reflective and PD intuitive parameters in Study 2.

Note: Standardized PD parameters are displayed; vertical bars represent 1 *SE*.

Table 4.

Descriptive statistics and significance values for standardized PD Parameters and Bipolar Scores between conditions in Studies 2-4.

Study 2: Cognitive Load						
	Control Condition	Cognitive Load Condition				
	M (SD)	M (SD)	t	p	d	
PD Intuitive	.38 (.67)	.24 (.59)	1.35	.179	.20	
PD Reflective	.23 (.93)	-.08 (.88)	2.12	.036	.35	
Bipolar	7.49 (3.45)	6.66 (3.30)	1.52	.130	.23	
Study 3: Favoring Intuition						
	Control Condition	Intuitive Condition				
	M (SD)	M (SD)	t	p	d	
PD Intuitive	.03 (1.00)	.13 (.81)	.70	.483	.11	
PD Reflective	.23 (.99)	-.14 (.94)	2.41	.017	.38	
Bipolar	7.25 (3.12)	6.07 (3.08)	2.41	.017	.38	
Study 4: Conflict Detection						
	Control Condition	Unconfident Condition	Reflective Condition			
	M (SD)	M (SD)	M (SD)	F	p	η_p^2
PD Intuitive	.14 (.87)	.01 (.94)	-.02 (.88)	.75	.474	.01
PD Reflective	-.11 (.98)	.04 (.91)	.07 (.91)	.84	.434	.01
Bipolar	7.87 (3.20)	8.13 (3.12)	8.13 (2.87)	.20	.817	< .01
Confidence	5.63 (.79)	5.26 (.94)	5.75 (.79)	7.35	.001	.06

Note: PD Intuitive and PD Reflective parameters are standardized.

In summary, cognitive load undermined the reflective parameter and diminished the PD reflective score while leaving the intuitive parameter intact. This suggests that the cognitive load manipulation may have interfered with the cognitive processes that was used in reflection but not used—or only minimally used—for intuitive judgments.

This could be interpreted as evidence that the reflective process occurs after the intuitive process. However, this could also be interpreted as evidence that the CRT does not accurately measure intuitive processes. As in Study 1, participants evinced less variation in their levels of intuition compared to reflection (see Table 2), further suggesting that perhaps people vary less on intuition and more on their ability to engage in reflection, as expected under serial models. However, these results are inconclusive on their own. To resolve this ambiguity, Study 3 attempts to manipulate the intuitive process independently.

Study 3

The purpose of Study 3 was to investigate whether the PD intuitive parameter could be manipulated independently of the PD reflective parameter. There are two competing hypotheses based on literature regarding parallel or serial dual-process models of cognition (e.g., Evans & Stanovich, 2013). First, if the two processes operate in parallel, then any process affecting intuition will not affect the reflective process, and vice-versa. Therefore, competing hypothesis 1, according to parallel models, is: when favoring intuitions under experimental constraints, the PD intuitive parameter will increase while leaving the PD reflective parameter intact. Alternatively, if the two processes operate in a stepwise, serial fashion, then any process enhancing intuition will instead undermine the reflective processes (because the reflective process is dependent upon intuition). Therefore, competing hypothesis 2, according to serial models, is: when favoring intuitions under experimental constraints, the PD intuitive parameter will be increased and the PD reflective parameter will simultaneously be decreased. It follows then that, when favoring intuitions, the relative intuitive-reflective parameter will be significantly decreased, indicating an increase in intuitive processing relative to reflection. The hypotheses, method, and target sample size were pre-registered on the Open Science Framework prior to data collection (https://osf.io/yg2nq/?view_only=1f96ebb7a2a044df988631aa6e0d18bc).

Participants

I recruited 164 undergraduates (125 female) in exchange for extra course credit. All subjects indicated they were native (or fluent in) English. Those who participated in Studies 1 or 2 were not eligible for participation here. Four participants were excluded because they did not answer all the questions, so a PD score could not be calculated for them. Thus, the final sample was 160 students (125 female). Participants had an average age of 19.98 ($SD = 1.18$) and were

57% white, 20% Asian, 8% black, 8% Hispanic, and 7% mixed or other. Four percent of student reported having earned a bachelor's degree.

Materials and Procedures

Upon arriving in the lab, participants were randomly assigned to one of two conditions where they responded to the same 28 logic problems as described in Studies 1 and 2. On incongruent trials, 39% were intuitive/incorrect, 52% were correct, and 10% were incorrect but non-intuitive. The full distribution of response types is presented in the supplementary materials (Table S1).

