UiT The Arctic University of Norway

Department of Psychology, Faculty of Health Sciences

Default, time-pressured and incentivized deliberate reasoning

Fozia Aweys Abukar

Master's thesis in psychology - May 2021

Acknowledgement

First and foremost, I would like to thank my supervisor, Professor Gerit Pfuhl. I am sincerely grateful and appreciative for the constant guidance and support, even when my motivation was very low. I would like to acknowledge that Gerit Pfuhl went above and beyond as a supervisor, especially in performing all the data analysis for this project. Thank you for your patience and for taking the time to explain things to me when it was difficult for me to understand or make sense of things.

I would like to thank the Department of Psychology at UiT for offering an open master's program with a wide range of opportunities that has allowed me to choose my field of interest. I am also grateful for Mikołaj Hernik and Jon-André Dalbakk, for following up and helping with any challenges.

Last but not least, I would like to thank my family and friends who have been incredibly supportive throughout this process. Thank you for pushing me and not letting me give up.

Thank you all!

Fozia A. Abukar Gerit Pfuhl

Default, time-pressured and incentivized deliberate reasoning

Fozia Aweys Abukar

PSY-3900

Master's thesis in psychology

May, 2021

UiT- The Arctic University of Norway

Abstract

The purpose of this study was to investigate the effect time-pressure and incentives have on deliberate reasoning. A person's tendency to engage in and enjoy processes that require reasoning, i.e., their need for cognition, and their algorithmic ability are factors that may moderate performance in reasoning tasks when they are under time-pressure or given monetary reward. Identifying how to increase deliberate reasoning is important in a great deal of societal challenges, e.g., immunity against fake news. This study was conducted online on Prolific in two sessions three weeks apart. The first day testing (baseline) was the same for all participants. In the second testing session, the reasoning items from the first session were counterbalanced and performed under two conditions - time-pressure and incentivized condition. Analysis from baseline showed that need for cognition and intelligence contribute to deliberate reasoning. There was a decrease in performance under time –pressure compared to baseline, and verbal ability seems to matter more than numerical ability. There was only little effect of incentive. Participants also reported how much effort they experienced during the reasoning tasks (measured with NASA Task Load Index), and the better the score in reasoning items, the less effortful the task was rated, and the higher a person's NFC was. The results are discussed within the two-system framework.

Abstrakt

Formålet med denne studien var å undersøke effekten tidspress og insentiver har på bevisst resonnement. En persons tendens til å engasjere seg i og nyte prosesser som krever resonnement, deres *need for cognition*, og deres algoritmiske evner er faktorer som kan moderere ytelsen i resonneringsoppgaver når de er under tidspress eller får økonomisk belønning. Å identifisere hvordan du kan øke bevisst resonnement er viktig i mange samfunnsutfordringer, for eksempel immunitet mot fake news. Denne studien ble utført online og i to økter med tre ukers mellomrom. Første testingen (baseline) var den samme for alle deltakere. I den andre økten ble resonnementene oppgavene fra første økt fordelt i to sett basert på vanskelighet og varighet og utført under to forhold - tidspress og insentiv forhold. Analyse fra baseline viste at need for cognition og intelligens bidrar til bevisst resonnement. Det ble funnet en reduksjon i opptreden under tidspress sammenlignet med baseline, og verbal evne ser ut til å ha mer betydning enn numerisk evne. Det var bare liten effekt av insentiv. Deltakerne rapporterte også hvor krevende de opplevde resonneringsoppgavene (målt med NASA Task Load Index), og det ble funnet at jo bedre opptreden, desto mindre anstrengende ble oppgaven vurdert, og jo høyere var en persons NFC. Resultatene diskuteres i lys av to-system teorien.

DEFAULT, TIME-PRESSURED AND INCENTIVIZED DELIBERATE REASONING

We are constantly making decisions and judgments throughout the day, whether it is concerning small decisions like the choice between coffee and tea, or big decisions like moving to a foreign country to pursue a dream job. On many occasions, it is essential to deliberate and think in a rational fashion to make an informed decision. Rationality is of great importance as many societal challenges require citizens to make careful and deliberate decisions. However, humans are said to be *cognitive misers* meaning that we have a tendency to find the easiest and fastest way to make decisions or solve a problem to preserve our limited cognitive resources (Neumann, 2015), and organisms generally tend to favor the option that demands least effort (Dunn, Lutes, & Risko, 2016; Hull, 1943; Kool, McGuire, Rosen, & Botvinick, 2010; Shenhav et al., 2017). Hull (1943) describes this in *"the law of least effort"*, which states that given the choice between multiple courses of actions, all leading to the same goal and reward, the one that requires the least effort will be preferred and chosen. In the event of selecting the more demanding course of acting, the expected value must be considered great (Dixon & Christoff, 2012; Shenhav et al., 2017).

How and if a person deliberates depends on numerous factors, for example, if a person enjoys engaging in cognitively effortful activities, if they are intrinsically or extrinsically motivated and their level of intellectual ability can all have an impact on their reasoning and deliberation (Cacioppo & Petty, 1982; Fleischhauer et al., 2010; Hill et al., 2013; Rucker & Petty, 2006). Additionally, nudges like rewards may appeal for some when performing deliberate reasoning tasks but not others, and time pressure may or may not hamper performance.

Dual-process model of thought

How we reason and make decisions has often been categorized as two different cognitive processes (Evans, 1984; Kahneman & Frederick, 2002; Pennycook, Fugelsang, & Koehler, 2015; Stanovich, West, & Toplak, 2011), referred to as System 1 and System 2. System 1 processing is fast, intuitive and automatic and requires less cognitive effort, where other available tools are used to process the information at hand in a simple manner, e.g. heuristics. System 2 is slow, deliberative and controlled and requires effortful thinking (Frank, Cohen, & Sanfey, 2009; Kahneman, 2003, 2011; Stanovich & West, 2000). Accordingly, there are situations where people evaluate available information, make decisions and judgements, in an effortless and easy way relying on their intuition. Other times decision-making and judgements are based on effortful deliberate reasoning, where all aspects of information is taken into consideration (Evans, 2008; Kruglanski & Gigerenzer, 2011). It is important to recognize that both processing systems work simultaneously and collaborate, and although system 1 depends on automatic processing, the resulting responses are not always less acceptable than those resulted from system 2 (Basel & Brühl, 2013).

Rationality and intelligence

To recognize the distinction between rationality and intelligence can provide insight in understanding how some persons can be irrational even though they are intelligent (Stanovich, 2016). Intelligence tests are frequently treated as if their measurements include all cognitive abilities, but it is strongly argued that they lack inclusion of components of rational thinking (Stanovich & West, 2014; Stanovich et al., 2011). Generally, it can be said that while intelligence tests measures maximum performance (i.e. the algorithmic mind) (Duckworth et al., 2011), rationality tests measure the reflective mind (Stanovich, 2009).

In a tripartite model of dual process thinking, system 2 processing is divided into the reflective mind and the algorithmic, in addition to the automatic process of system 1

(Stanovich, West, & Toplak, 2016). Stanovich et al. (2016) emphasize that differentiating system 2 processing into the algorithmic and reflective mind is fundamental in considering individual differences in rational thought, and also to separate rational thought from the cognitive abilities that are included in intelligence measures. To illustrate the distinction between these, we can say that maximal performance measures are measures of the algorithmic mind, while measures of typical performance assess the reflective mind (Sherman, Gawronski, & Trope, 2014).

Thinking dispositions

Individuals differ in their tendency to engage in processes of reasoning, also known as Need for Cognition (Cacioppo & Petty, 1982). Cacioppo & Petty (1982) described the Need for Cognition (NFC) as an attribute referring to people's tendency to not only participate in, but also find effortful tasks favorable. Individuals who have a high need for cognition pursue effort rather than avoiding it, while individuals with low need for cognition prefer to rely on heuristics and mental shortcuts to avoid exerting effort (Cacioppo, Petty, Feinstein, & Jarvis, 1996). The scale developed thereby is a reflection of individual differences in a person's motivation (intrinsic motivation) to involve themselves in and enjoy tasks that require cognitive effort or deliberate reasoning (Rudolph, Greiff, Strobel, & Preckel, 2018). NFC has therefore shown to be associated with multiple measures of ability, for example reasoning (Fleischhauer et al., 2010; Hill et al., 2013).

Individuals differ in e.g. cognitive style, motivation, need for cognition, intelligence and curiosity. These factors influence the thought processing strategy an individual adopts, i.e. how much cognitive effort is expended. Deliberate reasoning is a demanding and valued action, and variation in the cost of required cognitive effort and expected benefit is assumed to be an important predictor of effort exertion and effort avoidance, respectively (Kool et al., 2010; Kool, Shenhav, & Botvinick, 2017). Previous studies have found a relationship

between NFC and deliberate reasoning (Cacioppo & Petty, 1982), and NFC has also been shown to be associated with intelligence (Cacioppo & Petty, 1982; Fleischhauer et al., 2010; Hill et al., 2013). However, it is not known whether individuals with high NFC treat deliberate reasoning tasks as performance test, while individuals with low NFC will only perform well when an external incentive is given (e.g. money, or to pass an exam). Still, NFC and intelligence are positively related (Cacioppo & Petty, 1982; Cacioppo et al., 1996; Fleischhauer et al., 2010; Hill et al., 2013), but not NFC and working memory (Fleischhauer et al., 2010; Hill et al., 2013).

People differ in NFC, i.e. their level of motivation to get involved in and enjoy tasks that require effortful thinking (Hill et al., 2013). Since NFC measures the motivation one has to increase effort to execute cognitive processes that are perceived as costly, performance in an intelligence test can be predicted depending on whether one has high or low NFC. That means persons higher in NFC have shown higher performance in intelligence tests, and the other way around (Gonthier & Roulin, 2020; Hill et al., 2013; Sandra & Otto, 2018).

The role of motivation and cognitive ability on deliberate reasoning

There is a larger relationship between intrinsic motivation and NFC compared to extrinsic motivation (Amabile, Hill, Hennessey, & Tighe, 1994). Intrinsic motivation is a person's motivation to participate in an activity or a task mainly because one finds this act enjoyable, pleasing and interesting in itself (Amabile et al., 1994). In contrast, extrinsic motivation refers the motivation to perform or exert effort to a task in respons to something external, e.g. to receive reward or to impress someone (Amabile et al., 1994). Accordingly, individuals with high NFC are driven by intrinsic motivation, whereas actions of individuals with low NFC may be due to fulfilment of an extrinsic goal.

