5-2022

Do You Salt Your Soup: Investigating the Effect of Interference Control on the Cognitive Reflection Test

Matthew Lowrie
William & Mary

Do You Salt Your Soup:

Investigating the Effect of Interference Control on the Cognitive Reflection Test

A thesis submitted in partial fulfillment of the requirement
for the degree of Bachelor of Science in Psychological Sciences from
William & Mary

by

Matthew H. Lowrie

Accepted for Honors
(Honors)

Dr. Christopher Ball, Advisor

Dr. Harvey Langholtz

Dr. John Parman

Williamsburg, VA
May 2, 2022

Abstract:

The present study seeks to examine the role of interference control in solving the Cognitive Reflection Test (Frederick, 2005). Participants were given the CRT, CRT-2, and completed a novel adaptation of the Stop-Signal Task called the Change-Signal Task. The Change-Signal Task is similar to a stop-signal paradigm except that the participant must switch their response when a change-signal is present in the Change-Signal Task instead of withholding a response. This study found that interference control as assessed by the Change-Signal Task was important for determining performance on the CRT-2 but not for the CRT. Implications of these findings and interpretations are discussed before suggestions for further research.

Keywords: Cognitive reflection test, dual-process theory, response inhibition, interference control, Stop-Signal Task

Do You Salt Your Soup: Investigating the Effect of Interference Control on the Cognitive Reflection Test

When interviewing potential research assistants, the great American scientist and inventor Thomas Edison would order them soup for lunch. With salt and pepper on the table, Edison would observe at what point candidates seasoned their dish. Edison immediately disqualified anyone who seasoned their meal before tasting it on the basis that the candidate made too many unproven assumptions. Using this simple exercise Edison sought to evaluate the candidates' thought processes: were they instinctual or reasoned? Did the candidates assume or did they investigate? Although Edison did not have the same vocabulary to describe these cognitive processes as we do today, his soup test was an early attempt to determine whether an individual was predisposed to an intuitive, possibly faulty, thinking style or a rational, more analytical thinking style. The current study examines the contemporary approach to this assessment and analyzes the underlying cognitive processes determining the distinction.

Dual Process Theory

The set of theories consider to be dual-process theory have their origins in the 1970s and 1980s, relying heavily on the heuristics and biases movement spearheaded by Kahneman and Tversky (Evans, 1989; Tversky & Kahneman, 1974; Wason & Evans, 1975). The movement and subsequent dual-process theories have affected many disciplines, each giving it its own spin (Evans, 2008; Stanovich, 2004). Broadly, dual-process theory encompasses the idea that there are two pathways of higher-order processing in the brain, one that is fast and considered effortless, while the other is slower and computationally exhaustive (Kahneman, 2011). Stanovich and West (2000) describe the two processes using the terms System 1 and System 2. System 1 is fast and uses fewer cognitive resources, arriving at an answer without much or any conscious awareness.

System 1 is guided by heuristics and biases, which are fallible but useful mental shortcuts; it is our immediate, instinctual reaction to a stimulus. System 2 is a slower approach that uses more cognitive resources. It takes effort to activate and use but is more often correct than System 1. Evans (2008) suggested the terms be updated to Type 1 and Type 2 processing to highlight that these processes are not the activation of a single cognitive system, but a collection of cognitive processes. Evans and Stanovich (2013) define Type 1 processing by its autonomy. Type 1 processing happens rapidly and automatically at the presentation of the stimulus, without conscious attention or higher-order control systems. Type 2 processing is the opposite of those characteristics that make up Type 1. Type 2 is slow, computationally exhausting, and can require effort to engage. For example, if you salt your soup before tasting it you are said to have engaged in and be predisposed to Type 1 thinking whereas if you had tasted the soup first you would have engaged in Type 2 thinking.

Some researchers (e.g. Keren and Schuul 2009; Newstead, 2000) have suggested that the Type 1 and Type 2 distinctions are misleading, and that there is instead a continuum of processing styles for analytical thinking. Others (e.g. Kruglanksi and Gigerenzer, 2011; Osman, 2004) have gone even further and suggested that there is only a single pathway for analytical thinking. One possible explanation for the continuum and single pathway theories of analytical thinking is that individual differences arise from the differential strength of a component process; a difference in performance of the underlying cognitive mechanisms. Evans and Stanovich (2013) respond to these critiques by proposing a common understanding of dual-process theory and claiming the neuroscientific evidence supporting the dual-process distinction outweighs any contrary evidence. However, Evans and Stanovich (2013) do not convincingly rebuke these critiques and instead rely on the very framework in question. An alternate explanation for dual-process theory, that relies on individual differences in processing, can benefit from our current understanding of the relative strength of underlying cognitive mechanisms, such as response inhibition and working memory. The current study will attempt to demonstrate the impact of one such cognitive mechanism, interference control.

The Cognitive Reflection Test

The modern equivalent to Edison's soup test is the Cognitive Reflection Test, or CRT, a three-item free-response questionnaire that measures individual differences in modes of processing (Frederick, 2005). The three questions on the CRT (Figure 1) each reliably elicit an intuitive, incorrect answer which participants must overcome to determine the correct response. If a participant can overcome that initial response to answer correctly, then they are said to engage in cognitive reflection. Frederick defined cognitive reflection as "the ability or disposition to resist reporting the response that first comes to mind" (Frederick, 2005, p. 35). In terms of dual-process theory, someone who answers the initial, intuitive answer is thought to have engaged in Type 1 thinking. Meanwhile someone who answers correctly is thought to engage in Type 2 thinking that results from cognitive reflection. Frederick (2005) showed that performance on the CRT is related to other, more common tests of analytical thinking, such as the ACT, SAT, and Wonderlic Personnel Test. Although Frederick introduced the CRT as a measure of cognitive reflection, researchers have repeatedly used the CRT as a measure of much more.

(1) A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? *[Type 1/Intuitive = 10 cents; Type 2/Correct = 5 cents]*
(2) If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?*[Type 1/Intuitive = 100 minutes; Type 2/Correct = 5 minutes]*
(3) In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? *[Type 2/Correct = 47 days; Type 1/Intuitive = 24 days]*
(4)* If you're running a race and you pass the person in second place, what place are you in? [*Type 1/Intuitive = first; Type 2/Correct = second]*
(5)* A farmer had 15 sheep and all but 8 died. How many are left? *[Type 1/Intuitive answer = 7; Type 2/Correct answer = 8]*
(6)* Emily's father has three daughters. The first two are named April and May. What is the third daughter's name? *[Type1/Intuitive answer = June; Type 2/Correct answer = Emily]*
(7)* How many cubic feet of dirt are there in a hole that is 3'deep x 3' wide x 3' long? *[Type 1/Intuitive answer = 27; Type 2/Correct answer = 0]*

Figure 1. Questions from the Cognitive Reflection Test (Frederick, 2005); * indicates questions from the CRT-2 (Thomson and Oppenheimer, 2016).

The proliferation of CRT research since its introduction has been astounding. A search on Google Scholar revealed 5,342 citations for the article introducing the CRT (Frederick, 2005).[1] The reason for this proliferation is understandable; the CRT is short, has high face validity, and

[1] Retrieved April 22, 2022. This number likely includes duplicate articles.

correlates with a large and increasing number of other measures. For example, the CRT has frequently been used as a measure of cognitive abilities and a predictor of decision-making ability (Andersson et al. 2013; Besedes et al. 2012; Campitelli and Labollita 2010; Hoppe and Kusterer 2011; Moritz et al. 2013; Neyse et al. 2016; Oechssler et al. 2009). Oechssler et al. (2009) found that participants with low scores on the CRT were more likely to engage in the conjunction fallacy and more conservatively update probability assessments. Meanwhile, Neyse et al. (2016) found that high scorers on the CRT displayed better financial literacy. The CRT has even been used to predict strategic behavior in laboratory tasks (Brañas-Garza et al. 2019; Carpenter et al. 2013; Corgnet et al. 2015; Kiss et al. 2016) and social preferences (Corgnet et al. 2015; Cueva-Herrero et al 2016; Peysakhovic and Rand 2017; Ponti and Rodriguez-Lara 2015). Researchers also use the CRT as a catchall for individual differences in thinking and cognition. For instance, scores on the CRT have been correlated with risky choice behavior (Cokely & Kelley, 2009; Frederick, 2005), the ability to detect fake news (Pehlivanoglu et al. 2021), conspiratorial beliefs (Mikušková, 2021), COVID-19 precautionary behavior (Thoma et al. 2021), and religiosity (Freiden & Martini, 2021), among others. Much of this work has little to do with the concept of cognitive reflection. Instead, the CRT is used as a predictor and covariate, often in strictly correlational research, or as a stand-in for other concepts with which the CRT has been previously associated (Szaszi et al. 2017). As its impact has grown, so too has the need to understand the cognitive mechanisms underlying performance on the CRT. Without a strong understanding of these cognitive mechanisms, it is unclear what can be discerned from the multitude of studies using the CRT.

