



# The cognitive ability and working memory framework: Interpreting cognitive reflection test results in the domain of the cognitive experiential theory

Ayşegül Engin<sup>1</sup>

Accepted: 12 November 2020 / Published online: 24 December 2020

© The Author(s) 2020

## Abstract

The cognitive reflection test (hereafter, CRT) is a widely used tool in studies that deal with human decision-making and problem-solving. However, the interpretation of the results of the test remains an open discussion in extant literature. Despite the high predictive power, the characteristics of a decision-maker measured by the CRT remain unclear. This article suggests a novel and inclusive framework that relates the CRT to the cognitive experiential theory (hereafter, CET), which is a well-established dual-process theory. The framework explains the observed phenomena (e.g., individual decision-maker differences) using data from five laboratory experiments. The framework is inclusive because it provides possible explanations for the results in the published literature. The overall results support the argument that the CRT measures the efficiency of the resources of an individual's working memory. In contrast, the rationality experientiality inventory (hereafter, REI) (a measurement tool of the cognitive experiential theory) captures an individual's cognitive ability. The efficient use of the working memory leads to the full potential of decision-makers' cognitive abilities in analytical tasks, thereby indicating that *cognitive ability and working memory* form a viable framework for studies in the disciplines behavioral operational research (hereafter, OR) and decision sciences.

**Keywords** Cognitive ability and working memory framework · Cognitive reflection test · Cognitive experiential theory · Dual-process theory · Laboratory experiment · Behavioral OR

---

I want to thank Prof. Dr. Rudolf Vetschera for introducing me to research, motivating and inspiring talks, and believing in me.

---

✉ Ayşegül Engin  
ayseguel.engin@univie.ac.at

<sup>1</sup> Department of Business Decisions and Analytics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria

## 1 Introduction

In behavioral OR, a sub-discipline of OR that is concerned with decision-making practice and human problem-solving (Hämäläinen et al. 2013), the CRT is a popular tool. In their meta-study Branas-Garza et al. (2019) report 118 studies, which use the CRT across 21 countries with 44558 participants, thereby documenting the aforementioned scientific popularity. The CRT is a widely used measure for online studies and laboratory experiments because of its prediction power of an individual's ability to make rational decisions in a wide variety of contexts (Primi et al. 2016).

What remains unclear regarding the CRT is the source of its predictive power due to the lack of theoretical consensus. More specifically, the exact decision-maker traits that the tool measures remain an open question (Erceg and Bubić 2017). Until a theoretical link is established between the CRT and the related theories, the interpretations of the test will continue to remain unclear and the debate will remain open.

While the initial development of the test is associated with the domain of dual-process theories (Frederick 2005), the exact relationship with the dual-process framework remains to be investigated. Earlier literature suggests interpreting the results as intelligence (Gino and Ariely 2012), cognitive style (Pennycook et al. 2012), general mental ability (Thomson and Oppenheimer 2016; Toplak et al. 2014), and intuition (Alós-Ferrer and Hügelschäfer 2016) of individuals.

This paper suggests a new framework that places the measurement of the CRT in the context of the CET, which is an inclusive dual-process theory. It provides an encompassing explanation for the published results in the literature and a comprehensible conceptualization of the results of the CRT within the CET.

The remainder of the paper is structured in the following manner: Sect. 2 introduces the CET and presents the relevant literature for both the CET and the CRT. Section 3 presents the suggested novel framework in detail and formulates the hypotheses. Section 4 provides an insight into the data that is used for the analyses. Section 5 highlights the results, and Sect. 6 concludes the paper.

## 2 Literature

Among decision-making frameworks, dual-process theories constitute a significant pillar, with the main idea being the combined working of two decision-making processes, called type 1 and type 2 (for a review see Padilla et al. 2018). Initially, the idea of having merely two decision-making processes appears like an oversimplification (Evans 2008) and the lack of agreement on the definition of the two systems is discussed broadly in the literature (e.g., Marewski and Gigerenzer 2012; Evans 2008; Sloman 2002). Evans and Stanovich (2013) suggest a revised definition of the processes based on behavioral and neuroscience evidence, characterizing all processes requiring autonomy as type 1 and all processes requiring controlled attention as type 2. For the current paper, dual-process theories with the definition given by Evans and Stanovich (2013) are considered. The remainder of this section explains one of the dual-process theories as the main theoretical background for this study.

## 2.1 The cognitive experiential theory (CET)

The CET (formerly known as the cognitive experiential self-theory) is developed as an integrative personality theory, which incorporates the dual system perspective (i.e., the experiential and the rational systems) (Epstein 1973). It includes aspects from *self-theory*, *cognitive science*, *learning theory*, *theory of emotions*, and *psychoanalytic theory* (Epstein 2014). Because of its inclusive property, the CET delivers plausible explanations for situations that other theories declare as outliers, and therefore, do not provide an explanation for. Therefore, in the current paper, the CET is the choice of theoretical background. The relevant details of the CET are explained below.

### 2.1.1 Definition of the two systems

The CET utilizes an implicit self-theory to explain the adaptive nature of human behavior. On the one hand, it assumes that every individual has implicit beliefs regarding how they think and feel about themselves and their environment, which is automatically derived from their experiences. These implicit beliefs reside in the cognitive map of the individuals' brain and have an implicit, dynamic, and hierarchical structure, which is referred to as the *experiential system* (Epstein 1994). In the lingo of the general dual-process theories, the experiential system conducts type-1 processing. On the other hand, *the rational system* operates in accordance with the individual's logical reasoning and it performs type-2 processing. This system is majorly transmitted through culture, as it requires language for complex logical inference.

According to the CET, the experiential system has a biasing influence on the rational system, which is considered the primary source of systematic decision errors (Epstein et al. 1992). The CET argues that the main reason for the biasing influence is the different motives of the rational and the experiential systems. The rational system operates according to the reality principle, thereby leading to reflective and deliberative behavior. In comparison, the experiential system acts according to the hedonic principle, which can be considered a mixed blessing depending on the situation. Hence, the hedonic principle is the source for adaptive as well as maladaptive learning, where the latter has a biasing influence on the rational system (Epstein et al. 1992).