Factors that are important and influential regarding the tendency to engage in deliberate reasoning strategies, or elaborating on available and relevant information are

motivation and ability (Rucker & Petty, 2006). That is, a person must "want" and also be "able" to think systematically about an issue, in order for deliberation to occur. For example, if asked, "*Is И a valid Greek letter*?" you can easily answer correctly if you have the ability (i.e. learned Greek), if you haven't learned Greek not even high motivation will help answering correctly. However, if you are asked, "*Is A a valid Latin letter?*" a wrong answer may indicate lack of motivation (if you speak a language using the Latin writing system).

A study (Duckworth et al., 2011) indicated that motivation was higher among persons who had above average score in an IQ measure. This suggests that motivation is required to perform well on an IQ test as well as high intelligence, and that low motivation and low intelligence may lead to poorer IQ scores. In a recent study they investigated whether smarter persons reasoned more accurately because of their ability to deliberately correct a wrong intuition (they called this´the smart deliberator view´), or if their intuitions are more accurate to begin with (they called this ´the smart intuitor view´) (Raoelison, Thompson, & De Neys, 2020). Their results supported the latter view. However, it is also believed that responses that are usually thought to require deliberation, can also be achieved intuitivelly. Exposure, experience and practice of reasoning problems through for example education and everyday life contributes in automating that exact process, which will eventually be activated intuituvely when confronting similar problems. That means, the more it has been automized, the more it is probable that it will overcome a heuristic process (De Neys, 2012; Raoelison et al., 2020).

A person who has not been exposed to reasoning problems to the same degree, and/or has not had the opportunity to automatize reasoning, might still be able to put forth the correct response, but this will be more effortful and might take them more time as they have to override the heuristic response. For example, a person who engaged in reasoning activities as a child would built up a range of intuition, and would consequently do well on IQ tests in

their adult life, because they will probably give the right response intuitively without additional deliberation (Raoelison et al., 2020). This is not to say that persons lower in capacity do not have correct intuitions. In fact some participants lower in cognitive capacity in Raoelison and collegues study did show correct intuitions in their initial response, but this was rare (Raoelison et al., 2020). Nevertheless, studies demonstrate that an intuitiv correct response is more appearent in persons with higher cognitive capacity (Raoelison et al., 2020; Thompson & Johnson, 2014; Thompson, Pennycook, Trippas, & Evans, 2018). The smart intuitor view does not dismiss the role of deliberation and the concept of cognitive misers, but rather accentuates the idea that people high in cognitive capacity are not merely cognitive spenders who will perform well on cognitive tasks (Raoelison et al., 2020). They also argue that high capacity reasoners who arrive on the correct response intuitively in a short time frame might still engage in deliberation to rationalize their reasoning, as they might find this important (Raoelison et al., 2020).

Epistemic curiosity (i.e. curiosity or desire for knowledge) is also a construct that has been shown to be related to NFC (Jebb, Saef, Parrigon, & Woo, 2016; Mussel, 2010; Powell, Nettelbeck, & Burns, 2016). As NFC reflects an individuals intrinsic motivation to enjoy and engage in cognitively challenging tasks, it is understandable that this would be associated with epistemic curiosity; " the desire for knowledge that motivates to learning and solving intellectual problems" (Mussel, 2010). That is, a person who engages in effortful tasks and find these enjoyable (high in NFC), may also to some degree find the task interesting and want to investigate and understand the complex task.

Effort is valued (subjective and objective effort)

As deliberate processing is more demanding, it is also considered costly. The decision to exert effort to a task is affected by several factors, among others a person's evaluation of the value of options. A value-based framework of effort proposes that the subjective value of

each available option is evaluated in the process of decision-making, and the option that is considered to have more value is preferred (Dixon & Christoff, 2012; Schmidt, Lebreton, Cléry-Melin, Daunizeau, & Pessiglione, 2012).

The decision to exert cognitive effort is partially reflected in individual differences in evaluating the costs and benefits associated with cognitive processing and the expected benefits (Kool et al., 2010; Kool et al., 2017). Tasks involving high demand are considered costly, but when the cost of demand is outweighed by the subjective value of an expected benefit or reward, effort is more likely to be exerted (Kool et al., 2010; Kool et al., 2017). How much demand an individual experience in a task may be influenced by the individual's accessible cognitive resources (or ability). Individual differences in ability might have an impact on avoidance behavior (Kool et al., 2010). If a person has the skills that are necessary for performing a specific task, that task will not be perceived as demanding or effortful. For example, if you are fit, then running to the bus is not demanding or effortful. But, if you are not fit or are carrying a heavy bag, running to the bus will be effortful. Similarly, a cognitive task would not be particularly demanding for a person high in IQ. For a person with low IQ, or normal IQ but is burdened with ruminations or other mental disorder issues, cognitive tasks would be more effortful and demanding.

The more resources or cognitive capacity a person holds, the more likely it is for that person to engage in the demanding, deliberate and analytical process and, if needed, correct an intuition that is wrong (De Neys, 2006; Raoelison et al., 2020). A person with restricted cognitive resources evaluates exerting effort as more costly, and their decisions are steered towards a less demanding course of action (Sandra & Otto, 2018). Recently, Sandra and Otto (2018) showed that incentives can increase an individual´s expended cognitive effort. Persons with low need for cognition invested more cognitive effort in a task when the presented reward incentive was increased, but this increase in cognitive effort was not present in

persons who scored high in need for cognition. This may, in part, be explained by the differences in intrinsic motivation in high and low NFC individuals.

The effect of time pressure and incentive

In addition to an individual's intrinsic motivation, in some situations there could be external variables present that may have an impact on how information is processed, for example time-pressure and incentives (Kao, 2011). For instance, in a study examining the influence of NFC and time pressure on certain persuaded messages, Kao (2011) concluded that, given abundant amount of time to process information, persons who are motivated will possibly process this information in a systematic manner. On the other hand, if time is limited or there is little motivation, information will probably be processed in a heuristic manner (Kao, 2011). Other studies of the effect of time pressure in reasoning are in line with these findings (Evans & Curtis-Holmes, 2005; Raoelison et al., 2020), and show that time pressure led to an increase in heuristic processing and consequently lower percentage of correct answers, which in turn elevated belief biases.

Offering participants (performance contingent) monetary rewards results in different outcomes and to effect effort in various ways (Bonner & Sprinkle, 2002). For example, in a review of 74 experiments that provided no, low and high incentive for performance Camerer & Hogarth (1999) found that some studies with higher incentives had an effect on average performance, specifically in judgement and decision-making tasks, and also in prediction and recollecting items. However, they also found that in several studies it appeared incentives did not matter in task performance, which Camerer & Hogarth (1999) believed could be explained by participants intrinsic motivation to perform adequately regardless of the reward, or the task being too difficult, and more effort would not be helpful. Other times incentives were observed to be hurtful, for example when self-consciousness increases due to eagerness caused by incentives (e.g., anxiety when taking tests) (Camerer & Hogarth, 1999).

In sum, the effect of incentives on performance is disputed. Incentives do not necessarily improve performance but have an effect on the variance of the data (Camerer & Hogarth, 1999) as participants with little motivation seem to attempt doing better.

The present study

As reviewed above, NFC has been shown to be associated with intelligence (Cacioppo & Petty, 1982; Fleischhauer et al., 2010; Hill et al., 2013). And previous studies have also found a relationship between NFC and deliberate reasoning (Cacioppo & Petty, 1982) thought to be due to persons high in NFC engaging in and enjoying cognitively demanding tasks. Indeed, individuals with high NFC seem to perceive low subjective costs when engaging in cognitively demanding tasks and are less sensitive to incentives whereas individuals with low NFC deploy more cognitive control when monetary incentives are promised (Sandra and Otto, 2018). We therefore measure perceived subjective effort with the NASA task load index (N-TLX, Hart, 2006). We are also measuring verbal and numerical intelligence as proxy measurement of IQ and analytical ability, respectively. It might be that individuals with high NFC treat deliberate reasoning tasks as a performance test (like an IQ test) and perform at their maximal analytical and reflective capacity. If so, incentives may not increase performance. In contrast, individuals with low NFC might only perform at their maximal ability when an external incentive is given (Sandra and Otto, 2018).

The aim of this study is to investigate the effect time-pressure and incentives have on deliberate reasoning, controlling for an individual's need for cognition, their numerical and verbal intelligence and their perceived effort. By using a within-subject design and measuring NFC, numerical and verbal intelligence as well as subjective effort we assess under which condition best performance in rational thinking can be elucidated.

Research questions and hypotheses

Our first research question is about the effect of time pressure on deliberate reasoning.

1) Will persons with high algorithmic ability perform similar on deliberate reasoning tasks under time pressure than without time pressure (Raoelison et al., 2020)? And will persons with low algorithmic ability perform similar or worse on deliberate reasoning tasks under time pressure than without time pressure?

Our second research question is about the effect of incentives on deliberate reasoning.

2) Will persons with low need for cognition perform better on deliberate reasoning tasks when incentivized then when not incentivized, and will persons with high need for cognition perform similar on deliberate reasoning tasks in non-incentivized and incentivized conditions?

Our third research question looks at perceived effort.

3) Will perceived effort be lower for participants high in NFC?

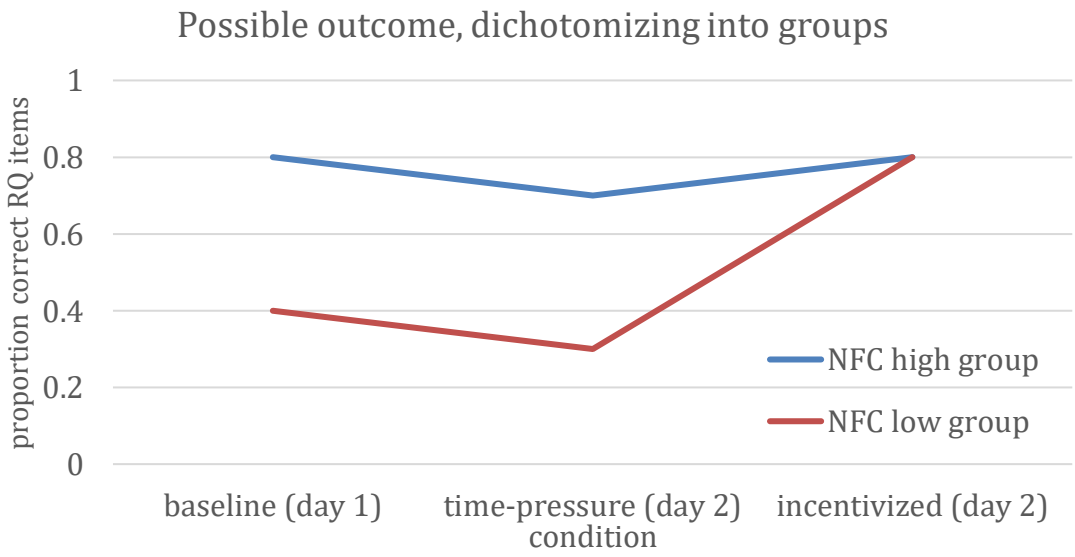

Figure 1: Predictions for the study

Method

Preregistration

This study was pre-registered on the Open Science Framework (http://osf.io/hv7pd/) prior to data creation.

