Cognition 236 (2023) 105440

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Is all mental effort equal? The role of cognitive demand-type on effort avoidance

Jake R. Embrey[a,*], Chris Donkin[b], Ben R. Newell[a]

[a] School of Psychology, UNSW, Sydney, Australia
[b] Ludwig Maximilian University of Munich, Germany

ARTICLE INFO

Keywords:
Mental effort
Cognitive effort
Opportunity cost
Effort aversion
Boredom

ABSTRACT

Humans are often termed "cognitive misers" for their aversion to mental effort. Both in and outside the laboratory people often show preferences for low-effort tasks and are willing to forgo financial reward to avoid more demanding alternatives. Mental effort, however, does not seem to be ubiquitously avoided: people play crosswords, board games, and read novels, all as forms of leisure. While such activities undoubtedly require effort, the type of cognitive demands they impose appear markedly different from the tasks typically used in psychological research on mental effort (e.g., N-Back, Stroop Task, vigilance tasks). We investigate the effect disparate demands, such as tasks which require problem solving (e.g., solve the missing number: 1, 3, 7, 15, 31,?) compared to those which require rule-implementation (e.g., N-Back task), have on people's aversion to or preference for increased mental effort. Across four experiments using three different tasks, and a mixture of online and lab-based settings, we find that aversion to effort remains largely stable regardless of the types of cognitive demands a task imposes. The results are discussed in terms of other factors that might induce the pursuit of mental effort over and above the type of cognitive demands imposed by a task.

1. Introduction

All those familiar with writing an opening sentence are well-aware of mental effort's aversive nature, in the same way all readers know the pains of parsing a poorly constructed one.

The idea that effort is aversive is intuitively appealing and has underpinned several significant psychological theories. Originally, such theories were concerned with aversion to physical effort (e.g., the law of less work, Hull, 1943), but when psychology's focus shifted towards unobservable cognitive mechanisms, the law of less work was similarly applied to cognition (Kahneman, 1973).

Following Inzlicht, Shenhav, and Olivola (2018) definition, we take effort to be the mediator between how well someone can perform on a task and how well they actually perform. Typically, when difficulty is held constant, performance increases as the effort exerted increases. For example, with minimal effort one might have a rough idea of how many times 21 goes into 351, but considerably more effort would be needed to obtain the precise answer.

The assumption that unnecessary bouts of mental effort are typically avoided has been invoked to explain a range of behaviours: simple strategy selection during arithmetic tasks (Baroody & Ginsburg, 1986); the role of heuristics and biases in decision making (Tversky & Kahneman, 1974); and human tendencies towards satisficing (Simon, 1955). Just as the financially thrifty are cautious with money, people in general seem unwilling to exert mental effort in the absence of sufficient reward. Consequently, Fiske and Taylor (1991) termed human beings "cognitive misers".

In this paper we focus on how effort aversion (or seeking) behaviours differ across tasks with differing cognitive demands. Explicitly, we ask whether people are less averse to increasing their exerted effort during tasks which require problem-solving (and have the potential for rule-discovery) as opposed to those which require maintained attention and rule-implementation.

1.1. Why is effort costly?

Shenhav et al. (2017) suggest cognitive theories of mental effort can be categorised into two types: opportunity-cost and intrinsic-cost models. Opportunity-cost accounts (Kurzban, Duckworth, Kable, & Myers, 2013; Otto & Daw, 2019) argue the *sense of effort* experienced when attending to a demanding task is a motivating force that drives agents away from unrewarding, yet effortful tasks, towards potentially

* Corresponding author.
E-mail address: j.embrey@unsw.edu.au (J.R. Embrey).

https://doi.org/10.1016/j.cognition.2023.105440
Received 9 June 2022; Received in revised form 2 March 2023; Accepted 10 March 2023

more rewarding alternatives. Analogous to how hunger drives us towards food, the sense of effort guides our behaviour towards rewarding activities. While similar in nature, intrinsic-cost accounts posit that there is a limit to the amount of effort one can allocate at a single time. Allocation of effort is therefore monitored by attributing an intrinsic cost to any effort that is exerted, which can then be balanced against any potential extrinsic rewards to determine whether a task is worthwhile. While both intrinsic- and opportunity-cost models assume the capacity for effort allocation is finite, they are markedly different from physical resource accounts (e.g., Muraven & Baumeister, 2000), which posit that a physical resource such as blood glucose underpins our ability to exert effort and the phenomenological cost of doing so (Gailliot & Baumeister, 2007). Recently, physical-resource theories of mental effort have undergone scrutiny (Kurzban, 2010; Kurzban, 2016) for failing to replicate and their inability to explain instances of sustained mental effort despite diminished blood glucose levels (e.g., after intense physical activity; Tomporowski, 2003).

1.2. Effort aversion in the laboratory

Many experiments examining mental effort have focused deliberately on a specific and well-defined aspect: cognitive control. Allocating cognitive control – that is, overriding more automatic processes to engage in deliberative information processing – is difficult and is the cognitive aspect necessary for adequate performance in a range of tasks (Shenhav et al., 2017). For example, the control required to suppress saying "green" when the stimulus "**GREEN**" is presented in red ink in a Stroop task (Cohen, Dunbar, & McClelland, 1990). As a result of this focus on the cost of control, much of the mental effort literature has used tasks where maintaining cognitive control is the key to task performance (e.g., N-Back, Stroop tasks); in the absence of time constraints, however, the rules and strategies of these tasks are not difficult to implement.

Kool, McGuire, Rosen, and Botvinick (2010) directly assessed whether people were averse to exerting mental effort, rather than invoking effort aversion as an explanatory tool itself (e.g., Baroody & Ginsburg, 1986; Tversky & Kahneman, 1974). Across five experiments, Kool and colleagues found that effort was avoided when external factors such as time and performance were controlled for, implying the exertion of effort itself carries an intrinsic cost. Kool et al. used *demand selection tasks,* which require participants to choose what they would like to do in a coming block of trials (represented by two decks of cards). Within the experiment, demand was defined by how frequently the tasks within a given block would change. For example, participants had to respond whether a number (between 1 and 9) was odd or even, or whether it was greater or <5. Here, participants showed strong preferences for options (i.e., decks of cards) which required less switching between the different rules. A result which is taken as indicating the effortful (and costly) nature of cognitive control required to switch from one decision rule (e.g., is the number odd or even) to another (e.g., is the number greater or <5).

Since Kool et al.' (2010) work, a multitude of experiments have further assessed effort aversion across a range of behavioural tasks (for review, see Kool & Botvinick, 2018): for example, people are willing to forgo financial reward in order to avoid harder difficulty levels in a working memory task (N-Back; Westbrook, Kester, & Braver, 2013); require financial incentive to maintain performance in a simple counting tasks (distinguishing between 7- and 9-sided polygons; Caplin, Csaba, Leahy, & Nov, 2020); and are willing to forgo reward to avoid implementing complex finite-state rules (Oprea, 2020).

Perhaps unsurprisingly, effort avoidance is ubiquitous in such tasks, with effort seeking rare even at the individual level (e.g., only two of 85 participants opted for a more demanding N-Back task when rewards were equal; Westbrook & Braver, 2015). Such overt effort avoidance however does not cohere with our experiences in day-to-day life where people regularly engage in effortful tasks for leisure. For example, playing video games, solving crosswords, or, more recently, solving the daily Wordle. While such activities may possess extrinsic rewards (e.g., social incentives to beat friends or improve skills), they are also phenomenologically enjoyable. Such activities are also markedly different from the tasks characteristically used in laboratory studies of mental effort. Furthermore, when playing such games (when mentally demanding tasks are enjoyable we call them games) participants choose difficulty levels which match their skill level rather than selecting the easiest difficulty (e.g., Baranes, Oudeyer, & Gottlieb, 2014). It is no coincidence game developers invest significant time and money ensuring their skill-based matchmaking algorithms work efficiently when online players are matched against one another (Graepel & Herbrich, 2006).

Research in a similar vein has been undertaken via work investigating positive 'flow' states (Csikszentmihalyi & Larson, 2014) which arise during bouts of cognitive exertion. Initially, theories of flow relied on anecdotes, such as the positive affect experienced by video gamers or athletes during bouts of intense focus (Csikszentmihalyi, 1975). More recently, however, researchers have aimed to induce states of flow in controlled environments using games such as PacMan and Tetris (for review, see Moller, Meier, & Wall, 2010). While participants in these experiments purportedly experienced flow-like states (measured by post-task self-report questionnaires) when difficulty levels matched their skillsets, participants were never required to explicitly choose which difficulty level they preferred to play (e.g., Keller & Bless, 2008; Rheinberg & Vollmeyer, 2003). It is therefore difficult to determine whether participants were genuinely less averse to increased effort in such tasks.

1.3. Additional influences: intrinsic incentives and boredom

Other than possible extrinsic rewards, demanding tasks which people leisurely partake in may also differ in terms of the effort required to perform them. As previously mentioned, psychological research typically employs tasks which require rule implementation and maintained attention. While leisurely tasks also require sufficient levels of attention, the rules for success are typically less clearly defined, and often require problem solving and the discovery of underlying rules in order to perform well. For example, although the formal rules of chess are well-defined, there is no precise strategy one can follow for guaranteed success. Similarly, there is no pre-defined strategies for success in most crosswords, boards games, or video games. This stands in contrast to typical experimental tasks such as the N-Back, where the rules (i.e., does the current symbol match the symbol from *N* turns ago) *and* strategy (i.e., maintain a string of the previous *N* symbols in working memory) required to succeed are well-defined.

A further difference between rule-implementation tasks and those which require rule- or strategy-discovery is the effect financial incentives have on people's performance. Osborn Popp, Newell, Bartels, and Gureckis (2022) find that while increased incentivisation can increase performance when a categorisation task (SHJ task; Shepard, Hovland, & Jenkins, 1961) requires rule-implementation, incentivisation bears no effect when such a task requires an individual to discover the underlying rules of the task. Similar findings have also been observed in tasks which require overcoming common cognitive biases (Enke et al., 2021), where participants are unable to solve problems such as those involving base rates, even when incentives are extraordinarily high (equivalent to $2350 USD). This stands in contrast to counting (Caplin et al., 2020) and repeated button-press tasks (DellaVigna & Pope, 2018) where increased incentives improve performance. People also report differences in the sense of effort experienced by disparate mental tasks such as those that require attending (e.g., counting the drips of a tap) compared to assessing (e.g., weighing up whether to purchase a banana or apple) (Robinson & Morsella, 2014). While the above findings do not necessarily imply exerting cognitive effort in abstract, problem-solving tasks is less (or more) aversive than attentionally demanding, rule-implementation tasks, it does suggest the type of thinking required to solve them is qualitatively different.

Similarly, the popularity of enjoyable, yet demanding games (e.g., chess) does not imply that increased effort is sought for its own sake in these instances. It does however raise questions about whether theories of mental effort can account for behaviours beyond those observed in the lab where the type of cognitive demands imposed are largely homogenous. Before theories of mental effort posit *why* effort is costly, it is of interest to assess whether aversion to effort is equal across varying types of demand.

Instances of effort seeking may also arise due to the alternative, yet effortful task being more aversive than a less effortful option. One such example is where the alternative is boring. Bench and Lench (2019) find that participants in high-boredom conditions are more likely to sensation seek than those in low-boredom alternatives. In some instances, participants are even willing to seek negative sensations (e.g., electric shocks, disgusting images) to alleviate boredom (Bench & Lench, 2019; Nederkoorn, Vancleef, Wilkenhöner, Claes, & Havermans, 2016). Furthermore, recent work by Wu, Ferguson, and Inzlicht (2022) shows that people would rather complete mathematical working-memory problems than a boring alternative – 'doing nothing'. In this case the cost of mental effort is not necessarily diminished, it is however less costly (or aversive) than a state of boredom and is therefore sought.