In the intuitive condition, participants were instructed to "Respond to each problem as quickly as possible. Please give the first answer that comes to mind, even if you think this answer is incorrect." Participants in the control condition were simply asked to "Respond to each problem to the best of your ability." After responding to all problems, participants completed demographics items, including the 6-item religiosity scale (Shariff, Cohen, & Norenzayan, 2008).

Results and Discussion

Zero-order correlations were examined first (see Table 5). The PD reflective and intuitive parameters correlated negatively, and the bipolar score was again correlated with both the PD reflective and intuitive parameters in the expected directions. The PD reflective parameter was negatively correlated with religiosity but this was weaker than the previous studies, $r = -.09$, $p = .252$. Similar to the previous studies, the PD intuitive parameter was not related to religiosity, $r = .01$, $p = .871$.

Table 5. Means, standard deviations, and correlations in Study 3.

Variable	<i>M</i>	<i>SD</i>	1	2	3	4
1. PD Reflective	-0.14	0.94				
2. PD Intuitive	0.13	0.81	-.36**			
3. Bipolar score	6.07	3.08	.90**	-.58**		
4. Incongruent time (seconds)	17.29	5.61	.02	-.08	.06	
5. Congruent time (seconds)	12.79	4.27	.29**	-.37**	.27*	.61**

Note. *M* and *SD* are used to represent mean and standard deviation, respectively;

* indicates $p < .05$. ** indicates $p < .01$.

To examine the main hypothesis, a t-test was used to compare the bipolar score and the PD intuitive and reflective scores between both conditions. As expected, the bipolar score was decreased in the intuitive condition compared to the control condition, indicating a decrease in reflective scores and a simultaneous increase in intuitive scores. Examination of the PD scores reveals that this was due to the PD reflective parameter being significantly decreased in the intuitive condition relative to the control condition. The PD intuitive score was higher in the intuitive condition relative to the control condition, but this increase was not significant. Standardized means and significance values are displayed in Table 4.

This finding is interpreted as indicating that the change in the bipolar score was due to a *decrease* in reflection rather than an increase in intuition. Thus, these results suggest that our intuition manipulation had a clear effect on the reflective process, rather than independently influencing the intuitive parameter. This finding is in line with other research failing to manipulate intuitive responses on the original CRT (Patel, Baker, & Scherer, 2018). Therefore, I interpret these results as 1) suggesting that the CRT does not contain information on intuition, and 2) as support for competing hypothesis 2, suggesting serial operation. Finally, raw means and variance are displayed in Table 2, again showing less variance on the PD intuitive scores compared to the PD reflective scores. This, coupled with the result of the main hypothesis, could

again suggest that the CRT is a poor measure of intuition and provide additional evidence against the parallel models.

Figure 2.

The effect of the favoring-intuition condition on PD reflective and PD intuitive parameters In Study 3.

Note: Standardized PD parameters are displayed; vertical bars represent 1 *SE*.

Study 4

Given that the results of Study 3 support the serial model, the purpose of Study 4 was to investigate the serial model further. Particularly, Pennycook et al (2015) posit that intuition leads to reflection by way of conflict detection. Thus, in this study, I hoped to subtly increase the likelihood of detecting a conflict without directly increasing reflection (e.g., time or devotion of cognitive resources).

It was hypothesized that a greater likelihood of detecting a conflict would result in a higher reflective parameter score. This should be differentiated from any increase in the reflective parameter that is observed from a standard reflection manipulation in which one simply spends more time reasoning. To test this, an experimental condition was created in which participants' confidence in their responses was decreased at the beginning of the survey. It was

expected that this reduced confidence would result in greater PD reflective scores, mediating the relation between condition and PD reflective scores within that condition but not in other conditions. This ‘unconfident’ condition was compared to a control condition and to a reflection condition where subjects were required to spend a longer amount of time on each problem in order to examine whether the effect of these manipulations on the PD reflective parameter can be distinguished from each other. It was also expected that the reflection and confidence manipulations would also result in increased PD reflective, but not PD intuitive, scores.

