Can we reliably measure cognitive effort? On the relation of implicit and explicit cognitive effort scales and tasks

Martin Jensen Mækelæ[a], Kristoffer Klevjer[a], Andrew Westbrook[b,c], Noah Eby[b], Gerit Pfuhl[a]

[a]Department of Psychology, UiT – The Arctic University of Norway, Norway

[b] Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, USA,

[c] Donders Institute for Brain, Cognition and Behaviour, Radboud University, Netherlands

Abstract

We tend to choose the action that is least demanding however measuring individual differences in the cost of cognitive effort can be elusive. We here report a set of five studies studying the relationship between four cognitive effort measures: the demand selection task, the cognitive effort discounting paradigm, a rationality battery to assess deliberate reasoning, and the Need for Cognition scale. We also measured working memory capacity in four studies. Subjective effort with the NASA task load index was measured in three studies. Need for Cognition was positively associated with effort spent in the cognitive effort discounting paradigm, and was also positively associated with deliberate reasoning, but there was no association with demand avoidance in the demand selection task. Working memory capacity was not related to Need for Cognition, demand avoidance or cognitive effort discounting but to deliberate reasoning. We conclude that the tasks may not measure the same latent cognitive effort construct. We discuss task-sensitivity in measuring cognitive effort, and the necessity to control for cognitive abilities.

Keywords: Cognitive control, mental effort, effort avoidance, effort discounting, effort cost, deliberate reasoning

Introduction

Laziness is build deep into our nature (Kahneman, 2011, p. 39)

People tend to choose the least demanding line of action, famously formulated as the "Law of least work" (Hull, 1943). Although originally applied to physical effort, it also applies to effort in the cognitive domain (Allport, 1954). The underlying assumption is that there is a cost associated with cognitive effort (Shenhav, Botvinick, & Cohen, 2013; Zénon, Solopchuk, & Pezzulo, 2019), although the nature of this cost is uncertain (Musslick, Cohen, & Shenhav, 2018). Indeed, brain imaging studies have shown that increased cognitive effort reduces activity in the reward network (Sayalı & Badre, 2019).

Proposed explanations for cognitive effort costs are e.g., resource limits (Baumeister, Bratslavsky, Muraven, & Tice, 1998; Bijleveld, Custers, & Aarts, 2009; Kahneman, 1973; Zénon et al., 2019), accumulation of toxic waste (Holroyd, 2016) or opportunity costs (Kurzban, Duckworth, Kable, & Myers, 2013; Zénon et al., 2019). Assertions of cognitive effort costs and minimization have been proposed to be implicated in a range of fields e.g., behavioral economics (Bonner & Sprinkle, 2002; Camerer & Hogarth, 1999), executive functions (Engelmann, Damaraju, Padmala, & Pessoa, 2009), linguistics (Kanwal, Smith, Culbertson, & Kirby, 2017), and judgment and decision-making (Shah & Oppenheimer, 2008).

There are well-established individual differences in the willingness to engage in effortful tasks. Those individual differences can be measured with the Need for Cognition scale (Cacioppo & Petty, 1982; Hussey & Hughes, 2020). Still, concerns about the reliability and validity of self-report, motivate the use of behavioral paradigms to complement self-report instruments (Paulhus & Vazire, 2007). Accordingly, a range of tasks have been developed to measure effort spent in a task. We here focus on cognitive effort, though physical and perceptual effort tasks have been developed too (for a review see e.g. (Horan et al., 2015; Reddy et al., 2015).

To gauge cognitive effort spent one can use tasks that have an intuitively wrong response and require overriding to answer correctly, as cognitive control is indeed effortful (Pennycook, Fugelsang, & Koehler, 2015; Toplak, West, & Stanovich, 2011, 2014; West, Toplak, & Stanovich, 2008). Another strand of research uses computerized tasks for measuring choices between cognitively more or less demanding options, respectively. Choice patterns are seen as an indication of preferences to avoid cognitive effort, e.g. (Gatzke-Kopp, Ram, Lydon-Staley, & DuPuis, 2018; Kool, McGuire, Rosen, & Botvinick, 2010; Pfuhl, 2012; ten Velden Hegelstad, Kreis, Tjelmeland, & Pfuhl, 2020; Westbrook, Kester, & Braver, 2013). These two approaches differ in numerous ways and it is unknown to what degree they are overlapping or if they are measuring the same "cognitive effort" construct. In this article, we present evidence that two out of three reviewed task paradigms correlate with Need for Cognition. We highlight strengths and weaknesses of the task paradigms, and identify the need for development of new tasks.

Task paradigms for measuring cognitive effort

Need for Cognition Scale: Individual trait differences in thinking disposition can be measured with the Need for Cognition Scale (NCS). The NCS is a self-reported measure of enjoying and engaging in cognitively demanding tasks (Cacioppo & Petty, 1982). The scale has good internal consistency, test- retest reliability, and measurement invariance (Hussey & Hughes, 2020). People who score high on Need for Cognition tend to seek out and engage with information to make sense of things and events. People who score low on Need for Cognition tend to use less demanding cognitive processes.

Rationality battery: Task performance on rational reasoning tasks (RQ) is an alternate way of measuring thinking disposition or "cognitive miserliness" (Stanovich, 2009; Toplak et al., 2014; Trippas, Pennycook, Verde, & Handley, 2015). Thinking disposition is proposed to be on a spectrum with one end being the preference for using computationally more demanding mechanisms for solving tasks, known as an analytic thinking disposition. On the other end of the spectrum is a preference for cognitive shortcuts namely an intuitive thinking disposition. An intuitive thinking disposition is prone to rely more on heuristics, which can serve to reduce cognitive effort (Shah & Oppenheimer, 2008). Task performance on rational reasoning tasks is dependent on using more cognitively demanding mechanisms, and avoid overreliance on heuristic responses. Indeed, suppression of intuitive but wrong answers requires cognitive control (Pennycook et al., 2015). Individual differences have early been noticed in tasks measuring deliberate reasoning (Frederick, 2005). Toplak, West and Stanovich (2011) showed that the cognitive reflection task, developed by Frederick, assesses both the ability and willingness to perform cognitive work.

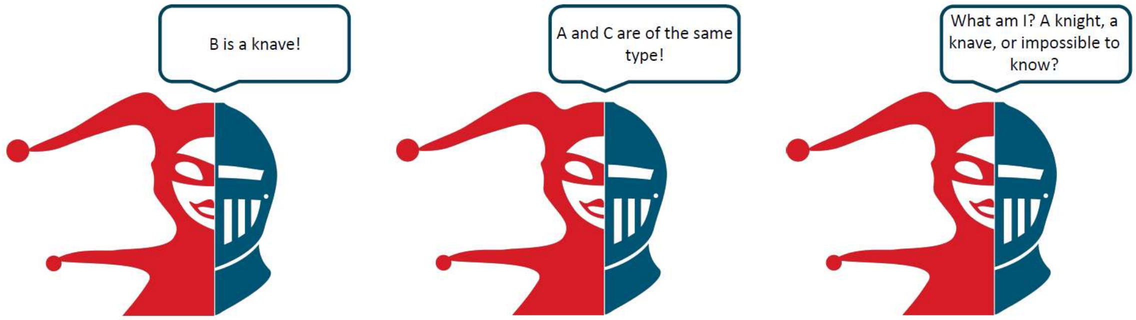

Figure 1. Example of a rational reasoning task. Imagine that there are three inhabitants of a fictitious country, A, B, and C, each of whom is a either a knight or a knave. Knights always tell the truth. Knaves always lie. Two people are said to be of the same type if they are both knights or both knaves. A and B make the following statements: A says: "B is a knave." B says: "A and C are of the same type." What is C?

Demand selection task: Evidence to support cognitive effort minimization or demand avoidance was shown with the demand selection task paradigm (DST) by (Kool et al., 2010). In this task, participants make either parity or magnitude judgements for numerical digits. Effort demands are manipulated by the frequency of task shifts and one line of action (high demand) has more frequent task shifts, thus increasing effort demand (Reddy, Reavis, Wynn, & Green, 2018). The DST can be considered an implicit measure of cognitive effort or demand avoidance as participants are not informed of the demands of the tasks, or given any incentive to choose high or low lines of action. However, a number of participants detect the demand manipulation and some evidence suggests this leads to increased effort avoidance (Gold et al., 2015).