The continuum of performance on the CRT provides evidence for the idea that there is not a strong distinction between Type 1 and Type 2 processes, but rather a continuum of processing ability as some critics of dual-process theory suggest (e.g. Keren & Schul, 2009). Frederick's original study only analyzes those participants who scored 0 or 3 on the CRT (Frederick, 2005). Frederick claimed that adding the analysis of the participants who score 1 or 2 provided no more conclusive evidence as to the power of the CRT. Participants in the middle fell somewhere in between the low and high scorers on the other measures used. That people are normally distributed across the CRT and its correlated measures again suggests that the CRT is flawed as a discrete sorting measure. If all the CRT problems rely on the same theoretical all-or-none function, then the number of correct answers should be bimodally distributed around 0 and 3. Instead, the plurality of participants fall in the low-middle, with the mean CRT score being 1.24 in the original study (Frederick 2005). The mean CRT score in the current study was 1.06. A meta-analysis of CRT research by Brañas-Garza et al. (2019) revealed a similar pattern. They found that a third of the population sampled completely lacked cognitive reflection, as measured by the CRT (a score of zero), while only 18% scored a perfect three. The remaining participants scored either one or two. According to the CRT this would indicate that only 18% of people are predisposed to Type 2 thinking. Data from the current study show a similar pattern (refer to Table 1). Only 13% of the participants in the present study scored perfectly on the CRT. This indicates that, according to the CRT, they cognitively reflected and engaged only in Type 2 thinking. A larger group of the participants (43%) scored zero and according to the CRT they were only engaged in Type 1 thinking. But an equally large group of participants (44%) scored either one or two on the CRT and this would suggest a mix of Type 1 and Type 2 thinking was involved. Consequently, a large number of participant responses and individual differences are being ignored if a researcher only examines participants who score 0 or 3. There is clearly no evidence from the current survey that relying on one mode of thinking does not rule out the possibility of utilizing the alternative style of thinking for different problems and contexts. For these reasons, the current study will analyze cognitive reflection test scores across the full range of possible scores.

Examining the construction of CRT questions can help us better understand how and why they work. For example, the bat and ball problem provides an intuitively appealing answer of 10 cents while the correct answer is actually 5 cents. If the ball costs 10 cents, then the bat would cost 10 cents plus one dollar, or $1.10 on its own. Together they would sum to $1.20. If the ball is five cents then the bat is $1.05 and together they sum to the requisite $1.10. The intuitive answer lures respondents by being readily available and making sense on first glance. Given how readily and strongly the intuitive response comes to mind, it is considered a prepotent response. These are responses that come to mind automatically and rely on Type 1 processing. Elicitation of prepotent responses to the CRT may not be solely a function of the question phrasing, but also a function of the numerical values used. Previous research suggests that more challenging numerical questions utilizing the same question framework fail to elicit a prepotent response with participants more likely to now give Type 2 answers. For example, Frederick (2005) points out that participants find the bat and ball problem much harder than an analogous problem which invites more effort and computation. The so-called "banana and bagel problem" follows the same structure as the bat and ball problem but differs in its ease of computation; there are no round numbers that can help produce a prepotent solution. The problem is as follows: "A banana and a bagel cost 37 cents. The banana costs 13 cents more than the bagel. How much does the bagel cost?"[2] That a problem with the same set up as the CRT problems does not elicit a Type 1 versus Type 2 conflict again raises questions about the mechanisms underlying the CRT. If the CRT is supposed to reliably sort people into two categories of thinking style, then adjusting the difficulty

[2] The correct answer is 12 cents. From Frederick (2005, p. 28)

of the question should not affect a person's inherent disposition of thinking. Instead this suggests that there is some mechanism that allows more participants to switch to Type 2 processing. Tweaking a CRT-style problem to make it more difficult separates participants into two groups: there are those who can solve the problem using Type 2 processes and those who answer incorrectly by guessing randomly or not answering at all. Making a problem more cognitively demanding therefore sorts participants better than relying on an intuitive response alone. The current study deploys a cognitive demand manipulation by having participants update their working memory as the task progresses. This increasing load is presumed to help better discriminate between participant groups.

Several researchers have attempted to duplicate the effects of the CRT with their own versions that claim to both correlate well with the CRT but also overcome some of the shortcomings of the original CRT. For example, males scored significantly higher on the CRT than females, something which Frederick and subsequent researchers have been unable to definitively explain (Cueva-Herrero et al. 2016; Frederick 2005; Holt et al. 2017; Hoppe & Kusterer 2011; Ring et al. 2016). However, this gender difference likely is due in part to males consistently demonstrating higher mathematical ability than females on tests (Benbow et al. 2000; Mau & Lynn 2010). There has been a large debate over whether the CRT is capturing numeracy, a person's mathematical ability. Some studies have found a moderate correlation between the CRT and numeracy (Finucane & Gullion, 2010; Liberali et al., 2012) while Welsh, Burns, and Delfabbro (2013) go further and claim that the CRT is only predictive of heuristic tasks that involve numeracy. Campitelli and Gerrans (2014) assessed whether the CRT actually tested for cognitive reflection or if it was merely a function of mathematical ability. Using several measures to test numeracy, syllogistic reasoning, and actively open-minded thinking, the researchers concluded that predictive models which included both an inhibition parameter and a mathematical parameter performed better than those with a mathematical parameter alone. These findings are consistent with other work which found that the CRT measures more than mathematical ability (Böckenholt, 2012; Campitelli & Labollita, 2010; Frederick, 2005; Liberali et al., 2011; Toplak et al., 2011). Importantly, Campitelli and Geranns (2014) show that we can create accurate predictive models of responses to the CRT, but that these models must include some form of inhibitory mechanism. This is an important conclusion that guided the design of the current study.

The CRT-2 (Thomson & Oppenheimer, 2016) is a recent example which was developed to overcome shortcomings with the original CRT. The CRT-2 (refer to Figure 1) attempts to eliminate the impact of repeated exposure, numeracy, and gender on performance. Thomson and Oppenheimer (2016) also found that participants produced fewer non-Type 1 errors on the CRT-2 than the CRT. This means that when participants got the question wrong, they were more likely to respond with the Type 1, intuitive response than a non-cued response. Additionally, scores on the CRT-2 have been highly correlated with scores on the CRT, with the two measures also displaying similar internal consistency (Thomson & Oppenheimer, 2016). Therefore, the CRT-2 is also used in the present study to also examine individual differences in cognitive reflection.