Again, these hypotheses, target sample size, and analysis procedures were pre-registered on the Open Science Framework prior to data collection

(https://osf.io/vf2pw/?view_only=4422ee3ce5724c21906b2da25b6f2523). There was one deviation from the registration: it was incorrectly noted that a sample size of 158 was needed, as in previous studies. However, applying the same effect size ($d = .45$) used to plan the previous studies to a study with three conditions (i.e. 79 per cell) results in a total sample size of 237. Thus, 245 subjects were recruited to account for possible exclusions.

Participants

I recruited 247 participants (175 female) from an undergraduate subject pool. All subjects indicated they were native (or fluent in) English. To increase the available pool, only those who participated in studies 2 or 3 were ineligible to participate in the present study. Participants had an average age of 20.10 ($SD = 1.39$) and were 44% white, 31% Asian, 8% black, 12% Hispanic, and 5% mixed or other. Six percent of student reported having earned a bachelor’s degree. Six participants were excluded: four showed no variance on the reflective parameter so PD scores could not be calculated and 2 experienced errors in the survey, resulting in a final sample of 241.

Materials and Procedure

Upon arriving in the lab, subjects were randomly assigned to one of three conditions (Reflective, Unconfident, and Control) where they responded to the same 28 problems used in the previous three studies. In the reflective condition, subjects were required to spend a minimum of 40 seconds on each problem; this was calculated by roughly doubling the average response time in the previous studies. In the unconfident condition, subjects are informed that many of the problems are trick questions, so it's likely that their response will be incorrect. This was intended to make participants wary of their intuitions without explicitly instructing them to reflect (i.e. spending more time, thinking more carefully, etc). In the control condition, participants were instructed to answer problems to the best of their ability.

After responding to each problem, participants were asked "How confident are you that your response here is completely correct?" and responded on a scale ranging from 1 (*not at all confident*) to 7 (*extremely confident*). These ratings were averaged into a confidence score across all items ($\alpha = .92$). Participants then responded to demographics, including the religiosity scale ($\alpha = .97$). I again calculated three scores using the same procedures as in previous studies: the PD reflective, PD intuitive, and a bipolar score. I also calculated the average time spent on each problem. On the incongruent trials, 39% of the total responses were intuitive/incorrect, 55% were correct, and 7% were incorrect but non-intuitive. The full distribution of response types is presented in the supplementary materials (Table S1).

Results

Zero-order correlations were examined first (see Table 6). The PD reflective and intuitive parameters correlated negatively, and the bipolar score correlated positively with the PD reflective parameter and negatively with the PD intuitive parameter.

Table 6. Means, standard deviations, and correlations in Study 4.

Variable	<i>M</i>	<i>SD</i>	1	2	3	4
1. PD Reflective	-0.11	0.98				
2. PD Intuitive	0.14	0.87	-.15			
3. Bipolar score	7.87	3.20	.89**	-.44**		
4. Incongruent time (seconds)	25.89	9.88	.10	-.10	.07	
5. Congruent time (seconds)	15.37	5.01	.31**	-.00	.21	.58**

Note. *M* and *SD* are used to represent mean and standard deviation, respectively;

* indicates $p < .05$. ** indicates $p < .01$.

As shown in Table 4, one-way ANOVA showed that average confidence differed across the three conditions. The pre-registered contrast comparing the unconfident condition to the other two conditions simultaneously was significant, $t(238) = 3.75$, $p < .001$. Post-hoc tests with no correction (least significant difference) show that control and reflection conditions were not significantly different ($p = .343$), but confidence was significantly lower in the unconfident condition compared to the control ($p = .005$) and the reflection ($p < .001$) conditions. However, omnibus ANOVAs showed that the conditions were not clearly related to either the PD reflective, $F(2, 238) = .84$, $p = .434$, $\eta_p^2 = .01$, or PD intuitive parameters, $F(2, 238) = .75$, $p = .474$, $\eta_p^2 = .01$.

The preregistered hypothesis also concerned a test of average confidence as a mediator between condition and the PD reflective parameter (see the supplementary materials), but I am not focusing on this due to the problems in interpreting cross-sectional mediation (Fiedler et al., 2018; Kline, 2015; Spencer et al., 2005; Trafimow, 2015). Instead, I provide an alternative analysis which allows for extremely similar conclusions without misleading interpretations.

The standardized PD R score was regressed onto average confidence, a dummy code for the reflective condition and a dummy code for the unconfident condition (control condition was the reference group) simultaneously. This analysis would be similar to the “step 3” of a

mediation model, examining which factors in the regression model ultimately explain unique variance in the PD reflective parameter.