Ethics

This research project has been ethically evaluated and approved by the Department of Psychology´s research ethics committee at UiT, The Arctic University of Norway (see Appendix C).

Participants

The sample size set for this study was 100 participants who complete both sessions of the study. We recruited 136 participants in total prior to exclusion of those who did not meet the inclusion criteria. Of the 136, 84 completed the surveys and were approved. 37 of the participants were women (aged 18- 49) and 46 were men (aged 18-64), one indicated as gender other. The experiment was conducted online, allowing people from several countries (24% from Poland, 13.5% Portugal, 12.5% Italy, 10% England, remaining were from 13 other countries) to participate.

Materials

The experiment was conducted in English and all instructions were in English. For a complete overview of all materials please see Appendix D-J.

Rationality Quotient (RQ) items. We used 14 items from the problem solving and reasoning literature:

- One item from Van Dooren et al., 2004.
- One item from Primi et al., 2016.
- Items two and three from Thomson and Oppenheimer (2016).
- Items 4-6 from Finucane and Gullion (2010).
- One item from Levesque (1986; 1989), as cited in Toplak & Stanovich (2002).
- One item from Wason and Brooks (1979).
- One item from West, Toplak and Stanovich (2008).
- One item from Kahneman and Tversky (1972).
- One item from Smullyan (1978), as cited in Toplak and Stanovich (2002)

- One item from the Wason selection task (Wason, 1966; as cited in Kornreich, Delle-Vigne, Brevers, Tecco, Campanella, Noël, Verbanck & Ermer, 2017).
- One item from Shafir (1994).

Although the degree of difficulty varied in these items, deliberate reasoning was required in all to reach the correct answer. The 14 items were sequentially and randomly presented to all participants. Participants had to provide an answer before proceeding to the next item. There were no time limits for answering the RQ items in session 1.

NASA Task Load Index (N-TLX). The N-TLX was developed by Hart and Staveland (1988; as cited in Hart, 2006). This is a multi-dimensional scale, which consists of six subscales. These six subscales represent a person´s experienced workload in a given task. On a scale ranging from 0 (very low) to 100 (very high), participants rate their perceived a) mental activity, b) physical activity, c) time pressure, d) frustration level, e) overall effort (both mental and physical), and f) satisfaction (Hart, 2006). The participants´ self-reported subjective workload was measured twice in session 1, following the RQ tasks and following the Berlin Numeracy Test and Word sum test. It was also measured twice in Session 2, following the seven RQ items in the time pressure condition, and following the seven RQ items in the reward condition.

Need for cognition scale (NFC). This scale was originally developed by Cacioppo & Petty (1982) and provides a measure of a person's tendency to enjoy, favor or engage in tasks that require cognitive effort, and their underlying intrinsic motivation.

We used the revised short version of this scale with 18 items (Cacioppo, Petty & Kao, 1984). At a five-point Likert scale (1= extremely uncharacteristic of me, 5= extremely characteristic of me), participants are asked to describe to which extent they agree with 18 statements. Half of these 18 statements assess low need for cognition, e.g., "I try to anticipate and avoid situations where there is a likely chance, I will have to think in depth about

something". The other half assesses high need for cognition, e.g., "I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought" (Cacioppo, Petty & Kao, 1984). Test scores can range from 18 to 90. For our data Cronbach's alpha = .868.

Berlin Numeracy Test (BNT). The Berlin Numeracy Test was designed by (Cokely, Galesic, Schulz, Ghazal, & Garcia-Retamero, 2012), and is built on previous works regarding statistical numeracy tools (e.g., Lipkus et al. 2001; Schwartz et al. 1997). This test consists of four mathematical questions and is a measure of one's ability to evaluate statistical numeracy and risk literacy (Cokely et al., 2012). This indicates one's ability to understand any operations of statistical and probabilistic computations, and the ability to make interpretation in an accurate manner and act based on the acquired information (Cokely et al., 2012). The BNT can be completed in several formats (e.g. with paper and pencil, computer adaptive, multiple choice), and for this study we used a computerized multiple-choice format.

The participants were shown one question at a time and the response was forced, so each question had to be answered before moving to the next. There were four answer options: one correct answer, two incorrect answer, and "none of the above" (Cokely et al. 2012). Participants were instructed to use paper and pen to make notes if they wished, but to not use a calculator. For each correct answer one point was awarded, and total score ranged from 0-4.

Wordsum test (WST). This test was used to measure an individual's verbal knowledge (Malhotra et al. 2007). The WST consists of ten words and the participants are to select the most suitable synonym for each word, given five alternative answers to choose from (four possible words, and "I don't know"). All ten items are presented at the same time and the participant must provide answers to all before continuing to the next part of the study. For each correct synonym one point was awarded, yielding a test score ranging from 0-10.

The 5-dimensional curiosity scale (5DC). The five-dimensional model of curiosity is used as a measure of an individual's motivations to avoid or approach new information or new experiences (Kashdan et al., 2018). The five dimensions in this scale are: joyous exploration, deprivation sensitivity, stress tolerance, social curiosity and thrill seeking (Kashdan et al., 2018). Collectively, these dimensions contain 25 question items, five items per dimension. We used only two dimensions for this study, joyous exploration and social curiosity. An example of an item from the joyous dimension would be "I view challenging situations as an opportunity to grow and learn" (Kashdan et al., 2018), and an example from the social curiosity dimension would be "When people quarrel, I like to know what's going on" (Kashdan et al., 2018).

The participants were informed to indicate the extent to which each statement accurately describes them on a scale from 1 (does not describe me at all) to 7 (completely describes me). Response for all statements was required in order to continue to the last part of session 1. An average joyous exploration score and an average social curiosity score was calculated. The data from this measure was not of main interest and will not be discussed further in this thesis.

Demographics. Collected demographic information was gender, age, and highest education on six levels (with five options: not completed high school, completed high school, bachelor's degree or equivalent, master's degree or equivalent, PhD or equivalent).

Filler task. In session 2 between the time pressure and incentivized condition participants had to perform a filler task, taking approximately 6-8 min. This task just served as a break and will not be described further, nor is any data of this task used in the analysis.

Design and recruitment

This experiment uses a mixed design where conditions are the repeated measurement and counterbalancing into group A and group B is the between-group factor. All participants

were exposed to all the conditions. The first session is referred to as baseline and was approximately 3 weeks before session 2.

Representative participants were recruited via Prolific Academic (www.prolific.co). Both recruitment and payment happened through this third part. Politics' policies emphasize that researchers and participants cannot be in personal contact, only through their assigned Prolific email address which ensures anonymity for both researcher and participant. Participants were redirected from this website to Qualtrics to access the tasks and questionnaires for this study. All data are saved and administered in Qualtrics. Participants received ca. £10 pounds after completion. Participation was voluntary and participants could withdraw their consent at any moment.

Procedure Session 1, baseline condition

Participants were invited through Prolific to take part. First, participants read about the purpose, data protection and open science practice and then provided their consent. Next, they were randomly shown the 14 RQ items. There were no time limits for answering those items. Thereafter they filled out the N-TLX, rating perceived effort for the rationality items. Then they filled out the NFC, then the BNT, WST, N-TLX for the BNT and WST, and finally the two subscales of the 5D.

Procedure session 2, time pressure and incentivized condition

Participants were invited back through the Prolific platform by sending emails with the link to session 2 to group A and B participants. Both groups started with the time-pressure condition. Time allowance was calculated by using 80% of the average response time from baseline. The minimum time was 10 seconds, the maximum time was 85 seconds. (See appendix D for set 1 and set 2).

Group A received set 1 whereas group B received set 2. They filled out the N-TLX. They then performed a filler task located on another platform (Jatos) and where then

redirected to the second part of the survey, where both groups received the other set with the instruction that better than baseline performance earns them additional money (group A received set 2, group B received set 1). They then again filled out the N-TLX and were thanked for participation.

Inclusion and exclusion criteria.

Participants had to be at least 18 years old and give consent before taking part in this study. Participants who had a word sum test score < 3 were not invited back to session 2, as this likely indicates low English proficiency and too low verbal ability to understand the items, including RQ items. Participant data was only analyzed if a participant had completed both session 1 testing and session 2 testing.

Sample size rational

We expect a large main effect of NFC on RQ, i.e. correlation of at least .4. We also expected a moderate effect of condition on RQ where the three conditions are 0.5 to 0.7 correlated with each other. Using a type II error of 20% and Type I of 5% G-power (using G*Power 3.1.9.4) (Faul, Erdfelder, Buchner, & Lang, 2009) estimated a sample size (effect size f=.15) of n=73 participants. Since we also measure numerical and verbal intelligence and use those as co-variates, we aimed for n=100. The sample size rational is also influenced by the available funding, as the participant fee set a limit of N=100.

Analysis

In both sessions, we recorded the average response time for the RQ items, percentage of correct responses, ratings in the N-TLX, overall score in the BNT, overall score in the WST, average score for joyous exploration subscale and social curiosity.

To create to equally difficult sets for session 2, we calculated the percentage of participants answering correctly for each item, as well as the response time. The sets were matched to yield similar difficulty and duration (Appendix B provides the two sets). We

counterbalanced the two sets, i.e. participants were randomized into Group A or Group B. Group A received set 1 (seven RQ items) as the time-pressure condition and set 2 (seven RQ items) as incentivized condition. Group B received set 2 as time-pressure condition and set 1 as incentivized condition. In both the time-pressure condition and the incentivized condition in session 2, we recorded percentage correct of the seven RQ items and average response time for the seven RQ items.

To compare the time-pressure to the baseline condition we calculated a difference score. A positive score means that more items got solved correctly in the baseline condition. To compare the incentivized to the baseline condition we calculated a difference score, positive scores mean more items solved correctly at baseline. We used rank correlations and repeated measure analysis of variances. JASP (JASP Team, 2020) was used for data analysis.

Results

Of the 136 invited participants, 6 had to be excluded due to a too low WST score, and 47 did not take part in or completed session 2. That left N=83 valid responses. Descriptive for the NFC, RQ, BNT and WST can be found in Table 1. Of the 83, 36 were female, 46 male and 1 other (Age M = 26.26, SD = 8.963, range from 18 to 64).

