BMC Psychology

RESEARCH | Article in Press

Been around the block: The sustained influence of efficacy on cognitive effort allocation over time

Willem Jacob Louter · S. Tabitha Steendam · Nicoleta Prutean · Joshua O. Eayrs · Nanne Kukkonen · Wim Notebaert · Ruth M. Krebs · Jan R. Wiersema · C. Nico Boehler

Received: 8 December 2025 / Accepted: 24 August 2026

Abstract

Background Modern accounts of cognitive effort posit that it is actively controlled based on a cost-benefit analysis, integrating (subjective) costs and possible rewards, which is furthermore moderated by efficacy (i.e., whether there is a clear relationship between successful task performance and reward). Although there is evidence that reward and efficacy interact, this relationship may not be stable over time, given that reward attenuates time-on-task effects while efficacy's relationship with time-on-task is still unclear.

Methods Thirty participants performed a Stroop task with a block design, with block-wise manipulation of efficacy (possible rewards were either contingent upon correct task performance or determined without regard to task performance) and reward (high or low). Behavioral measures (response time and accuracy), pupil data (tonic and phasic pupil size), and blink measures (rate and duration) were recorded and analyzed using a mixed-effects modelling approach, with each block split into two halves to assess time-on-task effects.

Results The behavioral data showed clear effects for efficacy, a pattern that was also largely mirrored in phasic pupil size, suggesting substantial effort allocation. In contrast, we found no clear impact of time-on-task or reward on behavioral data or phasic pupil size. Tonic pupil size and blink metrics (rate and duration) did suggest time-on-task effects, with reward and efficacy attenuating these in some cases. However, because these time-on-task effects diverged from behavioral outcomes and yielded theoretically ambiguous interaction patterns, they warrant cautious interpretation.

Conclusions These findings support the role of efficacy in effort-related cost-benefit calculations, even when efficacy information is sustained across relatively long task blocks. More broadly, the present study highlights the interpretive complexity of ocular measures for studying cognitive effort over time.

Keywords Expected value of control · Efficacy · Reward · Time-on-task · Cognitive effort

Willem Jacob Louter and S. Tabitha Steendam contributed equally to this work.

https://doi.org/10.1186/s40359-026-05453-x

Published online: 03 September 2026

Background

Every day, individuals engage in challenging activities to pursue specific goals, and alongside other factors such as skill level or experience, their performance in these tasks varies in accordance with the level of cognitive effort they invest. Cognitive effort, as delineated by Shenhav et al. [1], refers to the mental resources people mobilize to meet task demands and support performance. Effort is typically aversive, leading individuals to generally prefer tasks that are less cognitively demanding [2, 3] (but see [4, 5] for exceptions under which effort is sometimes seen as rewarding, such as for example individuals with a high need for cognition who tend to enjoy cognitively effortful activities). In addition, effort discounts the subjective value of rewards, meaning that people generally require greater rewards to choose more effortful options [3]. Together, these observations underscore the notion that allocating effort usually incurs a cost, prompting a cost-benefit analysis in which the anticipated effort is weighed against potential rewards.

The process of conducting this cost-benefit analysis can be explained using the expected value of control (EVC) theory, as proposed by Shenhav et al. [6]. The EVC model serves as a theoretical framework aimed at explaining how cognitive control resources are allocated, based on the expected value derived from this allocation. Central to the EVC calculation is the assumed goal of maximizing expected future rewards while avoiding unnecessary effort. The cost-benefit calculation involves several key components, including the anticipated rewards associated with allocating resources, the intrinsic cost of control itself (i.e., the required resources or effort), and the probability of obtaining the desired outcome when resources are allocated. A key factor in estimating the likelihood of obtaining a desired outcome is efficacy: the extent to which individuals expect their effort to influence that outcome. Importantly, the EVC model also incorporates information about the current state, encompassing task demands, processing capacity, and motivational state. In short, EVC theory proposes that control is allocated, and cognitive effort invested, to maximize expected net value by weighing anticipated rewards against associated costs.