As shown in Table 7, average confidence and the unconfident condition were more strongly correlated with the PD reflective parameter than the reflection condition. Further, only the unconfident condition resulted in higher PD reflective scores compared to the control condition. Considering the result above (the unconfident condition resulted in less confidence in responses) it appears that awareness of one's ability to respond accurately contributes to the PD reflective score. The same correlations were not significant in predicting the PD intuitive score.

Table 7. Regression model predicting standardized PD reflective and intuitive parameters with average confidence and condition in Study 4.

Predictor	PD Reflective		PD Intuitive	
	<i>beta</i>	95% CI	<i>beta</i>	95% CI
Average confidence	0.35**	[0.23, 0.48]	-0.03	[-0.16, 0.11]
Reflection condition	0.07	[-0.07, 0.20]	-0.08	[-0.23, 0.06]
Unconfident condition	0.15*	[0.00, 0.29]	-0.08	[-0.23, 0.07]

Note. * indicates $p < .05$. ** indicates $p < .01$.

It's also of note that controlling for confidence resulted in a slightly stronger relation between the unconfident condition and the PD reflective parameter (relative to the control condition). This could suggest that confidence uniquely accounts for some portion of the PD reflective score, as reported by previous research (e.g. Thompson et al., 2011) because the variance in these two parameters is redundant. Finally, as shown in Table 2, participants again displayed less variance on the PD intuitive parameter compared to the PD reflective parameter.

Despite an apparent lack of effectiveness in our experimental manipulation, this study does provide evidence towards our hypothesis that an increased sensitivity to possible

incongruence between the intuitive and correct responses results in greater levels of reflection. In the unconfident condition, lower levels of confidence were correlated with higher PD reflective scores. Yet, the PD reflective parameter was very similar between the unconfident and reflective conditions, suggesting each increased reflection relative to the control. Again, no effects were observed on the PD intuitive parameters. At the least, this again suggests that the reflective process occurs after the intuitive parameter, as predicted by serial models of operation (e.g. Evans & Stanovich, 2013).

General Discussion

Summary

The four studies here have examined whether the cognitive reflection test measures both intuition and reflection and, in the course of this examination, have offered some evidence for the two opposing models of cognition. Process dissociation was used to distinguish between the intuitive and reflective scores on the CRT and three of the studies attempted to manipulate each parameter individually. In summary, little evidence was found to suggest that the CRT measures intuition. Instead, it appears that the CRT measures reflection or something close to reflection, such as engagement.

The second main finding of this research concerns the parallel and serial models of cognition. Though the results of studies 2-4 do not provide strong support for the serial model, they do fail to provide any evidence for the parallel model—at least using the limited measurements available via the CRT. As a result, it appears that the CRT is unsuitable to provide direct tests of opposing models of cognitive engagement.

There are several reasons why I draw the above conclusions: 1) weak effects of manipulations, 2) an apparent lack of ability to influence the intuitive parameter, 3) the inability

to manipulate the bipolar score, and 4) extremely strong correlations between the PD reflective and bipolar scores. I discuss each below.

Weak overall effects. Across studies 2-4, manipulations of independent processes were weak. If the CRT were a precise and accurate measure of (at least) reflection, the manipulations used here should have resulted in fairly strong changes in the reflective parameter as seen in previous research (e.g., Conway & Gawronski, 2013; McPhetres et al., 2018). Additionally, the manipulations did not have any effect on the bipolar scores, possibly because the poles are confounded, but possibly because the CRT measure is imprecise. There are several possible reasons for this. For example, the CRT could be an imperfect measure of reflection and this may be exacerbated by confounding the measurements, or it may simply be that there is another process responsible for variation in reflective processes which was not directly targeted in these studies.

The inability to manipulate the intuitive parameter. Regarding the second point, the attempt to manipulate the intuitive parameter was unsuccessful. In Study 3, the manipulation was designed to increase intuitiveness, yet resulted in a decreased PD reflective parameter. This could have occurred for two reasons. One possibility is that people do not vary in their levels of intuitiveness and any attempt to manipulate intuition simply results in effects on reflection. Similarly, the CRT may not contain information about intuition (even if people do vary on intuitiveness) so the PD intuitive parameter is unable to be affected by any manipulation targeting this parameter. Another possibility is that reflection is dependent upon intuition, because intuition comes first temporally. That is, attempting to influence one's intuition would be like "kicking the legs out from under" reflection. The same possibilities exist when considering why the manipulations did not affect the bipolar score. Though this puzzle is not

solved in the present studies, this conundrum suggests that the CRT is not the proper tool to answer these questions.