Table 1: Descriptive statistics for RQ, NFC, BNT and WST.

	n	*M*	*SD*	*SE*
RQtotal	83	6.843	2.540	0.279
NFC	83	59.530	9.772	1.073
BNT	83	1.807	1.163	0.128
WST	83	6.807	1.721	0.189

Notes. n= number of participants, *M*= mean *score*, *SD*= standard deviation, *SE*= standard error.

First, we checked whether deliberate reasoning was associated with NFC, verbal and numerical ability. A linear regression explained 29.2% of the variance, $F(3,82) = 10.865$, $p < .001$. All three predictors were significant, NFC: $t = 2.955$, $p = .004$, BNT: $t = 2.002$, $p = .049$, WST: $t = 2.523$, $p = .014$. As figure 2 shows, the manipulation was successful, i.e.

participants had fewer correct items in the time-pressure condition. Given these results we proceeded with analyzing our research questions.

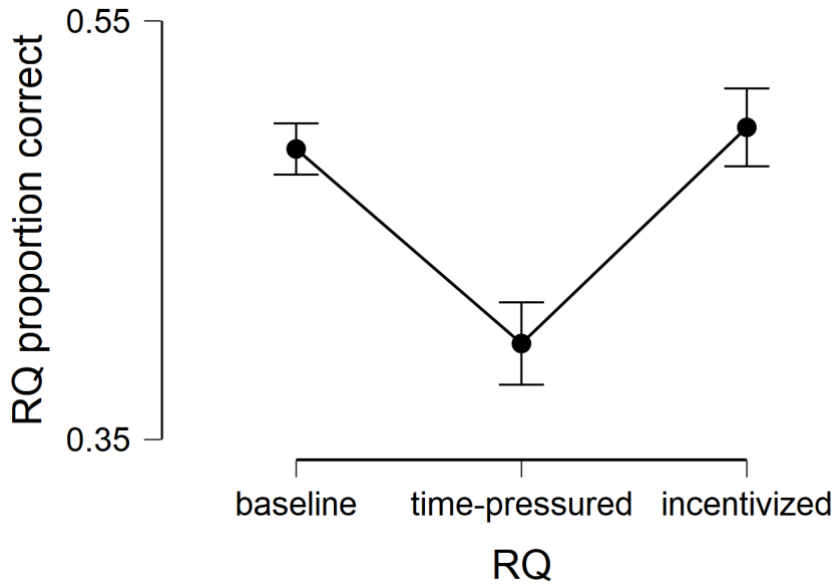

Figure 2: Average proportion of correct items in the three conditions for all 83 participants, error bars denote SEM

Does NFC affect performance under time pressure?

Regarding our first research question, we found some support. On average participants got 49.4% correct at baseline and 39.6% correct under time pressure for the same items (Figure 3).

Using a rmANOVA we found a main effect of session, i.e. proportion correctly answered items was lower in the time pressure than the baseline condition, $F(1,79) = 4.063$, $p = .047$, $\eta^2 = .014$. There was no main effect of NFC, $F(1,79) = 3.670$, $p = .059$, $\eta^2 = .043$, nor a main effect of WST, $F(1,79) < 1$ or BNT, $F(1,79) = 2.18$, $p = .144$, $\eta^2 = .025$. However, there was a significant interaction effect between session and WST, $F(1,79) = 4.357$, $p = .04$, $\eta^2 = .015$, but not for session and NFC, $F(1,79) = 2.93$, $p = .091$, $\eta^2 = .01$ or BNT, $F < 1$.

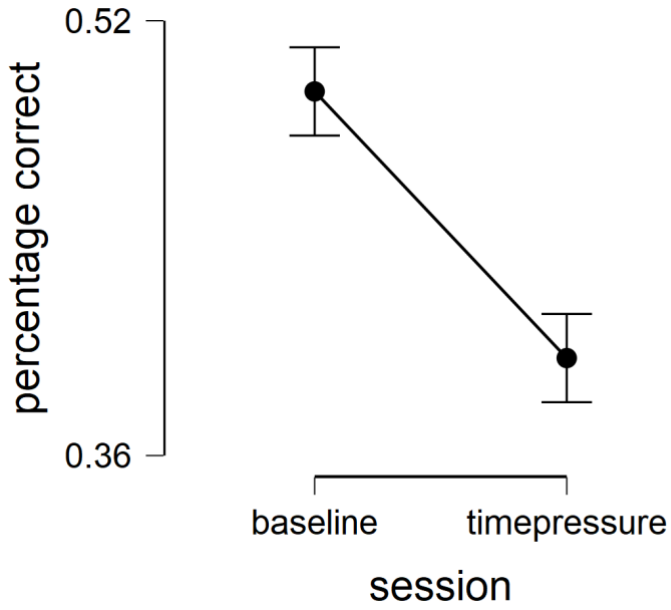

Figure 3: Average RQ score in the time pressure condition compared to baseline. Error bars denote SEM

The follow-up correlation analysis between WST and the difference score was significant, $\tau = .197$, 95% CI[.086, 1], $p = .012$, i.e. the better the verbal ability the better the performance under time pressure. A correlation between BNT and the difference score was not significant, $\tau = .03$, $p = .369$, 95% CI [-.089, 1]. A correlation between NFC and the difference score was not significant, $\tau = .079$, 95% CI[-.039, 1], $p = .169$. This partly confirms research question 1, i.e. verbal ability but not numerical ability influenced how well participants did under time pressure.

Does NFC affect performance under incentivized condition?

There was only a slight difference between baseline (48% correct) and the incentivized condition (50% correct), $F(1,79) = .002$, $p = .968$, $\eta^2 = 0$ (Figure 4). All three co-variates were significant, NFC: $F(1,79) = 5.184$, $p = .026$, $\eta^2 = .053$, BNT: $F(1,79) = 6.984$, $p = .01$, $\eta^2 = .072$, and WST: $F(1,79) = 5.889$, $p = .018$, $\eta^2 = .061$. There was no interaction with session, and all interaction terms had $F < 1$.

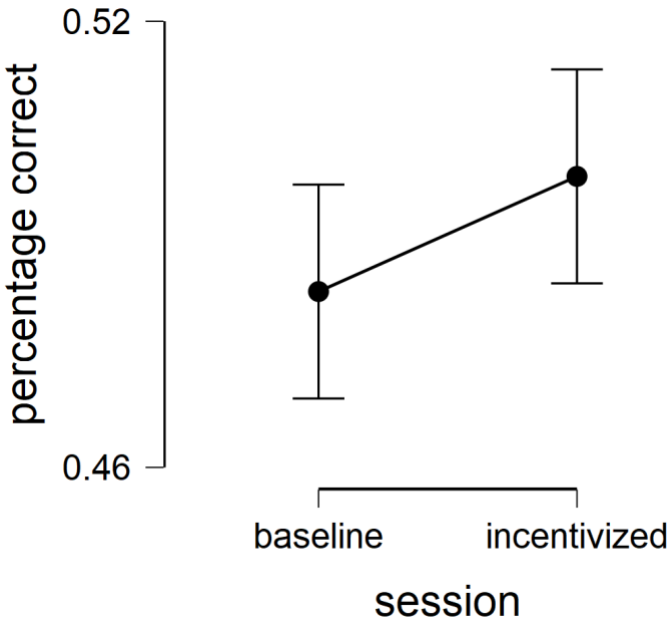

Figure 4: Average RQ score in the incentive condition compared to baseline. Error bars denote SEM

Is perceived effort lower in individuals high in Need for Cognition?

We used a linear regression for RQ at baseline with NFC, and the N-TLX subscales as predictors. The model was significant, $F(6,82) = 5.029$, $p < .001$, and explained 22.8% of the variance (adjusted R^2). NFC was a significant predictor for a person's RQ score, standardized estimate .429, $t = 4.313$, $p < .001$. Of the self-rated task load items, only perceived effort was a significant predictor, standardized estimate -.374, $t = -2.379$, $p = .02$. That is, the better the RQ score, the less effortful the task was rated, and the higher a person's NFC was. This supports the notion of lower perceived effort for deliberate reasoning tasks. We followed this up with a moderator analysis, i.e., would the relationship between NFC and RQ be moderated by perceived effort? The direct path between NFC and RQ was significant, Z= 4.324, p < .001 but the indirect path from NFC via perceived effort to RQ was not, Z = .552, p = .581.

Table 2: Unstandardized regression coefficient (B), standard error (SE), Standardized regression coefficient (β), t-values, p-values and confidence intervals for the five N-TLX items and NFC, outcome is RQ score.

						95% CI	
	B	*SE*	*β*	*t*	*p*	Lower	Upper
1 (Intercept)	1.400	1.703		0.822	0.414	-1.991	4.791
N-TLX1.mental	-0.007	0.016	-0.066	-0.452	0.653	-0.039	0.024
N-TLX2.physical	0.002	0.010	0.025	0.236	0.814	-0.017	0.022
N-TLX3.time	0.006	0.012	0.066	0.523	0.603	-0.018	0.030
N-TLX4.frustration	0.019	0.011	0.225	1.726	0.088	-0.003	0.042
N-TLX5.effort	-0.040	0.017	-0.374	-2.379	0.020	-0.074	-0.007
NFC	0.112	0.026	0.429	4.313	< .001	0.060	0.163

Note. There was no issue with collinearity.

Discussion

Our results show that deliberate reasoning is associated both with need for cognition and intelligence or algorithmic ability. We further found that verbal ability, not numerical ability, is associated with deliberate reasoning during time-pressure. Our baseline assessment, no time pressure, no incentivization, replicates previous findings showing that NFC and intelligence contribute to deliberate reasoning (Fleischhauer et al., 2010; Hill et al., 2013).

As predicted, there was a decrease in performance under time pressure compared to the baseline condition. We predicted that participants with low algorithmic ability and motivation (NFC) would perform worse under the time-pressure condition, and that participants high in algorithmic ability and high motivation (NFC) would perform similarly or slightly worse under time pressure. Results showed no main effect of NFC, verbal ability (WST) or numeric ability (BNT), but a significant interaction effect between session (baseline and time pressure condition) and verbal ability was revealed. This indicates that individuals

high in verbal ability, irrespective of their motivation and numerical ability, performed well during time-pressure, thereby partly confirming our hypothesis. It has previously been found that the BNT and cognitive reflection test (CRT) are related (Cokely et al., 2012), which did not play out in our data. A probable reason could be that we used more than just CRT items, we also used items that require more than numeracy to solve. For example, the knight and knave problem (Appendix C) involves no numbers.