This dynamic aspect of effort allocation is central to the present study. In many real-world contexts, people do not decide whether to invest effort only once. Rather, they must repeatedly evaluate whether continued effort remains worthwhile as fatigue, boredom, or declining task value emerge. From an EVC perspective, this implies that reward and efficacy should shape effort allocation not only at the start of a task, but also as the task unfolds over time. Previous work has separately examined the interplay between reward and efficacy [7–9], and between reward and time-on-task [10–12]. However, it remains unclear how reward, efficacy, and time-on-task jointly shape cognitive effort allocation. Specifically, do efficacy and reward motivate effort in a continuous fashion over time, or do these factors interact over time. For example, for low efficacy and high reward, it is conceivable that motivation either starts off and remains low, or that it starts off relatively high and then deteriorates with increasing realization that performance indeed is not related to outcome. Addressing this question is theoretically important because EVC theory assumes that effort allocation depends jointly on expected reward, efficacy, and current motivational state. Because behavioral results sometimes show little variation in performance during time-on-task experiments due to compensatory allocation of effort in order to maintain performance during the task [13], the present study addressed this question using not just behavioral performance, but also ocular psychophysiological measures of effort and fatigue, specifically pupil and blink measures.

While efficacy is traditionally studied as Bandura's concept of subjective self-belief [14, 15], it can also be experimentally manipulated by explicitly informing participants whether or not their performance determines reward attainment. For example, a study by Frömer et al. [7] showed that individuals integrate information regarding both efficacy and expected reward to gauge the anticipated value of control, assessing the potential benefits and costs associated with allocating effort in a given task. Frömer et al. [7] used a Stroop task where each Stroop stimulus was preceded by a cue containing information about efficacy (i.e., their potential reward for the trial would either be contingent on their accurate performance or determined completely independently from their performance) and reward levels. Results showed that participants were sensitive to both efficacy and

expected reward, with faster accurate response times when expecting high levels of both. The role of efficacy was further corroborated by Toobaei et al. [8], who used the same Stroop task to compare patients with depression and healthy controls. Their results showed that the effect of efficacy on accurate response times was significantly attenuated in the patient group compared to healthy controls. This suggests that efficacy may carry less weight in effort-related cost-benefit analyses in depression, consistent with evidence that patients with depression report lower levels of self-efficacy compared to healthy controls (e.g [16, 17]).

Alongside electrophysiological measures that converged with their behavioral findings, Frömer et al. [7] recorded pupil size, a well-established marker of cognitive effort [18–20], to investigate how reward and efficacy predictions affect task preparation. Specifically, Frömer et al. [7] measured the phasic pupil response (i.e., transient changes in pupil size in response to stimuli or cognitive events; see below for a more elaborate overview) to the informational cue preceding the Stroop target. Interestingly, counter to their expectations, they did not find any effect of reward on pupil size, and pupils were smaller instead of larger when efficacy was higher. However, Frömer et al. [7] focused their analysis on the phasic pupil response to the cue, while assuming a canonical pupillary response function that already starts to decline again before the moment of target presentation. This runs counter to research suggesting that proactive control is represented in the pupil as a gradual build-up [21]. Motivated by this, a recent follow-up by Eayrs et al. [9] employed a similar Stroop task with only incongruent trials, and instead analysed anticipatory pupil dilation by fitting a linear function to the mean pupil size during the 'ramping up' phase preceding the Stroop target presentation, and comparing their slopes. Behavioral results broadly replicated those from Frömer et al. [7]. Pupillometry results however showed faster anticipatory pupil dilation during the cue-target interval (i.e., steeper slopes) for high-reward cues compared to low-reward cues, in particular during high-efficacy trials. This interaction pattern then also continued during the subsequent target processing, as indicated by phasic pupil responses following target presentation. Taken together, these results from Eayrs et al. [9] seem largely in line with the expectations of the EVC model, and therefore suggest that pupil size can indeed be used to investigate how reward and efficacy affect cognitive effort allocation.

As mentioned above, the EVC model not only integrates reward and efficacy, but also information about the current motivational state. Time-on-task factors such as fatigue [22] and boredom [23] are important influences on this state. Mental fatigue refers to a psychobiological state of tiredness following prolonged engagement in cognitively demanding activities [24, 25], whereas boredom involves feeling unchallenged and perceiving activities as meaningless. As posited by Kurzban et al. [26], feelings of boredom and fatigue may reflect a decline in the perceived value of continued engagement, prompting individuals to disengage from the task. Typically, performance deteriorates as fatigue or boredom increases over time, though performance can also be kept stable through compensatory effort [13].