### 2.1.2 Comparison of the two systems

Since both rational and experiential systems have different motivators, they also differ in terms of characteristics. The cardinal difference between both systems is that the rational system is a verbal reasoning system, whereas the experiential system is an associative learning system. As a result of this cardinal difference, learning occurs differently in both systems. The rational system can learn by logical inference and reasoning. However, the experiential system learns through association, for which reinforcement through emotion is necessary, which is in line with the hedonic principle. Learned information is encoded in the experiential system primarily in the form of images and nonverbal forms; in comparison, the rational system stores information only verbally. As such, the experiential system works through associative connections and requires "*affect*" as a mediator for its learning process; in contrast, the rational

system operates with a cause-and-effect relationship, and it is affect-free. According to the CET, because of the cognitive resource efficiency of the experiential system, individuals primarily react associatively to cause and effect relationships. Subsequently, they adapt their initial interpretations with the intervention of the rational system if the initial interpretations are inadequate. If the rational system fails to intervene for various reasons or its capabilities are exceeded by the task (e.g., a 5-year-old child confronted with a calculus task), a maladaptive response occurs. In the literature, systematic maladaptiveness in decision-making is defined as biases (Epstein 2014).

Information processing is another aspect that differs between rational and experiential systems. The rational system analytically processes information at high cognitive resource costs; in contrast, this processing occurs effortlessly, automatically and holistically in the experiential system. Because of its holistic characteristic, the experiential system is less focused on details than the rational system, and thus, reacts more categorically. Even though analytical processing appears in the first step as the more effective one of the two systems, despite its cognitive costs, a variety of studies indicate that specific problems (i.e., problems requiring personal choices that contain a hedonic elements like choosing a decoration object like posters, complex tasks that exceed the processing capacity of the rational system) are addressed better by the experiential system (e.g., Dijksterhuis 2004; Wilson 2004; Wilson et al. 1993), thereby indicating that the decision dominance of one system over the other is strongly task-dependent. Wilson et al. (1993) conduct an experiment to study the relationship between reflection and post-decision satisfaction on a personal task. The authors argue that explicit reflection on hedonic or personal decisions raises awareness of a person's lack of consistency, thereby decreasing decision satisfaction. Dijksterhuis (2004) contributes to the discussion by finding evidence in favor of the decisions taken by the experiential system for multi-attribute decision problems. In this study, the main argument for the inferiority of the rational system is its low processing capacity. In his seminal work, Wilson (2004) describes the two systems mostly from a Freudian perspective, whereas the explanation for the information processing mechanisms between Wilson (2004) and Epstein (2014) overlap significantly.

The experiential system is outcome-focused compared to the process-oriented rational system. The former does not distinguish intentional outcomes from accidental ones, as opposed to the latter. For example, a young child judges a behavior only based on its outcome, whereas an older child takes the intent of the outcome into account as well. Further, compared with children, adults can behave more experientially with the knowledge that their behavior is irrational (e.g., tipping the messenger for delivering excellent news, even though the messenger does not affect the outcome of the news) (Epstein 2014; Epstein et al. 1992).

Another important aspect that differentiates between the two systems is their processing speed. The experiential system responds much more rapidly than the rational system; thus, it is almost always initiating subsequent interactions between the two systems. The rational system subsequently responds as a corrective system (Gilbert and Malone 1995). In their experimental paper Gilbert and Malone (1995) show that individuals decide on attributes before they take the situation into account (i.e., correspondence bias). The authors report that the primary reason for this behavior is an individual's lack of situational awareness. The correspondence bias can be remedied

through reflection, which requires resources from the working memory (i.e., increasing cognitive load). With reflection, the initial reaction produced by the experiential system is corrected by the rational system.

Further, changes in the experiential system require greater effort than those in the rational system. The reason for this difference is the trial-and-error learning that is typical to the experiential system. For this system, learning can only occur with increased repetition or intensity. The rational system only requires learning of the correct logical steps to mitigate its shortcomings (Epstein 2014).

While considering the working mechanisms of rational and the experiential systems, it is crucial to emphasize their parallel and bidirectionally interactive operations. The fast experiential system initiates the interaction and the slower rational system influences its outcomes as a corrective element. It is assumed that both systems contribute to all human behaviors in varying amounts. The difference in how much one system contributes to a specific decision task depends on personal characteristics and decision-making situations.

## 2.2 The cognitive reflection test (CRT)

The CRT is a widely used measure in economics and psychology literature. Recently, it has also been used in behavioral OR studies (e.g., Engin and Vetschera 2019). The test presents three open questions, with each of them having an intuitive wrong answer. However, if the respondents choose to reflect and take their time to consider, the correct answers to the questions can be found without complications. The number of correct answers provides the score to the respondents. Cueva et al. (2016) classify respondents that answer two or all questions correctly as *reflective* decision-makers, and respondents that have only one or no correct answers as *impulsive* decision-makers.

The original questions of the CRT are presented below:

- 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? Provide the answer in cents. [intuitive wrong answer: 10, correct answer: 5]
- If it takes five machines 5 minutes to make five widgets, how long would it take 100 machines to make 100 widgets? [intuitive wrong answer: 100, correct answer: 5]
- 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? [intuitive wrong answer: 24, correct answer: 47]

The three-item measure was first proposed by Frederick (2005) based on dual-process theories (e.g., Kahneman and Frederick 2002; Stanovich and West 2000; Sloman 1996; Epstein 1994). Because of the theoretical proximity of the CRT and dual-process theories, a high CRT score is argued to relate to high cognitive ability. Frederick (2005) finds a high correlation between CRT scores and other tests of analytic thinking. The CRT score is considered as a combination of cognitive ability and a reflective decision-maker characteristic (Toplak et al. 2014).

Despite its extensive usage (Branas-Garza et al. 2019), criticism is raised against the usage of the CRT and the interpretation of its results. The main arguments for the

criticism are the conciseness of the test (i.e., three items might not be sufficient to capture a decision-maker trait) and, the risk of memorized responses resulting from the popularity of the test, which might distort its results. Bialek and Pennycook (2018) respond to critics with an article that reveals that among six studies with a total of 2500 participants, repeated exposure to CRT did not significantly undermine the validity of the test. In addition, on the one hand, Blacksmith et al. (2019) provide further evidence on the validity of the CRT using item response trees, which is in line with the dual theory approach. On the other hand, they suggest that the CRT might not be interpreted as a test of general cognitive ability (Thomson and Oppenheimer 2016), cognitive style (Pennycook et al. 2012), or intelligence (Gino and Ariely 2012).