1.4. Overview of current work

The current set of experiments assess effort avoidance (and seeking) across three types of tasks. Specifically, one which is attentionally demanding and requires rule-implementation (N-Back task), and two which involve problem solving and/or rule-discovery (number sequence problems and anagrams). All four experiments involved several difficulty levels (measured by response time, accuracy, and self-report), equal reward structures, and gave participants an explicit choice between difficulty levels using the Cognitive Effort Discounting task (COG-ED) developed by Westbrook et al. (2013). Experiments 1–3 assessed people's aversion to increased difficulty (which requires increased effort) within a type of task: N-Back (Experiment 1), number sequence problems (Experiment 2) and anagrams (Experiment 3). Experiment 4 built on the first three by assessing whether effort avoidance differed when participants are given choices between different types of tasks (e.g., N-Back or anagrams) while still varying the difficulty level between the two offers.

1.5. The cognitive effort discounting paradigm (COG-ED)

First developed and implemented by Westbrook et al. (2013), the COG-ED task economises people's propensity to avoid effort by allowing an explicit financial cost to be placed on it. The original COG-ED was employed alongside a N-Back task containing six difficulty levels (1-Back through to 6-Back). After some initial exposure to the different N-Back levels, participants were offered a choice of what they would like to do next. This choice comes in the form of a discounting paradigm, which starts by presenting participants with the option of completing a 1-Back task for $1.00 or a higher effort option for $2.00 (e.g., 2-Back for $2.00). If participants chose the high effort option, the following choice would be between the 1-Back for $1.50 or the high effort option for $2.00 (e.g., 1-Back for $1.50 or 2-Back for $2.00). Alternatively, if the 1-Back was chosen, the next choice would be between the 1-Back for $0.50 or high effort for $2.00. Participants completed six choices for each comparison (1-Back compared to all higher levels) and the amount the offer changed (up or down depending on the previous choice) halved after each choice. The final value of the low effort option was taken as the indifference point – which is transformed to determine the proportion of reward participants were willing to forgo to avoid the more demanding task.

Westbrook et al. (2013) found that as the difficulty of the high effort option increased participants were willing to forgo greater amounts of reward. Furthermore, this willingness to forgo reward was not entirely accounted for by decrements in performance.

2. Experiments 1–3

Experiments 1–3 employed the COG-ED task developed by Westbrook et al. (2013) with some adjustments (which allowed for effort seeking) and used three tasks (one task per Experiment) which were designed to impose different types of demand on participants.

Experiment 1 used a N-Back task often used in psychological literature, but included a novel condition, 0-Back, designed to require minimal effort. We call our version of the 0-Back task novel because, though previous research has used versions of the N-back where participants were required to respond when a certain letter (e.g., 'X') appears on screen, we made the task even more monotonous in Experiment 1 by requiring participants to respond whenever *any* letter appeared. In a 0-Back, participants were required to respond each time any letter appeared on screen, they did not need to keep track of any previous letters displayed. The other N-Back levels used were 1-Back through to 5-Back.

Experiment 2 used number sequence problems (Fig. 1; NSPs) which came in three difficulty levels – Easy, Medium, and Hard. The aim when solving NSPs was to figure out the arithmetic pattern – starting from the top right segment and continuing clockwise around the circle – then input the missing number. All NSPs were developed by the authors and the difficulty of the problems (measured by response time, accuracy, and self-reported difficulty ratings) were determined in a separate norming experiment (Supplementary Materials, Experiment S1).

Experiment 3 used anagrams which also came in three difficulty levels – 3-letter, 5-letter, and 7-letter. All anagrams had a single solution and none of the solutions were abbreviations, slang, or pronouns. All of the letter scrambles were pre-determined (by the authors). The difficulty of the various anagrams (measured by response time, accuracy, and self-reported difficulty ratings) was assessed in a separate experiment (Supplementary Materials; Experiment S2).

The effortful aspect of Experiment 1 (N-back) is the attention required to maintain a list of the previous N letters in working memory; as the N value increases, the demands, and therefore effort required to perform well, significantly increase. The rule and strategy (maintaining a string of the previous N letters in mind) for the N-Back task was made clear to participants.

The effortful aspects of Experiments 2 and 3 could broadly be defined as problem solving. In Experiment 2, the problem was discovering the arithmetic pattern the numbers follow and then implementing this rule to figure out the missing number. Participants were told that each problem follows a specific pattern, but were not provided with any strategies or hints. In Experiment 3, verbal reasoning was required to unscramble the jumbled word and then input the answer. Furthermore, both tasks (NSPs and anagrams) were designed and used to mimic certain, though perhaps superficial, aspects of demanding tasks people play for fun: brain teasers and Scrabble (and Wordle), respectively.

Our aim across these three experiments was to assess how effort avoidance (or seeking) differed across the three tasks. Specifically, whether effort aversion differed between tasks which only required rule-implementation (N-Back) and those which required more abstract, problem-solving skills and rule-discovery (NSPs and anagrams). We expected participants to show less aversion to increased effort in NSP and anagram tasks than the N-Back task.

Secondly, we investigated whether the inclusion of a boring 0-Back task (in Experiment 1), and the simple nature of the problems from Easy and 3-letter sets in Experiments 2 and 3, would lead participants to voluntarily seek more difficult problems: if participants found 0-Back, Easy, and 3-letter problems boring, they may prefer to choose more difficult sets.

Experiment 4 extends this design by assessing effort avoidance across task types (e.g., choosing between a 1-Back task and 5-letter anagrams). This design allows us to test whether some types of demanding tasks (e.

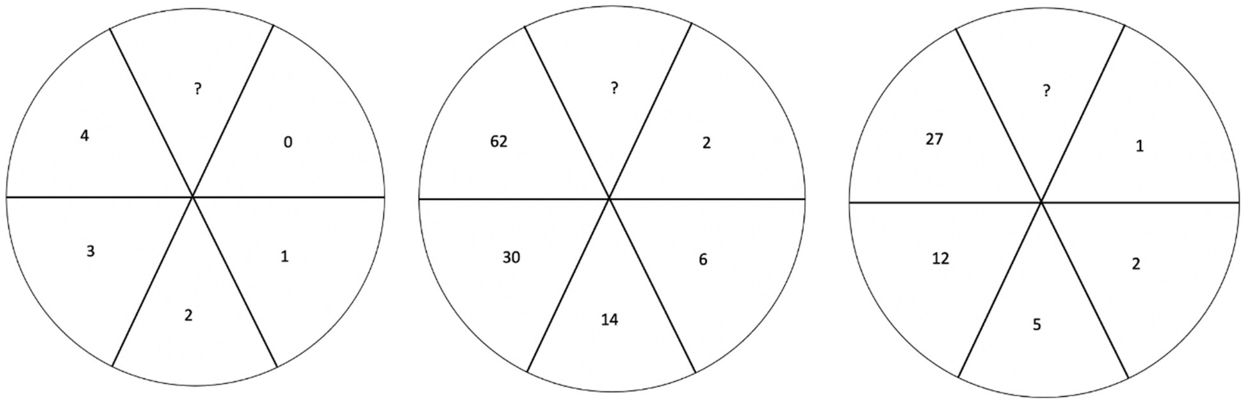

Fig. 1. Three examples of the number sequence problems from Experiment 2. Left is taken from the Easy set: the pattern is "plus-1" and the solution is 5. Centre is taken from the Medium set: the pattern is "double previous number plus-2" and the answer is 126. Right is taken from the Hard set: the pattern is "double, double plus-1, double plus-2…" and the answer is 58. Note there is occasionally more than one rule which can be used to solve a problem.

g., preferring anagrams to N-Back) were preferred over others despite differences in the effort required to complete them.

2.1. Method

2.1.1. Participants

A power analysis (δ = 0.8) was conducted to determine the sample size needed to detect a medium effect size (γ = 0.5). The results indicated 34 participants would be sufficient to detect an effect of this size (at α = 0.05) using a within-subjects design, we however aimed to sample 50 participants per experiment which was comparable to similar work (Kool et al., 2010; Westbrook et al., 2013).

Participants in Experiment 1 were 39 undergraduate[1] psychology students (M_{age} = 19.95 years; 26 females and 13 males) enrolled at the University of New South Wales. Participants were reimbursed with course credit and were paid a bonus depending on their choices in the COG-ED task (M = $5.11AUD).

Participants in Experiments 2 (N = 52; M_{age} = 26.29; 31 females, 20 males, 1 non-binary) and 3 (N = 49; M_{age} = 40.82; 17 female, 31 male, and 1 non-binary) completed the experiment via the Prolific platform. Participants were paid a flat rate (£3.25 in Experiment 2; £3.75 in Experiment 3) and were also paid a bonus depending on their choices in the COG-ED task (M = £1.54 in Experiment 2; M = £1.79 in Experiment 3).

All experiments were approved by the UNSW School of Psychology Ethics Committee (Approval number: HREAP 3477).

2.1.2. Materials

Experiment 1 was run in the UNSW Cognition lab on HP Desktop computers. The experiment was coded in jsPsych (De Leeuw, 2015) and JavaScript.

Experiments 2 and 3 were completed by participants on their own desktops or laptops (mobile phones were not permitted). The experiments were run online via a UNSW server and were coded in the same manner as Experiment 1.

2.1.3. Experiment 1: N-Back design and training phase

The N-Back task consisted of six difficulty levels (N = 0–5) and drew from 10 consonants (Z, X, C, V, B, N, R, P, T, S). There were 30 sequential trials (letters displayed) in each run. Each letter was displayed for 2000 ms and there was a post-trial gap of 1500 ms. Participants were required to press 'm' (during the 2000 ms trial) on their keyboard if the current letter matched the letter from N turns ago. During a 0-Back task letters were randomly drawn and participants were required to press 'm' for each letter. Each run lasted approximately 1-min and 45 s.

Performance feedback was also given after each run, although performance did not affect participants' bonus pay. Feedback was in the following form: "You correctly identified x of the X matching items. You incorrectly identified y of Y non-matching items. On average, you got z-percent correct."

On a single trial the chance of the current letter matching the one from N trials ago was 33% – excluding the 0-Back task where every trial was a target.[2] Each task (excluding the 0-Back) therefore had an approximate average target rate of 33% over each run. This was to ensure the experienced difficulty did not fundamentally differ between runs and participants. As a consequence, participants' average accuracy would be 67% if they did not respond at all during a run since 67% of trials, on average, were non-matches.

2.1.4. Experiment 2: NSP design and training phase

NSPs came in three sets – Easy, Medium, and Hard – which were based on results from the pilot experiment. The difficulty results ("How hard did you find the previous problem?"), answered on a 0–100 scale, for NSPs obtained in the pilot experiment (Supplementary Materials, Experiment S1) are as follows. Mean difficulty rating: Easy = *4.06* (range: 0*.73–6.27*), Medium = *33.89* (range: *23.00–41.88*), and Hard = *72.48* (range: *61.96–93.27)*.

In the training phase, participants completed two runs of each set (six in total; counterbalanced order) and each run lasted 3-min. After participants submitted an answer for a question, they were given feedback ("correct!" or "wrong!") before moving onto the next problem. The 3-min timer only included time spent on a problem, not time spent on the feedback screen. There was no minimum nor maximum time participants had to spend on a problem within a set, they were however encouraged to try and solve as many as they could. Furthermore, as in the N-Back task participants were not explicitly aware of how much time remained during a run of problems (i.e., there was no timer available).

To minimise any obvious demand effects, difficulty levels were not presented to participants as "Easy", "Medium", or "Hard". Instead,

[1] The number of participants for Experiment 1 was less than intended (i.e., 50) due to a nationwide Covid-19 lockdown in July 2021. The remainder of the experiments were therefore run online.