Response Inhibition

Response inhibition is the ability to suppress a prepotent or inappropriate behavior (Chambers et al., 2009; Schachar et al., 2007; van Velzen et al., 2014). Campitelli and Gerrans (2014) used a mathematical modeling approach to understanding the role that different cognitive mechanisms have in solving the CRT. Their research showed that models were most accurate when including a parameter for inhibiting the intuitive response combined with a numeracy parameter as compared to a numeracy parameter alone. Campitelli and Gerrans (2014) therefore concluded that response inhibition of the intuitive answer is critical for answering the CRT correctly. Although inhibition is frequently portrayed as a single process, it is best understood and evidenced as three interrelated processes (Brydges et al. 2012, 2013). For example, action suppression and action cancellation are two distinct components of response inhibition (Schachar et al., 2007). Action suppression allows for the response to continue after either more information is gathered or a response is prompted whereas action cancellation is the end result of that process wherein an initiated response does not occur. The third and most important aspect of inhibition is interference control, which is the ability to prevent interference from the competition of relevant and irrelevant stimuli or stimulus characteristics (Nigg, 2000). Schachar et al. (2007) suggest that interference control, action suppression, and action cancellation refer to distinct early, intermediate, and late processes of inhibition. Indeed these three components of response inhibition have been shown to have specific, but overlapping, neural activation (van Velzen et al., 2014). To understand the difference in these three components of inhibition, consider Edison's soup test. Interference control would play the role of mediating between the prepotent response of salting your soup before tasting it and the knowledge that you haven't tried the soup, allowing one response (salting or tasting) to continue. In one case, if you are someone who always salts their food first, the initial Type 1 response would likely continue. In another, perhaps you realize that you haven't tasted the food or that Edison is watching you. Now, interference control is balancing two more equally weighted stimuli and the decision is more difficult. If you had realized this after reaching for the salt, action suppression would slow you down and perhaps lead to action cancellation.

The CRT questions could work in a similar way to the example above. Participants must cognitively balance the initial Type 1 prepotent-response processing with competing signals provided by Type 2 processes involving a different answer. Therefore, interference control could be the underlying mechanism which modulates and selects the process that continues. One of the most famous examples of interference control tasks is the Stroop Task (Stroop, 1935). The Stroop Task measures the Stroop Effect, which is the phenomenon whereby participants show difficulty in naming the ink color of a color word when it is printed in a contrasting color (e.g., the word “red” printed in blue ink) compared to naming the ink color of a neutral stimulus printed in the same color (e.g., “XXXX” printed in blue ink). The printed word is automatically processed faster and in competition with the perception and naming of the ink color. This competition makes it difficult to respond correctly and participants with better interference control are better at naming the ink color without delay or error. Although the Stroop Task is a well-documented interference control task, it does not easily allow for difficulty manipulations.

However, other tasks have been developed within the response inhibition framework which do allow for difficulty manipulations. One of the simplest is the Go/No-Go task (Donders, 1868/1969; Sternberg, 1969). In this task, participants are presented with two distinct stimuli. They are instructed to respond to only one of the stimuli, usually by pressing a button as fast as possible. For the other stimulus, participants are told to refrain from responding when it is presented. Although it is easy to manipulate ‘when’ and ‘how’ the no-go stimulus is presented, the Go/No-Go task measures action suppression, not interference control (van Velzen et al., 2014). Logan et al. (1984) derived the Stop-Signal task from the Go/No-Go task with the purpose of measuring different inhibitory mechanisms. On all trials of the Stop-Signal task the participant is presented with one response stimulus. On a limited number of trials, participants are cued to refrain from responding by a so-called stop signal, a secondary audio stimulus that occurs after the presentation of the response stimulus. The stop signal indicates the participant should not respond, despite the continued presentation of the response stimulus. To manipulate difficulty on the Stop-Signal Task, a stepwise presentation is used to present the stop signal. When a participant provides a correct non-response, the time between the onset of the go stimulus and the stop signal increases. Much as with the Go/No-Go task, the Stop-Signal Task is best at measuring response inhibition in the latter two stages, suppression and cancellation, with the end result on stop signal trials being no response. Notably, however, the Stop-Signal task involves inhibiting a response in action, which increases the inhibitory load, or how difficult it is to inhibit the response, relative to the Go/No-Go task (Schachar et al., 2007). The CRT, meanwhile, requires a response for every question and is therefore conceptually different from these tasks. The participant must balance between competing Type 1 and Type 2 processes and allow one to continue as the final overt choice rather than withholding all responses entirely as in the Stop-signal task.

Present Study: Change-signal task

The experimental task in the present study is a novel adaptation of the Stop-Signal Task (Logan et al., 1984) that I have developed and will call the Change-Signal Task. In the Change-Signal Task, the setup is similar to a Stop-Signal Task but differs in how participants are required to respond when the secondary stimulus is active. Participants are presented with a visual stimulus on the screen (left or right arrow) that requires a specific type of response (left or right button press). However, on one-quarter of the trials, a short auditory cue is provided that requires the participant to now respond in the opposite direction (opposite button press) indicated by the visual stimulus (arrow direction). Whereas in a Stop-Signal Task signaled trials result in no response, signaled trials in the Change-Signal Task also require a response from the participant. The Change-Signal Task requires the participant on change trials to first inhibit the prepotent response (more frequent no-change trials) and then change to a different response, much like is required by the CRT.

The Change-Signal Task has three types of trials which are determined by the onset of the change-signal (the auditory cue) that tells the participant to switch their response. One type of trial has no change-signal and participants respond to the direction of an arrow (e.g. left arrow and left button press). For the early cue change-signal trials, the auditory cue occurs 100ms after the presentation of the arrow stimulus. This cue is heard by the participant as they are still perceptually processing the arrow stimulus and well before a response to this stimulus is being prepared in the motor cortex. A change in this response requirement would require less response inhibition for this reason. The third type of change-signal trial involves a 300ms delay from the presentation of the arrow stimulus on the screen and the presentation of the auditory cue. A delay of 300ms would allow the participant enough processing time to perceive the arrow stimulus on the screen and could begin forming a motor response that corresponds to this arrow direction. The change-signal cue would require response inhibition and interference control mechanisms to be activated to produce the correct response. To make sure that the participant does not simply delay their responses to the arrow stimulus until the auditory cue is heard or not heard, the participant is given 1000ms to complete the trial. This trial duration is consistent with trial durations used in other similar tasks (e.g. Cao & Cannon, 2021; Kok et al., 2003; Li, 2008;) and was selected based on pilot research using a range of trial durations.

Present Study: Under increased cognitive load

During the second experimental block of the Change-signal trials, participants engaged in a working memory task at the same time by maintaining a mental count of the number of “o”s (instead of “+”s) they see as fixation points across each block of trials. Increasing cognitive load in this way draws cognitive resources away from participants who may have otherwise performed well during the first experimental block. Stupple et al. (2013) found evidence that working memory capacity is a strong predictor of performance on the CRT, even stronger than the two reasoning tasks provided to their participants. Stupple et al. (2013) reasoned that the impact of individual working memory capacity in solving the CRT is profound. They further speculate that participants with lower working memory capacity may expend effort in solving the CRT problems, but lack the ability to hold the information in their head. Individuals with high working memory capacity, on the other hand, may find a relatively lower cognitive cost to solving a CRT problem and therefore be more willing to put in the effort. Presuming that the switching to Type 2, correct responses is modulated by interference control, then participants who perform better on the CRT should display better interference control, especially under increased cognitive load. On the Change-Signal Task this would be indicated by better overall performance on the more difficult 300ms change-signal condition of the dual-task experimental block.

Method

Participants

Seventy-two participants were recruited from undergraduate introductory psychology courses at a liberal arts university in the United States (54 female, 18 male) and received course credit for their participation. The mean age for females was 18.74, *SD* = .95 and was 19.38 for males, *SD* = 1.46. Overall, the mean age was 18.90, *SD* = 1.13. The research study was approved by the university's IRB and all participants provided informed consent before participating.

Apparatus

The critical reasoning questionnaires and the Change-Signal Task were completed by participants using a desktop computer with a 20-inch screen. The critical reasoning questionnaires were completed using the Qualtrics online survey platform. The Change-Signal Task was created and presented using SuperLab 6 (Cedrus Corp.). Participants entered their Change-Signal Task responses through an RB-830 CEDRUS response pad. The CEDRUS response pad provides millisecond accuracy when recording button presses. The bottom left and right buttons were used for the left and right button presses and were labeled with an L or R. The dual-task experimental block required participants to write the number of times they saw the letter 'o' appear as a fixation point. The participant wrote this number on the back of his/her consent form with a pen provided in the room. Participants were provided with a pair of Sony MDR-XD2000 stereo headphones for the Change-Signal Task with the volume set to conversational level.