The major open question in the CRT literature concerns the original conceptualization of CRT in the dual-process theories. In its first conceptualization, Frederick (2005) defines cognitive reflection as “*the ability or disposition to resist reporting the response that first comes to mind*”; its theoretical construct is embedded in the heuristic-analytic theory of reasoning (Evans 2006). Even though the heuristic-analytic theory can provide a dual process conceptualization, according to Blacksmith et al. (2019), it is an oversimplification of the original intent of the CRT. Another issue that leads to differing interpretations of CRT results is the insufficient specificity of the content domain (Erceg and Bubić 2017; Pennycook et al. 2016).

### 3 Research model

As evident from the previous section, CRT as a measurement tool is closely related to the CET. This section presents a framework to address the theoretical positioning of the CRT and CET.

#### 3.1 Cognitive ability and working memory framework

Previous works contributed to the theoretical grounding of CRT by distinguishing between CRT results and general mental ability (Blacksmith et al. 2019) and CRT results and faith in intuition scale (Alós-Ferrer and Hügelschäfer 2016). The latter emphasizes distinguishing the intuition concept in CET and CRT results. This paper adds to the discussion introduced in Alós-Ferrer and Hügelschäfer (2016) with a novel and integrative framework by investigating the contextual relationship between CRT as a measurement tool and CET as the theoretical domain from the dual-process theories. For clarity, two concepts are used in the subsequent sections of the article, as defined below:

- *Cognitive ability* refers to dual-processing skills. It has degrees—that is, a person can score high or low on rationality.
- *Working memory* is defined by Cowan (2017) as limited cognitive resources with multiple components for maintaining a limited amount of information for a limited amount of time. Existing theories emphasize the automatic response suppressing property of the working memory that aids in maintaining decision-relevant information but with limited capacity (for a review see Padilla et al. 2018).

Extracting from the descriptions of the CET and CRT (see Sect. 2), it is argued in the framework that CET is a measure for cognitive ability, and CRT is a measure for working memory efficiency. Even though CET is a theoretical concept and CRT is a measurement tool, they are related to each other because of the sequential processing of the two systems (Baron et al. 2015; Epstein 2014) (see a detailed description, and reasoning for the sequential information processing of both systems in Sect. 2.1.2). As Evans and Stanovich (2013) indicate, rational processing requires significant working memory. Hence, the cognitive ability cannot be utilized to its full extent with inefficient working memory usage at a theoretical level.

The proposed framework is based on Padilla et al. (2018) and Engin and Vetschera (2017). Figure 1 depicts the decision processes in abstract steps and with the example of the ball and bat question in the CRT (for the exact wording of the question, see Sect. 2.2). At this point, the reader must note that the proposed framework has similarities to the models in the decision-making with visualizations in the extant literature. Bearing in mind that analytical decision tasks are generally communicated to individuals through visual cues (i.e., text, numbers, graphs, tables) and do not contain other cues (i.e., auditory, haptic) that are relevant to the task, this framework can be considered as appropriate for analytical decision tasks. Further, it is important to remark that the model might not apply for hedonic decision tasks in its current format.

For the sake of clarity, let us first consider the abstract steps of the framework. An analytical decision-task is introduced to an individual with limited working memory. At this point, efficient working memory usage is assumed for the sake of argumentation. Therefore, the available working memory has sufficient resources. The external information of the task is perceived both by the rational and experiential systems. In the mental representation stage, the experiential system employs cognitive maps to ascer-

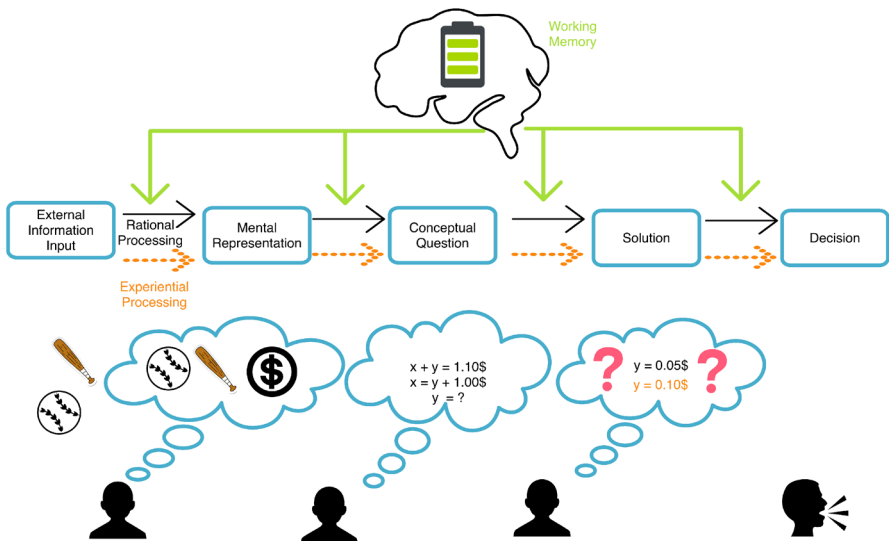


Fig. 1 Cognitive ability and working memory framework (based on Padilla et al. 2018; Engin and Vetschera 2017)



tain, if such a decision task has been experienced before. The rational system extracts decision-relevant information in a logical process for creating the mental representation using available working memory. Once the external information is translated to the mental representation of the task by the help of both systems, a conceptual question is generated. If a familiar task is recorded in the individual's cognitive map, the individual's experiential system can help the individual to extract the conceptual question of the decision task. Otherwise, learning takes place through experience (i.e., adaptive learning). Meanwhile the rational system continues to use the available cognitive resources and answers the conceptual question with a generated solution. In the case that a familiar task is discovered in the cognitive map, the experiential system can reach a solution, even faster than the rational system, because of its automated processes. In this case, the solution candidate/s will be considered by both systems and the decision is declared by the individual.

Now, let us consider the example case presented in the bottom half of Fig. 1 and assume that the individual uses his/her working memory inefficiently. Consequently, his/her cognitive resources can be depleted at any stage of this framework. Depending on the stage at which the working memory resources are exhausted, the individual can only rely on his/her experiential system. The example task depicted on the Fig. 1 is the first question of the CRT (for explicit task formulation see Sect. 2.2). Upon reading the task, both systems receive the task information. While the experiential system searches the cognitive map for any question related to the ball and bat, the rational system extracts the numerical information for creating an accurate mental model. If working memory is depleted in this stage, the individual may miss or misread an important detail associated with the task (e.g., misreading numbers). For the generation of the conceptual question, the experiential system functions as an aid, provided that it can find a sufficiently similar task stored in the cognitive map. In other cases it provides its best guess. A rational system with sufficient working memory will be able to solve the simple equation system. However, the exhaustion of working memory at this stage may cause an unnoticed calculation error. In the solution stage, the best guess of an experiential system, assuming it cannot remember the correct result, is 0.10\$, and the calculated result is 0.05\$. Thus, depending on the available working memory, the experiential system's impulsive wrong answer will be corrected before the declaration of the decision.