[2] This alteration to the 0-Back task meant that the target letter appeared on every trial, rather than with a 33% chance on any given trial. Participants could therefore predict with 100% accuracy whether the next letter would be a target or not. While this change may have made the task more monotonous and boring (as intended), it also adds a potential confound as predictability differs between the N-Back conditions.

participants were informed there were 3 different sets of number sequence problems: “Blue”, “Orange”, and “Red”. These colours corresponded to a particular difficulty which was randomised between participants.

2.1.5. Experiment 3: anagram design and training phase

Anagrams also came in three sets – 3-letter, 5-letter, and 7-letter. Analogous to NSPs these problems were tested in a pilot experiment to ensure the perceived difficulty levels were sufficiently different from one another. The difficulty results (“How hard did you find the previous problem?”), answered on a 0–100 scale, for anagrams obtained in the pilot experiment (Supplementary Materials, Experiment S2) are as follows. Mean difficulty rating: 3-letter = *6.63* (range: 0*.86–29.33*), 5-letter = *33.89* (range: *36.20–80.92*), and 7-letter = *63.89* (range: *24.27–88.89*).

The training phase was analogous to the NSP training phase. Furthermore, difficulty levels were also presented to participants as coloured sets – “Red”, “Blue”, or “Orange” – which corresponded to a particular difficulty level.

2.1.6. COG-ED design

We adapted the COG-ED task (Westbrook et al., 2013) to allow participants to prefer either the more effortful or less effortful option for each comparison. For Experiments 1–3 the first comparison was between a low-effort option for \$2.00 and a high-effort option for \$2.00. Whichever option was chosen first (either the low or high-effort option) reduced by \$1.00 on the next trial; this option then increased (decreased) by half of the last adjustment if the alternate (same option) was chosen; this titration occurred for seven trials and the final value (of the titrated option) was taken as the indifference point (Fig. 2).[3] It is important to note that whichever option was first chosen was the option which titrated over the remaining six choices, this was to ensure offers did not rise above \$2.00. Participants were therefore forgoing reward, rather than being offered greater rewards for doing a task they were averse to.

Comparisons in Experiment 1 were 0-Back vs 1- through to 5-Back and 1-Back vs 2- through to 5-Back (nine comparisons total) and were presented in a random order. Comparisons in Experiment 2 (and Experiment 3) were Easy (3-letter) vs Medium (5-letter), Easy (3-letter) vs Hard (7-letter), and Medium (5-letter) vs Hard (7-letter). In Experiment 1, all 0-Back comparisons were presented first (in a random order), followed by the 1-Back comparisons (in a random order). The presentation order for Experiments 2 and 3 was entirely randomised.

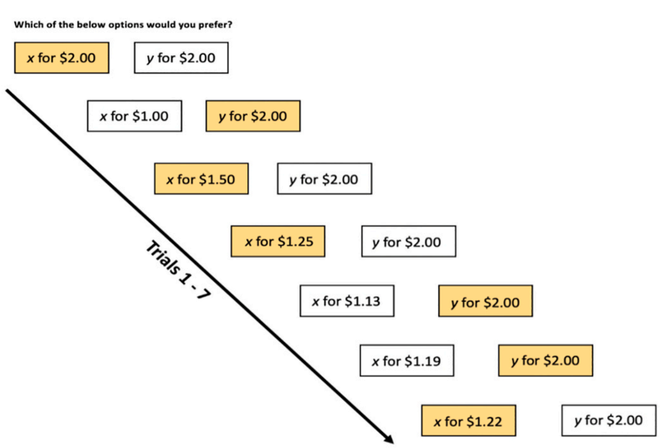

Fig. 2. A schematic of a COG-ED run of seven trials for a comparison between a low-effort and high-effort option (where *x* and *y* stand for a high- or low-effort option respectively; for example 1-Back or 2-Back). Highlighted options indicate which option was chosen on each trial. In this example the participant's indifference point would be \$1.22, indicating they are willing to forgo 39% of their potential reward in order to complete option *x*.

2.1.7. Procedure

The structure of all three experiments (with minor variations described below) was: training phase > COG-ED > run of randomly selected task for \$*X* > post-task questions.

Participants in all three Experiments were given instructions on the task they would need to complete and how to complete it (N-Back, NSPs, anagrams). Participants were also told that there was no minimum percent they needed to get correct in the task (or minimum number of problems they needed to solve) in order to be paid, but that participants would at least need to try (“not entering random letters every trial or browsing another tab”). It should however be noted that participants were paid a bonus regardless of performance (following precedent – see Westbrook et al., 2013).

Following the instructions participants completed the training phase of the experiment (described for each Experiment above).

After completing the training phase participants moved onto the COG-ED phase of the experiment. Participants were told they would be choosing which option they would “prefer to complete for the amount of money specified” and that one of their choices would be randomly selected at the end of the titration to be played out. In Experiment 1 this choice was between completing three runs of one level of the N-Back task (which lasted approximately 1 min and 45 s each). In Experiments 2 and 3 the choice was between completing a 3-min set of a particular difficulty (e.g., a set of Blue Anagrams).

Participants in Experiments 2 and 3 were also given reminders of the types of problems in each coloured set before starting the COG-ED task. For example, a problem from each coloured set of NSPs, or in Experiment 3 “Blue Set anagrams are 3-letters, Red Set anagrams are 5-letters, and Orange Set anagrams are 7-letters”. Participants were again reminded that there was no minimum percent correct they needed to achieve in order to be paid, although we would monitor their behaviour in order to check they were “trying”.

Once all titration trials were completed, in all three experiments one of a participant's choices from the titration (e.g., Medium NSPs for \$1.50) was randomly selected and they completed a run of the selected task for the specified amount of money.

In Experiment 1, after completing the final three runs of the randomly selected N-Back task participants were debriefed and provided a completion code. Participants in Experiments 2 and 3 were required to rate how difficult (on a 0–100 scale, where 0 is labelled “Easy!” and 100 is labelled “Hard!”) they found each coloured set before receiving their debrief; this was to check if participants confused the coloured sets (e.g., rating a Hard set as significantly easier than the Medium set).

3. Results experiments 1–3

3.1. Experiment 1 (N-Back)

3.1.1. Training phase

Performance for the N-Back task was measured by participants' accuracy in each run (Table 1). Participants' accuracy generally reduced as the difficulty level (i.e., N value) increased. The one exception to this was the 0-Back task ($M = 95.94\%$ correct); this discrepancy however was driven by four participants with comparatively low average accuracy in the 0-Back task (< 80%) who stopped responding for several consecutive trials. Each trial is a target trial in the 0-Back task, unlike 1–5-Back tasks where only 33% of trials are targets. Periods of inattentiveness therefore lead to greater impacts on accuracy in the 0-Back task compared to tasks where $N > 0$. When these four inattentive

[3] The changes in the offered value (to 3 d.p.) across the seven choices are \$1.00, 0.50, 0.25, 0.125, 0.062, 0.031, 0.015.

Table 1
Average accuracy per N-back level in the training phase of Experiment 1.

N-Back Level	Percent Correct	Standard Error
0	95.94	0.02
1	98.72	0.42
2	93.72	0.94
3	82.10	1.38
4	75.99	1.29
5	72.77	1.35

participants are excluded the 0-Back average rises to 99.29%.

3.1.2. COG-ED task

Indifference points for each comparison (and each participant) were converted to a subjective value (SV), where subjective value is indicative of the proportion of reward a participant was willing to forgo in order to avoid one of the options (low or high effort) in the comparison. This conversion was done by dividing the indifference point for each participant for each comparison by 2 (i.e., the amount offered for the alternative option – $2.00) and then subtracting 1. For example, a participant with an indifference point of $1.22 between a 1-Back and 2-Back task (as in the example in Fig. 2, if *x* and *y* were replaced with 1-Back and 2-Back, respectively) preferring 1-Back, would have an SV of −0.39, indicating they were willing to forgo 39% of the offered reward to avoid a 2-Back task.

A conversion was then performed on the indifference points of participants who preferred the more difficult option by inverting their SV for that comparison around 0. For example, if a participant had an indifference point of $1.22 for the 1-Back or 2-Back comparison, but preferred the 2-Back task, their SV would be +0.39. This conversion allows participants' preferences (i.e., do they prefer the easier or harder task), and the amount they were willing to forgo, to be deduced entirely from their SV: values between 0 and − 1 indicate participants were willing to forgo reward to do the easier option (and the proportion of reward they were willing to forgo) and values between 0 and + 1 indicate participants were willing to forgo reward to do the harder option (and the proportion of reward they were willing to forgo).

Performance discrepancy scores were also calculated by subtracting a participant's average accuracy in the harder option from their average accuracy in the easier option. For example, if a participant had an overall accuracy of 95% for 1-Back and 80% for 2-Back, they would have a performance discrepancy score of −15 for the 1-Back or 2-Back comparison.

On average, participants preferred the easier option from each comparison and were willing to forgo reward in order to avoid the more difficult alternative (Fig. 3). Exceptions to this were the 0-Back or 1-Back ($M = 0.02$, $t(38) = 0.62$, $p = .54$), 0-Back or 2-Back ($M = -0.07$, $t(38) = 0.99$, $p = .33$), and the 1-Back or 2-Back ($M = -0.09$, $t(38) = 2.17$, $p = .034$) where participants were, on average, indifferent between the two choices.[4]

Differences in subjective values across different comparisons were assessed using linear mixed modelling. Linear mixed models were run using the *lmer* package in R; parameters were estimated by maximising restricted log likelihood.

The baseline model contained only random intercepts (per participant) to predict a participant's SV for each comparison; subsequent models added the different comparison conditions (coded as a factor), performance discrepancy (continuous), with the fourth model containing both comparison and performance discrepancy as fixed factors. Random slopes were not added as each participant only had one SV per comparison (i.e., each comparison was asked once per participant in the COG-ED phase).

A summary of these models (including Akaike weights, or model probabilities, Wagenmakers & Farrell, 2004) and their comparisons is collated in Table 2. All models performed better than the baseline model which contained only random intercepts, with the full model (containing both comparison and performance discrepancy as fixed factors) performing better than the two which contained either comparison or performance discrepancy individually. These results indicate that comparison and performance discrepancy independently account for variance in SV.[5]

As with any decision-making paradigm it is plausible that individuals' responses are noisy and do not truly reflect their underlying choice preferences. This problem, however, is compounded in the COG-ED task due to the disproportionate influence of the first trial which wholly determines the sign of an individual's SV. To test the influence of noise on participants' SVs, we fit the data with a two-free-parameter model: one for an individual's tendency to seek effort and another which sets how deterministic the decision process is. The results of these model fits for all four Experiments are collated after the discussion of Experiment 4. To foreshadow, the results of the modelling is largely consistent with the behavioural data reported within each Results section.

3.2. Experiment 2 (Number sequence problems)

3.2.1. Training phase and difficulty ratings

Participants' average performance (percent correct and response time) as well as mean difficulty ratings (measured post-experiment) are collated in Table 3. Response times per problem, on average, increased as the set difficulty increased. As a result, participants completed fewer problems when executing 3-min blocks of the more difficult problems.

3.2.2. COG-ED phase

A proportion of participants across Experiments 2 and 3 (19.23% and 20.41%, respectively) gave responses in their post-task difficulty ratings that did not cohere with our pre-determined difficulty levels. For example, rating a set of Hard NSPs as easier than Easy NSPs. Such an answer could be taken as indicating participants had confused the coloured sets (Red, Blue, Orange). Removing these participants however has no substantial effect on the analysis or our theoretical interpretations (see Supplementary Materials, Section 4). We therefore report the results below without the removal of any participants.