Critical Reasoning Questionnaires

Participants completed Qualtrics versions of the CRT (Frederick, 2005) and the CRT-2 (Thomson & Oppenheimer, 2016). Questions (refer to Figure 1) were presented one at a time and participants were instructed to answer quickly and to the best of their ability without the use of any aid. After completing the three CRT questions, participants were asked to indicate how many questions they were familiar with on a scale of 0-3. The same was done after the CRT-2 on a scale from 0-4. Participants were unable to change their responses after advancing to the next question.

Change-Signal Task

Participants were first shown the instruction screen: "*You will be presented with a fixation cross followed by either a left-pointing or right-pointing arrow. Your task is to determine the direction of the arrows by pressing the corresponding button (L = Left and R = Right) on your response pad. However, if you hear a beep after the arrow is presented, press the opposite button. E.g. If you see a left arrow and then hear a beep, press the right button (R) instead of the left. Respond as quickly and accurately as possible. You must respond within 1 second. Press any button to begin.*" On the non-change trials, the participant was presented with a fixation cross in the center of the screen for 1000ms, presented as a plus sign (+) in Arial regular size 200. This fixation cross was followed by an arrow pointing either left or right for 1000ms. The arrows were all 14.5cm in length, 7cm tall at the head of the arrow, and 3.5cm tall in the body. All arrows were colored black and centered on the screen. If the arrow was pointing to the right, the participant was instructed to press the right button on their response pad. If the arrow was pointing left, the participant was instructed to press the left button. One-quarter of the trials in each block were change trials. Change trials involved a left or right arrow presented for either 100ms or 300ms before the presentation of a 100ms auditory cue called the change-signal. The arrow remained on the screen for the duration of the auditory stimulus and an additional 800ms (100ms change trials) or 600ms (300ms change trials) to allow again for 1000ms within which to respond. When the change-signal was present, this indicated to participants that they should respond in the direction opposite of that indicated by the arrow. A left arrow presented with a change-signal required the participant to press the right button, and if a right arrow was presented with a change-signal, then the participant pressed the left button. Refer to Figure 2 for schematic descriptions of the Change-Signal trial types.

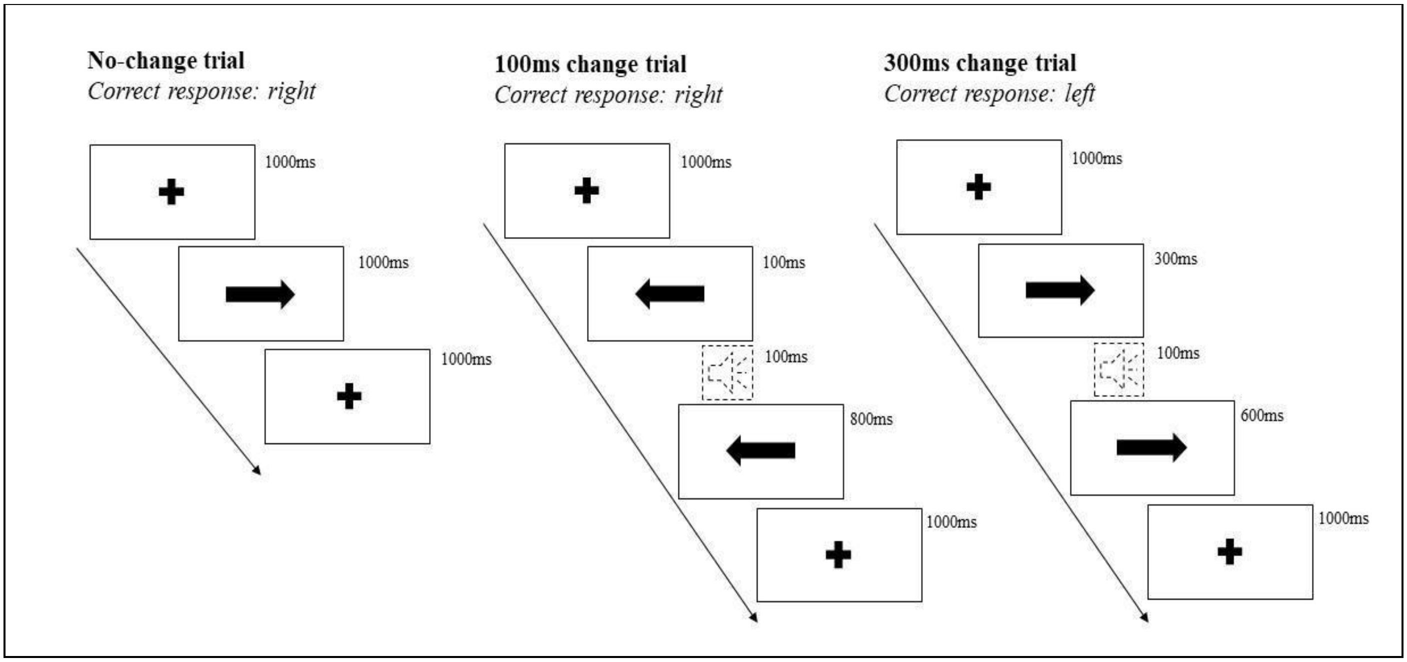

Figure 2. Change-Signal Task Diagram.

The first block of the Change-Signal task involved a practice block of 32 trials where participants were provided with feedback on incorrect answers. An error was designated as an incorrect button press or an inability to complete the correct response in 1000 ms. In the practice round there were 24 non-change trials and 8 change-signal trials split evenly between left and right required responses and across the two different signal cues (100ms, 300ms). The first experimental block consisted of 208 trials with 156 non-change trials and 52 change-signal trials split evenly between right and left responses. Additionally, the trials were split evenly between the 100ms and 300ms change-signal with 26 of each. The order of trials presentation was randomized for each block of trials. During the experimental blocks of trials, the participants did not receive error feedback and were interrupted by the computer after every fifty presentations for a short rest period. The participant resumed whenever he/she was ready to start again.

The second experimental block (dual-task) of trials involved the same task as the first experimental block of trials, except that on 78 of the trials the fixation cross was randomly replaced by a lowercase letter "o" as the fixation point, presented in Arial regular size 120, which appeared the same size as the fixation cross. Participants were required to perform dual-tasks by maintaining a count of the number of o's they saw as they performed the change-signal trials in this block. The trials in this block were again interrupted every fifty trials for the participants to now write down the number of o's they counted and to reset their mental count back to zero. Participants resumed this block of trials when they were ready to start again.

Procedure

Before the participants arrived, the experimenter set the sound level of the headphones for the experiment. Participants performed this study in person, and sat alone in a chair in a small experimental room with only the materials described. The door to the experimental room was closed for the duration of both the Change-Signal Task and critical reasoning measures with the participant instructed to alert the experimenter upon completion of either task. The order of presentation of the critical reasoning questionnaires and the Change-Signal Task was counterbalanced across participants. Once the participant had completed both tasks, he/she was provided debriefing information explaining the goals of the experiment.