Figure 1 depicts the parallel information processing of the systems. While one stream of literature in the dual-process theories advocates for parallel information processing (e.g., Evans and Stanovich 2013; Gigerenzer and Gaissmaier 2011), the idea is challenged by Epstein (2014) and Gilbert and Malone (1995). Epstein (2014) argues that the corrective interferences of the rational system are likelier to occur, (i.e., sequential processing) for efficiency reasons (see Sect. 2.1.2). The cognitive ability and working memory framework (depicted in Fig. 1) can incorporate both the idea of parallel processing given by Padilla et al. (2018), Evans and Stanovich (2013), Gigerenzer and Gaissmaier (2011) and that of sequential processing given by Epstein (2014). Strictly speaking, the latter causes the corrective inference of rational processing (illustrated as arrows in Fig. 1 before stages) to vanish (e.g., among the stages of mental representation, conceptual question, and solution). Nevertheless, this modification does not cause any change in the mechanism of the framework. The



only change that can occur is the effective usage of the working memory compared to that in parallel processing. For the sake of completeness, Fig. 1 illustrates the parallel processing variant. Nevertheless, in the remainder of the paper, sequential processing is assumed in line with the experimental evidence in the literature (Gilbert and Malone 1995).

### 3.2 Inventories and hypotheses

This section introduces the measurement tools required for the analysis and formulates the research question through hypotheses.

For the analyses, cognitive ability is measured by the REI, developed by Epstein et al. (1996). The inventory consists of 42 self-reported items regarding general statements regarding one's self (e.g., "I have a logical mind."). Individuals decide on the degree of agreement to the statements on a nine-point Likert scale. The inventory comprises four different constructs: rationality, emotionality, intuition, and imagination, where the experientiality score is the sum of the last three constructs.

It is argued that the efficient use of the working memory is measured by the results of the CRT. For interpreting the results, Cueva et al. (2016) introduces a scoring rule that is adopted in this article. He classifies individuals as *reflective* or *impulsive* decision-makers. Individuals who correctly reply to at least two questions of the original CRT are classified as reflective decision-makers, while individuals who commit more than one error on the test are classified as impulsive decision-makers. Cueva et al. (2016) interpret impulsive decision-makers as incapable of suppressing their immediate intuitive decisions. They report lower decision performance in judgment tasks from impulsive individuals. This interpretation of Cueva et al. (2016) overlaps with the properties of working memory (Shipstead et al. 2015; Zhu and Watts 2010; Kane et al. 2001; Engle et al. 1999), thereby making the aforementioned argument plausible.

*The cognitive ability and working memory framework* argues that impulsive decision-makers use their working memory inefficiently, thereby leading to quicker cognitive resource depletion compared to reflective decision-makers in analytical tasks. Because impulsive decision-makers lack available working memory their experiential system takes over the process and is more prone to committing judgment errors, as reported broadly in the literature (e.g., Barrafreem and Hausfeld 2019; Branas-Garza et al. 2019; Cueva et al. 2016; Alós-Ferrer and Ritschel 2018; Alós-Ferrer and Hügelschäfer 2016). The reason for this maladaptiveness is the motivation differences between the experiential and rational systems (see Sect. 2.1.1). Therefore, the inventories CRT and REI measure different processes, as formulated in hypothesis one:

**H1:** If the results of the CRT refer to the theoretical constructs in the CET—that is, the information processing systems of the dual-process theory—both the CRT and REI inventories measure the same decision-maker characteristics.

Epstein (2014) reports that the rationality and experientiality scores of the REI are reliable and highly stable measures. The original argument of Frederick (2005) is that the CRT is a measure of resisting the immediate answers that come to mind. Combining both arguments in *the cognitive ability and working memory framework*

following argument can be made: Assuming that individuals aim toward making reflective decisions in analytical tasks, their awareness about the efficiency of their current working memory can improve their future decision-making behavior. Because of its stable characteristic, a change in cognitive ability is more difficult than efficient utilization of working memory. In other words, the efficient usage of the working memory brings forth the full potential of cognitive ability. This argument leads to the following hypothesis:

**H2:** If the CRT and REI measure working memory efficiency and cognitive ability, respectively, rationality and experientiality scores:

- (a) cannot classify reflective/impulsive decision-maker characteristics but
- (b) can predict CRT scores.

Hence, an individual with a high cognitive ability (i.e., high rationality and experientiality scores) does not necessarily have to have an efficient working memory (i.e., high CRT scores).

## 4 Data

This section presents details on the data and exploratory analysis. All the calculations in this section and Sect. 5 are conducted using R (version 3.6.2) (R Core Team 2019) and Python (version 3.8) (Van Rossum and Drake 2009) languages.

The data for the analysis was collected during five different computerized lab experiments and contains 651 unique participants. Of these participants, 384 are female (age: mean = 24.26, std = 3.34) and 261 are male (age: mean = 24.27, std = 3.45). The same data collection procedures are employed for data collection for the CRT and REI. Participants are all recruited with in-class invitations during the courses in the Faculty of Business, Economics, and Statistics of University of Vienna. During all experiments, participants were not allowed to use any external aids, such as calculators and other web sites. In addition, because of the laboratory setting, participants were not allowed to communicate with each other or observe the screens of other participants.

Table 1 provides an overview of the CRT score of the participants. For each correct answer that they have provided, they receive one point. Thus, the CRT score varies between zero (all items are answered incorrectly) and three (all questions are answered correctly).

In Table 2, classification according to Cueva et al. (2016) (i.e., categorization of the individuals as reflective or impulsive decision-makers according to their CRT scores) is used to summarize the data.

Toplak et al. (2014) raise concerns regarding the validity of the original CRT questions because the inventory is published frequently. Even though Branas-Garza et al. (2019) report that multiple exposures to the CRT do not compromise its validity, original questions from Frederick (2005) are used with different numbers in the laboratory experiments as a preventive measure against memorized responses.