Participants, on average, preferred the easier option for all three comparisons and were willing to forgo reward in order to avoid the more difficult option (Fig. 4): Easy vs Medium ($M = -0.16$, $t(51) = 2.67$, $p = .010$), Easy vs Hard ($M = -0.42$, $t(51) = 7.20$, $p < .001$), and Medium vs Hard ($M = -0.27$ $t(51) = 3.96$, $p < .001$).

Analogous to Experiment 1 we assessed differences in individual SVs across comparisons using linear mixed modelling. We used three models in addition to a baseline model containing only random intercepts (per participant). The outputs from these models are shown in Table 4. While all models performed better than a baseline only model, they could not be significantly distinguished on their ability to predict variance in SV. Therefore, while it was clear that there were differences between the

[4] Bonferroni corrections were made for all nine comparisons therefore *p* values above 0.0055 were considered non-significant. Statistical values for the remaining six comparisons can be seen in the supplementary materials (section 3).

[5] Analogous to Kool et al. (2010) we intended to run an analysis assessing the preferences of participants with equivalent performance across difficulty levels. Such an analysis allows effort aversion to be more clearly delineated from error avoidance. For example, if a participant had equal performance across N levels 0–5 yet were still willing to forgo reward to avoid higher levels this would be clearly indicative of effort avoidance. However, across Experiments 1–3, at most four participants had average performance levels within 10% across difficulty levels. It was therefore not possible to run such an analysis. Further details are given in the Supplementary Materials (Section 4).

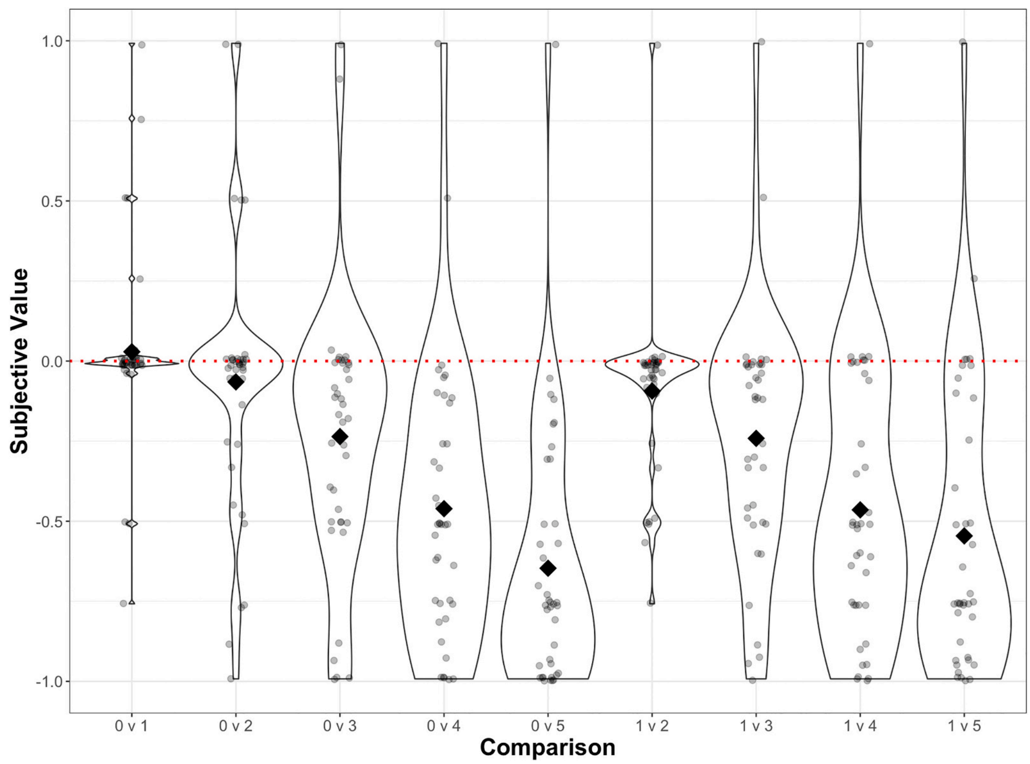

Fig. 3. Violin plots of COG-ED subjective value scores (SVs) across the nine comparisons for Experiment 1. Individual SVs are represented as transparent circles and mean SVs are represented as solid black diamonds. The y-intercept at 0 represents an SV indicative of indifference between the two options.

Table 2
Results from the four models used to predict SV in Experiment 1.

Model	AIC	Akaike Weights
$SV \sim 1 + (1 \mid subject)$	364.53	< 0.001
$SV \sim performance + (1 \mid subject)$	263.14	< 0.001
$SV \sim comparison + (1 \mid subject)$	214.27	0.151
$SV \sim performance + comparison + (1 \mid subject)$	210.81	0.849

Table 3
Mean percent correct per difficulty set in the training phase, mean rt per problem for each difficulty set, mean number of problems solved per 3-min run, and mean difficulty rating (0–100) per difficulty set, for Experiment 2. Standard errors in parentheses.

NSP Difficulty	Percent Correct	Mean RT	Mean Problems Solved	Mean Diff. Rating
Easy	97.53 (*0.36*)	10.53 s (*0.23*)	17.93	10.28 (*3.60*)
Medium	69.86 (*1.96*)	41.89 s (*1.83*)	5.33	45.58 (*4.02*)
Hard	18.73 (*2.15*)	73.66 s (*4.86*)	3.18	84.96 (*2.99*)

different comparison conditions, it is not clear whether comparison levels explain much variance in addition to performance discrepancies.

3.3. Experiment 3 (anagrams)

3.3.1. Training phase and difficulty ratings

Participants' average performance (percent correct and response time) as well as mean difficulty ratings (measured post-experiment) for Experiment 2 are collated in Table 5. Analogous to Experiment 2, participants completed fewer problems in the more difficult sets as their response times per problems were markedly slower.

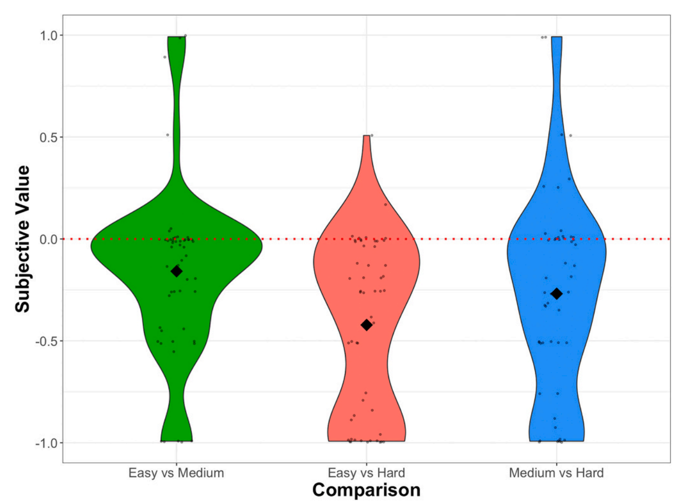

Fig. 4. Violin plots of participants' COG-ED subjective value scores (SVs) across the three comparisons for Experiment 2. Individual SVs are represented as transparent circles and mean SVs are represented as solid black diamonds. The y-intercept at 0 represents an SV indicative of indifference between the two options.

Table 4
Results from the four models used to predict SV in Experiment 2.

Model	AIC	Akaike Weights
$SV \sim 1 + (1 \mid subject)$	196.83	0.005
$SV \sim performance + (1 \mid subject)$	188.10	0.425
$SV \sim comparison + (1 \mid subject)$	188.93	0.281
$SV \sim performance + comparison + (1 \mid subject)$	188.87	0.289

Table 5
Mean percent correct per difficulty set in the training phase, mean rt per problem for each difficulty set, number of problems solved and mean difficulty rating (0–100) per difficulty set for Experiment 3. Standard errors in parentheses.

Anagram Type	Percent Correct	Mean RT	No. Problems Solved	Mean Diff. Rating
3-letter	95.43 (*0.34*)	4.68 s (*0.08*)	39.55	14.69 (*3.78*)
5-letter	73.45 (*1.00*)	18.05 s (*1.44*)	11.03	48.19 (*3.66*)
7-letter	57.43 (1.34)	33.55 s (*0.60*)	6.30	84.96 (*3.17*)

3.3.2. COG-ED phase

Participants, on average, were willing to forgo reward in order to avoid the more effortful option for all three comparisons (Fig. 5): 3-letter vs 5-letter ($M = -0.19$, $t(48) = 4.25$, $p < .001$), 3-letter vs 7-letter ($M = -0.26$, $t(48) = 4.73$, $p < .001$), 5-letter vs 7-letter ($M = -0.18$, $t(48) = 3.27$, $p = .002$).

As in Experiments 1 and 2, difference in SV was assessed using linear mixed modelling: three models containing performance discrepancy and comparison as fixed factors, in addition to a baseline model containing only random intercepts. The results from these models and their comparisons are collated in Table 6. The model containing only comparison as a fixed factor did not perform significantly better than baseline ($\chi 2(1) = 4.29$, $p = .11$), indicating that SVs did not significantly differ across the three comparison groups. However, models containing performance discrepancy were significantly better predictors of SV suggesting the difference between how well participants performed on two tasks was the primary driver of their choice preferences.

4. Discussion

The results of Experiments 1–3 correspond with prior work on mental effort (e.g., Kool et al., 2010; Oprea, 2020; Westbrook et al., 2013) in that more difficult tasks are generally avoided. Furthermore, the results of Experiments 1 and 2 cohere with Westbrook et al.' (2013) findings that participants are willing to forgo increased amounts of reward as the difficulty disparity between the two options grows larger.

In Experiment 1, both comparison and performance discrepancy were significant predictors of SV; indicating that participants were willing to forgo greater reward as the harder N-Back level increased and as their performance discrepancy between the two options increased. In Experiment 2, while models containing comparison and performance discrepancy performed better than a baseline model, the inclusion of both factors in a single model did not improve fit. Therefore, while SVs differed across the three comparison groups, this variance could be accounted for by participants' performance discrepancy (or the comparison level). Conversely, the results of Experiment 3 show no difference across the three comparison groups, suggesting participants were equally averse to more difficult tasks for all comparisons. Performance discrepancy however was still a significant predictor of SV, indicating the amount of reward participants were willing to forgo to avoid a task depended on how they performed at that task compared to the alternative task. We argue the homogeneity of SVs in Experiment 3 across the comparison groups was driven by the smaller differences in objective difficulty levels compared to Experiment 2. For example, mean accuracy and response times for Hard NSPs in Experiment 2 was 17% and 77 s compared to 57% and 33 s for 7-letter anagrams in Experiment 3. Put simply, 7-letter anagrams were not as difficult to solve as Hard NSPs, therefore participants' aversion to completing a set of them was comparatively less.

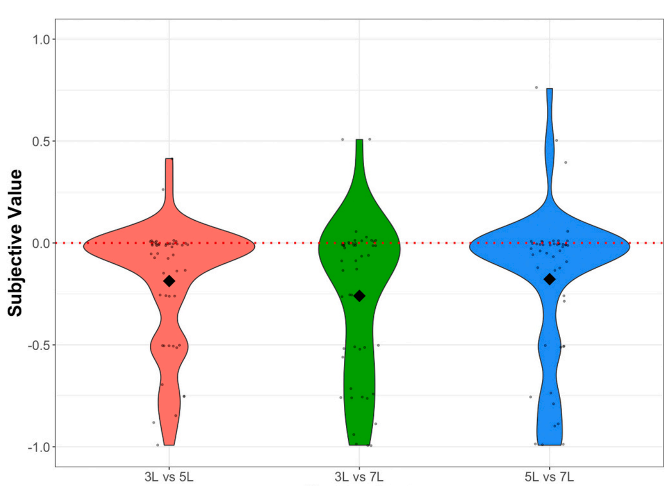

Fig. 5. Violin plots of participants' COG-ED subjective value scores across the three comparisons for Experiment 3. Individual SVs are represented as transparent circles and mean SVs are represented as solid black diamonds. The y-intercept at 0 represents an SV indicative of indifference between the two options.