Below, Figure 3 depicts a conceptual model of intuitive and reflective processes that allows for each of these possibilities.

Figure 3. Conceptual depiction of the time-series model of intuition and reflection.

In this diagram, reflective processes are depicted as being fully dependent on processes that occur prior to it in time: intuition and some process which comes between intuition and reflection. All people go through each of these stages in order, occurring stepwise over time. Thus, variance, only occurs at Stage 2 and afterwards. This also means that attempts at influencing *intuition* per se must result in effects on the processes that occur after intuition (though one might also be successful in affecting intuition).

As demonstrated in the four studies described here, people evince considerably less variance on intuition (Stage 1). Any variance at Stage 1 would be meaningless, it should always be small, and could be measurement error. Thus, the questions don't *measure* whether one does

or doesn't think of the intuitive response. Instead, the CRT questions must measure the processes that come after it in time. That is, whether people ultimately accept/evaluate the intuitive response, or detect a conflict (e.g., Pennycook et al., 2015) between the intuitive and reflective responses (or some other unknown process; Stage 2) and/or whether they go on to reflect on that option prior to arriving at a decision. Those who engage reflection have an extra stage (i.e., extra time) in between intuition and the decision.

An additional prediction made from this model is that people who engage in reflection should take longer to arrive at a decision compared to those who do not engage in reflection. This prediction is easily testable using the PD questions in the present studies. Indeed, the PD reflective parameter correlates positively with response time on *congruent trials* (where both the intuitive and reflective paths result in a correct decision) while the PD intuitive parameter remains uncorrelated (See Tables 1, 3, 5, and 6). This suggests that more reflection takes longer, even when both the intuitive and reflective processes arrive at the same conclusion (i.e. on congruent trials).

Of course, this also means that one could “intuitively” arrive at the correct answer as argued in previous research (e.g., Bago & de Neys, 2017). Indeed, there are many ways that formerly ‘reflective’ thought can become faster: one gets better at a challenging task through practice and repetition and it no longer requires the time or resources. However, intuition in those cases are still not measurable using the CRT because they cannot be distinguished from other intuitions or heuristics.

Strong similarity between the bipolar and PD reflective parameter. Third, across the four studies, the PD reflective parameter correlated very strongly with the bipolar scoring method, hovering around correlations of $\sim .90$. The strength of these correlations seems to

suggest there is little difference between the PD reflective and the bipolar scores. The PD intuitive parameter correlated around the range of .4, which suggests that it is not completely unrelated—that is, the score does not appear to be a complete artifact. Past research has reported correlations between the deliberative and bipolar scores in the range of .7-.8 (e.g., Friesdorf et al., 2015; Conway & Gawronski, 2013). However, in conjunction with the other interpretations just discussed, it does appear that the CRT measures reflection but not intuition. Still, it is unclear exactly what the PD intuitive parameter is measuring.

Of course, it could be argued that the problem lies within the PD procedure itself—that is, perhaps PD is insufficient to examine these processes. However, it's worth pointing to nearly two decades of past research using PD and the quad model to distinguish between various automatic and deliberative processes as evidence of its overall effectiveness (e.g. Conrey et al., 2005; Hense et al, 1995; Payne, 2001). Further, correlations between the two parameters, the bipolar scores, and other measures were consistent with expectations and past research (e.g. Friesdorf et al., 2015) suggesting that the PD procedure operated as it should. Thus, if the CRT actually measured automatic (intuition) and deliberative (reflection) processes, it seems very likely that PD would be able to isolate those processes effectively.

Theoretical Context

The conclusion that the CRT does not measure intuition is in line with other recent research on this topic. For example, Patel and colleagues (Patel, et al, 2018) offered participants a multiple-choice version of the CRT. They found that even when the response options did not include the incorrect intuitive response, participants still had difficulty solving the problems. Furthermore, the correlations between measures such as religiosity and moral judgments remained, suggesting that the CRT is more of a measure of numeracy and reflection than it is a

measure of intuition. Indeed, other researchers have found that the CRT correlates strongly with various other measures of cognitive ability (Brooks & Pui, 2010; Cokeley, Galesic, Schulz, Ghazal & Garcia-Retamero, 2012; Frederick, 2005; Peters et al., 2006; Peters, 2012; Weller et al., 2013; Zuckerman, Silberman & Hall, 2013). In particular, researchers have claimed that the CRT measures numeric ability more than any process resembling cognitive reflection (Sinayev & Peters, 2015). Other researchers (Stuppel et al., 2017) that the three items on the original CRT each measure different things, none of which are cognitive reflection. Still, some research has claimed that the CRT measures constructs unique from numeracy and cognitive ability (e.g., Liberali, Reyna, Furlan, Stein & Pardo, 2011).