The findings in this study further supports the findings from the heuristic literature that have shown an increase of using heuristics under time pressure when processing information (Suri & Monroe, 2003). Even highly motivated individuals given limited time show a greater likelihood that their information processing happens heuristically rather than systematically (Suri & Monroe, 2003). Although this finding is from studies on the processing of persuasive messages, a parallel can be draw in reasoning under time pressure.

This finding also reflects the well-known system 1 and system 2 processing of thought when deliberating and making decisions. Since heuristic processing is the fast and intuitive process that demands less cognitive effort, it makes sense that this would be applied under time pressure, i.e. since an initially demanding task becomes even more demanding to solve under this circumstance and relying on shortcuts would be the easier path. For persons with limited cognitive capacity the amount of demand required elevates when there is time pressure, and the more cognitive effort required, the less likely that deliberation will occur. Evan and Curtis-Holmes (2005), who showed that persons have less logical accuracy and more belief bias when time is limited, are also in support of the current findings. This is again in line with dual-process theory. Furthermore, De Neys (2006) also revealed the same results, though this study used task load and not time pressure per se.

As mentioned, motivation and ability are factors that have significant impact on whether or not deliberate reasoning strategies are expended. For participants lacking

understanding of English, i.e., they have some but not sufficient verbal ability, solving the reasoning items might have been easier when there was no time pressure and as this could give them a better opportunity to really understand the question first. Otherwise, their answer will more likely be rushed and their answers will more likely be based on heuristics. Given that they are motivated to solve the task and they have the ability to solve this kind of tasks, their lack of ability to understand the question will rapidly decrease their test score.

It is encouraging to compare our finding to that of Raoelison et al. (2020) who argue so well that people higher in cognitive capacity do not inevitably deliberate or reason better but perform better simply because their intuition is more likely to be correct, which means they do not have to spend time correcting a wrong intuition. Our result may be explained through this so-called "smart intuitor" view. Persons higher in cognitive capacity might not have experienced much effect of time-pressure, because they arrive at the answer right away, i.e., without having to correct an intuitively wrong answer before finely arriving at the correct one. This may also explain why we found no effect of numeracy.

There was only a slight effect of incentives in the current study. All co-variates had a significant effect and were predictors for how many items were answered correctly. The small difference in performance between baseline and incentive conditions may be due to our sample of participants. The participants are recruited from a website that helps university research studies, and because the reward for participation here is quite small, it is conceivable that participants are doing this for more than just the money, i.e., they may have an interest in research and want to make a helpful contribution. Furthermore, it is also conceivable that people on this website have participated in several studies and thus have been exposed to similar reasoning tasks, and perhaps even to the extent where they are able to give a somewhat intuitively correct answer (smart intuitor view). There were participants who reported that they were familiar with some of the items. Although it is likely to think that

prior knowledge might increase final scores on the reasoning items, it has been shown that previous knowledge of the CRT items does not decrease its predictive validity (Bialek, Pennycook, 2008). From six studies with a sample of approximately 2500, Bialek and Pennybrook (2008) did not find an instance where the predictive power of the CRT decreased.

Our finding supports to a degree what Camerer & Hogarth (1999) reviewed. In some cases, incentives do not increase performance either because the task is too hard and added effort would not help, or because the participant is not motivated by the incentive. In our case it can be rationalized those participants are rather driven by an intrinsic motivation to just do well. With this in mind, a value-based framework of effort may further offer an explanation for these results. As the value-based framework states, individuals differ in subjective evaluation of the costs and benefits of engaging in deliberation (Kool et al, 2010; Kool et al., 2017). In light of this framework, it can be argued that our participants utilized cognitive effort as a result of their intrinsic motivation. Moreover, when the intrinsic motivation to engage in cognitively demanding tasks is high, a person will also have the needed ability to do them. Motivation did also prove to be a predictor on how well participants did on the reasoning items, particularly without time pressure. This corroborates the findings of Sandra and Otto (2018) who also found that persons higher in intrinsic motivation were not as responsive to monetary incentives as those with lower intrinsic motivation.

How much workload the participants experienced in the reasoning tasks (N-TLX) was measured at baseline without time-pressure because they are most naïve in this condition. As predicted, perceived effort was lower for persons higher in Need for cognition. Out of the five items in the N-TLX, it was only perceived effort that was significant. NFC as a covariate was also significant. That is, persons with higher NFC, had better RQ score, and also perceived the tasks as less effortful. This finding is in accordance with Mækelæ and Pfuhl (2019) who also found a relationship between perceived effort and deliberate reasoning.

Limitations

There are several limitations for our study. One possible limitation is the time limit we used in the time pressure condition. We use the average total time spent on solving the 14 RQ items in session 1, divide it by 2 and subtracted 20% from it for a whole number value (in minutes). Other reduction times might give slightly different results.

Another possible limitation regards the question of if we used enough RQ items. These selected items are regularly debated to be suitable measures of deliberate reasoning (e.g., Toplak & Stanovich, 2002; West, Toplak & Stanovich, 2008; Finucane & Gullion, 2010), but more items could provide even more accuracy. However, we wanted to keep the study as short as possible while including all the necessary tests, because of our budget limit and also to get a larger sample.

Further, it was pre-registered and rationalized that the study needed 100 participants, which we did not achieve as we had a total of 84 who took part in all conditions. As for most research a larger sample size would be more sufficient and provide more precision.

Conclusion

There are numerous factors that influence people's willingness to exert effort and deliberate in cognitively challenging situations and tasks. In this study we examined if time-pressure and incentives would have an impact on deliberate reasoning, and if the participants need for cognition, their verbal and numeric abilities and how difficult they perceived the reasoning tasks also influenced their scores. Our findings show that how well you do without time-pressure depends mostly on your level of NFC and your intelligence, with verbal intelligence being a stronger predictor than numerical intelligence. This is the main finding and the new contribution of this study. Despite these findings, caution must be taken in making conclusions. There are several factors which may have had an impact on the results and further investigations are necessary.

References

Amabile, T., M., Hill, K., G., Hennessey, B., A., & Tighe, E., M. (1994). The Work Preference Inventory: Assessing intrinsic and extrinsic motivational orientations. *Journal of Personality and Social Psychology, 66*(5), 950-967. doi:10.1037/0022-3514.66.5.950

Basel, J. S., & Brühl, R. (2013). Rationality and dual process models of reasoning in managerial cognition and decision making. *European Management Journal, 31*(6), 745-754. doi:10.1016/j.emj.2013.07.004

Bialek, M., & Pennycook, G. (2018). The cognitive reflection test is robust to multiple exposures. *Behavior Research Methods, 50(5),* 1953-1959. doi:10.3758/s13428-017-0963-x

Bonner, S. E., & Sprinkle, G. B. (2002). The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research. *Accounting, Organizations and Society, 27*(4), 303-345. doi:10.1016/S0361-3682(01)00052-6

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*(1), 116-131. doi:https://doi.org/10.1037/0022-3514.42.1.116

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48(3),* 306–307. doi:10.1207/s15327752jpa4803_13

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional Differences in Cognitive Motivation. *Psychological Bulletin*, *119*(2), 197–253. doi:10.1037/0033-2909.119.2.197

Camerer, C. F., & Hogarth, R. M. (1999). The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework. *Journal of Risk and Uncertainty, 19*(1/3), 7–42. doi:10.1023/A:1007850605129

Cokely, E. T., Galesic, M., Schulz, E.,, Ghazal, S., & Garcia-Retamero, R. (2012). Measuring Risk Literacy: The Berlin Numeracy Test. *Judgment and Decision Making*, *7*(1), 25–47

De Neys, W. (2006). Automatic–Heuristic and Executive–Analytic Processing during Reasoning: Chronometric and Dual-Task Considerations. *Quarterly Journal of Experimental Psychology*, *59*(6), 1070-1100. doi: 10.1080/02724980543000123

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

Dixon, M. L., & Christoff, K. (2012). The Decision to Engage Cognitive Control Is Driven by Expected Reward-Value: Neural and Behavioral Evidence. *PloS one, 7*(*12),* E51637 doi:10.1371/journal.pone.0051637

Duckworth, A. L., Quinn, P. D., Lynam, D. R., Loeber, R., & Stouthamer-Loeber, M. (2011). Role of test motivation in intelligence testing. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(19), 7716–7720. doi:10.1073/pnas.1018601108

Dunn, T. L., Lutes, D. J., & Risko, E. F. (2016). Metacognitive evaluation in the avoidance of demand. *Journal of experimental psychology. Human perception and performance*, *42*(9), 1372–1387. doi:10.1037/xhp0000236

Evans, J. S. B. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology*, *75*(4), 451-468. doi: https://doi.org/10.1111/j.2044-8295.1984.tb01915.x

Evans, J. S. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*, *59*(1), 255-278. doi:10.1146/annurev.psych.59.103006.093629

Evans, J. S. B. T., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & reasoning, 11*(4), 382-389. doi:10.1080/13546780542000005

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods, 41*(4), 1149-1160.

Finucane, M. L., & Gullion, C. M. (2010). Developing a tool for measuring the decision-making competence of older adults. Psychology and Aging, 25(2), 271-288. doi:10.1037/a0019106

Fleischhauer, M., Enge, S., Brocke, B., Ullrich, J., Strobel, A., & Strobel, A. (2010). Same or different? Clarifying the relationship of need for cognition to personality and intelligence. *Personality and Social Psychology Bulletin*, *36*(1), 82-96. doi:10.1177/0146167209351886

Frank, M. J., Cohen, M. X., & Sanfey, A. G. (2009). Multiple systems in decision making: A neurocomputational perspective. *Current Directions in Psychological Science : a Journal of the American Psychological Society*, *18*(2), 73–77. doi:10.1111/j.1467-8721.2009.01612.x

Gonthier, C., & Roulin, J.-L. (2020). Intraindividual strategy shifts in Raven's matrices, and their dependence on working memory capacity and need for cognition. *Journal of Experimental Psychology: General, 149*(3), 564-579. doi:10.1037/xge0000660

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 904-908. doi:10.1177/154193120605000909

Hill, B. D., Foster, J. D., Elliott, E. M., Shelton, J. T., McCain, J., & Gouvier, W. D. (2013). Need for cognition is related to higher general intelligence, fluid intelligence, and

crystallized intelligence, but not working memory. *Journal of Research in Personality, 47*(1), 22-25. Doi:10.1016/j.jrp.2012.11.001

Hull, C. L. (1943). *Principles of behavior* (Vol. 422): <new York: Appleton-century-crofts.