Table 6
Results from the four models used to predict SV in Experiment 3.

Model	AIC	Akaike Weights
$SV \sim 1 + (1 \mid subject)$	63.16	0.044
$SV \sim performance + (1 \mid subject)$	57.43	0.774
$SV \sim comparison + (1 \mid subject)$	62.87	0.051
$SV \sim performance + comparison + (1 \mid subject)$	60.97	0.131

While there were differences between the results of the three Experiments, we find no strong evidence to suggest that effort avoidance (or seeking) is modulated by the type of demand imposed by a task (e.g., problem solving compared to rule implementation). Conversely, we find that participants, on average, are willing to forgo reward in order to avoid harder difficulty levels (and therefore the effort required to complete them) across all task types used here. While a proportion of participants indicated they preferred more difficult tasks when rewards were equal (with some willing to forgo reward to perform more difficult tasks), these participants were in the minority in Experiments 2 and 3 (see Supplementary Materials S5 for an analysis of Experiment 1–4). In Experiment 1, the majority of participants preferred the more effortful option at least once, though the majority of these preferences occurred when the easier task was the boring 0-Back (e.g., 0-Back or 1-Back and 0-Back or 2-Back). Furthermore, it is worth noting that participants had nine opportunities to prefer the harder option in Experiment 1 (as there are nine comparisons), whereas there were only three opportunities to do so in Experiments 2 and 3. Given the fleeting nature of apparent 'effort seeking' behaviour, such instances may arise simply due to noise in participants' COG-ED responses.

Moreover, we also find little evidence that participants, on average, *preferred* effortful tasks when the alternative was boring such as the 0-Back task, Easy NSPs or 3-letter anagrams. At most, participants were indifferent between options when the alternative was intended to be boring (e.g., 0-Back), and indifference only occurred when the harder alternative was not exceedingly difficult (up to a 2-Back task [93.72% accuracy]). When participants were given a choice between the 0-Back and noticeably harder levels (e.g., 4-Back and 5-Back, which had average accuracies of 75.99% and 72.77%, respectively), they showed significant aversions to increased difficulty. While Easy NSPs and 3-letter anagrams were also intended to be boring given the ease in solving them (97.53% and 95.43% average accuracy, respectively), participants on average were averse to completing more difficult NSPs and anagrams, instead opting for the easy alternatives. It would be of interest to assess whether such effort aversion persists in difficulty levels equivalent to the 0-Back task for both NSPs and anagrams (i.e., retyping an unscrambled anagram, or retyping the last number in a solved NSP).

Overall, while we find no evidence that the propensity to avoid (or seek) effort differs when difficulty varies within a task type (e.g., choosing between different sets of NSPs), this does not rule out the possibility that some types of tasks (which impose different demands) are preferred over others, even when there is a discrepancy in the effort required between them. If the effort required for tasks which involve abstract problem solving is less aversive than that required for rule implementation type tasks, people may prefer difficult levels of the former over easier levels of the latter. For example, preferring to complete a 3-min set of Medium NSPs than a 3-min 1-Back task. Experiment 4 aimed to address this possibility by directly comparing N-Back tasks, NSPs and anagrams via a COG-ED task.

5. Experiment 4

Experiment 4 assessed whether some types of demanding tasks were preferred over others despite discrepancies in their difficulty level. Specifically, whether people would be willing to perform more difficult iterations of tasks which required problem solving (e.g., Medium anagrams) over easier, rule-implementation tasks (e.g., 1-Back). If such a preference was exhibited it would suggest that increases in effort in the absence of extrinsic reward are not ubiquitously avoided, even in experimental settings. Furthermore, that the type of demands imposed by a task are important in determining how aversive (or rewarding) it may be.

To simplify the design, Experiment 4 contained three types of tasks (N-Back, NSPs, and anagrams) each of which had only two difficulty levels. For the N-Back tasks these levels were 1-Back and 3-Back, for NSPs they were Easy and Medium problems, and for the anagrams 3-letter and 5-letter word strings. These difficulty levels were selected for their similar accuracies and difficulty ratings (for NSPs and anagrams) across task types. Additionally, participants were asked after each block in the training phase of Experiment 4 to rate each set – "How effortful did you find the previous task?" – on a 0–100 scale.

For all six comparisons in the COG-ED phase of the Experiment, no two options had equal difficulties. Instead, the difficulty, and presumably the effort required to perform the task, was always greater for one of the options. This was because we were interested in whether participants' aversion to effort depended on the type of task, not whether people prefer solving N-Back, number sequence problems, or anagrams, all else equal.

5.1. Method

5.1.1. Participants

Participants in Experiment 4 ($N = 50$; $M_{\text{age}} = 43.48$; 22 females and 28 males) enrolled via the Prolific platform and were paid a flat rate of £5.63. Participants were also paid a bonus payment dependent on their choices in the COG-ED task ($M = £1.77$).

5.1.2. Materials

The same as Experiments 2 and 3.

5.1.3. Training phase design

Similar to Experiments 1–3, participants had two training runs of each of the six different tasks present in the task (two N-Back, two NSPs, and two anagrams) – 12 training runs in total. Training runs were presented to participants in 3-min blocks.

For the N-Back task, there were 51 sequential trials (letters displayed) as opposed to 30 in Experiment 1 in order for the run to last 3-min. All other aspects of the N-Back task remained the same. NSP and anagram blocks were the same as in Experiments 2 and 3. NSP and anagram sets were also presented to participants as colours: NSPs were either "Orange" or "Green, and anagrams either "Red" or "Blue", randomised between participants.

Participants only received feedback at the end of the run, unlike Experiments 2 and 3 where participants received feedback after solving each problem. For the N-Back task it was in the form: "You correctly identified x of the X matching items. You incorrectly identified y of Y non-matching items. On average, you got z-percent correct." For NSPs and anagrams, it was in the form "You solved *n* out of N problems correctly. On average you got *x* percent correct." This adjustment was made so there were no differences in feedback between the types of tasks.

Presentation order of the runs was randomised between task type (i.e., N-Back, NSPs, anagrams), although the difficulties were nested so that different difficulties of the same task type were presented adjacently (e.g., 3-Back then 1-Back). Furthermore, participants encountered one training run of each task type and difficulty before encountering the same again.

5.1.4. COG-ED design

The COG-ED design consisted of six comparisons: 1-Back or Medium NSPs, 1-Back or 5-letter anagrams, 3-Back or Easy NSPs, 3-Back or Easy anagrams, Easy NSPs or Medium anagrams, and Medium NSPs or Easy anagrams. All comparisons were presented in a random order. The remainder of the COG-ED design was analogous to Experiments 1–3.

5.1.5. Procedure

The procedure was the same as Experiments 1–3 except for the COG-ED comparisons outlined above.

6. Results

6.1. Training phase, effort and difficulty ratings

Performance for all task types and difficulty levels is collated in Table 7. Average 'effort' self-reports (made after each run in the training phase) and average difficulty self-report ratings are collated in Fig. 6. Accuracy (percent correct) and mean effort rating scores were such that participants, on average, found the more difficult version of each task more effortful than the easier versions. Participants self-reported difficulty ratings also aligned with their accuracy in the training phase. While both the effort and difficulty ratings differed across the tasks, particularly among those of 'Medium' difficulty, there was no overlap between tasks labelled Easy and those labelled Medium.

Table 7
Accuracy, response time, and no. problems solved for all tasks and difficulties in Experiment 4. Standard error in parentheses.

Task Type	Difficulty	Percent Correct	Mean RT	No. Problems Solved
N-Back	1-Back	96.36 (*0.82*)	*NA*	*NA*
N-Back	3-Back	82.75 (*1.19*)	*NA*	*NA*
NSP	Easy	96.69 (*1.13*)	6.54 s (*0.07*)	28.22
NSP	Medium	78.11 (*2.41*)	23.87 s (*0.80*)	8.51
Anagrams	3-letter	97.38 (*0.38*)	4.29 s (*0.14*)	44.15
Anagrams	5-letter	77.75 (*2.30*)	17.59 s (*0.68*)	11.43

Note. Participants had 2.0 s to respond to each stimulus during N-Back tasks and the number of stimuli presented in each run was constant (51).

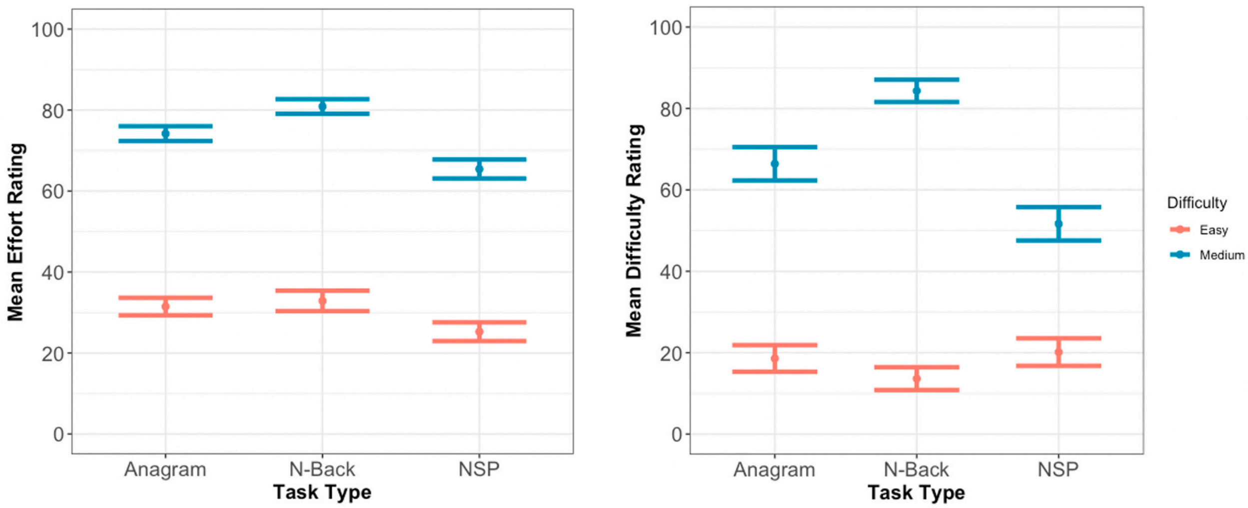

Fig. 6. Left: Mean effort ratings for each task and difficulty level. Effort ratings were made on a 0–100 scale after each run in the training phase. Right: Mean difficulty ratings made for each task and difficulty level. Ratings were made on a 0–100 scale at the end of the experiment. Note. 3-letter and 5-letter anagrams are labelled as Easy and Medium, respectively.