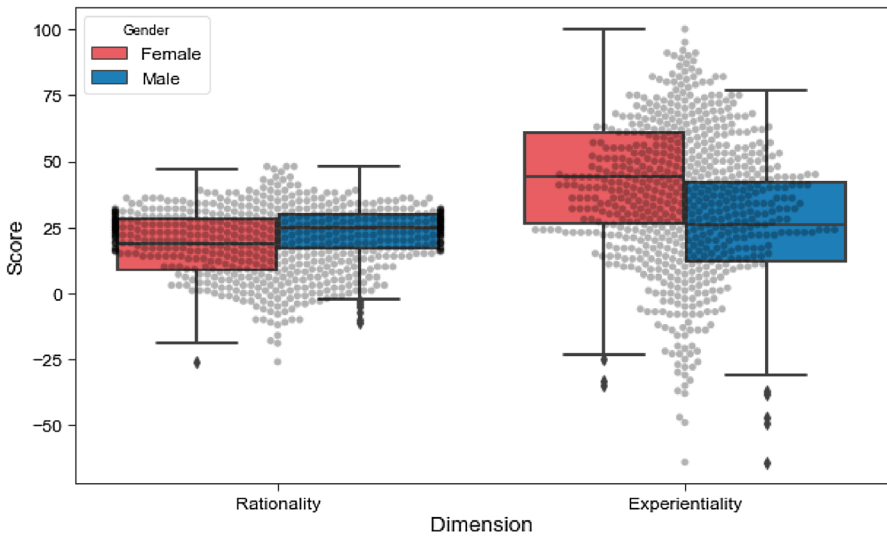
Figure 2 presents the dispersion of the rationality and experientiality scores among genders. On average, females score slightly lower than males on the rationality score but significantly higher on the experientiality score. Given that the data set includes individuals who have self-selected themselves for the business and economics studies,

**Table 1** Frequencies for the CRT scores

CRT score	Gender	Count	Percent	Cumulative count	Cumulative percent
0	Female	197	0.30	197	0.30
1	Female	84	0.13	281	0.43
2	Male	80	0.12	361	0.55
2	Female	68	0.10	429	0.65
1	Male	68	0.11	497	0.76
0	Male	62	0.09	559	0.85
3	Male	57	0.09	616	0.94
3	Female	35	0.06	651	1

**Table 2** Frequencies for the decision-maker type

Decision-maker type	Gender	Count	Percent	Cumulative count	Cumulative percent
Impulsive	Female	254	0.39	254	0.39
Reflective	Male	138	0.21	392	0.60
Reflective	Female	130	0.20	522	0.80
Impulsive	Male	129	0.20	651	1



**Fig. 2** Distribution of the REI scores according to gender

comparable rationality scores are expected. Further, gender bias in the CRT scores is a frequently reported topic in the literature. Bosch-Domènech et al. (2014) and Shaywitz et al. (1995) argue that the physiological differences of the brain according to gender can explain the bias in the results toward males (see, e.g., Holt et al. 2017; Cueva et al.

2016; Hoppe and Kusterer 2011). The gender bias in the experientiality score is in line with the physiological differences reported in the literature.

The next section presents the analyses and the associated results related to the proposed hypotheses in Sect. 3.2.

## 5 Results and discussion

In order to investigate if the CRT and REI measure the same theoretical cognitive construct (*H1*), first, a correlation table is reported. The upper portion of the Table 3 presents the correlations between CRT scores and the main dimensions of the REI inventory. The lower part of the Table 3 presents the correlations between the CRT scores and the sub-dimensions of the experientiality scores. Both correlation tables are tested using two-sided Spearman correlations with step-down methods (Bonferroni adjustments are utilized).

In Table 3, it is evident that significant but low correlation exists between CRT scores and REI dimensions, which rejects *H1* and supports the suggested framework described in Sect. 3.1. The effect sizes reported in Cohen's *d* and for more intuitive interpretation common language effect size (CLES), for all the reported correlations are high (Fritz et al. 2012; McGraw and Wong 1992).

Before continuing to the analysis for *H2*, let us explicitly distinguish between *classifiers* and *predictors* for technical clarification. Classifiers are the F-measure in the statistical binary classification analysis (Powers 2011) or the correct classification ratio of the dependent variable by a regression model. Predictors include predictor variables in the regression analysis (McCullagh 1980).

Examining the correlation coefficients alone for *H2(a)* might be insufficient. As is known, correlation does not detect any non-linear relationship among entities and it is only defined for the numerical variables. Hence, it will not serve as an adequate method for CRT classification (i.e., reflective/impulsive). For these reasons, calculating the predictive power score between CRT classification and the relevant variables is a better-suited method. The predictive power score is an asymmetric, data-type-agnostic score for predictive relationships between two columns, ranging from 0 to 1. A score of 0 implies that the feature column *x* cannot predict the target column *y* better than a naive baseline model. A score of 1 implies that the feature column *x* can perfectly predict the target column *y* given the model. In the calculations, the decision tree is

**Table 3** Correlations between REI and CRT—main dimensions

X	Y	Correlation coeff.	CI 95%	<i>p</i> value	Cohen's <i>d</i>	CLES
Rationality	CRT	0.310	[0.24, 0.38]	0.000	2.198	0.919
Experientiality	CRT	-0.180	[-0.25, -0.1]	0.000	1.838	0.903
Intuition	CRT	-0.180	[-0.25, -0.1]	0.000	1.079	0.796
Emotionality	CRT	-0.160	[-0.23, -0.08]	0.000	0.92	0.747
Imagination	CRT	-0.085	[-0.16, -0.01]	0.090	2.048	0.914

**Table 4** Predictive power scores for the decision-maker type as the target

Feature	PP score	Baseline score	Model score
Rationality	0.084	0.435	0.483
Experientiality	0.161	0.436	0.527
Intuition	0.193	0.436	0.545
Emotionality	0.080	0.436	0.481
Imagination	0.102	0.436	0.493

used as the learning algorithm. Using this method will also show whether there are any asymmetrical patterns in the data. For the predictive power score calculations, *ppscore* (Wetschoreck et al. 2020) package is used, which is a predictive power score implementation for Python language.