Cognitive effort discounting paradigm: (Westbrook et al., 2013) were able to quantify the individual differences in effort costs with the Cognitive Effort Discounting Paradigm (COG-ED). In this paradigm, participants make repeated choices between performing a low demand working memory task (1-back) for a small reward, or performing a high demand working memory task for a larger reward (n-back, n being 2, 3, 4, 5, or 6). The reward for the low demand task is titrated in response to participant's choices and to find a subjective indifference point between the low demand and high demand option. The COG-ED thereby quantifies the subjective monetary discounting due to cognitive effort costs across multiple demand levels. Given that task load levels and offer amounts are all explicit, the COG-ED is an explicit cognitive effort measure. Participants experience the effort demand for each load level prior to making choices between explicit monetary offers.

Figure 2: Illustration of the demand selection task (left) and cognitive effort discounting task (right)

Algorithmic ability: In the COG-ED paradigm, subjective cognitive effort may depend on one's cognitive ability. A person with very good working memory capacity and high general cognitive abilities (e.g. as assessed with an IQ test) will perform better on the n-back task. To control for effects of cognitive ability participants are instructed that they will be required to achieve at least 80% of their own performance for the task levels they decide among in the choice phase of the COG-ED.

The DST, in contrast, should depend less on working memory capacity, given that the parity and magnitude judgements are relatively simple and are unlikely to tax working memory – although DST choices have been related to measures of cognitive flexibility (switch costs [Kool et al 2010]. Although commonly employed rational reasoning items are only nominally demanding, and choices are made without time pressure, rational reasoning does correlate positively with cognitive ability measures (Mækelæ, Moritz, & Pfuhl, 2018; Toplak et al., 2011, 2014; West et al., 2008). We therefore controlled for cognitive ability using either the n-back performance score (d' from signal detection theory) or the operation span score.

An outstanding question is whether these four tasks are related. A positive correlation between all four tasks would imply that they all index task-invariant cognitive effort. Another possible outcome is that only tasks measuring intrinsic motivation to spend cognitive effort correlate. Then, we would expect a positive correlation between the rationality tasks, DST, and NFC, but not for COG-ED. If cognitive effort requires some explicit signaling of demand, then both the NCS and COG-ED should be positively correlated. It is also possible that the rationality tasks are related, but the DST should not be related to any of the other tasks.

Previous work has already shown a positive relationship between COG-ED and NCS, and separately, correlations between rationality and NCS (Thomson & Oppenheimer, 2016; Toplak et al., 2014; West et al., 2008; Westbrook et al., 2013). However, it is not known whether COG-ED is related to rational reasoning performance or whether the propensity towards controlled deliberation is related to the subjective cost of cognitive effort (but see Shenhav et al. (2017) on a relationship between CRT and time costs). It is also not known whether cognitive demand avoidance in the DST relates to the subjective value of cognitive effort in the COG-ED. For the DST, there is also, to our knowledge, no report how demand avoidance relates to the NCS and rational reasoning. Table 1 summaries possible scenarios.

Table 1.

Possible relationship between the four measurements

Predictions	Pattern of results
All tasks measure cognitive effort	Positive correlation between NCS & COG-ED, NCS & DST, NCS & RQ, COG-ED & DST, COG-ED & RQ, DST & RQ
Intrinsic motivation	Positive correlation only between NCS, DST & RQ, weak or no correlation of COG-ED with NCS, DST & RQ
Demand signaled	Positive correlation only between NCS, COG-ED, & RQ, but no relationship of DST with NCS, COG-ED & RQ
Approach motivation	Positive correlation between NCS & COG-ED, none or negative correlation between COG-ED & DST
Cost of cognitive control	Positive correlation between RQ & COG-ED, RQ & DST, COG-ED & DST, but none of them with NCS
Task-specific	Positive correlation between NCS & RQ, and between COG-ED & DST, or no relationship between any of the tasks

Finally, subjective effort, which may deviate from objective effort (e.g. (Kreis, Moritz, & Pfuhl, 2020), can be assessed with the NASA task load index (N-TLX) (Hart & Staveland, 1988). Westbrook et al. (2013) found increasing subjective ratings of mental and physical effort, temporal demand, failure rate, effort demand and frustration for increasing working memory load levels in the COG-ED. We here report five studies from two independent labs investigating the relationship between the DST, COG-ED, NCS, and rational reasoning. In addition, for three studies we present N-TLX ratings for each task paradigm, as well as test –

retest reliability for NCS and DST. Lastly, we present the results from all studies pooled together. We report two-sided and non-corrected p-values using Spearman's rank coefficient.

Study 1: Relationship between cognitive demand avoidance, cognitive effort discounting and need for cognition

Human behavior is driven by both intrinsic and extrinsic motivation (Ryan & Deci, 2000). The COG-ED paradigm relies on external rewards and explicit task demands, whereas the DST aims to measure internal avoidance of demand without incentives or information about task demands (Table 1). Both the COG-ED and DST have been used as measures of cognitive effort in clinical and developmental research (Chevalier, 2018; Culbreth, Westbrook, & Barch, 2016; Gold et al., 2015; Patzelt, Kool, Millner, & Gershman, 2019; Westbrook et al., 2013; Westbrook et al., 2020). However, it is not known whether those two task measure the same latent construct.

We expected a positive correlation (replication of Westbrook and Braver, 2013) between the COG-ED and NCS. We also expected that individuals scoring high on the NCS should show reduced cognitive demand avoidance in the DST, and that demand avoidance is negatively related to the subjective value of cognitive effort in the COG-ED, i.e. the more a person prefers the low switching cue the more she discounts offers to perform more demanding n-back levels in COG-ED.

Methods

Participants

All participants were undergraduate students at Washington University in St. Louis, USA (N=76, 50 female). The mean age was 21.2 (range 18 to 32 years). All study procedures including informed consent were approved by the Institutional Review Board at Washington University in Saint Louis.

Materials

Effort discounting. The COG-ED task was administered through E-prime 2.0 (Psychology Software Tools, Inc., Sharpsburg, PA). The task started with a practice phase of the n-back task (Owen, McMillan, Laird, & Bullmore, 2005). Participants played all load levels (n= 2 – 6) for three runs. All runs consisted of 64 items (consonants, presented in Courier New font, font size 24). Items were presented on screen for 1.5 seconds, during which participants could respond. After 1.5 seconds the items were replaced by a fixation cross. The inter-trial interval was 3.5 seconds. Participants were given feedback about % of targets and % of non-targets correct. Feedback of "Good job!" was given if both scores were above 50%

or “Please try harder!” if not. In the discounting procedure participants were offered to play n = 1 for a small reward or n < 1 for a larger reward. Participants were offered six choices for each load level. The amount for the higher offer (n > 1) was always $2. The reward amount for the lower offer (n = 1) started at $1 and was adjusted up if participants chose the high offer, and was adjusted down if participants chose the low offer. Each time a choice was made, the reward amount was adjusted to half as much as on the previous choice. After the last choice (six choices in total), the amount was adjusted to $0.015. The final amount was taken as participants' subjective indifference point. Participants played five load levels and made six choices for each level, yielding 30 choices in total. The subjective value of the Average Indifference point (AIP) across all load levels is the cognitive effort discounting measure used for between-subjects analyses. To ensure choices reflected participants' preferences, they were told that one of the choices would be selected for them to repeat 10 more times and they would be paid for each repetition. Further, they were told that payment was contingent on maintaining effort, but not on performance. Effort would be monitored by “behavioral clues”. All participants completed their randomly selected offer four times, and were paid the associated amount.

Cognitive demand avoidance. We used an exact replication of Experiment 3 in Kool et al. (2010). The task was administered on a computer, using MatLab 2018a (The MathWorks, MATLAB, Version 9.4, 2018), with Psychophysics Toolbox 3 extension (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). The task starts with a training phase where participants complete two different tasks. Participants are presented with a number (between 1 and 9, excluding 5). The number can be either blue or yellow. The color of the number signaled the task required on that trial. If the number is blue, participants must decide if the number is higher or lower than 5. If the number is yellow then participants must decide if the number is odd or even. Participants indicate their choice by clicking on the right or left side of a computer mouse. During the training phase (60 trials), participants received feedback on their performance. None of the participants had to redo the training phase. In the main task, participants see two colorful balls on screen (they appear along an invisible circle at an angular distance of 45 degrees). The location of the balls changes between runs, but is stable throughout a run. Participants must sample from each option, but are told they can stay with one if they develop a preference. There are eight runs with 75 trials in each run (600 in total). There is one high demand option (ball) where the task switches with a probability of 0.9, and there is a low demand option where the probability of task switch is 0.1. Task instructions were available in paper format in case participants forgot the rules. Demand avoidance is quantified in terms of the proportion selection of the low demand decks – thus a demand avoidant participant would score between .5 and 1.