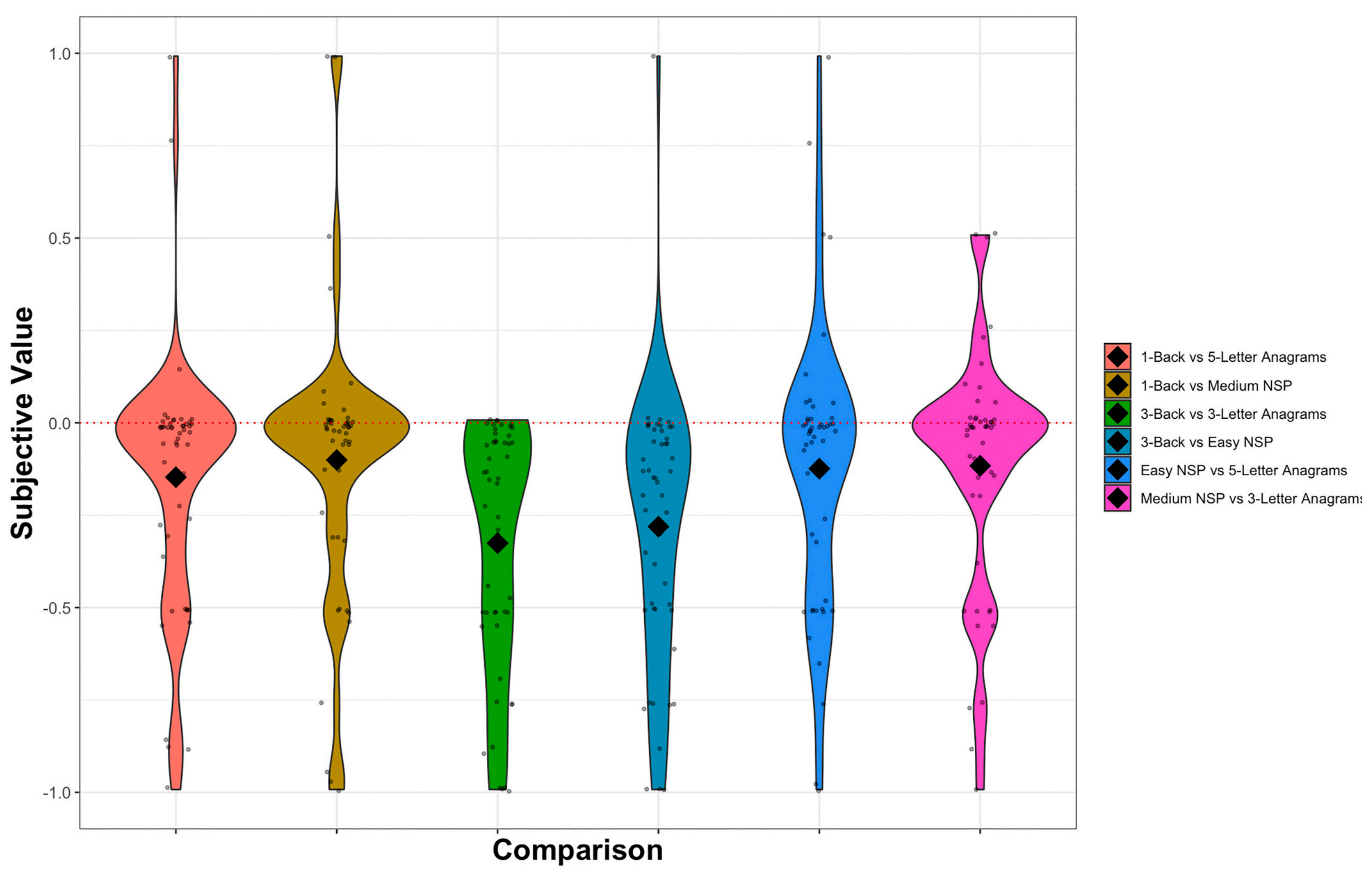

Fig. 7. Violin plots of participants' COG-ED subjective value scores (SVs) across the six comparisons for Experiment 4. Individual SVs are represented as transparent circles and mean SVs are represented as solid black diamonds. SVs between −1 and 0 indicate participants were willing to forgo reward to avoid the more difficult option, SVs between 0 and + 1 indicate participants were willing to forgo reward to avoid the easier option. The y-intercept at 0 represents an SV indicative of indifference between the two options.

Table 8
Results from mixed models used to predict SV in Experiment 4.

Model	AIC	Akaike Weights
SV ~ 1 + (1 | subject)	234.19	< 0.001
SV ~ performance + (1 | subject)	210.07	< 0.001
SV ~ comparison + (1 | subject)	222.34	< 0.001
SV ~ effort + (1 | subject)	224.23	<0.001
SV ~ performance + comparison + (1 | subject)	190.06	0.655
SV ~ performance + comparison + effort + (1 | subject)	191.34	0.345

Note. Further models reported in Supplementary Materials (Table S7d).

6.2. COG-ED phase

Individual and mean subjective values (SV) for the six COG-ED comparisons in Experiment 4 are represented in Fig. 7. *t*-tests were run for each comparison to assess whether participants, on average, were willing to forgo reward in order to avoid the more difficult option.

After correction,[6] only three of the comparisons indicated participants were willing to forgo reward in order to avoid the more difficult alternative: 1-Back or 5-letter anagrams ($M = -0.15$, $t(49) = 2.95$, $p = .005$), 3-Back or 3-letter anagrams ($M = -0.33$, $t(49) = 6.90$, $p < .001$), and 3-Back or Easy NSPs ($M = -0.28$, $t(49) = 5.47$, $p < .001$). The remaining three comparisons – 1-Back or Medium NSPs, Easy NSPs or 5-letter anagrams, and 5-letter anagrams or Medium NSPs – all had subjective values which were not significantly different from 0 (statistics contained in Supplementary Materials, section 3); though it is worth noting that in the cases where no significant difference was observed, the trend was towards effort aversion (their mean SVs sat noticeably below 0).

We again used linear mixed modelling to assess how SV differed across comparisons (factor) and the influence participants' performance (measured in the training phase; coded numerically) had on their SVs. Performance discrepancy was calculated for each participant for each comparison by subtracting their mean performance in the typically easier option (e.g., 1-Back) from their mean performance in the typically harder option (e.g., Medium NSPs). In addition, we also included a model containing subjective effort ratings (obtained after each block in the training phase) as a fixed factor and incorporated these ratings into the full model to assess the relationship between participants' sense of effort and their SVs.

The results from the single fixed factor models and the best performing model are reported in Table 8 (additional mixed models are included in Tables S7d). All models performed better than a baseline only model, with comparison, performance discrepancy, and subjective effort ratings significantly improving model fit (individually). The conjunction of all above factors into a single model further improved the model's performance, however, this improvement was not greater than a model containing only performance discrepancy and comparison (assessed by Akaike weights; see Table 8).

7. Discussion

Following the results of the first three experiments, Experiment 4 also fails to find consistent or strong evidence that people voluntarily seek out effortful tasks, at best showing either indifference or intermittent, but inconsistent effort seeking.

It is however notable that the extent to which people avoid the more difficult task is tempered here compared to Experiments 1–3. For instance, only three of the six comparisons had average SVs significantly less than zero (i.e., the point of indifference between the two options). This stands in contrast to Experiments 1–3 where SVs were significantly less than zero for all comparisons except those where the alternative was exceptionally easy (e.g., the 0-Back task). Furthermore, the modelling analysis we report in the Model-Based Analysis section also suggests that people tend to be indifferent between tasks for multiple comparisons in Experiment 4. A result, which in conjunction with participants' SVs, suggests people were less averse to harder difficulty levels than they were in prior experiments where difficulty was manipulated within a task.

One interpretation of these findings is that the effort required for rule-discovery tasks is less aversive than when a task requires repetitive rule-implementation. However, given the multitude of differences between the tasks (e.g., the number of trials, the responses required, and the time taken to solve a problem), generally noisy responses, and the lack of overt effort seeking observed, we caution against interpreting the results in only this way. Furthermore, the mean SVs are approximately equivalent between comparisons where the harder option required rule-discovery and the easier rule-implementation (i.e., 1-Back or Medium NSP and 1-Back or 5-Letter Anagram) and when both tasks required rule-discovery (i.e., Medium NSPs or 3-Letter Anagrams and Easy NSPs or 5-Letter Anagrams). If rule-discovery tasks were inherently less aversive than tasks which require repeated rule-implementation we would expect most of the effort seeking to have occurred in the former comparisons (1-Back or Medium NSP and 1-Back or 5-Letter Anagram) as opposed to the latter (Medium NSPs or 3-Letter Anagrams and Easy NSPs or 5-Letter Anagrams); SVs between these comparison groups however are roughly equal.

Overall, while the problem-solving (or rule-discovery) tasks impose different types of demands on participants compared to attentionally demanding, rule-implementation-like tasks, people were generally still averse to more effortful tasks. Put simply, contrary to our initial hypothesis, increased difficulty and effort was generally avoided irrespective of the types of cognitive demands a task imposed or the type of mental computation required to perform the task.

8. Model-based analysis

The conclusions we have drawn from the COG-ED task thus far assume that participants' responses (and therefore SVs) are perfectly in line with their underlying preferences. In this section we show that our conclusions generally hold when we relax this assumption. We developed, fit and compared two models, differing in terms of how noise (related to participants' choices in the COG-ED task) was implemented. In general, the conclusions from the model we present below is similar to the statistical analyses above.

The model assumes that participants have a preference for each comparison k, γ_k, that represents how much more they prefer the harder task over the easier task for each comparison. This preference is multiplied by the ratio of the offer to do the harder task, $\$X_H$, and the offer to do the easier task, $\$X_E$, to produce an overall tendency to pick the harder task, $\theta_k = \gamma_k \frac{X_H}{X_E}$. We call θ_k a tendency because the choice comes from a softmax process, such that the probability of choosing the harder task on decision i of a COG-ED trial in condition k is $p(H_{ik}) = \frac{e^{b_i \theta_k}}{e^{b_i \theta_k} + e^{b_i}}$ if $\theta_k > 1$ and $p(H_{ik}) = \frac{e^{b_i}}{e^{b_i} + e^{b_i \theta_k^{-1}}}$, if $\theta_k < 1$. This dependence on θ_k simply ensures that the model is symmetric with regards to easy and hard tasks. Here, b_i controls how deterministic the decision process is. In this model we assume that it decreases linearly with the number of decisions that a participant makes (i.e., $b_i = b/_i$). The details for the model where we freely estimate the noise parameter b across decisions (i.e., choices 1–7 in the COG-ED phase) is presented in the Supplementary Materials (Section 8).

To compare the fit of both models we used Leave-One-Out Cross-Validation (LOO-CV; Vehtari et al., 2017/2019) via the *loo* package in *R*. Models were compared based on their expected log pointwise predictive density (ELPD). For all four experiments, the model with a diminishing noise parameter (outlined above) provided a better account of the data

[6] Bonferroni corrections were made for all *t*-test comparisons, therefore the nominal *p*-value for Experiment 4 was 0.008.