Table 4 summarizes the predictive power scores, where the REI dimensions function as features, and the decision-maker type acts as the target. For the sake of completeness, the sub-dimensions of experientiality are also presented. Different dimensions of the REI refer to the metric scores obtained by the participants and the decision-maker type refers to the binary classifier (reflective/ impulsive) calculated from the participants' CRT scores. The REI scores are used as features for the decision-maker type classification to identify if  $H2(a)$  holds. Note that the predictive power score is calculated separately for each feature. Because of the binary nature of the variable decision-maker type, the predictive power score computes the weighted F1 score as the underlying evaluation model, which can be interpreted as a weighted average of the precision and recall, where a score of one denotes the best value and zero denotes the worst value. As a baseline score, a weighted F1 score for a naive model is calculated, which always predicts the most common class for the decision-maker type. The predictive power scores presented in Table 4 are the result of the following normalization in equation 1:

$$PPS = (F1_{model} - F1_{naive}) / (1 - F1_{naive}) \quad (1)$$

Table 4 indicates that the REI scores fail to classify the decision-maker type correctly, thereby providing a strong indication that  $H2(a)$  does not hold. It can be argued that selecting the target variable as the decision-maker type rather than the CRT score is crude. For this reason, Table 5 depicts predictive power scores calculated for the CRT score as the target. The features and the underlying model in this calculation are the same as those, in Table 4.

Table 5 indicates that selecting the CRT score as an alternative target variable only improves the predictive power score of the rationality feature, whereas the predictive power score of the experientiality and, consequently, its sub-dimensions decreases. Further, the predictive power of the rationality dimension in the REI inventory can be considered a consequence of the mathematical emphasis of CRT questions. According to the predictive power score analyses, there is evidence that  $H2(a)$  does not hold.

For analyzing if REI scores can predict the CRT score, ordinal logistic regressions are reported. *MASS* (Venables and Ripley 2002) package of R is used for the regression calculations. In the regressions, gender is considered as a control variable that considers

**Table 5** Predictive power scores for the CRT score as target

Feature	PP Score	Baseline score	Model score
Rationality	0.167	0.226	0.355
Experientiality	0.114	0.226	0.315
Intuition	0.161	0.226	0.351
Emotionality	0.110	0.226	0.312
Imagination	0.080	0.226	0.289

the bias in the data set presented in Sect. 4. Table 6 presents the summary statistics for Model 1 (which only takes rationality and experientiality scores as well as gender as predictors), and Model 2 (with the rationality, sub-dimensions of the experientiality score, and gender variables). Confidence intervals at the 0.95 level are depicted in parentheses. Both regressions are run with 60% of the data set.

Model 1 clearly indicates that the rationality score and gender variables are highly significant. Holding everything else constant, an increase in the rationality score by one increases the expected value of the CRT score in log odds by 0.036, which translates to a probability of 0.51. Similarly, the male participants have an increased expected value of CRT scores in log odds by 0.779 (translated to probability 0.68) compared to females. Model 2 presents the same analysis with the experientiality score sub-

**Table 6** Ordinal logistic regression

	<i>Dependent variable</i>	
	CRT Score	
	Model 1	Model 2
Rationality Score	0.036***	0.036***
Experientiality Score	-0.006 (-0.013,0.001)	
Intuition Score		-0.024** (-0.045,-0.004)
Emotionality Score		0.007 (-0.012,0.026)
Imagination Score		-0.001 (-0.020,0.018)
Gender (Male)	0.779*** (0.380,1.178)	0.859*** (0.449,1.270)
Intercepts:		
0—1	0.211	0.257
1—2	1.281***	1.336***
2—3	2.874***	2.940***
AIC	997.409	997.635
Observations	390	390

Note: <sup>+</sup> $p < 0.1$ ; \* $p < 0.05$ ; \*\* $p < 0.01$ , \*\*\* $p < 0.001$

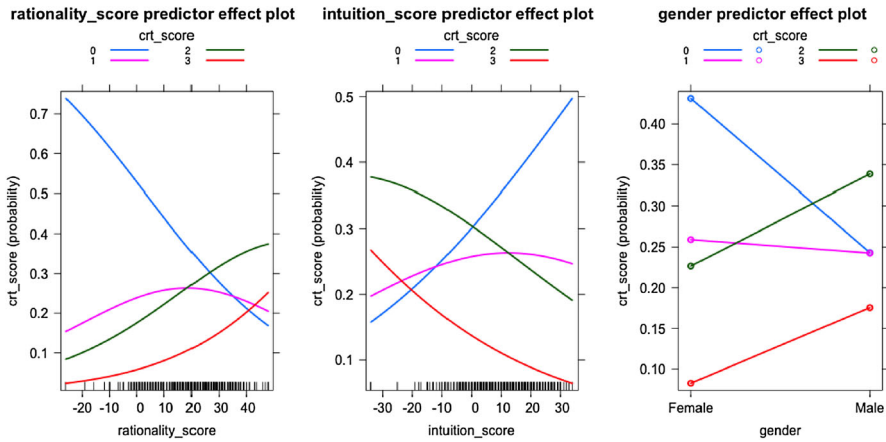


Fig. 3 Effect plots

dimensions, thereby providing the additional information that the intuition score is also significant with the coefficient -0.024 (translated to a probability of 0.49).

Plots in Fig. 3 visualize the above-described effects for the variables of interest—namely, rationality score, intuition score, and gender. A low rationality score is associated with a high probability of scoring low in the CRT. The opposite is observed for the intuition score. As seen in the literature (for the references, see Sect. 4), there is a higher probability for males to obtain high CRT scores, whereas the opposite is true for females.

The takeaway from the regression analysis is that rationality and intuition scores are predictors for the CRT score, thereby supporting *H2(b)*. In addition, gender is also a predictor for the CRT score.

After building and interpreting the ordinal logistic regression models, the next step is model evaluation. For the evaluation, Model 2 is used for the remaining part of the data set (40%) that has not been used in the model-building phase. For the model evaluation, the confusion matrix and the misclassification error are calculated and presented in Table 7.

The confusion matrix presents the performance of Model 2, contrasting the actual CRT scores with the predicted CRT scores by Model 2. The correctly identified cases by the regression model are given on the diagonal of the confusion matrix (in bold).

Table 7 Confusion matrix for the regression model 2

CRT Score	Predicted CRT Score			
	0	1	2	3
0	<b>94</b>	0	23	0
1	38	<b>0</b>	21	2
2	24	0	<b>18</b>	1
3	7	0	30	<b>3</b>

Misclassification error: 0.56



It is evident that the model poorly identifies CRT scores larger than 0. It is possible to calculate the misclassification error using the confusion matrix, which is 0.56 for regression Model 2. From the coefficients of the presented regression and model evaluation, the following conclusions can be obtained:

- predictive power scores for the rationality and experientiality (and the sub-dimensions of experientiality) scores and the regression model evaluation provide evidence for rejecting  $H2(a)$ ,
- the rationality score and the intuition score from the REI inventory are significant predictors with the expected direction for the CRT score, thereby supporting  $H2(b)$ ,
- gender is a significant predictor for the CRT score.