Results

Individual differences in critical reasoning

The mean score of the seventy-two participants on the CRT was 1.06 (*SD*=1.09) and the mean score on the CRT-2 was 2.64 (*SD* = .99). Table 1 provides the frequency of correct responses for each questionnaire. An independent samples t-test found a significant difference between the scores of males (*M* = 1.56, *SD* = 1.09) and females (*M* = .89, *SD* = 1.06) on the

CRT, $t(70) = 2.29$, $p = .02$. An independent samples t-test revealed no significant difference in the score of males (M =2.83, SD = .98) and females (M = 2.57, SD = 1.00) on the CRT-2, $t(70) = .95$, $p = .34$. The CRT and CRT-2 scores in this study comprise only Type 2, correct responses (e.g. five cents for the bat and ball problem). A Pearson correlation found that scores on the CRT were correlated with scores on the CRT-2, $r(72) = .31$, $p < .001$. Performance on the CRT questions was correlated with self-reported familiarity with the CRT questions, $r(72) = .24$, $p = .02$. Performance on the CRT-2 was correlated with familiarity with the CRT-2 questions, $r(72) = .37$, $p < .001$. For both the CRT and CRT-2, participants were split into two groups (low score versus high score) on the basis of the number of Type 2 answers provided (refer to Table 1). The cutoff for group membership was a score of zero for the CRT ($n_{low} = 31$, $n_{high} = 41$) and a score of three for the CRT-2 ($n_{low} = 25$, $n_{high} = 47$). As can be seen from Table 1, by sorting participants into only two groups, a third of the participants were not assigned group membership that matched their CRT and CRT-2 scores.

Table 1

Distribution of Correct (Type 2) Answers for the CRT and CRT-2

		CRT-2						
		0	1	2	3	4	Total	Percentage
CRT	0	1	9	6	11	4	31	43%
	1	0	0	5	9	2	16	22%
	2	0	0	2	11	2	15	22%
	3	1	0	1	4	4	10	13%
Total		2	9	14	35	12	72	100%
Percentage		3%	12%	19%	49%	17%	100%	

Change-Signal Task

Two three-way ANOVAs were conducted with the three independent variables being: change-signal condition (no change, 100ms change, 300ms change), cognitive load (single-task, dual-task), and critical reasoning group (low, high). Performance on the Change-Signal Task was calculated as the percentage of Type 2, correct responses for each condition as incorrect responses could come from either a late response or pressing the incorrect button. Percentages were calculated because there were 156 no-change trials, twenty-six 100ms-change trials, and twenty-six 300ms-change trials in each experimental block of trials. Statistical analyses were conducted with the full data set of participants and again with outliers (i.e., task performance outside 1.5*interquartile range) removed. As no differences were found for the findings that resulted from these two analyses, the analysis involving the full data set is provided below.

The first three-way ANOVA using the CRT groups found a significant main effect for the cognitive load, $F(1, 70) = 30.64$, $p < .001$, $\eta_p^2 = .30$, with participants performing much better on the single-task experimental block (*M* = 78.42, *SD* = 15.17) compared to the dual-task experimental block (*M* = 71.40, *SD* = 15.58). A significant main effect was also found for the change-signal delay, $F(2, 140) = 187.91$, $p < .001$, $\eta_p^2 = .73$, with participants performing progressively worse from the no-change-signal condition (*M* = 96.03, *SD* = 2.87) to the 100ms change-signal delay (*M* = 79.00, *SD* = 22.97) and finally to the 300ms change-signal delay (*M* = 49.71, *SD* = 22.41). The between subjects effect for the CRT groups was not significant, $F(1, 70) = 1.19$, $p = .27$. The interaction between the cognitive load and the change-signal delay was significant, $F(2, 140) = 9.95$, $p < .001$, $\eta_p^2 = .12$, and is displayed in Figure 3. This interaction revealed that performance was most challenging when the extra cognitive load (dual-task) was provided for the 300ms change-signal condition. There was no significant interaction found for CRT groups and cognitive load, $F(1, 70) = 0.93$, $p = .38$ and no significant interaction found for CRT groups and change-signal delay, $F(2, 140) = 1.65$, $p = .19$. The three-way interaction involving the CRT groups, cognitive load, and change-signal delay variables was also not significant, $F(2, 140) = 0.44$, $p = .65$.

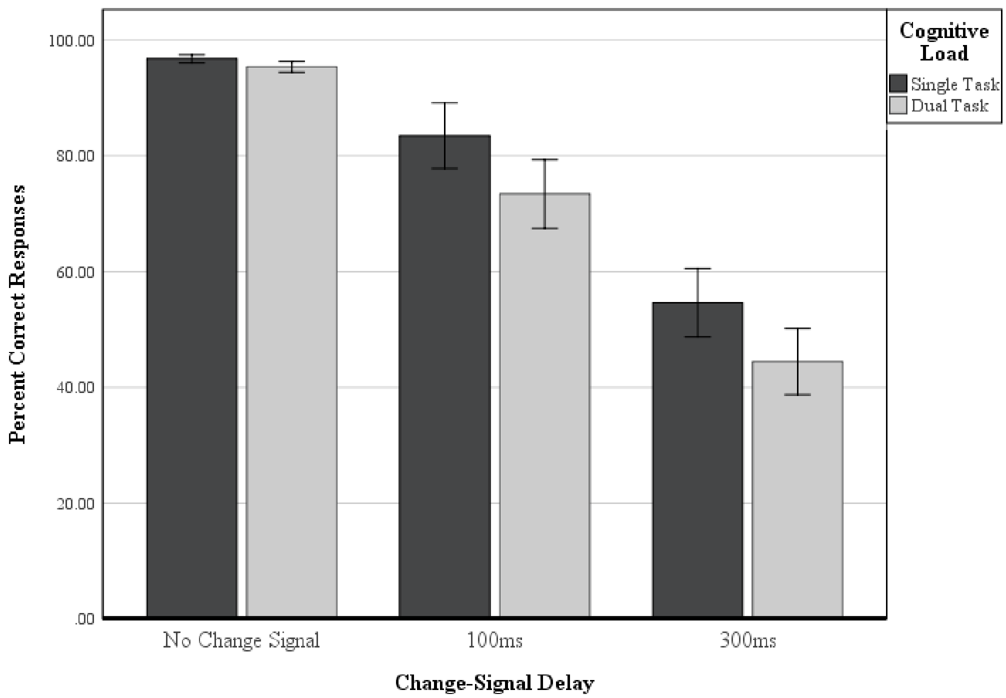

Figure 3. Interaction of cognitive load and cue condition. *Note*. Error bars are 95% CI.

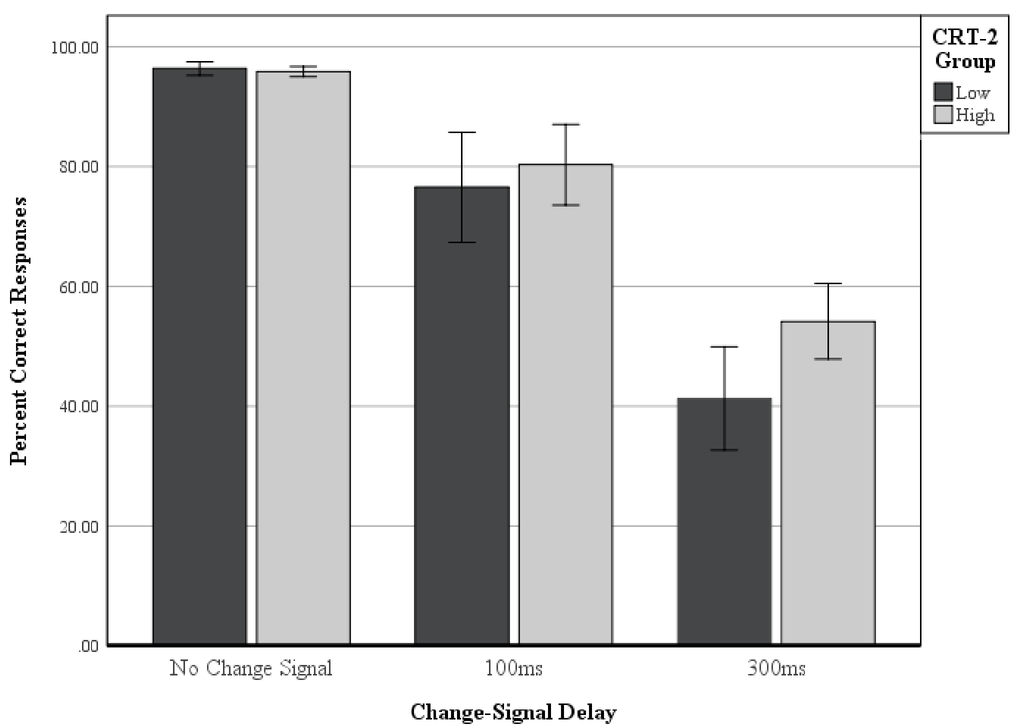

Figure 4. Interaction between the change-signal delay and CRT-2 groups. *Note*. Error bars are 95% CI.