Thinking disposition is measured with the 18-item Need for Cognition scale (Cacioppo, Petty, & Feng Kao, 1984). The Need for Cognition scale measures a person's tendency to engage in and enjoy cognitively effortful activity. An example item is *"I prefer complex to simple problems".* The 18 items are rated on a 5 point Likert scale from 1 = "*Extremely uncharacteristic of me*" to 5 = "*Extremely characteristic of me*". Total score range 18 – 90.

Working Memory capacity is measured with the operation span (O-Span) task by (Turner & Engle, 1989). In the O-Span task, participants see relatively simple math equations presented together with a random word, this is an equation – word pair. These equation - word pairs are presented in series of differing length, here between 1 and 5 equation - word pairs were used. After each equation, the participants must solve the equation and remember the word. At the end of a trial participants must recall all the words in the series sequentially, in the order they appeared. Each length of series (1 – 5) had five trials. Trials were presented in randomized order. Trials with all words correct were scored according to the length of the series (e.g. a two equation - word pair series is scored as 2). Trials with any incorrect answers were scored as 0. Total scores ranged from 0 – 75.

In addition we used the d' from the n-back task (average from level 1 – 6).

Procedure

Participant were paid 10$ per hour for their participation, and they could earn additional money based on their choice in the COG-ED. Participants received their payment at the end of the testing session. Testing was completed individually at the University of Washington. Order of the tasks were; DST, COG-ED, Need for Cognition, O-Span. Usual participation time was approximately two hours. Kendall's τ, a rank-order correlation coefficient, was applied for correlation analysis, as the rank ordered relationship was of interest. Statistical analysis was carried out in JASP (JASP Team, 2019, version 0.11.1).

Results

Descriptive statistics for the tasks can be found in Table 2. Neither cognitive effort discounting in COGED, nor low demand preference in the DST were normally distributed.

Table 2.

Descriptive statistics for tasks in Study 1. N=76

	NCS	COG-ED AIP	COG-ED d’	DST-HDP	O-SPAN
Mean	65.145	1.459	2.057	0.423	51.816
SD	11.300	0.438	0.697	0.165	16.805
Skewness	-0.328	-0.361	-1.613	0.331	-0.961
Kurtosis	-0.323	-1.101	6.749	1.745	0.634
Minimum	35	0.534	-1.163	0.053	0
Maximum	86	1.984	3.434	0.947	75

Note. NCS = Need for Cognition Scale. COG-ED AIP = Cognitive Effort Discounting Paradigm - Average indifference point. COG-ED d’ = Cognitive Effort Discounting Paradigm d’. DST-HDP = Demand Selection Task – High demand preference. O-Span = Operation span task.

The COG-ED paradigm showed significantly larger monetary discounting with increasing load levels $F(4, 380) = 24.109$, $p < .001$, $\eta^2 = .202$. For the DST, accuracy was high (Mean = 0.94, $SD = 0.07$). The median rate of high (low) demand choice was 0.42 (0.58). A Z-test showed that demand avoidance was not significantly different from 0.50 in the DST (Z = 0.674, $p < 0.5$, d = .077). As expected, cognitive effort discounting was positively related to a person’s NCS score, τ = .169, p = .035. There was no relationship between effort discounting and demand avoidance, ρ = .042, p = .595. There was also no relationship between demand avoidance and NCS, ρ = .002, p = .979. None of the tasks related to cognitive ability as assessed with the O-Span task, all τ < .01, all p > .2, nor did, τ = .101, p = .2. Working memory capacity in the n-back task (d’) did not relate to any of the three effort measures (DST: τ = .043, p = .584, COGED: τ = .069, p = 387, NCS: ρ = -.024, p = .757. O-Span and d’ were not associated either.

Discussion

We tested if cognitive effort as measured by effort discounting in COG-ED and demand avoidance in DST were related. We did not find any relationship between these measures. This is surprising as both measures are supposed to measure cognitive effort, and have been used as such. However, there are quite distinct differences between the two paradigms, COG-ED is explicitly signaling effort and effort spent is motivated by external rewards. In contrast, the effort spent in the DST relies on intrinsic motivation. Further, we did not find any relationship between the DST and NCS. This was surprising as both measures should rely more on internal motivation to use cognitive effort. We did replicate the finding that effort discounting in COG-ED is related to NCS. Further, we found that working memory capacity measured by O-SPAN was not related to any effort measure.

Study 2: Cognitive effort discounting is not related to rational reasoning but both are related to Need for Cognition

Rational reasoning is dependent on cognitive (algorithmic) ability and thinking disposition (Stanovich, 2013). Individuals high in NCS tend to do better on rational reasoning tasks. Individuals high in NCS also report lower subjective cognitive effort costs in COG-ED (Westbrook and Braver, 2013). Since NCS relates to both COG-ED and rational reasoning, one may expect also a positive relationship between COG-ED and rational reasoning. Accordingly, we next investigated the relationship between effort discounting in COG-ED, rational reasoning and NCS in a student sample. We again included a working memory measure to control for individual differences in cognitive ability.

Methods

Participants

All participants were undergraduate students at Washington University in St. Louis. They were all healthy, with no neurological disorders, and not taking any psychoactive medication. The sample consisted of 82 participants (54 female, 28 male). The mean age was 21.2 (range 18 to 32 years). All study procedures including informed consent were approved by the Institutional Review Board at Washington University in Saint Louis.

Materials

Working Memory capacity. O-Span task and d' from COG-ED, identical to study 1.

Thinking disposition. NCS, identical to study 1.

Effort discounting. COG-ED, identical to study 1.

Rational reasoning. We used the 18 items scale from Toplak et al. (2011). The reasoning problems included probabilistic thinking, scientific reasoning, theory justification, and hypothetical thought. In addition, the original 3-item Cognitive Reflection Test was included, measuring individual differences in detecting errors and overriding an initial intuitive response (Frederick, 2005). The remaining 15 items were problems from the heuristics and biases literature: two-sample size problems, two gamblers fallacy problems, regression to the mean, a base rate problem, a covariation detection problem, one Bayesian reasoning problem, one conjunction fallacy problem, a denominator neglect problem, a methodological reasoning problem, a probability matching problem, a sunk cost fallacy problem, one outcome bias problem, and a framing problem. Correct answers were scored as 1, incorrect as 0. Total composite score ranged between 0 and 18.

Procedure

The procedure was identical to study 1, only the DST was replaced with the rational reasoning problems, i.e. the order of the tasks were: rational reasoning problems, COG-ED, Need for Cognition, O-SPAN. Usual participation time was approximately two hours. Spearman's rank-order correlation coefficient was applied for correlation analysis, as the rank ordered relationship was of interest. Statistical analysis was carried out in JASP (JASP Team, 2019, version 0.11.1).

Results

One participant was removed due to scoring zero on multiple tests. Results do not change if this participant is included in the analysis. Table 3 shows that the sample had a similar O-Span, NCS and COG-ED than the sample in study 1. Most participants got more than half of the rationality items correct.

Table 3.

Descriptive statistics for tasks in Study 2. N=82

	NCS	RQ	COG-ED AIP	COG-ED d'	O-SPAN
Mean	66.561	10.037	1.584	1.679	50.780
SD	9.821	4.041	0.421	0.759	18.565
Skewness	-0.638	-0.130	-1.637	-2.600	-0.788
Kurtosis	0.112	-0.754	2.930	10.812	-0.102
Minimum	36	1	0.063	-2.424	0
Maximum	84	18	1.969	2.760	75

Note. NCS = Need for Cognition Scale. RQ = Rational reasoning tasks. COG-ED AIP = Cognitive Effort Discounting Paradigm - Average indifference point. COG-ED d' = Cognitive Effort Discounting Paradigm discriminability score d'. O-SPAN = Operation span task.

The COG-ED paradigm showed significantly larger monetary discounting with increasing load levels $F(4, 240) = 24.734$, $p < .001$, $\eta^2 = .191$.

We found a non-significant but positive correlation between cognitive effort discounting and Need for Cognition, τ = .116, p = .146, between Need for Cognition and rational reasoning, τ = .147, p = .062; and between cognitive effort discounting and rational reasoning, τ = .151, p = .062. O-Span was positively correlated with rational reasoning, τ = .209, p = .008, and non-significantly with cognitive effort discounting, τ = .096, p = .226; but not with Need for Cognition, τ = .006, p = .422. Finally, working memory capacity in the n-back task was positively correlated with the O-Span score, τ = .153, p = .045, with rational reasoning, $\tau = .347$, $p < .001$, with effort discounting, $\tau = .198$, $p = .012$, but not with Need for Cognition, $\tau = .114$, $p = .062$.