Overall, the results show that rationality and intuition scores are predictors (see  $H2(b)$ ) but not classifiers (see  $H2(a)$ ) for the reflectivity/impulsivity of decision-makers. These results reveal that the CRT and REI are measures for different theoretical constructs in the dual-process framework. Further, observing that high scores from the REI are not necessarily accompanied by high CRT scores indicates evidence for the proposed framework in Sect. 3.1.

Additional results regarding gender bias are in line with the literature, but, with an interpretation difference. Table 2 indicates a lack of gender bias, whereas Fig. 2 plots the gender differences in the cognitive ability, particularly in the experientiality score. Figure 4 presents the gender bias in the sub-dimensions of the experientiality scores in greater detail. In light of these results, one can argue that the gender bias present in the studies that employ the CRT originates from individuals' cognitive ability but not necessarily from inefficiencies related to working memory.

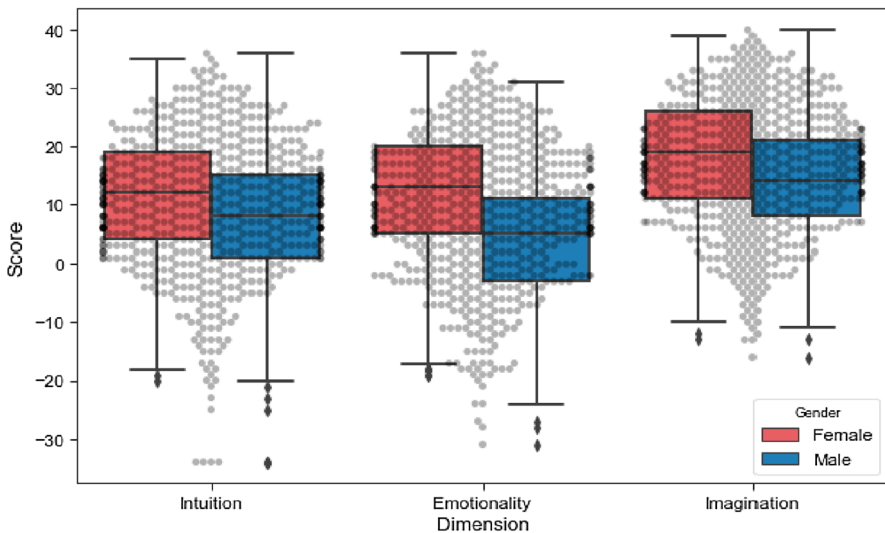


Fig. 4 Distribution of the sub-dimensions of experientiality according to gender

## 6 Conclusion

This paper contributes to the open discussion on the theoretical construct of the CRT by proposing a novel framework. The framework enables a disentangling of the relationship between the cognitive ability (rationality and experientiality) and the efficient/inefficient use of working memory (reflectivity/impulsivity). The findings suggest that the efficient use of working memory can enable a decision-maker to use his/her full cognitive potential in analytical tasks. Establishing the theoretical relationship between CRT results and the CET provides a valid contribution for measuring individual differences in decision-maker characteristics through the *framework of cognitive ability and working memory*. While this new interpretation does not contradict the more general interpretation of the test (i.e., as an overall measure of a subject's ability to resist the first response that comes to mind), it disentangles which decision process is actually at work. Therefore, this research indicates the importance of capturing both decision-maker characteristics (CRT and REI results) as relevant measures to understand the drivers underlying human decisions.

As indicated in the literature (see Sect. 2.1.2), the dominance of one system over the other can only be argued within a specific task context. Therefore, it must be emphasized that the framework is constructed with analytical decision-making tasks in mind. For non-analytical decision tasks, the framework may not be applicable in its current form. This aspect could be addressed in future research.

Another aspect that requires attention is that cognitive resources get depreciated as a result of inefficient usage and also by external stressors like time pressure. While the effects of this study are tested without external stressors, it remains an exciting topic for future research to investigate the effects of various taxing strategies on participants' cognitive resources. Further, testing the framework among participants from different age groups can be another interesting research direction. The experiential system benefits significantly from associative learning. If there is a systematic dominance of one system among the other one due to experience, it is worth exploring in future research.

**Funding** Open access funding provided by University of Vienna.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Alós-Ferrer C, Hügelschäfer S (2016) Faith in intuition and cognitive reflection. *J Behav Exp Econ* 64:61–70
- Alós-Ferrer C, Ritschel A (2018) The reinforcement heuristic in normal form games. *J Econ Behav Organ* 152:224–234