Jebb, A. T., Saef, R., Parrigon, S., & Woo, S. E. (2016). The need for cognition: Key concepts, assessment, and role in educational outcomes. In *Psychosocial Skills and School Systems in the 21st Century* (pp. 115-132). Springer International Publishing. doi: 10.1007/978-3-319-28606-8_5

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*(9), 697-720. doi:10.1037/0003-066X.58.9.697

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and biases: The Psychology of Intuitive Judgment* (pp. 49-81). Cambridge University Press. doi:10.1017/CBO9780511808098.004

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive psychology, 3*(3), 430-454. doi:10.1016/0010-0285(72)90016-3

Kao, D. T. (2011). Message sidedness in advertising: The moderating roles of need for cognition and time pressure in persuasion. *Scandinavian Journal of Psychology, 52*(4), 329-340. doi: 10.1111/j.1467-9450.2011.00882.x

Kashdan, T. B., Stiksma, M. C., Disabato, D. J., McKnight, P. E., Bekier, J., Kaji, J. & Lazarus, R. (2018). The five-dimensional curiosity scale: Capturing the bandwidth of curiosity and identifying four unique subgroups of curious people. *Journal of Research in Personality, 73*, 130-149. doi:10.1016/j.jrp.2017.11.011

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General, 139*(4), 665-682. doi:10.1037/a0020198

Kool, W., Shenhav, A., & Botvinick, M. M. (2017). Cognitive control as cost-benefit decision making. In *The Wiley Handbook of Cognitive Control* (pp. 167–189). John Wiley & Sons, Ltd. doi:10.1002/9781118920497.ch10

Kornreich, C., Delle-Vigne, D., Brevers, D., Tecco, J., Campanella, S., Noël, X., Verbanck, P., & Ermer, E. (2017). Conditional Reasoning in Schizophrenic Patients. *Evolutionary Psychology*. doi:10.1177/1474704917721713

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. *Psychological Review, 118*(1), 97-109. doi:10.1037/a0020762

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). Gen- eral performance on a numeracy scale among highly- educated samples. *Medical Decision Making, 21*, 37– 44. doi:10.1177/0272989X0102100105

Mækelæ, M.J., & Pfuhl, G. (2019). Correction: Deliberate reasoning is not affected by language. PloS One, 14(3), e0213609–e0213609. doi:10.1371/journal.pone.0213609

Malhotra, N., Krosnick, J. A., & Haertel, E. (2007). The psychometric properties of the GSS Wordsum vocabulary test. *GSS Methodological Report, 11*, 1-63.

Mussel, P. (2010). Epistemic curiosity and related constructs: Lacking evidence of discriminant validity. *Personality and Individual Differences, 49*(5), 506-510. doi:10.1016/j.paid.2010.05.014

Neumann, F. (2015). A Cognitive Psychology Perspective on Individual Knowledge Creation. In *Analyzing and Modeling Interdisciplinary Product Development* (pp. 79-87): Springer. doi:10.1007/978-3-658-11092-5_4

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive psychology, 80*, 34-72. doi:10.1016/j.cogpsych.2015.05.001

Powell, C., Nettelbeck, T., & Burns, N. R. (2016). Deconstructing intellectual curiosity. *Personality and Individual Differences*, *95*, 147–151. doi:10.1016/j.paid.2016.02.037

Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., & Hamilton, J. (2016). The development and testing of a new version of the cognitive reflection test applying item response theory (IRT). *Journal of Behavioral Decision Making*, *29*(5), 453-469. doi:10.1002/bdm.1883

Raoelison, M., Thompson, V. A., & De Neys, W. (2020). The smart intuitor: Cognitive capacity predicts intuitive rather than deliberate thinking. *Cognition, 204*, 104381. doi:10.1016/j.cognition.2020.104381

Rucker, D. D., & Petty, R. E. (2006). Increasing the effectiveness of communications to consumers: Recommendations based on elaboration likelihood and attitude certainty perspectives. *Journal of Public Policy & Marketing, 25*(1), 39-52. doi:10.1509/jppm.25.1.39

Rudolph, J., Greiff, S., Strobel, A., & Preckel, F. (2018). Understanding the link between need for cognition and complex problem solving. *Contemporary Educational Psychology*, *55*, 53–62. doi:10.1016/j.cedpsych.2018.08.001

Sandra, D. A., & Otto, A. R. (2018). Cognitive capacity limitations and Need for Cognition differentially predict reward-induced cognitive effort expenditure. *Cognition, 172*, 101-106. doi:10.1016/j.cognition.2017.12.004

Schmidt, L., Lebreton, M., Cléry-Melin, M.-L., Daunizeau, J., & Pessiglione, M. (2012). Neural mechanisms underlying motivation of mental versus physical effort. *PLoS Biology, 10*(2), E1001266. doi:10.1371/journal.pbio.1001266

Shafir, E. (1994). Uncertainty and the difficulty of thinking through disjunctions. *Cognition, 50*(1), 403-430. doi:10.1016/0010-0277(94)90038-8

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience, 40*, 99-124.

Sherman, J. W., Gawronski, B., & Trope, Y. (2014). *Dual-process theories of the social mind*: Guilford Publications.

Stanovich, K. E. (2009). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory? *In two minds: Dual processes and beyond*, 55-88. doi:10.1093/acprof:oso/9780199230167.003.0003

Stanovich, K. E. (2016). The comprehensive assessment of rational thinking. *Educational Psychologist, 51*(1), 23-34. doi:10.1080/00461520.2015.1125787

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *The Behavioral and Brain Sciences, 23*(5), 645-665. doi:10.1017/S0140525X00003435

Stanovich, K. E., & West, R. F. (2014). The Assessment of Rational Thinking:IQ ≠ RQ. *Teaching of Psychology, 41*(3), 265-271. doi:10.1177/0098628314537988

Stanovich, K. E., West, R. F., & Toplak, M. E. (2011). Intelligence and rationality. In R. J. Sternberg & S. B. Kaufman (Eds.), Cambridge handbooks in psychology. The Cambridge handbook of intelligence (p. 784–826). Cambridge University Press. doi:10.1017/CBO9780511977244.040

Stanovich, K. E., West, R. F., & Toplak, M. E. (2016). The rationality quotient: Toward a test of rational thinking: MIT press.

Suri, R., & Monroe, K. B. (2003). The Effects of Time Constraints on Consumers' Judgments of Prices and Products. *Journal of Consumer Research, 30*(1), 92-104. doi:10.1086/374696

Thompson, V. A., & Johnson, S. C. (2014). Conflict, metacognition, and analytic thinking. *Thinking & reasoning, 20*(2), 215-244. doi:10.1080/13546783.2013.869763

Thompson, V. A., Pennycook, G., Trippas, D., & Evans, J. S. B. (2018). Do smart people have better intuitions? *Journal of Experimental Psychology. General, 147*(7), 945-961. doi:1037/xge0000457

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

Toplak, M. E., & Stanovich, K. E. (2002). The domain specificity and generality of disjunctive reasoning: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology*, *94*(1), 197. doi:10.1037/0022-0663.94.1.197

Van Dooren, W., De Bock, D., Hessels, A., Janssens, D., & Verschaffel, L. (2004). Students' Overreliance on Proportionality: Evidence from Primary School Pupils Solving Arithmetic Word Problems. *International Group for the Psychology of Mathematics Education*.

Wason, P. C., & Brooks, P. G. (1979). THOG: The anatomy of a problem. *Psychological Research, 41*(1), 79-90. doi:10.1007/BF00309425

West, R. F., Toplak, M. E., & Stanovich, K. E. (2008). Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions. *Journal of Educational Psychology, 100*(4), 930-941. doi:10.1037/a0012842

Appendix A

Consent Form

Who can participate?
All over the age of 18 and fluent in English.

What is the study about?
This study looks at human decision-making. In two sessions, separated by approximately 2 months, we are interested in how you solve a set of given tasks and whether there is any change in your decision-making between the two sessions. We hope you are willing to participate in both sessions. Payment will occur when you have completed both sessions, and you can receive up to £10.

Session 1: You will fill out and answer a survey composed of some short reasoning tasks, a word game and a few questionnaires. This will take ca. 35-40 minutes to complete.

Session 2: This is very similar to session 1 but instead of a word game there is a short information sampling task. Testing should take less than 30 min, however some participants may prefer to spend a bit more time.

What information will be collected?
The information collected are: demographic information (gender, age and educational background), response time and task performance. Participants who take part in both sessions have to provide their assigned Prolific ID to be invited back for participation in session 2. Contact between the participant and the project manager can only happen within Politics' email system. As Prolific emails and IDs consists of 24 randomized alphanumerical characters, confidentiality is guaranteed.

What will happen to the collected data?
Collected information will only be used as described for the purpose of this experiment. All information will be processed without name or any identification number, or other direct information that will identify you. You will complete the surveys using your unique Prolific ID - which is removed from any reports - and which means that you will not be identified in the data analysis and reports, or possible publication of the results.
You have the right to request insight to the information collected about you. You may also request information about how you did on the tasks. In both these instances, you can contact the researcher through Prolific. The researcher can find your information by using your Prolific ID. If you want to withdraw you can request deletion of all your data.
Anonymized data will be made available on the Open Science Framework, as we adhere to Open Science Policies. The project managers of this project are responsible for ensuring that information about participants is treated securely.

Timeframe for the study
Data collection is carried out from December 2020 to February 2021.

Possible benefits and disadvantages
There are no direct benefits other than the payment you receive via Prolific. There are no disadvantages of taking part.

Voluntary participation and possibility to withdraw consent (opt-out)
Participation in the study is voluntary. If you wish to take part in this study, you will need to give informed consent by clicking "Yes, I am willing to participate" on this page. You can, at any given time and without giving a reason, withdraw from participation. However, Prolifics´ policy indicates that you can withdraw from the study (1) when you are active on the study and (2) after the study has been submitted and is still awaiting review. This means that you can demand that your information is erased, and your test results deleted from the study, unless they have already been reviewed and analyzed. If you decide to withdraw from participation you will not receive payment. If you have any questions regarding this, you can contact Prolific https://participant-help.prolific.co/hc/en-gb.

Transfer of information to others
We adhere to Open Science, which means that we share anonymized data with other researchers to improve the quality of scientific research by e.g. replicate our analysis or build on this data for other studies. By agreeing to participate in the study, you are also consenting that the anonymized data is made available in a data repository. The data this refers to only applies to unidentifiable data from the task performances.

Finance
You will receive at least £8,20 after completing both sessions and have the opportunity to earn up to approximately £10.