Discussion

We replicated that Need for Cognition was positively related to rational reasoning, and that cognitive effort discounting was positively related to Need for Cognition. However, the relationship was small and non-significant using two-sided testing. In line with previous research (West et al., 2008), we found that operation span and working memory capacity, both proxies of one's algorithmic ability, were positively related to rational reasoning. We found also that operation span was related to cognitive effort discounting. This is not surprising as the n-back task should be easier the higher the working memory capacity. However, contrary to our hypothesis we did not find that cognitive effort discounting was related to rational reasoning. We did also not find a relationship between operation span or working memory capacity and Need for Cognition, replicating also study 1's findings. This is similar to previous research showing that Need for Cognition is related to general intelligence but not working memory (Hill et al., 2013).

Study 3: Cognitive effort discounting is not related to rational reasoning and Need for Cognition

Study 3 is in independent conceptual replication of study 2. Study 3 uses n-back levels 1 – 4 (Not 1 – 6 as in Study 2) in the COG-ED. Rational reasoning was measured using a different set of tasks, although some tasks were similar. Working memory capacity was assessed with performance (d') on the n-back task. All students were psychology students, who were familiar with rational reasoning tasks. In addition, we included the N-TLX to assess effort and demand differences for the COG-ED and rational reasoning.

Methods

Participants

All participants were undergraduate psychology students at UiT, The Arctic University of Norway. The total sample consisted of 100 participants (62 female, 25 male, 13 unknown) whereof 65 completed both sessions. The mean age was 22.6 (range 20 to 38 years). The study was approved by the internal review board at the Department of Psychology, UiT, The Arctic University of Norway.

Materials

Thinking disposition. identical to study 1 and 2, implemented in Qualtrics.

Effort discounting. We used the COG-ED, but only levels 1-4, and implemented in Inquisit (Millisecond, Software). The first phase of this task consisted of five runs per n-back level (2, 3, & 4), each run with 5 target trials (response would be correct), and 10+N non-target trials (response would be incorrect) in a pseudo-random sequence. Each trial lasted 2.5 s, and in each trial participants were presented with a stimuli (one of 20 consonants, centered white letters on a black screen, sans-serif font) for 0.5 s, followed by a black screen for 2.0 s, and during the 2 seconds had to either respond (press 'A' on the keyboard) or not respond. After each run, the participants were presented with a summary feedback of their accuracy, and after the last run on each n-back level they were presented with a level summary. The second phase consisted of three blocks, 1-back vs. 2-back, 1-back vs. 3-back, and 1-back vs. 4-back, presented in a pseudo-random order across participants. Each block had six runs in which the participants chose between a 1-back task or n-back task. The tasks themselves were equal to the n-back task described above. Discounting was identical to study 1 and 2. Importantly, participants were informed that they would not receive money.

Working memory capacity. We used the d' from the n-back task, average of the 2-back, 3-back and 4-back performance.

Rational reasoning. Measured with 14 items from the heuristics and biases literature. We used items 2-7 from the Cognitive Reflection Test (Toplak et al., 2014), one fully disjunctive reasoning problem "the marriage problem" (Levesque, 1986), one probability matching task (Koehler and James, 2010), one probability estimation task "the bus problem" (Teigen & Keren, 2007), one making sense of medical results problem (Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, and Woloshin, 2007), one Bayesian reasoning problem from (Toplak, Liu, Macpherson, Toneatto, and Stanovich, 2007) adapted from (Beyth-Marom and Fischhoff, 1983), one covariation detection problem (Stanovich and West, 1998), one knight and knave problem (Smullyan, 1978), one conditional reasoning problem (Lehman, Lempert, and Nisbett, 1988). Correct answers were scored as 1, incorrect as 0. Total composite score range between 0 and 14. The task was implemented in Qualtrics (Qualtrics, Provo UT).

Nasa Task Load index. Subjective effort was measured with the NASA task load index. Effort is self-assessed and scored along six dimensions: mental, physical, temporal, performance, effort, frustration. Participants rated their effort on a 0–100 scale in steps of 5 (range 0–21) where 0 was "very low" and 21 was "very high". Mental demand is reported from the scales of the NASA task load index. The task was implemented in Qualtrics and used both after the COG-ED and after the rational reasoning tasks.

Procedure

Testing took place over two separate sessions in small groups in a computer pool. Students took part for course credit and received no monetary compensation. Students could choose to partake in one session only. On day 1, 82 student took part, and were tested on the COG-ED. On day 2, approximately 3 weeks later, 84 students took the RQ including the N-$TLX_{rational}$, and NCS. 65 students took part in both test sessions. Participants could withdraw or indicate on the consent form that they do not permit to use their data for research, which was once the case. Each session took approximately 1 hour.

Results

The NCS and effort discounting was similar to study 1 and 2. Rational reasoning score was high, most answered more than half of the items correctly.

Table 4.

Descriptive Statistics for study 3. N=102 (for NCS N=84, for COG-ED N=82)

	NCS	RQ	RQ N-TLX	COG-ED AIP	COG-ED d'	COG-ED N-TLX
Mean	61.726	5.080	67.200	1.176	2.081	73.125
SD	9.152	2.485	14.946	0.555	0.658	16.037
Skewness	-0.278	0.229	-0.502	-0.238	-0.195	-1.003
Kurtosis	-0.532	-0.858	1.117	-0.788	0.337	1.809
Minimum	42.000	1.000	15.000	0.016	0.389	15.000
Maximum	81.000	11.000	100.000	1.984	3.552	100.000

Note. NCS = Need for Cognition Scale. RQ = Rational reasoning. RQ N-TLX = Rational reasoning task - Nasa task load index, subscale mental demand. COG-ED AIP = Cognitive Effort Discounting Paradigm - Average indifference point. COG-ED d' = Cognitive Effort Discounting Paradigm d prime. COG-ED N-TLX = Cognitive Effort Discounting Paradigm - Nasa task load index, subscale mental demand.

We did not find any statistical significant association between effort discounting and NCS, $\tau = .032$, $p = .716$, effort discounting and rational reasoning, $\tau = .006$, $p = .943$, rational reasoning and NCS, $\tau = .028$, $p = .723$, rational reasoning and d', $\tau = .131$, $p = .102$, or NCS and d', $\tau = -.017$, $p = .841$. Similar to study 1 we did also not find a relationship between effort discounting and d', $\tau = .046$, $p = .552$.

Perceived mental demand in the COG-ED task was high, $M = 73.1$, $SD = 16.0$; and also high for rational reasoning, $M = 67.2$, $SD = 14.9$. However, there was no association between perceived mental demand and cognitive discounting, τ = -.133, p = .098 or rational reasoning and perceived mental demand, τ = .041, p = .581.

Discussion.

In contrast to study 1 and 2 we could not replicate a positive relationship between NCS and cognitive effort discounting, nor between rational reasoning and NCS. Our participants were naïve to the rational reasoning items used but not to such tasks per se, as they were taught about heuristics and biases in a previous course. Another explanation for the discrepancy of cognitive effort and NCS is that there was no monetary incentive in the COG-ED. Choice was more driven by intrinsic motivation than a prospect of earning a reward. We thought this would make it closer to the rational reasoning task and NCS. In study 4 we controlled for that.

Study 4: Cognitive effort discounting, demand avoidance and rational reasoning

Study 1 found no relationship between DST and COG-ED and between DST and NCS. Study 3 found no relationship between COG-ED and NCS, and between COG-ED and rational reasoning. However, we do not know whether DST and rational reasoning are related. Since both "rely" on intrinsic motivation, it is plausible that demand avoidance in the DST is related to rational reasoning.

In addition to DST and NCS, we included a rational reasoning task on Day 1, and COG-ED on Day 2. We also report task demands and effort measured with N-TLX for all tasks.

Methods

Participants

Participants were healthy adults (N= 40, 27 female), age range 18 – 37 years, most were students at UiT, The Arctic University of Norway, three were full-time workers, three were high school students All participants completed both testing sessions. The study was approved by the internal review board at the Department of Psychology, UiT, The Arctic University of Norway.

Materials

Cognitive demand avoidance. Measured with the Demand Selection Task (DST) from Kool et al. (2010). The task is an exact replication of Experiment 3 in Kool et al. (2010), implemented in Matlab 2018a. The task was completed on both testing days to assess the stability of cognitive demand avoidance.

Thinking disposition. Identical to study 1-3, implemented in Qualtrics (Qualtrics, Provo UT). The task was completed in both testing sessions to assess the stability of Need for Cognition.

Rational reasoning. Identical to study 3.

Effort discounting. Identical to study 3.

Working memory. Identical to study 3.