Empirical studies have shown that time-on-task effects can be attenuated by reward. For example, Esterman and colleagues showed that reward can reduce performance decrements over time when it is implemented in a way that helps maintain motivation [10]. Consistent with this, Garner et al. [11] found that a time-based reward, where the duration of the experiment is contingent upon correct task performance, reduced the performance decline over time so that participants that were rewarded showed more stable performance than participants that were not rewarded. Herlambang et al. [12] provided further evidence using a within-subject design that alternated no-reward blocks with reward blocks, in which participants received a monetary reward for accurate responses and an additional reward for responses that were both correct and fast. Herlambang et al. [12] found that although participants' subjective fatigue ratings increased linearly throughout the experiment (i.e., across both non-reward and reward blocks), participants' subjective reports of effort allocation showed that more effort was exerted in the reward blocks. Importantly, this difference was also reflected in the behavioral data, where they observed clear time-on-task effects in the no-reward blocks, but not in the reward blocks. Together, these findings suggest that reward can counteract performance decrements over time, particularly when it sustains motivation or effort allocation. Similar reward effects counteracting fatigue have also been observed in other domains, such as in working memory [27, 28] and response-conflict tasks [29].

The interplay between reward and time-on-task effects is hence well-documented, as is the relationship between reward and efficacy [7–9]. However, as stated, the combined influence of reward, efficacy, and time-on-task remains unclear. As mentioned before, to address this question, the present study combined behavioral measures with psychophysiological indices of effort and fatigue. This is important, because previous research has suggested that behavioral measures sometimes show little variation in performance during time-on-task experiments due to compensatory allocation of effort in order to maintain performance during the task [13], so that solely relying on behavioral results can be ambiguous. Therefore, besides recording behavioral responses, we also employed a psychophysiological approach and measured both phasic and tonic pupil size, as well as blink rate and blink duration. As further motivation behind this choice, we provide a brief overview of these measures below.

Phasic pupil responses, which refer to transient changes in pupil size in response to stimuli or cognitive events, increase as more effort is exerted [18–20]. In addition, phasic pupil size has been shown to decrease as time-on-task increases [27, 28, 30], reflecting problems with maintaining effort over time. Tonic pupil size, in contrast, reflects sustained cognitive processes rather than transient ones [31]. Nevertheless, tonic pupil size similarly also increases with increased cognitive effort [32], and tonic pupil size has also been shown to decrease with time-on-task [27, 33, 34]. Where the phasic pupil response represents the momentary allocation of cognitive resources to a single stimulus, tonic pupil size can represent sustained effort allocated to the broader task or context, but is also generally linked to a broader concept of arousal, which is under various influences [35]. Notably, both tonic and phasic pupil size have also been found to increase in response to higher rewards [36–38].

Eye blink information has long been proposed as a valid measure of task duration-related effects [39]. For example, studies have described the relationship between blink rate and (mental) fatigue [30, 40–43], where blink rates increase as fatigue increases with time on task. Research has also shown that blink rates are increased during episodes of mind wandering [44–46]. Blink duration has been found to show similar response patterns to blink rate, with blink duration increasing with increased fatigue [30, 42, 43], as well as during mind wandering [44, 47, 48]. In addition, a review by Cori et al. [49] concluded that blink duration is a robust measure of drowsiness.

In the present study, we build upon the earlier work by Frömer et al. [7] and Eayrs et al. [9] with an additional focus on the potential interaction with time-on-task, which has thus far not been investigated. Specifically, we adapted the setup from Eayrs et al. [9] which, in turn, was based on Frömer et al. [7], to a counterbalanced block design in order to investigate how time-on-task effects interact with reward and efficacy. In line with EVC theory, we expected both efficacy and reward to influence task performance. Specifically, we predicted higher accuracy and faster response times in high-efficacy than in low-efficacy blocks, and in high-reward than in low-reward blocks. We further expected efficacy and reward to interact, such that performance would be best when both efficacy and reward were high, and worst when both were low. We expected the physiological measures to show a corresponding effort-related pattern. For both tonic and phasic pupil size, this was expected to take the form of larger pupil sizes in high-efficacy and high-reward blocks, with the largest pupil sizes when both factors were high. For blink measures, we expected the reverse directional pattern, with lower blink rates and shorter blink durations in high-efficacy and high-reward blocks, and the lowest blink rates and shortest blink durations when both factors were high. To examine time-on-task effects, we additionally compared performance and physiological measures between the first and second half of each block. We expected fatigue-related changes over time, reflected in poorer performance, smaller pupil sizes, and more frequent and longer blinks in the second half of blocks relative to the first half. Finally, we expected these time-on-task effects to be attenuated in high-efficacy and high-reward blocks compared with low-efficacy and low-reward blocks, with the most stability when both efficacy and reward were high.