Project managers
The Department of Psychology at UiT The Arctic University of Norway is responsible for the implementation of this project. Please contact the project managers if you have any questions.
Fozia A. Abukar, UiT The Arctic University of Norway, Tromsø (+47 98130832; fab003@uit.no) or
Prof. Gerit Pfuhl, UiT The Arctic University of Norway, Tromsø (+47 776 46 276; gerit.pfuhl@uit.no)

I AM WILLING TO PARTICIPATE IN THE RESEARCH PROJECT
1. I affirm that I have read and understood the study information.
2. I know that I can contact the corresponding investigator anytime if I have questions about the study (fab003@uit.no).
3. I know that participation in this study is voluntary and that I can withdraw my consent without providing reasons. I am aware that deletion of my data is possible until data collection is completed.
4. I declare that I participate voluntarily in this study.

Are you willing to participate in the research project?

- o Yes, I am willing to participate; all of the statements 1-4 apply to me.
- o No, I am not willing to participate; one or more statement does not apply to me.

Appendix B

IRB- Ethics Application

Default, time-pressured and incentivized deliberate reasoning
Application for ethical approval by the Institutional Ethics Committee, IPS, UiT The Arctic University of Norway

• **Project title:** Default, time-pressured and incentivized deliberate reasoning
• **Investigators:**

1. Name: Fozia Aweys Abukar

Academic degree: Bachelors

Email: fab003@uit.no

Phone: +47 98130832

2. Name: Gerit Pfuhl

Academic degree: PhD

Position: Professor

Workplace: Department of Psychology, UiT

Email: gerit.pfuhl@uit.no

Phone: +47 77646276

• **Expected starting date of the project:** 01. 10. 2020
• **Expected ending date of the project:** 30. 06. 2021
• **Are collaborators from other institutions involved in the project?** No.
• **Is the project related to other research projects already approved by an ethical committee?** Yes
• **Is the project part of an education or doctorate?** Yes.
• **Does the project involve drug testing?** No.
• **Does the project involve collecting new health-related data?** No.
• **Does the project involve collecting biological material?** No.
• **Number of research participants:** We plan to recruit at least 100 participants.
• **Recruitment of research participants:** Via prolific for the main study, via snowballing for pilot testing
• **Will written consent be obtained from all participants?** Yes.
• **Inclusion criteria:** signed informed consent, participants must be 18 years +
• **Exclusion criteria:** Psychology students who have knowledge of the concepts in this study to be excluded. i.e. first-year psychology students who have yet not learned about system 1 and system 2 processes can also participate.
• **Describe how participants, the society and/or the scientific community might benefit from the results of the research project.**

Rationality is, in contrast to intelligence, a skill we all can acquire. Individual differences in deliberate reasoning are known, but these are based on unrestricted and un-incentivized testing. Individual differences may lessen or increase under time-pressure or reward condition. This is important as many societal challenges require citizens to make careful decisions. Nudges like rewards may appeal for some but not others, time pressure may hamper or not performance. By using a within-subject design and measuring “need for

cognition" a stable personality trait, we hope to identify under which condition best performance in rational thinking can be elucidated.

• Describe the potential disadvantages of participating in the research project. What measures will be taken to minimize the impact of these factors?

There are no expected disadvantages of participating in this research project.

• Fees for project manager / co-workers: None.

• Compensation for research participants: Approximately 120 NOK per participant.

• Any conflicts of interest for the project manager/co-workers: None

• Are there restrictions on publication of results of the project? No

• In what form will personally identifiable information and collected data be used and kept?

No personally identifiable information will be collected. Collected data will not be saved in any way that will identify the participants.

• Plan for publishing the results and/or the obtained information and potential further use of results, data, biological material.

The results shall be published in a peer-reviewed journal.

• Describe the academic and scientific rationale for the selection of data collection:

The purpose of this research project is to investigate the effect time-pressure and incentives have on deliberate reasoning. A participant's "Need for cognition" may moderate the performance under the default, time-pressure and reward condition. Identifying how to increase deliberate reasoning is important for societal challenges, e.g. immunity against fake news.

• Summary of the project:

Background

Individuals differ in their tendency to engage in processes of reasoning, which can reliably be measured with the "Need for cognition" scale (NFC, Cacioppo & Petty, 1982). Individuals who have a high need for cognition pursue effort rather than avoiding it, while individuals with low need for cognition prefer to rely on heuristics and mental shortcuts to avoid exerting effort (Cacioppo, Petty, Feinstein & Jarvis, 1996).

Previous studies have found a relationship between NFC and deliberate reasoning (Cacioppo & Petty, 1982), and NFC has also been shown to be associated with intelligence (Cacioppo & Petty, 1982; Fleischhauer et al., 2010; Hill et al., 2013). However, it is not known whether individuals with high NFC treat deliberate reasoning tasks as a performance test (like an IQ test) and perform at their maximal analytical and reflective capacity. In contrast, individuals with low NFC might only perform at their maximal ability when an external incentive is given, e.g. monetary rewards, or achieving a good grade (Sandra and Otto, 2018).

Methods

The experiment will be conducted online and in English. This will be a within-subject experiment, and all participants will be exposed to all the conditions, however the items are counterbalanced on session 2.

We will be using Prolific, a database specialized for recruiting representative samples. Both recruitment and payment will happen through this third part. Prolific´s policies emphasize that

researchers and participants cannot be in personal contact, only through their assigned Prolific email address which ensures anonymity for both researcher and participant.

Before starting the task, a consent form will be provided with the purpose of the study, and data management including confidentiality explained. It will also be made clear that participation is voluntary and that participants can abort at any time.

The participants will perform two similar studies; the first in Autumn 2020, the second in early. The first testing is identical for all. The second testing will administer the same order of condition but will counterbalance the items; i.e. the 14 RQ items from session 1 will be split in two sets with equal difficulty.

Session 1 testing material:

1. 14 RQ items from the reasoning and heuristic literature (e.g. Mækelæ & Pfuhl, 2019).
2. N-TLX (Hart, 2006)
3. NFC scale (Cacioppo & Petty, 1982).
4. Berlin Numeracy Task, 4 items (Cokely et al., 2012).
5. Word sum test, 10 items
6. The 5-dimensional curiosity scale, using 2 out of 5 dimensions (Joyous exploration, social curiosity) (Kashdan et al., 2018)

Session 2 testing material:

1. Half of the RQ items from session 1 with time-pressure, randomly receiving set 1 or set 2 (group, counterbalancing)
2. Dice task (a novel information sampling task) – not analysed here
3. Half of the RQ items with reward, complementary set
4. N-TLX

Testing on session 1 is expected to last around 30 min, testing on session 2 may last 20-30 min depending on how long a person spends in the incentive condition.

Appendix C

Ethics application approval

UiT Norges arktiske universitet

Institutt for psykologi
Arkiv ref.: 2017/1912
Dato: 16.09.2020
offl. § 26,4

Prof Gerit Pfuhl
Dept of Psychology
UiT

Ethical evaluation of research project

Dear colleague,
Your research project:
« Default, time-pressured and incentivized deliberate reasoning »

has been ethically evaluated by the Department of Psychology's internal research ethics committee based on received information.Your project, has been exempted from full review and thus approved by the Department of Psychology's Research Ethics Committee.

Sincerely yours, on behalf of the Committee

Martin Eisemann, Professor

Appendix D

RQ items

The name of the items and the answers were not shown to participants.

Introduction:

You will now answer some questions and solve some tasks. Answer to the best of your ability and select the answer that fits you best. Some tasks will be difficult, others will be easier, do your best to solve them.

Item 1: (CRT2_H7) from Van Dooren et al, 2004.

Ellen and Kim are running around a track. They run equally fast, but Ellen started later. When Ellen has run 5 laps, Kim has run 15 laps. When Ellen has run 30 laps, how many has Kim run? ____ laps.

Correct answer: 40

Item 2: (CRT2_ H6 from Thomson and Oppenheimer (2016).

A farmer had 15 sheep and all but 8 died. How many are left?

Correct answer = 8

Item 3: (CRT2_H5) from Thomson and Oppenheimer (2016).

Emily's father has three daughters. The first two are named April and May. What is the third daughter's name? _____

Correct answer: Emily

Item 4: (CRT2_H4) from Finucane and Gullion (2010).

If it takes 2 nurses 2minutes to measure the blood pressure of 2 patients, how long would it take 200 nurses to measure the blood pressure of 200 patients? _____minutes.

Correct answer: 2

Item 5: (CRT2_H3) from Finucane and Gullion (2010).

Soup and salad cost $5.50 in total. The soup costs a dollar more than the salad. How much does the salad cost? _____$

Correct answer: 2.25

Item 6: (CRT2_H2) from Finucane and Gullion (2010).
Sally is making sun tea. Every hour, the concentration of the tea doubles. If it takes 6hours for the tea to be ready, how long would it take for the tea to reach half of the final concentration?
_____ hours.
Correct answer: 5

Item 7: (CRT2_H1) from Primi et al., 2016.
In an athletics team, tall members are three times more likely to win a medal than short members. This year the team has won 60 medals so far. How many of these have been won by short athletes?
Correct answer = 15

Item 8: (CRT2_nH1) from Levesque (1986;1989), as cited in Toplak & Stanovich (2002).
There are 5 blocks in a stack. The blocks are either green or not green. The second block from the top is green, and the fourth block from the top is not green. Is there a green block directly on top of a non-green block?
Yes (correct answer)
No
Cannot be determined

Item 9: (CRT2_nH2) from Wason and Brooks (1979).
Imagine you are given 4 shapes with different colours: a red diamond, a blue diamond, a red circle, and a blue circle. You are to assume that I have written down one of the colors (blue or red) and one of the shapes (diamond or circle).Now read the following rule carefully: If, and only if, any of the designs includes either the color I have written down, or the shape I have written down, but not both, then it is called a "THOG." I will tell you that the Blue Diamond is a THOG. Is there another THOG?
Yes (correct answer)
No
Cannot be determined

Item 10: (CRT2_nH3) from West, Toplak and Stanovich (2008).
It is known that 1 dollar out of every 10,000 is counterfeit. Imagine a money-changing machine that rejects real dollar bills 5 out of every 100 times when it changes money.