Procedure

All participants were tested individually at UiT – The Arctic University of Norway. All participants completed a second testing session between 4 and 8 week after the first testing session. Day 1 task order was; DST, Rational reasoning, Bullshit receptivity scale (Pennycook et al., 2015, Mækelæ, Moritz, Pfuhl, 2018, not included in analysis), NCS, Effort expenditure for rewards task (EEfRT, Treadway, Buckholtz, Schwartzman, Lambert, & Zald, 2009, not included in analysis) and N-TLX$_{EEfRT}$ (not included). Day 2 task order was; DST, NCS, Handgrip (Xu et al., 2014, not included in analysis), COG-ED, and N-TLX$_{COGED}$. Participants received a voucher with a fixed amount of 20$ for participation, plus between 5$ and 15$ depending on task performance in the COG-ED and EEfRT.

Analysis

We calculated internal consistency of the DST by using the eight individual blocks per participant. Similarly, we calculated internal consistency for the NCS by using the individual 18 items per participant. All other analysis are as for study 1-3.

Results

Two participants were excluded due to accuracy below 0.49 on the DST. Including the participants in the analysis did not change the main results substantially. Descriptive statistics for the tasks in Study 3 can be found in Table 5.

Table 5.

Descriptive statistics for tasks in Study 4. N=40.

	DST-HDP Day 1	DST-HDP Day 2	NCS Day 1	NCS Day 2	COG-ED AIP	COG-ED d'	RQ
Mean	0.47	0.46	62	65	1.16	2.37	7
SD	0.08	0.09	10.11	11.5	0.513	0.3	2.7
Minimum	0.34	0.25	38	34	0.48	1.15	1
Maximum	0.78	0.69	81	81	1.98	3.76	11

Note. DST-HDP = Demand Selection Task – High demand preference. NCS = Need for Cognition Scale. RQ = Rational reasoning tasks. COG-ED AIP = Cognitive Effort Discounting Paradigm - Average indifference point. COG-ED d' = Cognitive Effort Discounting Paradigm d'.

For the Demand Selection Task, accuracy was high on both Day 1 (Mean = 0.98, *SD* = 0.02) and Day 2 (Mean = 0.96, *SD* = 0.08). The median rate of high demand choice was 0.49 for Day 1 and Day 2. Demand avoidance was not different from .5, neither on day Day1 (Z = .21, p = .834, d = .034) nor on Day 2 (Z = .238, p = .812, d = .039).

The COG-ED paradigm showed significant increases in monetary discounting with increasing load levels $F(2, 114) = 4.432$, $p = 0.014$, $\eta^2 = .072$.

DST High demand preference was positively correlated with NCS ($\tau = .372$, $p = 0.001$), and rational reasoning ($\tau = .259$, $p = 0.031$). NCS was positively correlated with rational reasoning on Day 1 ($\tau = .343$, $p = 0.004$). Demand avoidance on Day 2 was not significantly related to cognitive effort discounting assessed on Day 2 ($\rho = .207$, $p = 0.078$). Cognitive effort discounting was not related to NCS ($\tau = .062$, p = .592), or to rational reasoning ($\tau = .012$, p = .919).

Need for Cognition had good internal consistency on Day 1 (Cronbach's α = .83) and Day 2 (Cronbach's α = .89), and good reliability across the two testing sessions ($\rho = .823$, $p < .001$). Demand avoidance showed acceptable internal consistency on Day 1 (Cronbach's α = .71), but poor internal consistency on Day 2 (Cronbach's α = .52), and poor reliability across the two testing sessions ($\rho = .537$, $p < .001$). On Day 2 demand avoidance was not related to Need for Cognition ($\rho = .165$, p = .16).

Perceived mental demand as assessed with the N-TLX was larger for COG-ED (*M* = 79.4, *SD* = 15.7) than DST (day 1: 52.1, SD = 22; day 2: 50.1, SD = 23.5) and Rational reasoning (M = 48.8, SD = 24.5). N-TLX did not correlate with demand avoidance (τ = -.144, p = .218) or effort discounting (τ = .112, p = .345), or rational reasoning (τ = -.198, p = .095).

Discussion.

We assessed the reliability and internal consistency of the DST and NCS. Reliability and internal consistency of the DST was inferior to the reliability and internal consistency of the NCS. This suggests that demand avoidance and possibly cognitive effort is task-specific and state dependent. Demand avoidance was related to rational reasoning when tested on the same day, but not when demand avoidance was tested four weeks later. We replicated (study 1) that demand avoidance was not related to effort discounting in the COG-ED. Further, we replicated the results from Study 2 and 3 that COG-ED did not correlate with rational reasoning. Further, we could not replicate that cognitive effort discounting relates to Need for Cognition. Need for Cognition showed good reliability, and was related to rational reasoning on both days. It is probable that both changes in state and trait differences influence rational reasoning. It should be noted that our sample was small and random variation can influence results from a single study.

Study 5: Rational Reasoning is related to Need for Cognition, but not Cognitive Demand Avoidance

In study 4, we found a relationship between cognitive demand avoidance, rational reasoning, and Need for Cognition, but not between cognitive effort discounting and rational reasoning and Need for Cognition. Since reliability of the demand avoidance task was low, replicability had to be assessed. We therefore conducted a follow up study with the Demand Selection Task, rational reasoning and Need for Cognition.

Method

Participants

All participants were students (non-psychology) at UiT, The Arctic University of Norway (N=45, 27 female), mean age was 23.35 (range 18 and 37). The study was approved by the internal review board at the Department of Psychology, UiT, The Arctic University of Norway.

Materials

Rational reasoning. identical to study 3. Item order was randomized.

Thinking disposition. identical to study 1-4.

Cognitive demand avoidance. identical to Study 4.

Procedure

Participants were tested individually at UiT, The Arctic University of Norway. All participants received a voucher worth 40 dollars after completing two days of testing[1]. Each testing session lasted approximately between 1.5 and 2 hours, depending on participants speed. Task order was; Demand Selection Task, rational reasoning task, Need for Cognition, Teleological reasoning with pupillometry (not reported here).

Results

The Need for Cognition score was similar as in study 1-4. On average half of the items in the rational reasoning task were solved correctly, again similar to the sample from study 4, see Table 6. The results from the Demand Selection Task show that participants avoided cognitive demand (Median = 0.468). Demand preference was not significantly different from 0.50 (Z test, Z = .37, $p < 0.711$, d = .055). Accuracy on the demand selection task was high (Mean = 0.97, *SD* = 0.04).

Table 6.

Descriptive Statistics for study 5. N=45

	NCS	DST-HDP	RQ	DST N-TLX	RQ N-TLX
Mean	63.356	0.445	6.822	45.556	79.111
Std. Deviation	9.773	0.122	2.605	24.500	18.868
Skewness	-0.310	-0.194	0.107	0.180	-1.088
Kurtosis	0.943	2.745	-0.779	-0.853	1.043
Minimum	33	0.083	2	0	25
Maximum	83	0.832	13	100	100

Note. NCS = Need for Cognition Scale. DST-HDP = Demand Selection Task – High demand preference. RQ = Rational reasoning tasks. DST N-TLX = Demand Selection Task – Nasa task load index, mental demand subscale. RQ N-TLX = Rational reasoning – Nasa task load index, mental demand subscale.

The Demand Selection Task was not correlated with rational reasoning ($\tau = .059$, $p = 0.587$), or NCS ($\tau = -.115$, $p = 0.272$). Rational reasoning was correlated with NCS ($\tau = .225$, $p = 0.04$). Dichotomizing participants into avoiders and seekers (Sayalı & Badre, 2019) yielded also no significant association for NCS and RQ (logistic regression, p > .25).

Similar to study 4, perceived mental effort was rated lower in the DST (M = 45.6, SD = 24.5) than for rational reasoning (M = 79.2, SD = 18.9). However, N-TLX rating was not

[1] Day 2 involved eye-tracking, not reported here.

related to actual performance, for rational reasoning: ρ = -.147, p = .194, for DST: τ = -.016, p =.883.

Discussion

The relationship between DST and rational reasoning did not replicate, and as found in study 1, there was no relationship between DST and NCS in this sample. We did replicate the finding that rational reasoning and Need for Cognition was related, and that the DST is perceived as less demanding than rational reasoning items.

Study 6: combining studies 1-5

Method

For the meta-analysis we transformed all values into z-scores. Note that these correlations are not independent (same participants) for some of the meta-analytical effects, e.g. cognitive effort tasks and NCS. We were interested in a) relationship for each of the three cognitive effort tasks to the NCS; b) the relationship among those three tasks; and c) the relationship of those tasks to cognitive ability. We used d' from the COG-ED (study 1, 2, 3, 4) as a measure of cognitive ability.