A three-way ANOVA using the CRT-2 groups revealed a significant main effect for cognitive load, $F(1, 70) = 27.41$, $p < .001$, $\eta_p^2 = .28$ with participants performing better on the single-task experimental block ($M = 78.43$, $SD = 15.17$) compared to the dual-task experimental block ($M = 71.40$, $SD = 15.58$). A significant main effect was also found for the change-signal delay, $F(2, 140) = 194.24$, $p < .001$, $\eta_p^2 = .73$, with participants again performing progressively worse from the no-change-signal condition ($M = 96.03$, $SD = 2.87$) to the 100ms change-signal delay ($M = 79.00$, $SD = 22.98$) and finally the 300ms change-signal delay ($M = 49.71$, $SD = 22.41$). The between subjects effect for the CRT-2 groups was not significant, F(1, 70) = 2.32, $p = .13$. The interaction between the cognitive load and change-signal delay was again significant, $F(2, 140) = 9.14$, $p < .001$, $\eta_p^2 = .11$). There was no significant interaction found between CRT-2 groups and cognitive load, $F(1, 70) = .06$, $p = .80$. However, for this analysis, a significant interaction was found between CRT-2 groups and the change-signal delay, $F(2, 140) = 3.78$, $p = .025$, $\eta_p^2 = .05$; refer to Figure 3. This interaction revealed that the only difference between the two CRT-2 groups was for the 300ms change-signal trials (regardless of cognitive load changes). This result highlights the difficulty participants in the low CRT-2 group found when inhibiting the pre-potent response later into the response preparation processes. This finding is important insofar as it shows a difference in the strength of interference control between the two CRT-2 groups. Meanwhile, the three-way interaction for CRT-2 groups, cognitive load, and change-signal delay was not significant, $F(2, 140) = .24$, $p = .78$.

Discussion

The findings of the current study support the task design of the Change-Signal Task developed by the author. Participants found the Change-Signal Task successively more difficult to perform correctly in the 1000 ms provided as the trials varied from the no-change, 100ms, 300ms conditions, and even more so when adding an extra load on working memory during the dual-task condition. The main hypothesis that participants who performed worse on the CRT would perform significantly worse on the Change-Signal Task, especially during the dual-task block of trials, was not supported. However, the CRT-2 findings revealed a significant performance difference between the two groups as evidenced by the difference in performance during the 300ms change-signal condition. The high CRT-2 group performed significantly better than the low-scoring CRT-2 group for this condition, regardless of cognitive load imposed. Interference control demand at the 300ms condition was highest for the low CRT-2 participants because they had to balance the initiation of a response (arrow response) with new information to inhibit that response and form a new motor response (i.e., opposite button press). The high-scoring CRT-2 participants appear to demonstrate greater interference control in the current study. These findings need to be replicated and perhaps further testing of the Change-Signal Task would reveal a delay manipulation that can better discriminate between the two groups in both the CRT and CRT-2. The 100ms and 300ms delays were chosen based on the presumed cognitive processing speed of the visual stimulus and corresponding response found during pilot research of this task. A further analysis of the time course of stimulus perception and response could help determine the most appropriate delay signals, while also incorporating EEG recording of the perceptual, motor, and interference control processing occurring during each trial to determine optimal manipulation of this task variable.

Although this experiment found a significant result favoring interference control as a determinant in CRT-2 scores, the effect was not particularly large and did not show the predicted stronger effect during the dual task condition. Additionally, no such result was found for the CRT. While this could be explained by claiming that the CRT relies more on numeracy, that position would be complicated by the inclusion of the CRT-2 and its weak effects. The CRT-2 was included as a way to remove the impact of numeracy while still measuring cognitive reflection. It is therefore likely that interference control plays an important, but not exclusive role in determining the Type 1 and Type 2 distinction. Evans (2008) reconceptualized the terms System 1 and System 2 into Type 1 and Type 2 to emphasize the idea that these distinctions are the result of several interrelated processes. The current study has only examined the impact of one cognitive mechanism, that of interference control, and it is likely other related mechanisms could also explain individual differences in critical reflection measures. While the CRT and CRT-2 have been consistently correlated with other measures of analytical thinking such as the ACT, SAT, and various intelligence tests, the two CRTs do not have the same reliability and validity as those more well-established tests. This mild effect reported for interference control may instead be a product of the low reliability and validity of the two CRT measures to distinguish Type 1 and Type 2 processing groups.

Scores on the CRT and CRT-2 are often interpreted using three models of dual-process theories: default-intervention model, parallel-processing model, and the hybrid model. The default-intervention model posits that intuitive responses arise from Type 1 processes, which occur first and automatically, but that Type 2 processes can intervene and correct (Evans & Stanovich, 2013; Kahneman, 2011). In the default-intervention model, these processes are distinct. Type 2 is a late-stage, corrective process that needs to be effortfully engaged to replace the intuitive response. Meanwhile, the parallel-processing model holds that Type 1 and Type 2 processing occur simultaneously and compete for control over the response (Sloman 1996; 2014). Again, Type 1 responses must be inhibited for a correct response. However, in the parallel-processing model, Type 2 processes can signal the correct response without fully overriding the intuitive response. Owing to this simultaneity of activation, the parallel-processing model is seen as an early-stage intervention model (Cokely & Kelley 2009). Similarly, the hybrid model assumes that the Type 1 and Type 2 processes occur simultaneously, but this model emphasizes that Type 1 processes can sometimes lead to an intuitive but correct solution (De Neys, 2012, 2014; Pennycook et al., 2015; Thompson et al., 2011). Hybrid model interpretations predict that Type 1 and Type 2 processes have the ability to detect conflict between responses as the problem solving processes continue. Since a Type 1 response is often prepotent and therefore stronger, it can be difficult to overcome the response even if a conflict is signaled. Therefore, participants can experience conflict without it directing them towards the correct answer.

Several recent studies have examined the validity of the three dual-process models in explaining response behavior and performance on the CRT. Travers et al. (2016) used a mouse-tracking experiment where participants were presented with a multiple-choice version of the CRT on a computer screen. Correct participants were first drawn to the Type 1 option before selecting the correct, Type 2 option whereas the opposite was not true for incorrect respondents. This finding is in favor of the default-intervention model with Travers and colleagues reasoning that late-stage, Type 2 processes were deployed to override the Type 1 response. Meanwhile, a think-aloud protocol found that participants followed a hybrid model approach (Szaszi et al., 2017). However, these studies relied on the recording of overt responses from participants that may not capture the covert or implicit processing involved with intuitive responses (Crutcher, 1994; Glöckner & Betsch, 2008; Glöckner & Herbold, 2011). Purcell et al. (2022) use a more rigorous eye-tracking method with a multiple-choice version of the CRT. This study found that correct participants did not consider the intuitive response more than the other incorrect responses. Instead, participants considered the correct, Type 2 choice the most without first considering the intuitive, Type 1 choice. These results provide support for the hybrid model of dual-process theory by showing that participants respond to the stronger process, as evidenced by how much they considered the corresponding answer. Importantly, the hybrid model holds that Type 1 and Type 2 processes both occur simultaneously and signal responses, leading to conflict.