However, it always rejects bills when they are counterfeit. If this machine rejects your dollar bill, what is the probability (expressed as a percentage ranging from 0% to 100%) that your bill is counterfeit? Choose the best answer.
Less than 1% (correct answer)
About 5%
About 50%
About 95%
More than 95%

Item 11: (CRT2_nH4) from Kahneman and Tversky (1972).
A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. As you know, about 50% of all babies are boys. However, the exact percentage varies from day to day. Sometimes it may be higher than 50%, sometimes lower. For a period of 1 year, each hospital recorded the days on which more than 60% of the babies born were boys. Which hospital do you think recorded more such days?
The larger hospital
The smaller hospital (correct answer)
About the same (that is, within 5% of each other)

Item 12: (CRT2_nH5) from Smullyan (1978), as cited in Toplak and Stanovich (2002).
Imagine that there are two inhabitants of a fictitious country: A and B, each of whom is a either a knight or a knave. Knights always tell the truth. Knaves always lie.
Person A says: "We are both knaves."

(CRT2_nH51)
What is person A?
A knight
A knave (correct answer)
Cannot be determined

(CRT2_nH52)
What is person B?
A knight (correct answer)

A knave
Cannot be determined

Item 13: (CRT2_nH6) from Wason, 1966; as cited in Kornreich et al., 2017.
Teenagers who don't have their own cars usually end up borrowing their parents' cars. In return for the privilege of borrowing the car, the Goldstein's have given their kids the rule, "If you borrow my car, then you have to fill up the tank with gas."
You are interested in seeing whether any of the Goldstein teenagers broke this rule.
These cards represent four of the Goldstein teenagers. Each card represents one teenager. One side of the card tells whether or not a teenager has borrowed the parents' car on a particular day, and the other side tells whether or not that teenager filled up the tank with gas on that day.
Which of the following cards would you definitely need to turn over to see if any of these teenagers are breaking their parents' rule:
"If you borrow my car, then you have to fill up the tank with gas."
Don't turn over any more cards than are absolutely necessary.
Borrowed car (correct answer)
Did not borrow car
Filled up tank with gas
Did not fill up tank with gas (correct answer)

Item 14: (CRT2_nH7) from Shafir (1994).
Imagine that in front of you are two boxes. Inside each of the boxes is a ball that is equally likely to be either white, blue, or purple. You are now offered to play one of the following two games of chance:
Game 1: You guess the color of the ball in the left-hand box. You win 50 dollars if you were right, and nothing if you were wrong.
Game 2: You choose to open both boxes. You win 50 dollars if the balls in the boxes are the same color, and nothing if they are different colors.
Which game would you prefer to play?
Game 1
Game 2
The games are equal (correct answer)

Appendix E

14 RQ items divided into two groups in session 2 time-pressure condition. Average response time at baseline and time allowed in session 2.

Set 1:

Items	RT baseline	Auto- advanced time (rounded)
CRT_H2	50.14	40
CRT_H5	18.96	10
CRT_H6	17.39	10
CRT_nH2	87.63	75
CRT_nH4	66.59	50
CRT_nH6	103.78	85
CRT_nH7	69.85	65

Set 2:

Items	RT baseline	Auto- advanced time (rounded)
CRT_H1	58.8	40
CRT_H3	83	55
CRT_H4	34.25	20
CRT_H7	48.97	30
CRT_nH1	43.27	35
CRT_nH3	85.08	65
CRT_nH5	69.1	55

Appendix F

Need for Cognition scale

Short version 18 items. Reverse scored items: 3-5, 7-9, 12, 16, 17.

Introduction:

Below are some questions about how you normally work, like to approach tasks, and how you are prepared for the outcome of various events. Give your honest and candid opinion. There is no right or wrong answer. It is important that you specify what you normally do - what is typical for you. Tick the box that best describes the extent to which you agree with the statements below.

	very strong disagreement (1)	Somewhat uncharacteristic of me (2)	Uncertain (3)	Somewhat characteristic of me (4)	very strong agreement (5)
I would prefer complex to simple problems. (1)	○	○	○	○	○
I like to have the responsibility of handling a situation that requires a lot of thinking. (2)	○	○	○	○	○
Thinking is not my idea of fun. (3)	○	○	○	○	○
I would rather do something that requires little thought than something that is sure to challenge my thinking abilities. (4)	○	○	○	○	○
I find satisfaction in deliberating hard and for long hours. (5)	○	○	○	○	○
I find satisfaction in deliberating hard and for long hours. (6)	○	○	○	○	○
I only think as hard as I have to. (7)	○	○	○	○	○
I prefer to think about small, daily projects to long-term ones. (8)	○	○	○	○	○
I like tasks that require little thought once I've learned them. (9)	○	○	○	○	○
The idea of relying on thought to make my way to the top appeals to me. (10)	○	○	○	○	○
I really enjoy a task that involves coming up with new solutions to problems. (11)	○	○	○	○	○

Learning new ways to think doesn't excite me very much. (12)	○	○	○	○	○
I prefer my life to be filled with puzzles that I must solve. (13)	○	○	○	○	○
The notion of thinking abstractly is appealing to me. (14)	○	○	○	○	○
I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought. (15)	○	○	○	○	○
I feel relief rather than satisfaction after completing a task that required a lot of mental effort. (16)	○	○	○	○	○
It's enough for me that something gets the job done; I don't care how or why it works. (17)	○	○	○	○	○
I usually end up deliberating about issues even when they do not affect me personally. (18)	○	○	○	○	○

Appendix G

Berlin Numeracy Test

Please answer the following questions. Do not use a calculator but feel free to use the space available for notes (i.e., scratch paper).

BNT1
Out of 1,000 people in a small town 500 are members of a choir. Out of these 500 members in the choir 100 are men. Out of the 500 inhabitants that are not in the choir 300 are men. What is the probability that a randomly drawn man is a member of the choir?
(Please indicate the probability in percent). _____%

- ○ 10% (0)
- ○ 25% (1)
- ○ 40% (0)
- ○ None of the above (0)

BNT2
Imagine we are throwing a five-sided die 50 times. On average, out of these 50 throws how many times would this five-sided die show an odd number (1, 3 or 5)? ____ out of 50 throws.

- ○ 5 out of 50 throws (0)
- ○ 25 out of 50 throws (0)
- ○ 30 out of 50 throws (1)
- ○ None of the above (0)

BNT3
In a forest 20% of mushrooms are red, 50% brown and 30% white. A red mushroom is poisonous with a probability of 20%. A mushroom that is not red is poisonous with a

probability of 5%. What is the probability that a poisonous mushroom in the forest is red?_____%

- 4% (0)
- 20% (0)
- 50% (1)
- None of the above (0)

BNT4

Imagine we are throwing a loaded die (6 sides). The probability that the die shows a 6 is twice as high as the probability of each of the other numbers. On average, out of these 70 throws how many times would the die show the number 6? _____ out of 70 throws.

- 20 out of 70 throws (1)
- 23 out of 70 throws (0)
- 35 out of 70 throws (0)
- None of the above (0)

Appendix H

Wordsum Test

We want to know how people guess words them do not know or do not use so frequently. At the bottom there are a few words. You know some of them, and some words you might not know. At the top is the word in capital letters e.g. beast and then there are five other words: afraid - words - large - animal - separate.

In this case you would choose "animal", since it is closer to beast than any of the other words.

WST1
SPACE

- ○ school (0)
- ○ noon (0)
- ○ captain (0)
- ○ room (1)
- ○ board (0)
- ○ do not know (0)

WST2
BROADEN

- ○ efface (0)
- ○ make level (0)
- ○ elapse (0)
- ○ embroider (0)
- ○ widen (1)
- ○ do not know (0)

WST3
EMANATE

- ○ populate (0)
- ○ free (0)
- ○ prominent (0)
- ○ rival (0)
- ○ come (1)
- ○ do not know (0)

WST4
EDIBLE

- ○ auspicious (0)
- ○ eligible (0)
- ○ fit to eat (1)
- ○ sagacious (0)
- ○ able to speak (0)
- ○ do not know (0)

WST5
ANIMOSITY

- ○ hatred (1)
- ○ animation (0)
- ○ disobedience (0)
- ○ diversity (0)
- ○ friendship (0)
- ○ do not know (0)

WST6
PACT

- ○ puissance (0)
- ○ remonstrance (0)
- ○ agreement (1)
- ○ skillet (0)
- ○ pressure (0)
- ○ do not know (0)

WST7
CLOISTERED

- ○ miniature (0)
- ○ bunched (1)
- ○ arched (0)
- ○ malady (0)
- ○ secluded (0)
- ○ do not know (0)

WST8
CAPRICE

- ○ value (0)
- ○ a star (0)
- ○ grimace (0)
- ○ whim (1)
- ○ inducement (0)
- ○ do not know (0)

WST9
ACCUSTOM

- ○ disappoint (0)
- ○ customary (0)
- ○ encounter (0)
- ○ get used to (1)
- ○ business (0)
- ○ do not know (0)

WST10
ALLUSION

- ○ reference (1)
- ○ dream (0)
- ○ eulogy (0)
- ○ illusion (0)
- ○ aria (0)
- ○ do not know (0)

Appendix I

NASA Task Load Index

Indicate from 0 = Very low, to 100 = very high

0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100

Mental Demand - How much mental and perceptual activity was required? Was the task easy or demanding, simple or complex?	
Physical demand - How much physical activity was required? Was the task easy or demanding, slack or strenuous?	
Temporal demand - How much time pressure did you feel due to the pace at which the tasks or task elements occurred? Was the pace slow or rapid?	
Frustration level - How irritated, stressed, and annoyed versus content, relaxed, and complacent did you feel during the task?	
Effort - How hard did you have to work (mentally and physically) to accomplish your level of performance?	

Please rate your overall performance, from 0 = not successful / satisfied at all to 100 = full success / fully satisfied

	0 10 20 30 40 50 60 70 80 90 100
How successful were you in performing the task?	
How satisfied were you with your performance?	

Appendix J

5 Dimensional Curiosity Scale

Below are statements people often use to describe themselves. Please use the scale below to indicate c. There are no right or wrong answers.

1 = Does not describe me at all to 7 = Completely describes me.

	Does not describe me at all.	2	3	4	5	6	Completely describes me.
I view challenging situations as an opportunity to grow and learn.	○	○	○	○	○	○	
I am always looking for experiences that challenge how I think about myself and the world.	○	○	○	○	○	○	
I seek out situations where it is likely that I will have to think in depth about something.	○	○	○	○	○	○	
I enjoy learning about subjects that are unfamiliar to me.	○	○	○	○	○	○	
I find it fascinating to learn new information.	○	○	○	○	○	○	
I like to learn about the habits of others.	○	○	○	○	○	○	
I like finding out why people behave the way they do.	○	○	○	○	○	○	
When other people are having a conversation, I like to find out what it's about.	○	○	○	○	○	○	
When around other people, I like listening to their conversations.	○	○	○	○	○	○	
When people quarrel, I like to know what's going on.	○	○	○	○	○	○