Finally, to assess the contribution of thinking disposition and cognitive ability on cognitive effort, we used linear regression with NCS and d' as predictors for effort discounting (COG-ED), demand avoidance (DST) and rational reasoning.

Results

Do the DST, COG-ED and RQ relate to the NCS?

The effect size of the relationship between Need for Cognition and rational reasoning and cognitive effort discounting was very similar and around .2. However, there was no consistent relationship between NCS and demand avoidance.

Figure 3-5 show the forest plot for the meta-analysis. NCS has a reliable positive relationship with rational reasoning but note the heterogeneity. Figure 4 shows for a positive relationship of NCS with effort discounting, and homogeneity. Figure 5 shows no consistent association between the NCS and demand avoidance, and moderate heterogeneity.

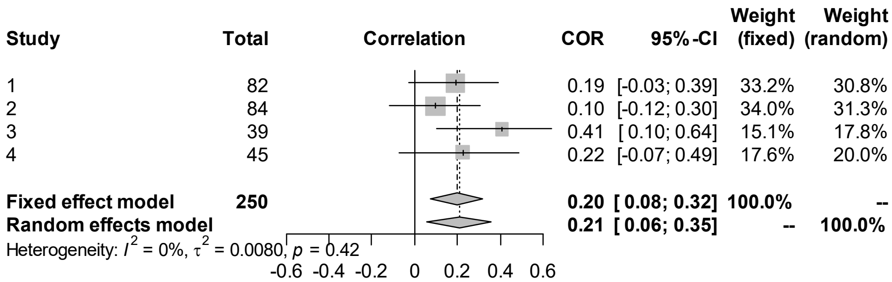

Figure 3: Need for Cognition and rational reasoning, data from study 2-5

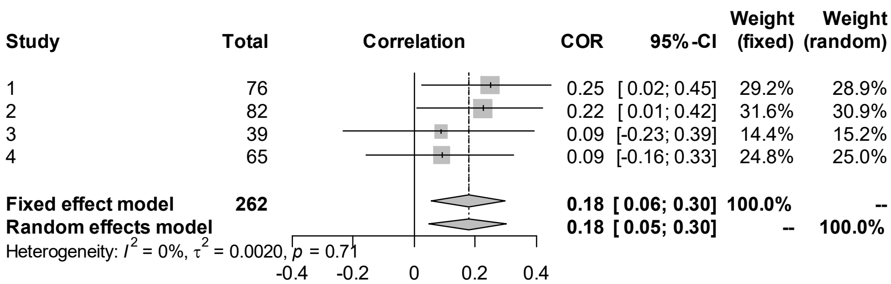

Figure 4: Need for Cognition and cognitive effort discounting, data from study 1-4

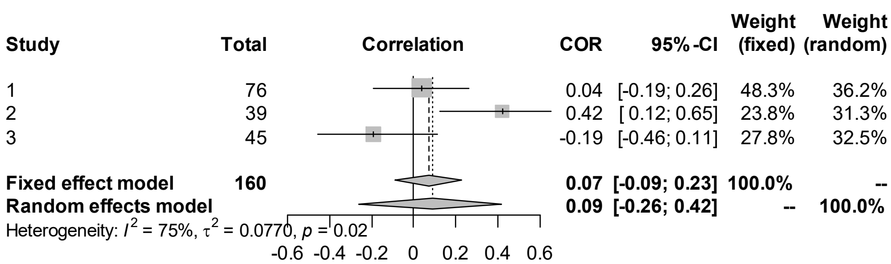

Figure 5: Need for Cognition and demand avoidance, data from study 1, 4-5

Is there a relationship among the three cognitive tasks? Based on the three studies measuring both cognitive effort discounting and rational reasoning, we find no relationship between those two tasks, r(201) = .019, p = .79; nor do we find a relationship between cognitive effort discounting and demand avoidance, r(115) = .148, p = .115. There was also no relationship between rational reasoning and demand avoidance, r(84) = .127, p=.248 (Figure 6).

Is there a relationship of those tasks with d' from the COG-ED? Cognitive ability (d') was not related to Need for Cognition (r = .054, p = .389), demand avoidance (r = .15, p = .113), and cognitive effort discounting (r = .112, p = .063), but was positively related to rational reasoning (*r* = .299, p = < .001).

Finally, we ran linear regressions with rational reasoning and cognitive effort discounting as outcome and Need for Cognition and Cognitive ability (d') as predictors. The model for the rational reasoning task explained 11.1% of the variance, $F(2, 186) = 11.47$, $p < .001$; and both predictors were significant, NCS: $t = 2.477$, $\beta = .173$, $p = .014$; d': $t = 3.918$, $\beta = .273$, $p < .001$. The model for cognitive effort discounting explained 5.1% of the variance, $F(2, 259) = 6.898$, $p < .001$. Both predictors were significant, NCS: $t = 2.95$, $\beta = .18$, $p =.003$; d': $t = 2.092$, $\beta = .127$, $p = .037$.

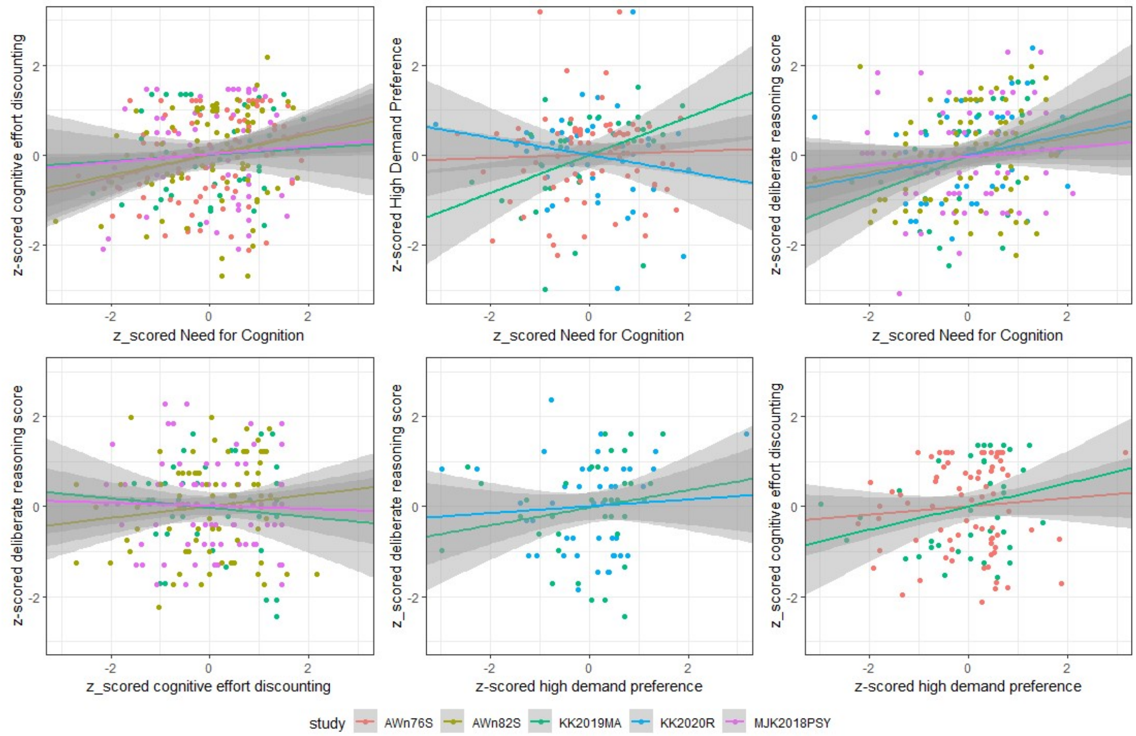

Figure 6: relationship between z-scored (per study) task indices, Top row: Correlation of COGED with NCS, DST with NCS and RQ with NCS. Bottom row: Correlation between RQ and COGED, RQ and DST, and COGED and DST. Study 1 = Awn76S, Study 2 = Awn82S, Study 3 = MJK2018PSY, Study 4 = KK2019MA, Study 5 = KK2019R

Table 7 provides a brief overview of the characteristics of the four cognitive effort measures included in the study.

Table 7.

Characteristics of the four tasks used to measure cognitive effort.