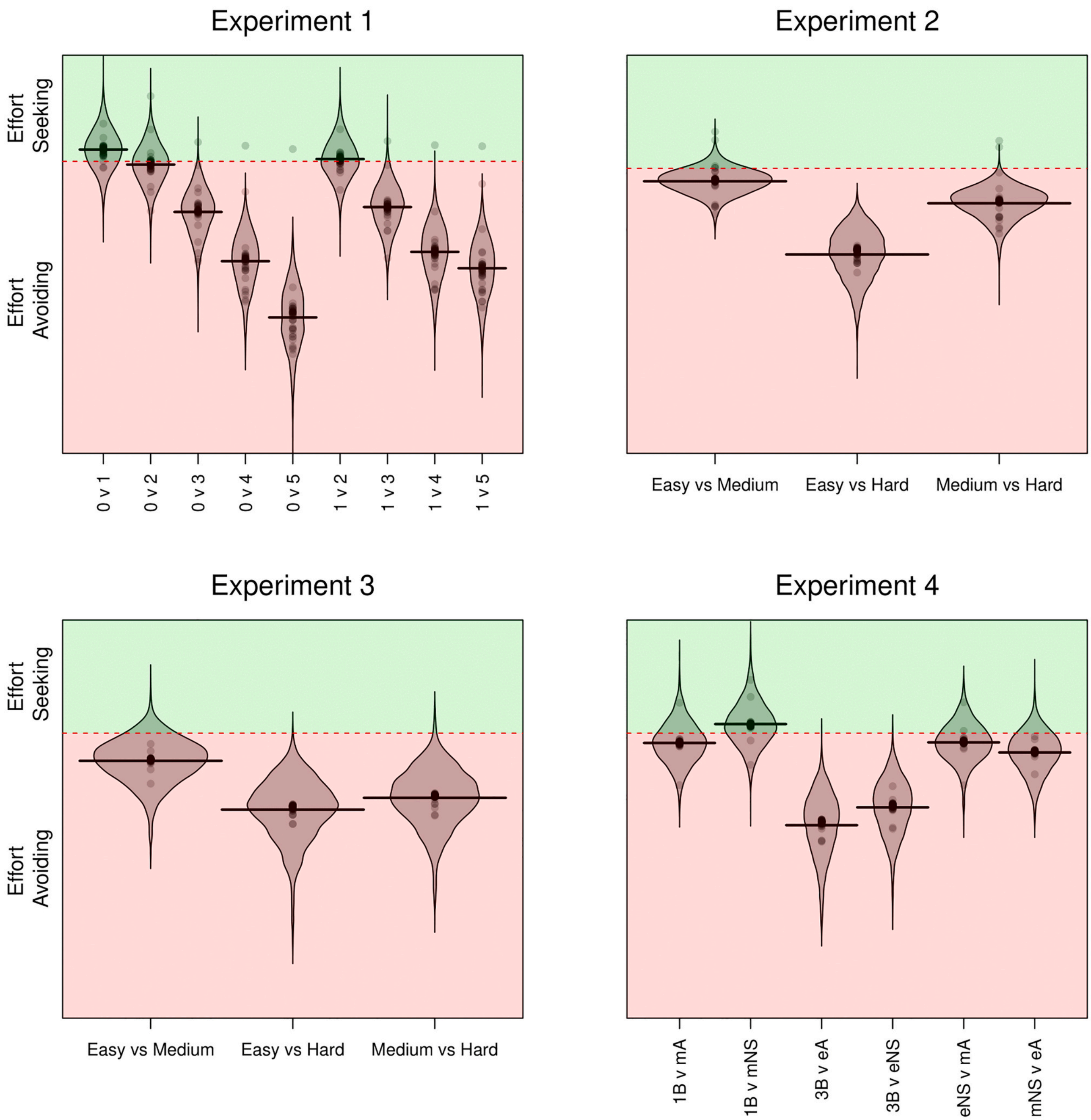

Fig. 8. Model-based inferred preference for the harder task for each comparison in Experiments 1 to 4. The violin plots show the group-level posterior distributions of logγ, with the horizontal line showing the group-level median. The circles show the median values of individual-participant logγ. The dotted red line represents indifference between the two offered tasks. For Experiment 1, the number on the x-axis indicates the values of N in the N-Back task. For Experiment 4, 1B and 3B represent 1-Back and 3-Back, respectively. Similarly, eA and mA show 3-letter (i.e., easy) and 5-letter (i.e., hard) anagram tasks, and eNS and mNS show easy and medium NSPs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(i.e., lower LOO-IC scores) than the model with a constant noise parameter. We therefore focus on the model with a diminishing noise parameter for the remainder of this section. See Supplementary Materials Section 8 for details of the model comparison and the parameter estimate details.

Since γ is a ratio, in Fig. 8 we plot logγ for each experiment so that positive and negative values represent a preference for the more and less effortful tasks, respectively. The overall pattern of preferences, especially those within each experiment, follows roughly the same pattern as was observed in the earlier statistical analyses. Indeed, we observe that this analysis also sees the same conditions having a relative indifference between more or less effortful tasks (e.g., 0-Back or 1-Back; 1-Back or Medium NSPs), and that participants are otherwise generally happy to forgo some amount of money to avoid the effortful tasks. Not all participants, however, are categorised as being consistently effort averse. Analogous to the SVs obtained in the COG-ED, the parameter estimates for the above model suggest that at least some participants are effort seeking (or indifferent), particularly in conditions where the difficulty disparity between tasks is small (e.g., 1-Back or 2-Back). However, the level of uncertainty around individual-participant parameters prohibits any strict classification of individuals as either effort seeking or avoiding.

It is worth noting that the amount of effort seeking we infer from the parameter estimates depend on the assumptions we make about noise in the decision process. In the model presented here, we assume that responses become less deterministic with each decision in each COG-ED phase. Such an assumption is consistent with the fact decisions in the COG-ED task become decreasingly consequential as more are made as well as people's increased experience with the choice paradigm. This assumption, however, also serves the functional purpose of reducing the influence of later decisions on participants' overall effort preferences (i.e., γ). Without such an assumption, we worry the model mischaracterises individuals who have a slight preference for one task but are otherwise indifferent – for example, preferring the 1-Back over the 2-Back when offered rewards are equal, but being unwilling to forgo any money to avoid either task. Such participants will choose one task in the first decision and the alternative in every subsequent decision – generating an SV of $\sim$0. If the model does not weight choices in the COG-ED phase differently, then it will infer an overall preference for the option chosen six times, rather than the option chosen once (see the alternative model in Supplementary Materials).

The above modelling analysis attempts to explicate and account for the effect that noise in the decision process may have on participants' SVs in the COG-ED. While a novel approach, it provides a possible way of accounting for noise in people's choices (particularly on the first trial of the COG-ED phase) and a complementary way of assessing people's aversion to increased effort in conjunction with SV. For the current experiments, the similarities in the results between the two analyses (i.e., frequentist statistical analysis and hierarchical Bayesian modelling) bolsters our confidence in the conclusions drawn.

9. General discussion

Psychological research on mental effort has predominantly used rule-implementation type tasks which bear little similarity to the types of tasks people partake in for fun (e.g., Sudoku, crosswords, chess). These four Experiments aimed to broaden the scope of mental effort research to determine whether the mental demands of a task modulate peoples' propensity to avoid effort. To this end, we used three types of tasks: the N-Back task which is frequently used in effort research (e.g., Westbrook et al., 2013), NSPs which we designed to mimic brain teaser games, and anagram solving which is a feature in many popular games (e.g., Scrabble). These tasks differ substantially in the type of demand they place on people; the N-Back requires maintained attention and the implementation of a simple rule (i.e., does the current letter match the letter from *N* turns ago), compared to NSPs and anagrams where success requires abstract, problem solving and rule discovery for success.

We also aimed to assess whether people would be willing to increase their effort (in the absence of extrinsic rewards) when the alternative was exceedingly simple, verging on boring across all types of tasks (e.g., "Solve the missing number: 1 2 3 4 5 _" or "Press 'm' each time a letter appears on screen).

Experiments 1–3 found participants, in the aggregate, were unwilling to increase their effort by choosing harder difficulties regardless of the type of task (N-Back, NSPs, or anagrams). Conversely, on average, participants were willing to forgo a proportion of their potential reward in order to complete easier alternatives. Moreover, in Experiments 1 and 2, people were willing to forgo greater reward as the discrepancy in difficulty (and therefore effort required) increased between the two options. Effort aversion, however, was not observed when the easier option was boring (i.e., 0-Back) and the harder option was not exceptionally difficult (i.e., 1-Back or 2-Back). When the choice was between simple/boring tasks (e.g., 0-Back, Easy NSPs) and hard tasks (e.g., 5-Back, Hard NSPs), however, participants showed strong preferences for the easier, less effortful options. This suggests that instances where effort (Wu et al., 2022) or other types of typically negative sensations are sought (electric shock; Bench & Lench, 2019) to escape a sense of boredom, the boredom-inducing alternative must be exceedingly boring (i.e., doing nothing; Wu et al., 2022).

Experiment 4 built on these results and assessed whether aversion to increased difficulty persisted when participants had a choice *between* task types (e.g., N-Back or NSPs). While our hypothesis was that people may be more willing to increase their effort by choosing harder problem-solving tasks, as opposed to more mundane, rule-implementation tasks, we found people people's preferences for easier, less effortful tasks was only marginally affected by whether the harder alternative required rule-implementation or more abstract types of reasoning. Participants were less averse to increased difficulty and effort than Experiments 1–3, but did not, on average, exhibit any strong preferences for harder tasks.

Generally, our results, while not supportive of our hypothesis that effort aversion may be modulated by the type of demands imposed by a task, are congruent with the mental effort literature at large in that people are unwilling to exert increased effort in the absence of extrinsic reward (for review, see Kool & Botvinick, 2018). However, given the fact that people do seek out mental challenges for the sake of it in everyday life, these findings beg the question as to what intrinsic qualities a task must possess for increased effort to be sought when extrinsic rewards are unclear or non-existent.

One candidate is the information potentially gained during a task, which was found to modulate effort avoidance in a demand selection task (Devine & Otto, 2022). In the current set of experiments, participants received feedback in the form of 'correct' or 'wrong' for NSPs and anagrams, and summative accuracy feedback for the N-Back. When they are correct, participants gain information as to what the correct numerical sequence or rearrangement of the letters were. When they are incorrect, however, they gain no insight as to the correct solution, nor the strategy that would lead them towards it. Therefore, in our tasks, if participants willingly chose to complete harder tasks (which they typically performed worse at) they were forgoing the amount of information (i.e., the correct answer) they could potentially gain. Furthermore, the amount of information gained in NSP and anagram tasks was influenced by how many problems a person solved in a set, with fewer problems typically being solved as set difficulty increased. A solution to this issue would be to fix the number of problems per set, but since easier problems can be solved significantly faster, participants could finish the experiment sooner.

The idea that effort exertion is modulated by the presence of information also coheres with current computational theories of effort aversion (e.g., Kurzban et al., 2013; Shenhav et al., 2017), in that information is the reward individuals typically seek when they exert effort towards a task. For example, from an opportunity cost perspective of mental effort (i.e., where the decision to allocate effort is considered in respect to the probable opportunity costs and potential rewards), it is unsurprising participants are unwilling to voluntarily increase their effort given the probability of gaining information (an intrinsic reward) is minimal.

More specifically, Devine and Otto (2022) found that the willingness to exert effort in a demand selection task was modulated by the availability of task relevant, non-instrumental information (i.e., a progress bar). Specifically, they found that participants were more willing to increase their effort (i.e., choose a higher demand option) when that choice resulted in a progress bar being shown and the alternative (e.g., low demand option) offered no progress bar. If task relevant information modulates effort avoidance, the presence of full-feedback (e.g., the correct answer and solution to the problem) may reduce the costly nature of engaging in more effortful activities; in such a paradigm, if an individual is incorrect, information is still received as to what the correct solution was, unlike the current set of experiments where solutions remain unknown to participants who do not correctly solve them. Anecdotally, it is hard to imagine people would enjoy cognitively demanding activities such as Wordle, sudoku, and crosswords if the correct solutions were only available to those who correctly solved the puzzle.

Alternatively, it may be that instances of supposed effort seeking outside of controlled laboratory environments are not driven by factors intrinsic to the task. For example, people solving brain teasers may believe that solving such tasks will improve their cognitive ability (as many such games purport to do); those solving crosswords may be wanting to expand their vocabulary or general knowledge; or, simply, such games may be rewarding due to the social incentive of trying to beat friends and other players. Future work could investigate the latter claim with a task which provides social comparative feedback (e.g., you performed better than *X*% of people) when effort is sought (e.g., choosing Medium NSPs) as opposed to when it is avoided (e.g., Easy NSPs).