- Baron J, Scott S, Fincher K, Metz SE (2015) Why does the cognitive reflection test (sometimes) predict utilitarian moral judgment (and other things)? *J Appl Res Memory Cogn* 4(3):265–284
- Barraffrem K, Hausfeld J (2019) Tracing risky decisions for oneself and others: the role of intuition and deliberation. *J Econ Psychol* 77:102188
- Bialek M, Pennycook G (2018) The cognitive reflection test is robust to multiple exposures. *Behav Res methods* 50(5):1953–1959
- Blacksmith N, Yang Y, Behrend TS, Ruark GA (2019) Assessing the validity of inferences from scores on the cognitive reflection test. *J Behav Dec Mak* 32(5):599–612
- Bosch-Domènech A, Branas-Garza P, Espín AM (2014) Can exposure to prenatal sex hormones (2d: 4d) predict cognitive reflection? *Psychoneuroendocrinology* 43:1–10
- Branas-Garza P, Kujal P, Lenkei B (2019) Cognitive reflection test: Whom, how, when. *J Behav Exp Econ* 82:101455
- Cowan N (2017) The many faces of working memory and short-term storage. *Psychon Bull Rev* 24(4):1158–1170
- Cueva C, Iturbe-Ormaetxe I, Mata-Pérez E, Ponti G, Sartarelli M, Yu H, Zhukova V (2016) Cognitive (ir) reflection: new experimental evidence. *J Behav Exp Econ* 64:81–93
- Dijksterhuis A (2004) Think different: the merits of unconscious thought in preference development and decision making. *J Personal Social Psychol* 87(5):586
- Engin A, Vetschera R (2017) Information representation in decision making: the impact of cognitive style and depletion effects. *Decis Support Syst* 103:94–103
- Engin A, Vetschera R (2019) Optimistic overconfidence in electronic reverse auctions. *Electron Comm Res Appl* 35:100842
- Engle RW, Kane MJ, Tuholski SW (1999) Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex
- Epstein S (1973) The self-concept revisited: or a theory of a theory. *Am Psychol* 28(5):404
- Epstein S (1994) Integration of the cognitive and the psychodynamic unconscious. *Am Psychol* 49(8):709
- Epstein S (2014) *Cognitive-experiential theory: an integrative theory of personality*. Oxford University Press, Oxford
- Epstein S, Lipson A, Holstein C, Huh E (1992) Irrational reactions to negative outcomes: evidence for two conceptual systems. *J Personal Soc Psychol* 62(2):328
- Epstein S, Pacini R, Denes-Raj V, Heier H (1996) Individual differences in intuitive-experiential and analytical-rational thinking styles. *J Personal Soc Psychol* 71(2):390–405
- Erceg N, Bubić A (2017) One test, five scoring procedures: different ways of approaching the cognitive reflection test. *J Cogn Psychol* 29(3):381–392
- Evans JSB (2006) The heuristic-analytic theory of reasoning: extension and evaluation. *Psychon Bull Rev* 13(3):378–395
- Evans JSB (2008) Dual-processing accounts of reasoning, judgment, and social cognition. *Annu Rev Psychol* 59:255–278
- Evans JSB, Stanovich KE (2013) Dual-process theories of higher cognition: advancing the debate. *Perspect Psychol Sci* 8(3):223–241
- Frederick S (2005) Cognitive reflection and decision making. *J Econ Perspect* 19(4):25–42
- Fritz CO, Morris PE, Richler JJ (2012) Effect size estimates: current use, calculations, and interpretation. *J Exp Psychol Gen* 141(1):2
- Gigerenzer G, Gaissmaier W (2011) Heuristic decision making. *Annu Rev Psychol* 62:451–482
- Gilbert DT, Malone PS (1995) The correspondence bias. *Psychol Bull* 117(1):21
- Gino F, Ariely D (2012) The dark side of creativity: original thinkers can be more dishonest. *J Personal Soc Psychol* 102(3):445
- Hämäläinen RP, Luoma J, Saarinen E (2013) On the importance of behavioral operational research: the case of understanding and communicating about dynamic systems. *Eur J Oper Res* 228(3):623–634
- Holt CA, Porzio M, Song MY (2017) Price bubbles, gender, and expectations in experimental asset markets. *Eur Econ Rev* 100:72–94
- Hoppe EI, Kusterer DJ (2011) Behavioral biases and cognitive reflection. *Econ Lett* 110(2):97–100
- Kahneman D, Frederick S (2002) Representativeness revisited: attribute substitution in intuitive judgment. *Heuristics Biases Psychol Intuit Judgm* 49:81
- Kane MJ, Bleckley MK, Conway AR, Engle RW (2001) A controlled-attention view of working-memory capacity. *J Exp Psychol General* 130(2):169

- Marewski JN, Gigerenzer G (2012) Heuristic decision making in medicine. *Dialogues Clin Neurosci* 14(1):77
- McCullagh P (1980) Regression models for ordinal data. *J R Stat Soc Series B (Methodological)* 42(2):109–127
- McGraw KO, Wong SP (1992) A common language effect size statistic. *Psychol Bull* 111(2):361
- Padilla LM, Creem-Regehr SH, Hegarty M, Stefanucci JK (2018) Decision making with visualizations: a cognitive framework across disciplines. *Cogn Res Princ Implic* 3(1):29
- Pennycook G, Cheyne JA, Seli P, Koehler DJ, Fugelsang JA (2012) Analytic cognitive style predicts religious and paranormal belief. *Cognition* 123(3):335–346
- Pennycook G, Cheyne JA, Koehler DJ, Fugelsang JA (2016) Is the cognitive reflection test a measure of both reflection and intuition? *Behav Res Methods* 48(1):341–348
- Powers DM (2020) Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv preprint [arXiv:2010.16061](https://arxiv.org/abs/2010.16061)
- Primi C, Morsanyi K, Chiesi F, Donati MA, Hamilton J (2016) The development and testing of a new version of the cognitive reflection test applying item response theory (IRT). *J Behav Dec Mak* 29(5):453–469
- R Core Team (2019) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Shaywitz BA, Shaywitz SE, Pugh KR, Constable RT, Skudlarski P, Fulbright RK, Bronen RA, Fletcher JM, Shankweiler DP, Katz L et al (1995) Sex differences in the functional organization of the brain for language. *Nature* 373(6515):607–609
- Shipstead Z, Harrison TL, Engle RW (2015) Working memory capacity and the scope and control of attention. *Atten Percep Psychophys* 77(6):1863–1880
- Sloman SA (1996) The empirical case for two systems of reasoning. *Psychol Bull* 119(1):3
- Sloman SA (2002) Two systems of reasoning. In: Gilovich T, Griffin D, Kahneman D (eds) *Heuristics and biases: the psychology of intuitive judgment*. Cambridge University Press, pp 379–396. [https://scholar.google.com/scholar?hl=en&as\\_sdt=0%2C5&q=Sloman+SA+%282002%29+Two+systems+of+reasoning&btnG=](https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Sloman+SA+%282002%29+Two+systems+of+reasoning&btnG=)
- Stanovich KE, West RF (2000) Individual differences in reasoning: implications for the rationality debate? *Behav Brain Sci* 23(5):645–665
- Thomson KS, Oppenheimer DM (2016) Investigating an alternate form of the cognitive reflection test. *Judgm Decis Mak* 11(1):99
- Toplak ME, West RF, Stanovich KE (2014) Assessing miserly information processing: an expansion of the cognitive reflection test. *Think Reason* 20(2):147–168
- Van Rossum G, Drake FL (2009) Python 3 reference manual. CreateSpace, Scotts Valley, CA
- Venables WN, Ripley BD (2002) *Modern applied statistics with S*, 4th edn. Springer, New York. <http://www.stats.ox.ac.uk/pub/MASS4>, ISBN 0-387-95457-0
- Wetschoreck F, Krabel T, Krishnamurthy S (2020) 8080labs/ppscore: zenodo release. <https://doi.org/10.5281/zenodo.4091345>
- Wilson TD (2004) *Strangers to ourselves*. Harvard University Press, Harvard
- Wilson TD, Lisle DJ, Schooler JW, Hodges SD, Klaaren KJ, LaFleur SJ (1993) Introspecting about reasons can reduce post-choice satisfaction. *Personal Soc Psychol Bull* 19(3):331–339
- Zhu B, Watts SA (2010) Visualization of network concepts: the impact of working memory capacity differences. *Inf Syst Res* 21(2):327–344