Methods

Participants

A total of 39 participants were recruited from the student population of Ghent University via Sona Systems (https://www.sona-systems.com/), an online participant recruitment platform, between February 20 and April 16, 2024. Each participant received a base payment of 15 euros, plus an additional reward which was contingent on fast and accurate responses (detailed below). On average, participants earned an extra 5.01 euros. Eligibility criteria were being aged between 18 and 35 years, Dutch-speaking, and having normal or corrected-to-normal vision. Nine participants were excluded from all analyses due to missing data (3 participants), or failure to reach the required accuracy level of 70% during the practice phase (6 participants), resulting in a final sample size of 30 participants (24 identified as female, 6 identified as male). Age ranged from 18 to 33 years old (M=21.07, SD=3.05). All participants provided written informed consent before the study.

We did not perform a sensitivity analysis or an a-priori power analysis for the present study. Our sample size is similar to those used in Frömer et al. [7] and Eayrs et al. [9], who each reported 37 participants in their pupillometry analyses. The present design also yields a similar number of observations per condition, with 4,860 observations (324 trials per block, split into halves, x 30 participants), where both Frömer et al. [7] and Eayrs et al. [9] each reported a total of 22,200 observations across four conditions (eight mixed blocks of 75 trials), yielding 5,550 observations per condition. Our number of observations per condition also aligns with the general recommendation by Brysbaert and Stevens [50], who, based on simulation studies, suggest that at least 1,600 observations per condition are needed to detect small effects (Cohen's d=0.1–0.2) in repeated-measures designs using mixed-effects models.

Procedure

The task was an adjusted version of the Stroop task used by Eayrs et al. [9]. An illustration of the structure of the task is presented in Fig. 1. The Stroop stimuli consisted of the Dutch color words for "red," "purple," "green," and "blue" ("rood," "paars," "groen," and "blauw," respectively). Stimuli were presented against a grey background with an RGB value of (176, 176, 176), and were always presented incongruently in one of the remaining three font colors: red, green, blue, or purple, with RGB values of (255, 0, 0), (0, 128, 0), (0, 0, 255), and (128, 0, 128), respectively. Thus, the font color never matched the semantic meaning of the color word. In line with previous work [7, 9], the luminance of these colors was not equated. Yet, crucially, font colors were balanced across reward and efficacy conditions, such that each word-color combination occurred equally often. The stimuli were

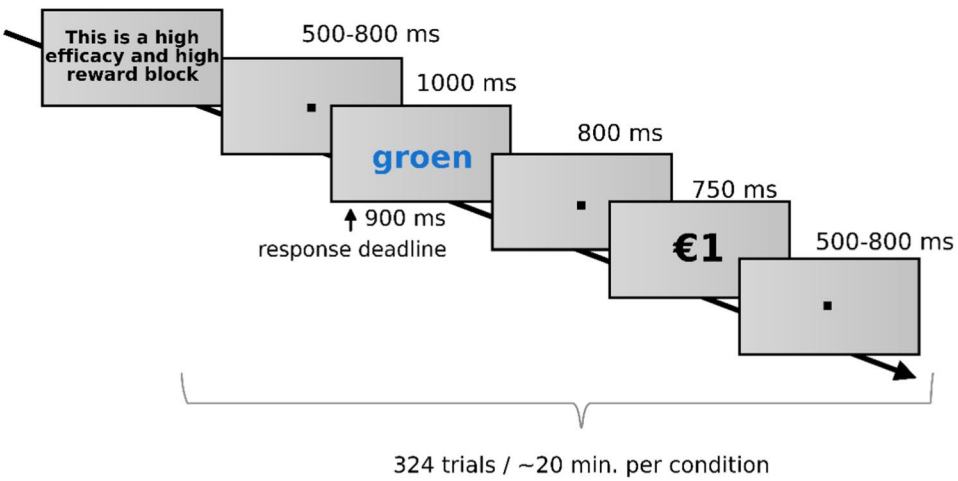

Fig. 1 Illustration of the Stroop task. Stimuli were always presented incongruently in one of the other three font colors ("groen" translates to green). Experimental sessions consisted of four conditions (combinations of efficacy/no efficacy and high/low reward) presented in a block design