Furthermore, when increased effort is preferred as opposed to avoided (e.g., playing board games), people are voluntarily seeking out effort – that is, the effort is self-initiated. This stands in contrast to lab experiments where a participant's primary goal is to earn money (on Prolific and mTurk) or receive course credits; in the latter scenario, engaging with an effortful task is simply a means to an end. In support of such anecdotal examples, research by Hockey and colleagues (Hockey, 2011; Hockey & Earle, 2006) suggest controllability is a determining factor in how effortful tasks feel and whether mental fatigue develops. Specifically, they found that fatigue (measured by subjective experience ratings and decrements in performance) in a simulated work environment was more pronounced when participants had less control over the schedule of the tasks they had to complete (Hockey & Earle, 2006). When participants had autonomy over how (and in what order) they completed a set of tasks, the sense of effort was reduced even when the workload necessary for success was high.

Adjacently, a cynical interpretation may be that the people we refer to in effort-seeking anecdotes (e.g., those who play board games and crosswords for fun) are simply part of the small population of effort seekers we observe in experimental settings, as opposed to there being something inherently enjoyable (or less aversive) to the mental demands imposed by such tasks outside of the laboratory. However, such an interpretation likely places too much importance on tasks in which mental effort conforms to what psychology researchers deem effortful 'cognition'. For example, researchers may not consider watching one's favourite soap opera to be a mentally effortful task, but it is unlikely a coincidence that the most popular and long-lasting shows require the viewer to construct, maintain, and constantly update an elaborate and complex mental model of the relationships between characters.

One limitation of the approach used here is the inability to reliably manipulate the amount of mental effort required (and subsequently exerted) by a task. In Experiments 1–3, as the difficulty level increased, the effort required presumably increased – as indicated by performance decrements. Delineating effort from task difficulty level, however, is challenging and leads to questions as to whether participants here are avoiding effort or merely tasks in which they perform worse – especially considering the strong predictive value of performance in the mixed model analyses. While previous work has succeeded in delineating performance and effort (see Kool et al., 2010) by providing participants with sufficient training so that performance across tasks was equal, such an approach would be counterproductive to our aims here. For example, if participants had enough training as to complete Hard NSPs with similar accuracy as Easy NSPs (see footnote 4 and Supplementary Materials), it is arguable the rule-discovery aspect of the task would be reduced due to participants' vast exposure to such problems during the training phase (since we used a fixed number of generative rules to create the NSPs). Moreover, given the depletive nature of people's willingness to exert effort over time in the absence of reward (e.g., Wu et al., 2022), we were concerned about the effect substantial training may have on peoples' choice preferences. Another possible solution would be to remove feedback altogether, although in the rule-discovery tasks used here participants could self-generate feedback. For instance, it is obvious if you have correctly solved the anagram or NSP.

Relatedly, the relationship between difficulty and performance is less clear when comparing across different tasks. For example, while participants have lower accuracy in Medium NSPs and 5-letter anagrams, compared to a 1-Back task, we merely assume the effort required for the former tasks is greater. It is unclear however whether quantitative and self-report measures of effort can be equated across tasks which impose distinctly different types of demands. For example, can the effort required to maintain a string of 3 letters in working memory (as in a 3-Back task) be equated with the abstract mathematical reasoning required to solve the missing number in the sequence "2 5 11 23 47 _", even if accuracy and average self-reported effort ratings are equal. Focusing on cognitive control (e.g., Shenhav et al., 2017) circumvents this issue, however it limits the range of tasks which can be employed to assess the phenomenon of mental effort.

The difficulty in equating performance across different tasks may also contribute to the decreased effort aversion observed in some comparisons in Experiment 4 – for example, the two 1-Back comparisons. One reason for the increased willingness to engage with more difficult tasks here may be the inherent difficulties in comparing performance across tasks. While block feedback was given in the form "On average, you got *X*-percent correct" for all task types, it is unclear how to compare performance feedback between two entirely different tasks, especially from a participant's perspective (see Stewart, Chater, & Brown, 2006). Given the strong influence performance has on participants' aversion to more difficult tasks, this difficulty in comparing performance across tasks may have reduced participants' apparent effort avoidance.

A similar, yet separate limitation is the measurement of effort itself. Here, we use difficulty (measured by accuracy) as a proxy for required effort – as the difficulty increases so too does the effort required in order to perform the task. However, there is little stopping a participant from trying less (i.e., exerting less effort) in harder difficulties as performance was not incentivised in our experiments. The use of physiological measures of effort exertion such as heart rate (Clay, Mlynski, Korb, Goschke, & Job, 2022) and pupil dilation (van der Wel & van Steenbergen, 2018) may therefore be useful additions to future work. In Experiment 4 we asked participants to rate how effortful they found each task to attain a more direct measure of effort (compared to inferring effort from accuracy and response times). Presumably, however, participants respond to these questions in accordance with their *sense of effort* experienced during the task, rather than as a metric of how they deployed their cognitive faculties. While the sense of effort arguably tracks how demanding a task is given the reward (Kurzban, 2016), it is plausible that a sense of effort may divorce from the amount of effort objectively exerted (e.g., during so-called flow states; Csikszentmihalyi & Larson, 2014).

In summary, the four experiments documented here broaden our understanding of mental effort avoidance and the domains in which it is observed. Using an effort discounting task (Westbrook et al., 2013) we find that effort avoidance remains consistent across tasks which impose different types of demands. Rather than the type of task driving people's preferences, here, we find their willingness to engage with a task is primarily driven by how much effort is required and how much reward is likely to be gained.

CRediT authorship contribution statement

Jake R. Embrey: Conceptualization, Methodology, Software, Investigation, Formal analysis, Validation, Visualization, Data curation, Writing – original draft. **Chris Donkin:** Conceptualization, Methodology, Formal analysis, Writing – review & editing. **Ben R. Newell:** Conceptualization, Methodology, Writing – review & editing, Resources, Supervision, Project administration, Funding acquisition.

Data availability

The experiment code and relevant behavioural data can be found at the following link: https://osf.io/ydsau/?view_only=3de5b358275c420f86df48a9b96e720f

Acknowledgments

We would like to thank all members of the UNSW Cognition Lab for their feedback on this research when presented at lab meetings and in general discussions. We would also like to thank the reviewers and editor for their constructive and insightful feedback on the original manuscript which significantly improved the final work. Funding was provided by the Australian Research Council (DP190101076).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2023.105440.

References

Baranes, A. F., Oudeyer, P. Y., & Gottlieb, J. (2014). The effects of task difficulty, novelty and the size of the search space on intrinsically motivated exploration. *Frontiers in Neuroscience, 8*, 317.

Baroody, A. J., & Ginsburg, H. P. (1986). The relationship between initial meaningful and mechanical knowledge of arithmetic. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 75–112). Lawrence Erlbaum Associates, Inc.

Bench, S. W., & Lench, H. C. (2019). Boredom as a seeking state: Boredom prompts the pursuit of novel (even negative) experiences. *Emotion, 19*(2), 242–254.

Caplin, A., Csaba, D., Leahy, J., & Nov, O. (2020). Rational inattention, competitive supply, and psychometrics. *The Quarterly Journal of Economics, 135*(3), 1681–1724.

Clay, G., Mlynski, C., Korb, F. M., Goschke, T., & Job, V. (2022). Rewarding cognitive effort increases the intrinsic value of mental labor. *Proceedings of the National Academy of Sciences, 119*(5), Article e2111785119.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97*(3), 332.

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety* (1st ed.). San Francisco: Jossey-Bass.

Csikszentmihalyi, M., & Larson, R. (2014). Validity and reliability of the experience-sampling method. In *Flow and the foundations of positive psychology* (pp. 35–54). Dordrecht: Springer.

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods, 47*(1), 1–12.

DellaVigna, S., & Pope, D. (2018). What motivates effort? Evidence and expert forecasts. *The Review of Economic Studies, 85*(2), 1029–1069.

Devine, S., & Otto, A. R. (2022). Information about task progress modulates cognitive demand avoidance. *Cognition, 225*, Article 105107.

Enke, B., Gneezy, U., Hall, B., Martin, D. C., Nelidov, V., Offerman, T., & van de Ven, J. (2021). *Cognitive biases: Mistakes or missing stakes?* (Tech. Rep.). National Bureau of Economic Research

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. Mcgraw-Hill Book Company.

Gailliot, M. T., & Baumeister, R. F. (2007). The physiology of willpower: Linking blood glucose to self-control. *Personality and Social Psychology Review, 11*(4), 303–327.

Graepel, T., & Herbrich, R. (2006). *Ranking and matchmaking. 25* p. 34). Game Developer Magazine.

Hockey, G. R. J. (2011). A motivational control theory of cognitive fatigue. In *Cognitive fatigue: Multidisciplinary perspectives on current research and future applications* (pp. 167–187).

Hockey, G. R. J., & Earle, F. (2006). Control over the scheduling of simulated office work reduces the impact of workload on mental fatigue and task performance. *Journal of Experimental Psychology: Applied, 12*(1), 50.

Hull, C. L. (1943). The problem of intervening variables in molar behavior theory. *Psychological Review, 50*(3), 273.

Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The effort paradox: Effort is both costly and valued. *Trends in Cognitive Sciences, 22*(4), 337–349.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Keller, J., & Bless, H. (2008). Flow and regulatory compatibility: An experimental approach to the flow model of intrinsic motivation. *Personality and Social Psychology Bulletin, 34*(2), 196–209.

Kool, W., & Botvinick, M. (2018). Mental labour. *Nature Human Behaviour, 2*(12), 899–908.

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.

Kurzban, R. (2010). Does the brain consume additional glucose during self-control tasks? *Evolutionary Psychology, 8*(2), 244–259.

Kurzban, R. (2016). The sense of effort. *Current Opinion in Psychology, 7*, 67–70.

Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences, 36*(6), 661–679.

Moller, A. C., Meier, B. P., & Wall, R. D. (2010). *Developing an experimental induction of flow: Effortless action in the lab. Effortless attention: A new perspective in the cognitive science of attention and action* (pp. 191–204).

Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin, 126*(2), 247.

Nederkoorn, C., Vancleef, L., Wilkenhöner, A., Claes, L., & Havermans, R. C. (2016). Self-inflicted pain out of boredom. *Psychiatry Research, 237*, 127–132.

Oprea, R. (2020). What makes a rule complex? *American Economic Review, 110*(12), 3913–3951.

Osborn Popp, P. J., Newell, B. R., Bartels, D. M., & Gureckis, T. M. (2022). Rule discovery performance unchanged by incentives. In *Proceedings of the annual meeting of the cognitive science society (Vol. 44, no. 44)*.

Otto, A. R., & Daw, N. D. (2019). The opportunity cost of time modulates cognitive effort. *Neuropsychologia, 123*, 92–105.

Rheinberg, F., & Vollmeyer, R. (2003). Flow-Erleben in einem Computerspiel unter experimentell variierten Bedingungen. [flow experience in a computer game under experimentally varied conditions]. *Zeitschift f ü r Psychologie, 221*, 161–170.

Robinson, M. M., & Morsella, E. (2014). The subjective effort of everyday mental tasks: Attending, assessing, and choosing. *Motivation and Emotion, 38*(6), 832–843.

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.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied, 75*(13), 1.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics, 69*(1), 99–118.

Stewart, N., Chater, N., & Brown, G. D. (2006). Decision by sampling. *Cognitive Psychology, 53*(1), 1–26.

Tomporowski, P. D. (2003). Effects of acute bouts of exercise on cognition. *Acta Psychologica, 112*(3), 297–324.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science, 185*(4157), 1124–1131.

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review, 11*, 192–196.

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review, 25*(6).

Westbrook, A., & Braver, T. S. (2015). Cognitive effort: A neuroeconomic approach. *Cognitive, Affective, & Behavioral Neuroscience, 15*(2), 395–415.

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), Article e68210.

Wu, R., Ferguson, A. M., & Inzlicht, M. (2022). Do humans prefer cognitive effort over doing nothing? *Journal of Experimental Psychology: General* (online).