Although the experimental design utilized in the current study did not explicitly test different models of dual-process theory, the findings may indirectly support the hybrid model by providing a mechanism for modulating the two processes. As noted above, hybrid models of the CRT and CRT-2 hold that the Type 1 and Type 2 processes occur in parallel, which can generate conflict between the two responses until one process wins out. For example, a Type 1 response to the bat and ball problem would be 10 cents, which would find itself in conflict with the Type 2 response of 5 cents because those two responses are signaled in parallel. This holds true even if the responses are not fully formed. Eventually, one process will continue over the other. Despite recent studies, it is unclear which cognitive mechanisms modulate conflict between the Type 1 and Type 2 processes. The present study suggests that interference control is likely part of that conflict mediation process as an early-stage intervention. High-scorers on the CRT-2 performed significantly better in the 300ms change-signal condition (refer to Figure 4), and this change was cued before a response was even initiated on average in the no-change trials – the mean response time for participants on no-change trials was 519.10 milliseconds (*SD* = 83.62). This indicates that individual differences in interference control are important for determining an individual's ability to answer the CRT-2 correctly. Comparisons of response times between change-signal conditions was not possible because the change-signal was varied systematically. Additionally, it is not entirely clear what the response time would be indicative of if evaluated.

Limitations

The present study included only 72 participants who were sorted into two groups based on their CRT and CRT-2 scores. Collecting more data would enable the groupings to be based on a continuum of four groups for the CRT and five groups for the CRT-2. A more precise grouping could provide a stronger test of the effects of interference control under cognitive load (dual task) than was available with the current data set. The participants in the current study do not represent the general population as the same factors that help someone reach college may also help them in answering the CRT or Change-Signal Task. Additionally, 54 out of our 72 (75%) participants were females and a significant gender difference in CRT measures was found in the current study with males outperforming females. This may affect the results of this study by lowering the average CRT score compared to the population. Indeed, the average CRT score in this study was 1.06 while Frederick (2005) initially reported the mean score to be 1.24. Additionally, the meta-analysis by Brañas-Garza et al. (2019) found that students performed better than non-students. This student effect may somewhat offset the effect of oversampling females in the current study. However, to what degree this is true is unknown. A student effect may also increase CRT scores overall for participants in this study. Further, some research has also suggested that there is a difference in neural activation between males and females on Stop-Signal Tasks (Li et al., 2006), which is the task upon which the Change-Signal Task design was based. Which is to say that any gender difference in scores seen on the CRT may have also been realized as gender differences in performance of the Change-Signal Task. While Li and colleagues (2006) show that these activational differences do not seem to affect performance on the Stop-Signal Task, it does perhaps provide an explanation to the differences seen on the CRT. A biased sample such as is in this study may restrict the degree to which we can assess individual differences and be able to generalize to the broader population. Future studies should aim for a more even distribution of males and females. Given the small number of participants and lack of experimental manipulation for gender, the present study does not include a reliable measurement of gender differences on the Change-Signal Task. Future research should look at any possibility of performance differences on the CRT as a result of differential neural activation.

Future research

The current study has shown that interference control may be important for modulating responses on the CRT-2 and therefore likely a factor in determining individual differences in reasoning and problem solving (critical reflection). Future research could focus on the neural basis for this cognitive mechanism. Mapping the neural networks involved in interference control would lead to a better understanding of dual-process models. Electroencephalogram (EEG) studies have provided some information about the areas of the brain involved in inhibitory and decision-making processes. For example, van Vugt et al. (2012) found large individual differences in decision making at the neural level as recorded by the EEG. Future research utilizing the Change-Signal Task may be able to use these neural signatures to develop a better understanding of the neural differences between Type 1 and 2 processing, as well as more specific testing of the different dual process models.

Conclusion

The present study has shown support for the response inhibition mechanism of interference control as a partial determinate of performance on the CRT-2. No significant support was shown for performance on the CRT. However, with modifications to the Change-Signal task, a significant result may appear as evidenced by similar patterns for participant performance of CRT and CRT-2 groups on the Change-Signal Task. The experimental task, the Change-Signal

Task, performed as expected. Delays in the change-signal made it much more difficult for participants to respond correctly. For the CRT-2, this shows that interference control, the mechanism by which competing stimuli are balanced, plays a role in correct performance. Given the weak effects, however, it is likely that interference control is one of many processes that contribute to the individual differences seen in the CRT and CRT-2. Future research should look towards identifying neural correlates of interference control to see if differential activation can explain performance on analytical reasoning measures such as the CRT and CRT-2.

References

Andersson, O., Tyran, J. R., Wengström, E., & Holm, H. J. (2013). Risk Aversion Relates to Cognitive Ability: Fact or Fiction? *IFN Working Paper No. 964*.

Benbow, C. P., Lubinski, D., Shea, D. L., & Eftekhari-Sanjani, H. (2000). Sex differences in mathematical reasoning ability: Their status 20 years later. *Psychological Science*, *11*(474-480).

Besedes, T., Deck, C., Sarangi, S., & Shor, M. (2012). Decision-making Strategies and Performance among Seniors. *Journal of Economic Behavior and Organization*, *81*(2), 524-533.

Böckenholt, U. (2012). The cognitive-miser response model: Testing for intuitive and deliberate reasoning. *Psychometrika*, *77*, 388–399.

Brañas-Garza, P., Kujal, P., & Lenkei, B. (2019). Cognitive reflection test: Whom, how, when. *Journal of Behavioral and Experimental Economics*, *82*.

Brydges, C. R., Anderson, M., Reid, C. L., & Fox, A. M. (2013). Maturation of Cognitive Control: Delineating Response Inhibition and Interference Suppression. *PLoS ONE*, *8*(7), e69826.

Brydges, C. R., Klunies-Ross, K., Clohessy, M., Lo, Z. L., Nguyen, A., & Rousset, C. (2012). Dissociable Components of Cognitive Control: An Event-Related Potential (ERP) Study of Response Inhibition and Interference Suppression. *PLoS ONE*, *7*(3), e34482.

Campitelli, G., & Gerrans, P. (2014). Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. *Mem Cogn*, *42*, 434–447. https://doi.org/10.3758/s13421-013-0367-9

Campitelli, G., & Labollita, M. (2010). Correlations of cognitive reflection with judgments and choices. *Judgment and Decision Making*, *5*(3), 182-191.

Cao, H., & Cannon, T. D. (2021). Distinct and temporally associated neural mechanisms underlying concurrent, postsuccess, and posterror cognitive controls: Evidence from a stop-signal task. *Human Brain Mapping*, *42*, 2677-2690.

Carpenter, J., Graham, M., & Wolf, J. (2013). Cognitive ability and strategic sophistication. *Games and Economic Behavior*, *80*(1), 115-130.

Chambers, C. D., Garavan, H., & Bellgrove, M. A. (2009). Insights into the neural basis of response inhibition from cognitive and clinical neuroscience. *Neuroscience & Biobehavioral Reviews*, *33*(5), 631-646. https://doi.org/10.1016/j.neubiorev.2008.08.016

Cokely, E. T., & Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgement and Decision Making*, *4*(1), 20-33.

Corgnet, B., Espín, A. M., & Hernán-González, R. (2015). The cognitive basis of social behavior: cognitive reflection overrides antisocial but not always prosocial motives. *Behavioral Neuroscience*, *9*, 1662-5153.

Crutcher, R. J. (1994). Telling what we know: The use of verbal report methodologies in psychological research. *Psychological Science*, *5*(5), 241.

Cueva-Herrero, C., Iturbe-Ormaetxe, I., Mata-Prez, E., Ponti, G., Yu, H., & Zhukova, V. (2016). Cognitive (Ir)reflection: New Experimental Evidence. *Journal of Behavioral and Experimental Economics*, *64*.

De Neys, W. (2012). Bias and Conflict: A Case for Logical Intuitions. *Perspectives on Psychological Science*, *7*(1), 28-38. doi:10.1177/1745691611429354

De Neys, W. (2014). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking & Reasoning*, *20*(2), 169–187. https://doi.org/10.1080/13546783.2013.854725

Donders, F. C. (1868/1969). Over de snelheid van psychische processen [On the speed of mental processes] (W. Koster, Trans.). *Attention and Performance*, *II*, 412– 431.