Task	Demand selection task	Cognitive effort discounting	Rational reasoning tasks	Need for Cognition scale
Duration	With instructions: 25 min	80 min for 6 levels, 30-40 min for 3-4 levels	often below 20 min (15-40 min)	5 min
Confounder	Color preference	Reward sensitivity	Gender differences	none

Test-retest reliability	Low	Unknown	Good	Good
Intrinsic motivation	Yes	No if monetary incentives given	Yes	Yes
Control for algorithmic ability	No	Yes	No	N/A
Suitable for fMRI/pupillometry	Yes	Yes	Yes	No
Patient vs control difference	No (Gold et al., 2015; Patzelt et al., 2019)	Yes, e.g. (Culbreth et al., 2016), (Chang et al., 2020)	Yes, e.g., (Mækelæ et al., 2018) (Lewton, Ashwin, & Brosnan, 2019)	Puveendrakumaran et al. 2020; Broyd et al. 2019, Brosnan et al 2017
Developmental trajectories	None yet	Yes, (Chevalier, 2018)		

General Discussion

The willingness with which we engage in cognitively demanding activities has been operationalized with a range of measurement tools. Already in the 1980s a questionnaire was developed, the Need for Cognition scale. This scale is a reliable and valid trait measure of cognitive effort (Bruinsma & Crutzen, 2018; Cacioppo & Petty, 1982; Hussey & Hughes, 2020) and also in our studies showed very good internal consistency and test-retest reliability. The samples in the five studies reported here were also very similar in their Need for Cognition, suggesting that they come from a similar population.

Through 5 studies we found that NCS was positively related to rational reasoning, replicating previous findings (Thomson & Oppenheimer, 2016; Toplak et al., 2014; West et al., 2008). Similarly, we found a positive relationship between cognitive effort discounting measured with the COG-ED and NCS. However, there was no strong relationship between rational reasoning and cognitive effort discounting. Demand avoidance, as measured with the DST, was not related to NCS, cognitive effort discounting, or to rational reasoning.

Cognitive effort remains an elusive concept to capture. Indeed, cognitive effort may depend on differences in cognitive ability, intrinsic- and extrinsic motivation, reward sensitivity, task automaticity, and effort costs (Musslick, Cohen & Shenhav, 2018). In addition, researchers should be mindful of differences in state and trait, as well as differences in subjective and objective effort. Consistent with recent computational modeling work, individual difference correlations across tasks are weak – indeed, e.g. they may be absent as we found for the DST. Thus, multiple tasks should be considered for estimating the latent construct of sensitivity to cognitive effort costs. We highlight strengths and weaknesses of the tasks and address the intricate interplay of cognitive effort with motivation, cognitive ability and alike below.

Rational reasoning items measure cognitive effort

Rational reasoning tasks are a convenient, fast, implicit measure of a tendency to expend cognitive effort and successfully engage in deliberate reasoning. The tasks implicitly measure intrinsic motivation to solve novel tasks and the ability to expend necessary cognitive effort. However, individuals may be affected by attitudes toward problem-solving tasks, and it is hard to know why some spend more effort on the tasks. Rational reasoning tasks are most likely assessing multiple facets of reasoning and decision-making, and we do not propose that they are measuring one underlying uniform construct. However, performance across rational reasoning tasks are moderately correlated and individual differences in task performance seem to be stable. Performance on these tasks seem to be contingent on a combination of sufficient cognitive ability and a willingness to expend cognitive effort through task engagement, as a consistent correlation with NCS corroborates (Stanovich, 2009; Toplak et al., 2011; 2014; West et al., 2008). However, at least for the cognitive reflection test, performance may not be dependent on more effort. Individuals high in cognitive ability may simply have better intuitions (Raoelison, Thompson, & De Neys, 2020). Still, some of the tasks require conflict detection and overriding, while others require assessment of multiple options or reasoning fully through all options. Thus, successful performance is also dependent on sufficient cognitive effort expenditure. The rational reasoning battery differed in subjective ratings of mental demand in Study 4 and 5, which may be due to differences in the total task set completed during each study. There is a plethora of reasoning items, allowing to adjust the length and difficulty of the task set. The tasks are suitable for multiple exposures and task performance is linked to multiple real world outcomes (Pennycook & Rand, 2019). The tasks are suitable for multiple exposures (Bialek & Pennycook, 2018; Woike, 2019). They can be combined with measures of intelligence, working memory and numeracy, to investigate separate effects. In addition they are adaptable for fMRI, EEG, eye-tracking etc. Researchers should be mindful that performance is dependent on sufficient analytical and reflective abilities yet to be properly defined (Stanovich, 2018).

Cognitive effort discounting measures cognitive effort

The COG-ED is a behavioral economic approach to assess cognitive effort discounting of monetary rewards. It is a useful tool for explicitly assessing cognitive effort expenditure and assess cognitive effort costs. The task was subjectively rated as the most mentally demanding task in our studies. COG-ED is based on the n-back, a well-established working memory paradigm with parametrically varying cognitive load. Thus, a strength of the COG-ED paradigm is that performance level is adjusted to a participant's ability and performance in the practice phase. In addition, the measure can provide an estimate of analytical ability, which is convenient as this allows for correction of analytic ability which is a confounding variable with most cognitive effort measures. Similarly, Hegelstad et al. (2020) controlled for cognitive ability in a visual search and effort task, finding similar relative effort despite different visual memory accuracy. COG-ED might be influenced by individual differences in reward sensitivity, as individuals high in Need for Cognition are less sensitive to rewards (Sandra and Otto, 2018). This underlines the importance of disentangling intrinsic and extrinsic motivation. The measure is suitable for use with fMRI, EEG, Eye-tracking, PET, and pharmacology.

Demand avoidance does not measure cognitive effort

The implicit nature of the DST makes it appealing, however it is also its weakness as the task is subject to choices being influenced by factors such as side- and color preferences. The DST showed low test-retest reliability and was not perceived as very effortful. The task yielded very low demand avoidance in our samples. The task was not related to NCS, COG-ED, or rational reasoning. As such, it is hard to tell what the task is actually measuring. The validity of the task is questionable, and if it cannot reliably show demand avoidance we question if it is suitable as a measure of individual differences in demand avoidance. For future studies we recommend to use a modification of the DST, varying the effort level by changing the frequency of rule changes between rounds (Sayalı & Badre, 2019).

Limitations

Firstly, our samples are mostly students, cautioning any generalizability beyond young, healthy, well-educated participants. Secondly, only one study used all four tasks. Thirdly, test-retest reliability was only assessed for the demand selection task and NCS but not for rational reasoning or cognitive effort discounting. Fourthly, subjective effort of the task was not related to task performance on any of the tasks. Individual differences in cognitive abilities, liking of challenges, curiosity and metacognition may explain the discrepancy between performance and subjective difficulty ratings (Norman et al., 2019). Fifth, researchers should be mindful of effects of tasks performed prior to effort tasks and effects due to exhaustion and boredom.

Conclusion

We did not find that the three effort tasks measured the same latent construct of cognitive effort. However, both the COG-ED and rational reasoning task set have good associations with the NCS. Consistent with recent computational modeling work, individual difference correlations across tasks are weak – indeed, e.g. they may be absent as we found for the DST. As both the DST and COG-ED are used frequently as measures of cognitive effort our findings have large implications for interpretations of previous findings. If the two tasks are measuring different constructs then research on one task should not be interpreted as applying to the other task. Lastly, this highlights the need for developing new behavioral paradigms for measuring internal motivation to expend cognitive effort. Thus, multiple tasks should be considered for estimating the latent construct of sensitivity to cognitive effort costs.