presented at central fixation in random order. Each Stroop stimulus was presented for 1000 ms and was followed by a blank screen with a fixation dot presented for 800 ms. Participants were instructed to respond as quickly and accurately as possible by pressing the keys corresponding to the font color using the middle and index fingers of both hands: left middle finger on D, left index finger on F, right index finger on J, and right middle finger on K. Color-key mappings were counterbalanced across participants. We followed the previous studies by Frömer et al. [7] and Eayrs et al. [9] and used a fixed response deadline that was held constant across all participants. Specifically, responses had to be accurate and faster than 900 ms to qualify for a potential reward (the ultimate reward was determined through an additional trial-selection procedure detailed below). Finally, a feedback screen was presented for 750 ms, indicating the potentially earned reward, rather than explicitly stating whether the response was correct or incorrect. In high-efficacy blocks, participants received "€0.1" or "€1" for correct and fast responses in low- and high-reward trials, respectively, and "€0" for incorrect or slow responses. In low-efficacy blocks, reward outcomes were not contingent on trial-level performance but were matched in frequency to the corresponding high-efficacy blocks. The feedback screen was followed by a second fixation dot with a jittered presentation duration between 500 and 800 ms. The task was programmed using OpenSesame 3.3.14 [51].

Before the main task, participants completed a practice phase. The first practice block included 12 trials (three repetitions of each font color) to familiarize participants with the Stroop task. Afterwards, detailed instructions were presented in which the efficacy and reward structure were explained. Another practice block consisting of 24 trials followed, this time simulating all aspects of the main task, including receiving information about efficacy and reward (the practice block was always high-efficacy/high-reward), and feedback regarding their accuracy and reward rate. Participants had to achieve at least 70% accuracy in one of these training blocks to proceed to the main task.

Experimental sessions consisted of four conditions (combinations of efficacy/no efficacy and high/low reward) presented in a block design, with each block starting with the presentation of information regarding the efficacy and reward conditions: "This is a high/low efficacy and high/low reward block. Your potential reward depends on your performance/is determined at random. Trials can be rewarded with €1/€0.1". Participants completed all four conditions in a counterbalanced order (determined using a Latin square design), with a one-minute break between conditions to minimize carryover effects. After exclusions, three conditions had 8 participants each, and one condition had 6 participants. To check for any effect of block order, all analyses were re-run with counterbalancing condition included as a nuisance variable; this had no impact on the results, and there was never a significant effect of counterbalancing condition. Each condition involved 324 trials, resulting in a duration of approximately 20 min per condition. The experimenter was present in the room but shielded from the view of the participants.

Due to the block-wise manipulation of efficacy and reward in the present study, our reward probability calculation for low-efficacy trials differed from the one used by Eayrs et al. [9] and Frömer et al. [7], who used a rolling window of 10 previous trials to calculate the reward probability. For low-efficacy trials in our study, whether a specific trial was potentially rewarded or not was independent of performance and was based on the reward probability in high-efficacy blocks to match reward rates between conditions. If the experiment started with a low-efficacy block, the accuracy from the last practice block (where participants had to reach an accuracy of 70% or more) was used. As explained above, for high-efficacy trials, only correct responses within the 900 ms deadline were potentially rewarded. At the end of the experiment, 10 trials were randomly selected across all conditions to determine the total bonus reward, meaning rewards theoretically ranged from 0 to 10 Euros. The reward procedure was explained to the participants before they started the main task.