Evans, J. S. B. T. (1989). *Bias in human reasoning: Causes and consequences.* Erlbaum.

Evans, J. S. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*, *59*, 255-78. 10.1146/annurev.psych.59.103006.093629

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, *8*(3), 223-241. https://doi.org/10.1177/1745691612460685

Finucane, M. L., & Guillon, C. M. (2010). Developing a tool for measuring the decision-making competence of older adults. *Psychology and Aging*, *25*(2), 271-288.

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, *19*(4), 25-42.

Freiden, E., & Martini, L. A. (2021). An Analytic Cognitive Style Negatively Predicts a More Literal but Not a More Symbolic Religiosity Type. *The International Journal for the Psychology of Religion*, *32*(1), 31-52.

Glöckner, A., & Betsch, T. (2008). Multiple-reason decision making based on automatic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1055–1075.

Glöckner, A., & Betsch, T. (2011). Glöckner, A., & Herbold, A.-K. (2011). An eye-tracking study on information processing in risky decisions: Evidence for compensatory strategies based on automatic processes. *Journal of Behavioral Decision Making*, *24*(1), 71–98.

Holt, C., Porzio, M., & Song, M. Y. (2017). Price bubbles, gender, and expectations in experimental asset markets. *European Economic Review*, *100*(C), 72-94.

Hoppe, E. I., & Kusterer, D. J. (2011). Behavioral biases and cognitive reflection. *Economics Letters*, *110*, 97-100.

Kahneman, D. (2011). *Thinking Fast and Slow*. Farrar, Straus and Giroux.

Keren, G., & Schul, Y. (2009). Two Is Not Always Better Than One: A Critical Evaluation of Two-System Theories. *Perspectives on Psychological Science*, *4*(6), 533-550. 10.1111/j.1745-6924.2009.01164.x

Kiss, H. J., Rodriguez-Lara, I., & Rosa-García, A. (2016). Think Twice Before Running! Bank Runs and Cognitive Abilities. *Journal of Behavioral and Experimental Economics*, *64*(C), 12-19.

Kok, A., Ramautar, J. R., De Ruiter, M. B., Band, G. P. H., & Ridderinkhof, K. R. (2003). ERP components associated with successful and unsuccessful stopping in a stop-signal task. *Psychophysiology*, *41*(1), 9-20.

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberative judgements are based on common principles. *Psychological Review*, *118*, 97–109.

Li, C. R. (2008). Sub-Cortical Processes of Motor Response Inhibition during a Stop Signal Task. *Neuroimage*, *41*(4), 1352-1363.

Li, C. R., Huang, C., Constable, R. T., & Sinha, R. (2006). Gender differences in the neural correlates of response inhibition during a stop signal task. *NeuroImage*, *32*, 1918-1929.

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.

Liberali, J. M., Reyna, V. F., Furlan, S., Stein, M. L., & Pardo, T. S. (2011). Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment. *Journal of Behavioral Decision Making*, *25*, 361–381.

Logan, G. D., Cowan, W. B., & Davis, K. A. (1984). On the ability to inhibit simple and choice reaction time responses: A model and a method. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(2), 276–291.

Mau, W. C., & Lynn, R. (2010). Gender differences in homework and test scores in Mathematics, Reading and Science at tenth and twelfth grade. *Psychology, Evolution & Gender*, *2*(2), 119-125.

Mikušková, E. B. (2021). The Analytic Cognitive Style Predicts Conspiracy Beliefs: Replication Study on a Non-student Sample. *Studia Psychologica*, *63*(2), 190–203.

Moritz, B., Hill, A. V., & Donohue, K. (2013). Individual Differences in the Newsvendor Problem: Behavior and Cognitive Reflection. *Journal of Operations Management*, *31*(1-2), 72-85.

Newstead, S. E. (2000). Are there two different kinds of thinking? *Behavioral and Brain Sciences*, *23*, 690–691.

Neyse, L., Bosworth, S., Ring, P., & Schmidt, U. (2016). Overconfidence, Incentives and Digit Ratio. *Scientific Reports*, *6*, 23294.

Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin*, *126*(2), 220-246.

Oechssler, J., Roider, A., & Schmitz, P. (2009). Cognitive Abilities and Behavioral Biases. *Journal of Economic Behavior and Organization*, *72*(1), 147-152.

Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review*, *11*, 988–1010.

Pehlivanoglu, D., Lin, T., Deceus, F., Heemskerk, A., Ebner, N. C., & Cahill, B. S. (2021). The role of analytical reasoning and source credibility on the evaluation of real and fake full-length news articles. *Cogn. Research*, *6*, 24.

Pennycook, G., Fugelsang, J. A., & Kohler, D. J. (2015). What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology*, *80*, 34–72.

Ponti, G., & Rodriguez-Lara, I. (2015). Social preferences and cognitive reflection: evidence from a dictator game experiment. *Frontiers in Behavioral Neuroscience*, *19*, 1662-5153.

Purcell, Z. A., Howarth, S., & Wastell, C. A. (2022). Eye tracking and the cognitive reflection test: Evidence for intuitive correct responding and uncertain heuristic responding. *Mem Cogn*, *50*, 348–365.

Ring, P., Neyse, L., David-Barrett, T., & Schmidt, U. (2016). Gender Differences in Performance Predictions: Evidence from the Cognitive Reflection Test. *Frontiers in Psychology*, *7*, 1680.

Schachar, R., Logan, G. D., Robaey, P., Chin, S., Ickowicz, A., & Barr, C. (2007). Restraint and Cancellation: Multiple Inhibition Deficits in Attention Deficit Hyperactivity Disorder. *Journal of Abnormal Child Psychology*, *35*, 229-238.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.

Sloman, S. A. (2014). Two systems of reasoning: An update. In *Dual-process theories of the social mind* (pp. 69–79). The Guilford Press.

Stanovich, K. E. (2004). *The robot's rebellion: Finding meaning the age of Darwin*. University of Chicago Press.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*, 645–726. doi:10.1017/S0140525X00003435

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, *36*, 276 –315.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–661.

Stupple, E., Gale, M., & Richmond, C. (2013). Stupple, E., Gale, M., & Richmond, C. (2013). Working memory, cognitive miserliness and logic as predictors of performance on the cognitive reflection test. *Proceedings of the Annual Meeting of the Cognitive Science society*, *35*, 1396-1401.

Szaszi, B., Szollozi, A., Palfi, B., & Aczel, B. (2017). The cognitive reflection test revisited: Exploring the ways individuals solve the test. *Thinking and Reasoning*, *23*(3), 207–234.

Thoma, V., Weiss-Cohen, L., Filkuková, P., & Ayton, P. (2021). Cognitive Predictors of Precautionary Behavior During the COVID-19 Pandemic. *Frontiers in Psychology*, *12*, 1664-1078.

Thompson, V. A., Prowse Turner, J. A., & 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), 99–11.

Toplak, M. F., West, R. F., & Stanovich, K. E. (2011). The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*, 1275–1289. doi:10.3758/s13421-011-0104-1

Travers, E., Rolison, J. J., & Feeney, A. (2016). The time course of conflict on the Cognitive Reflection Test. *Cognition*, *150*, 109-118.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.

van Velzen, L. S., Vriend, C., de Wit, S. J., & van den Heuvel, O. A. (2014). Response inhibition and interference control in obsessive–compulsive spectrum disorders. *Frontiers in Human Neuroscience*, *8*, 1662-5161.

van Vugt, M. K., Simen, P., Nystrom, L. E., Holmes, P., & Cohen, J. D. (2012). EEG oscillations reveal neural correlates of evidence accumulation. *Frontiers in Neuroscience*, *6*(106).

Wason, P. C., & Evans, J. S. B. T. (1975). Dual processes in reasoning? *Cognition*, *3*, 141–154.

Welsh, M., Burns, N., & Delfabbro, P. (2013). The Cognitive Reflection Test: How much more than numerical ability. *In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.)*, *Proceedings of the 35th annual conference of the cognitive science society*, 1396–1401.