The present findings also have implications for how the CRT is interpreted. As mentioned previously, researchers often interpret the CRT as describing *more intuitive* or *more reflective* participants (e.g. Pennycook et al., 2015; Shenhav et al., 2012). However, in the present studies it was found that religiosity correlated only with the PD reflective parameter and not with the PD intuitive parameter. This pattern was found in Studies 1 and 2, and the correlation was weaker and non-significant in Studies 3 and 4. Thus one interpretation of this is that religious people are not more intuitive but are instead less reflective. However, as noted by multiple authors previously, it is also possible that the CRT measures effort rather than cognitive ability (e.g. Evans, 2011; Frederick, 2005; Toplak et al., 2014; Patel et al., 2018). Further, reflection can be many things: cognitive decoupling, verification, correction, or simply time (e.g., Evans, 2019). While the present study does not clarify whether the CRT measures reflection or numeracy or engagement (or something else), the results elucidated by the PD procedure here do clarify that the correct interpretation of the CRT is not one focused on intuition.

Final Comments

These results both complicate and clarify several issues with the CRT and with dual-process models of cognition. While these studies do not conclusively offer support for the serial/default-interventionist models of cognitive engagement, they do suggest a lack of evidence for parallel models. Further, they suggest that the CRT is not the right tool to investigate such questions given that the measure does not appear to capture intuition. However, it remains unclear what the PD intuitive parameter represents.

Moving forward, there are a few suggestions for researchers interested in using the CRT. First, it seems that researchers should consider alternative measures of cognitive reflection. It's worth noting here that, in Study 1, the modified CRT questions were strongly related to the original questions, but they did not correlate perfectly, suggesting that people did not perceive them as the same problems. This also suggests that modified versions of the questions may be useful in determining what processes, if any, the CRT actually captures. Thus, one solution to the overuse of the CRT may be to simply reword some of the original items. Second, it is recommended that researchers are cognizant of the phrasing used when describing intuitions as inferred from the CRT, because these studies provide no evidence that the CRT actually described one's level of intuition. Further, whether one can be more or less intuitive also depends on assumptions about the nature and relation between the two processes. Finally, these problems point to a need for measures of intuition. In my view, how intuitions lead us astray or in the right direction are still interesting and important questions for psychologists to consider. We just need to devise a good way of measuring something on which everyone is expected to score highly.

Footnotes.

1. The final two items of the faith in intuition scale (*I have a very good sense of rhythm*, and *I am good at visualizing things*) were removed because alpha reliability levels were increased from .68 to .77 without them.

References

- Bago, B., & De Neys, W. (2017). Fast logic?: Examining the time course assumption of dual process theory. *Cognition*, 158, 90-109.
- Bago, B., & De Neys, W. (2019a). Advancing the specification of dual process models of higher cognition: a critical test of the hybrid model view. *Thinking & Reasoning*, 1-30.
- Bago, B., & De Neys, W. (2019b). The smart System 1: Evidence for the intuitive nature of correct responding on the bat-and-ball problem. *Thinking & Reasoning*, 25(3), 257-299.
- Barbey, A. K., & Sloman, S. A. (2007). Base-rate respect: From ecological rationality to dual processes. *Behavioral and Brain Sciences*, 30(3), 241-254.
- Bialek, M., & Pennycook, G. (2017). The Cognitive Reflection Test is robust to multiple exposures. *Behavior research methods*, 1-7.
- Brooks, M. E., & Pui, S. Y. (2010). Are individual differences in numeracy unique from general mental ability? A closer look at a common measure of numeracy. *Individual Differences Research*, 8(4), 257-265.
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., ... & Altmejd, A. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, 2(9), 637.
- Cohen, A. B., Shariff, A. F., & Hill, P. C. (2008). The accessibility of religious beliefs. *Journal of Research in Personality*, 42(6), 1408-1417.

- Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin numeracy test. *Judgment and Decision Making*, 7(1), 25-47.
- Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: the quad model of implicit task performance. *Journal of personality and social psychology*, 89(4), 469.
- Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: the quad model of implicit task performance. *Journal of personality and social psychology*, 89(4), 469.
- Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: a process dissociation approach. *Journal of personality and social psychology*, 104(2), 216.
- Deppe, K. D., Gonzalez, F. J., Neiman, J. L., Jacobs, C., Pahlke, J., Smith, K. B., & Hibbing, J. R. (2015). Reflective liberals and intuitive conservatives: A look at the Cognitive Reflection Test and ideology. *Judgment & Decision Making*, 10(4).
- Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American psychologist*, 49(8), 709.
- Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive–experiential and analytical–rational thinking styles. *Journal of personality and social psychology*, 71(2), 390.
- Evans, J. S. B. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning*, 13(4), 321-339.
- Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review Psychology*, 59, 255-278. Evans, J. S. B. (2011). Dual-process theories of

- reasoning: Contemporary issues and developmental applications. *Developmental Review*, 31(2), 86-102.
- Evans, J. S. B. (2019). Reflections on reflection: the nature and function of type 2 processes in dual-process theories of reasoning. *Thinking & Reasoning*, 1-33.
- Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science*, 8(3), 223-241.
- Farias, M., Mulukom, V., Kahane, G., Kreplin, U., Joyce, A., Soares, P., ... & Möttönen, R. (2017). Supernatural belief is not modulated by intuitive thinking style or cognitive inhibition. *Scientific Reports*, 7(1), 15100.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2), 175-191.
- Fiedler, K., Harris, C., & Schott, M. (2018). Unwarranted inferences from statistical mediation tests—An analysis of articles published in 2015. *Journal of Experimental Social Psychology*, 75, 95-102.
- Finucane, M. L., & Gullion, C. M. (2010). Developing a tool for measuring the decision-making competence of older adults. *Psychology and aging*, 25(2), 271.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, 19(4), 25-42.
- Friesdorf, R., Conway, P., & Gawronski, B. (2015). Gender differences in responses to moral dilemmas: a process dissociation analysis. *Personality and Social Psychology Bulletin*, 41(5), 696-713.

- Gervais, W. M., & Norenzayan, A. (2012). Analytic thinking promotes religious disbelief. *Science*, 336(6080), 493-496.
- Handley, S. J., Newstead, S. E., & Trippas, D. (2011). Logic, beliefs, and instruction: A test of the default interventionist account of belief bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(1), 28.
- Hayes, A. F. (2017). My macros and code for SPSS and SAS. URL: <http://afhayes.com/spss-sas-andmplus-macros-and-code.html> (visited on 26/01/2017).
- Hense, R. L., Penner, L. A., & Nelson, D. L. (1995). Implicit memory for age stereotypes. *Social Cognition*, 13(4), 399-415.
- Hense, R. L., Penner, L. A., & Nelson, D. L. (1995). Implicit memory for age stereotypes. *Social Cognition*, 13(4), 399-415.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of memory and language*, 30(5), 513-541.
- Kahneman, D., & Egan, P. (2011). *Thinking, fast and slow* (Vol. 1). New York: Farrar, Straus and Giroux.
- Kahneman, D., & Slovic, P. (1982). Amos Tversky, eds. 1982.
- Kline, R. B. (2015). The mediation myth. *Basic and Applied Social Psychology*, 37(4), 202-213.
- 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(4), 361-381.

- 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.
- McCall, R. (2017). Can you pass the world's shortest IQ test? It's just three questions long, but few can get them all right. Accessed September 29, 2018 from:
<http://www.iflscience.com>.
- McPhetres, J., Conway, P., Hughes, J. S., & Zuckerman, M. (2018). Reflecting on God's will: Reflective processing contributes to religious peoples' deontological dilemma responses. *Journal of Experimental Social Psychology, 79*, 301-314.
- Meyer, A., Zhou, E., & Shane, F. (2018). The non-effects of repeated exposure to the Cognitive Reflection Test. Judgment and Decision making.
- Newman, I. R., Gibb, M., & Thompson, V. A. (2017). Rule-based reasoning is fast and belief-based reasoning can be slow: Challenging current explanations of belief-bias and base-rate neglect. *Journal of experimental psychology: learning, memory, and cognition, 43*(7), 1154.
- Neys, W. D. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological science, 17*(5), 428-433.
- Neys, W. D. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological science, 17*(5), 428-433.
- Patel, N. (2017). *The Cognitive Reflection Test: a measure of intuition/reflection, numeracy, and insight problem solving, and the implications for understanding real-world judgments and beliefs* (Doctoral dissertation, University of Missouri--Columbia).

- Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of personality and social psychology*, 81(2), 181.
- Payne, B. K., & Bishara, A. J. (2009). An integrative review of process dissociation and related models in social cognition. *European Review of Social Psychology*, 20(1), 272-314.
- Pennycook, G. (2017). A perspective on the theoretical foundation of dual process models. In *Dual Process Theory 2.0* (pp. 13-35). Routledge.
- Pennycook, G., & Thompson, V. A. (2012). Reasoning with base rates is routine, relatively effortless, and context dependent. *Psychonomic Bulletin & Review*, 19(3), 528-534.
- Pennycook, G., Cheyne, J. A., Koehler, D. J., & Fugelsang, J. A. (2016). Is the cognitive reflection test a measure of both reflection and intuition?. *Behavior Research Methods*, 48(1), 341-348.
- Pennycook, G., Cheyne, J. A., Koehler, D. J., & Fugelsang, J. A. (2016). Is the cognitive reflection test a measure of both reflection and intuition?. *Behavior Research Methods*, 48(1), 341-348.
- Pennycook, G., Cheyne, J. A., Seli, P., Koehler, D. J., & Fugelsang, J. A. (2012). Analytic cognitive style predicts religious and paranormal belief. *Cognition*, 123(3), 335-346.
- Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). Everyday consequences of analytic thinking. *Current Directions in Psychological Science*, 24(6), 425-432.
- Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology*, 80, 34-72.
- Pennycook, G., Ross, R. M., Koehler, D. J., & Fugelsang, J. A. (2016). Atheists and agnostics are more reflective than religious believers: Four empirical studies and a meta-analysis. *PloS one*, 11(4), e0153039.

- Peters, E. (2012). Beyond comprehension the role of numeracy in judgments and decisions. *Current Directions in Psychological Science*, 21(1), 31-35
- Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science*, 17(5), 407-413.
- Raoelison, M., & De Neys, W. (2019). Do we de-bias ourselves?: The impact of repeated presentation on the bat-and-ball problem. *Judgment and Decision making*.
- Sanchez, C., Sundermeier, B., Gray, K., & Calin-Jageman, R. J. (2017). Direct replication of Gervais & Norenzayan (2012): No evidence that analytic thinking decreases religious belief. *PloS one*, 12(2), e0172636.
- Shenhav, A., Rand, D. G., & Greene, J. D. (2012). Divine intuition: Cognitive style influences belief in God. *Journal of Experimental Psychology: General*, 141(3), 423.
- Sinayev, A., & Peters, E. (2015). Cognitive reflection vs. calculation in decision making. *Frontiers in Psychology*, 6, 532.
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological bulletin*, 119(1), 3.
- Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and social psychology review*, 4(2), 108-131.
- Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of personality and social psychology*, 89(6), 845.
- Stanovich, K. E., & Toplak, M. E. (2012). Defining features versus incidental correlates of Type 1 and Type 2 processing. *Mind & Society*, 11(1), 3-13.

- Stuppel, E. J., Pitchford, M., Ball, L. J., Hunt, T. E., & Steel, R. (2017). Slower is not always better: Response-time evidence clarifies the limited role of miserly information processing in the Cognitive Reflection Test. *PloS one*, *12*(11), e0186404.
- Thompson, V. A., Turner, J. A. P., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive psychology*, *63*(3), 107-140.
- Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, *11*(1), 99.
- 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.
- Trafimow, D. (2015). Introduction to the special issue on mediation analyses: What if planetary scientists used mediation analysis to infer causation?. *Basic and Applied Social Psychology*.
- Travers, E., Rolison, J. J., & Feeney, A. (2016). The time course of conflict on the Cognitive Reflection Test. *Cognition*, *150*, 109-118.
- Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J., & Peters, E. (2013). Development and testing of an abbreviated numeracy scale: A Rasch analysis approach. *Journal of Behavioral Decision Making*, *26*(2), 198-212.
- 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.

Zuckerman, M., Silberman, J., & Hall, J. A. (2013). The relation between intelligence and religiosity: A meta-analysis and some proposed explanations. *Personality and Social Psychology Review*, 17(4), 325-354.