References

Allport, G. W. (1954). *The nature of prejudice*. Oxford, England: Addison-Wesley.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *J Pers Soc Psychol, 74*(5), 1252-1265. doi:10.1037/0022-3514.74.5.1252

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

Bijleveld, E., Custers, R., & Aarts, H. (2009). The unconscious eye opener: pupil dilation reveals strategic recruitment of resources upon presentation of subliminal reward cues. *Psychol Sci, 20*(11), 1313-1315. doi:10.1111/j.1467-9280.2009.02443.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:https://doi.org/10.1016/S0361-3682(01)00052-6

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433-436. doi:10.1163/156856897X00357

Bruinsma, J., & Crutzen, R. (2018). A longitudinal study on the stability of the need for cognition. *Personality and Individual Differences, 127*, 151-161. doi:https://doi.org/10.1016/j.paid.2018.02.001

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *J Pers Soc Psychol, 42*(1), 116-131. doi:10.1037/0022-3514.42.1.116

Cacioppo, J. T., Petty, R. E., & Feng Kao, C. (1984). The Efficient Assessment of Need for Cognition. *Journal of Personality Assessment, 48*(3), 306-307. doi:10.1207/s15327752jpa4803_13

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), 7-42. doi:10.1023/A:1007850605129

Chang, W. C., Westbrook, A., Strauss, G. P., Chu, A. O. K., Chong, C. S. Y., Siu, C. M. W., . . . Chen, E. Y. H. (2020). Abnormal cognitive effort allocation and its association with amotivation in

first-episode psychosis. *Psychological Medicine, 50*(15), 2599-2609. doi:10.1017/S0033291719002769
Chevalier, N. (2018). Willing to Think Hard? The Subjective Value of Cognitive Effort in Children. *Child Development, 89*(4), 1283-1295. doi:https://doi.org/10.1111/cdev.12805
Culbreth, A., Westbrook, A., & Barch, D. (2016). Negative symptoms are associated with an increased subjective cost of cognitive effort. *J Abnorm Psychol, 125*(4), 528-536. doi:10.1037/abn0000153
Engelmann, J. B., Damaraju, E., Padmala, S., & Pessoa, L. (2009). Combined effects of attention and motivation on visual task performance: transient and sustained motivational effects. *Front Hum Neurosci, 3*, 4. doi:10.3389/neuro.09.004.2009
Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives, 19*(4), 25-42. doi:doi: 10.1257/089533005775196732
Gatzke-Kopp, L. M., Ram, N., Lydon-Staley, D. M., & DuPuis, D. (2018). Children's Sensitivity to Cost and Reward in Decision Making across Distinct Domains of Probability, Effort, and Delay. *Journal of Behavioral Decision Making, 31*(1), 12-24. doi:doi:10.1002/bdm.2038
Gold, J. M., Kool, W., Botvinick, M. M., Hubzin, L., August, S., & Waltz, J. A. (2015). Cognitive effort avoidance and detection in people with schizophrenia. *Cogn Affect Behav Neurosci, 15*(1), 145-154. doi:10.3758/s13415-014-0308-5
Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in Psychology* (Vol. 52, pp. 139-183): North-Holland.
Holroyd, C. B. (2016). The waste disposal problem of effortful control. In *Motivation and cognitive control.* (pp. 235-260). New York, NY, US: Routledge/Taylor & Francis Group.
Horan, W. P., Reddy, L. F., Barch, D. M., Buchanan, R. W., Dunayevich, E., Gold, J. M., . . . Green, M. F. (2015). Effort-Based Decision-Making Paradigms for Clinical Trials in Schizophrenia: Part 2-External Validity and Correlates. *Schizophr Bull, 41*(5), 1055-1065. doi:10.1093/schbul/sbv090
Hull, C. L. (1943). *Principles of behavior: an introduction to behavior theory*. Oxford, England: Appleton-Century.
Hussey, I., & Hughes, S. (2020). Hidden Invalidity Among 15 Commonly Used Measures in Social and Personality Psychology. *Advances in Methods and Practices in Psychological Science, 3*(2), 166-184. doi:10.1177/2515245919882903
Kahneman, D. (1973). *Attention and Effort*: Englewood Cliffs, NJ: Prentice-Hall.
Kanwal, J., Smith, K., Culbertson, J., & Kirby, S. (2017). Zipf's Law of Abbreviation and the Principle of Least Effort: Language users optimise a miniature lexicon for efficient communication. *Cognition, 165*, 45-52. doi:https://doi.org/10.1016/j.cognition.2017.05.001
Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoolbox-3. *Perception 36 ECVP Abstract Supplement*.
Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *J Exp Psychol Gen, 139*(4), 665-682. doi:10.1037/a0020198
Kreis, I., Moritz, S., & Pfuhl, G. (2020). Objective versus subjective effort in schizophrenia. *Frontiers in Psychology, 11*, 1469.
Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behav Brain Sci, 36*(6), 661-679. doi:10.1017/s0140525x12003196
Lewton, M., Ashwin, C., & Brosnan, M. (2019). Syllogistic reasoning reveals reduced bias in people with higher autistic-like traits from the general population. *Autism, 23*(5), 1311-1321. doi:10.1177/1362361318808779
Musslick, S., Cohen, J. D., & Shenhav, A. (2018). *Estimating the costs of cognitive control from task performance: theoretical validation and potential pitfalls*. Proceedings of the 40th Annual Meeting of the Cognitive Science Society
Mækelæ, M. J., Moritz, S., & Pfuhl, G. (2018). Are psychotic experiences related to poorer reflective reasoning? *Frontiers in Psychology, 9*, 122.
Norman, E., Pfuhl, G., Sæle, R. G., Svartdal, F., Låg, T., & Dahl, T. I. (2019). Metacognition in psychology. *Review of General Psychology, 23*(4), 403-424.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Hum Brain Mapp, 25*(1), 46-59. doi:10.1002/hbm.20131

Patzelt, E. H., Kool, W., Millner, A. J., & Gershman, S. J. (2019). The transdiagnostic structure of mental effort avoidance. *Scientific Reports, 9*(1), 1689. doi:10.1038/s41598-018-37802-1

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In *Handbook of research methods in personality psychology.* (pp. 224-239). New York, NY, US: The Guilford Press.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis, 10*(4), 437-442.

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

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

Pfuhl, G. (2012). Two strings to choose from: do ravens pull the easier one? *Anim Cogn, 15*(4), 549-557. doi:10.1007/s10071-012-0483-0

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

Reddy, L. F., Horan, W. P., Barch, D. M., Buchanan, R. W., Dunayevich, E., Gold, J. M., . . . Green, M. F. (2015). Effort-Based Decision-Making Paradigms for Clinical Trials in Schizophrenia: Part 1-Psychometric Characteristics of 5 Paradigms. *Schizophr Bull, 41*(5), 1045-1054. doi:10.1093/schbul/sbv089

Reddy, L. F., Reavis, E. A., Wynn, J. K., & Green, M. F. (2018). Pupillary responses to a cognitive effort task in schizophrenia. *Schizophr Res, 199*, 53-57. doi:10.1016/j.schres.2018.03.005

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemp Educ Psychol, 25*(1), 54-67. doi:10.1006/ceps.1999.1020

Sayalı, C., & Badre, D. (2019). Neural systems of cognitive demand avoidance. *Neuropsychologia, 123*, 41-54. doi:https://doi.org/10.1016/j.neuropsychologia.2018.06.016

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin, 134*(2), 207-222. doi:10.1037/0033-2909.134.2.207

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron, 79*(2), 217-240. doi:10.1016/j.neuron.2013.07.007

Stanovich, K. E. (2009). *What intelligence tests miss. The psychology of rational thought*: Yale University Press.

Stanovich, K. E. (2018). Miserliness in human cognition: the interaction of detection, override and mindware. *Thinking & Reasoning, 24*(4), 423-444. doi:10.1080/13546783.2018.1459314

ten Velden Hegelstad, W., Kreis, I., Tjelmeland, H., & Pfuhl, G. (2020). Psychosis and psychotic-like symptoms affect cognitive abilities but not motivation in a foraging task. *Frontiers in Psychology, 11*.

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making, 11*(1), 99-113. Retrieved from https://EconPapers.repec.org/RePEc:jdm:journl:v:11:y:2016:i:1:p:99-113

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition, 39*(7), 1275. doi:10.3758/s13421-011-0104-1

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning, 20*(2), 147-168. doi:10.1080/13546783.2013.844729

Trippas, D., Pennycook, G., Verde, M. F., & Handley, S. J. (2015). Better but still biased: Analytic cognitive style and belief bias. *Thinking & Reasoning, 21*(4), 431-445. doi:10.1080/13546783.2015.1016450

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*(2), 127-154. doi:https://doi.org/10.1016/0749-596X(89)90040-5

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

Westbrook, A., Kester, D., & Braver, T. S. (2013). What is the subjective cost of cognitive effort? Load, trait, and aging effects revealed by economic preference. *PLOS ONE, 8*(7), e68210. doi:10.1371/journal.pone.0068210

Westbrook, A., van den Bosch, R., Määttä, J. I., Hofmans, L., Papadopetraki, D., Cools, R., & Frank, M. J. (2020). Dopamine promotes cognitive effort by biasing the benefits versus costs of cognitive work. *Science, 367*(6484), 1362-1366. doi:10.1126/science.aaz5891

Woike, J. K. (2019). Upon Repeated Reflection: Consequences of Frequent Exposure to the Cognitive Reflection Test for Mechanical Turk Participants. *Front Psychol, 10*, 2646. doi:10.3389/fpsyg.2019.02646

Zénon, A., Solopchuk, O., & Pezzulo, G. (2019). An information-theoretic perspective on the costs of cognition. *Neuropsychologia, 123*, 5-18. doi:https://doi.org/10.1016/j.neuropsychologia.2018.09.013