Pupillometry

Pupil data were collected using an SR Research EyeLink 1000 Plus eye tracking device with a 1000 Hz sampling rate. Participants were tested in a dimly-lit room, seated approximately 90 cm from the screen (with a resolution of 1920 × 1080 and a screen refresh rate of 240 Hz), with their heads placed in a chin rest to stabilize head position. The lens of the eyetracker was positioned 60 cm from the eye. A nine-point calibration was performed

at the beginning of each task block. Data from both eyes were collected and pre-processed using the "pypillometry" package [52] in Python. For each participant, data from both eyes were imported and checked for quality through visual inspection, and where possible we used the average pupil size across both eyes for our analyses. For four participants, data quality for one eye was poor due to technical issues, and we only used the data from the other eye. The continuous data were first low-pass filtered using a Butterworth filter with a cutoff frequency of 9 Hz. Blinks were detected using velocity thresholds following the procedure taken from Mathôt [53], with velocity parameters optimized for each individual participant. Temporally adjacent blinks were merged, with the required time of clean signal between blinks again optimized for each participant (to minimize interpolation artifacts). A cubic spline interpolation function [53] was used to reconstruct data affected by blinks. Lastly, the data were downsampled to 500 Hz, and absolute (non-baselined) pupil size for each trial was exported to RStudio (v2023.9.1.494; [54]) for further analysis.

Statistics

Analyses were conducted using R 4.2.3 [55] with a mixed effects modelling approach using the "glmer" function from the "lme4" package (v1.1-32; [56]). We used the "emmeans" package [57] for pairwise comparison of significant interactions. Predictors included efficacy (high/low), reward (high/low), and block half (first/second), as well as their interactions. For each analysis, the maximal converging random effect structure was determined using the "buildmer" package (v2.11; [58]). In case no random effect structure converged, we resorted to the use of repeated-measures ANOVA for analysis instead, using the "afex" package [59]. All predictors were within-subject and sum-to-zero coded.

For accuracy, the maximal converging random effect structure included random slopes and intercepts for each participant and for efficacy, reward, block half, and the efficacy x reward, and efficacy x block half interactions. For RT, the maximal converging random effect structure included random slopes and intercepts for each participant, and for all predictors and their interactions. In addition to these separate analyses of RT and accuracy, we also analysed our behavioral data in a fashion that controls for speed-accuracy trade-offs using the Balanced Integration Score (BIS; [60]). The BIS is computed as the difference in standardized mean correct response times and standardized proportion correct responses, meaning RT and accuracy are integrated into a single score with equal weights. Because of this equal weighting, BIS has been found to be relatively insensitive to speed-accuracy trade-offs, even when compared to, for example, the often used Inverse Efficiency Score [61], and has been shown to successfully capture "real" effects that are not a product of speed-accuracy trade-offs [62]. For the BIS, no random-effects structure converged because the BIS scores are computed from *z*-transformed reaction time and accuracy measures, which substantially attenuate between-subject variability, thus making it difficult for the model to reliably estimate random effects. For this reason, we resorted to a $2 \times 2 \times 2$ repeated measures ANOVA, with efficacy (high, low), reward (high, low), and block half (first, second) as factors.

For tonic pupil size, we used the average pupil size in the 500 ms preceding stimulus presentation. Additionally, we looked at the target-related phasic pupil size by using the average baseline-corrected data from a window from 400 to 1500 ms post-stimulus, after baseline-correcting the data with regard to the 500 ms pre-stimulus baseline. This time-window was chosen agnostic to any condition differences, aiming to both cover a typical time-range for phasic pupil responses, but to also include an early initial local peak in the pupil response (see Fig. 3), which was clearly visible. For both tonic and phasic pupil size, the maximal converging random effect structure included random slopes and intercepts for each participant and for all of our predictors and their interactions, except for the three-way interaction.

Blink data from the entire duration of each experimental block were extracted from the raw data files obtained from the EyeLink 1000, where blinks are defined as events in which, due to eyelid occlusion, the pupil size is extremely small, or the pupil signal is lost or distorted. Following recommendations from Hollander and Huette [47], we visually inspected our data, and excluded blink durations shorter than 50 ms and longer than 500 ms from the analyses (Hollander and Huette [47] also state that signal loss shorter than 85 ms is unlikely to be a

blink. However, occlusion has been shown to last between around 40 ms and 200 ms [63], and can of course last longer for voluntary blinks). Our time-range was selected agnostic to any potential condition differences. For each participant, we averaged the number of blinks and the blink durations within each condition across both eyes, separately for the first and second half of each experimental block. For both blink rate and blink duration, the maximal converging random effects structure included random slopes and intercepts for each participant, as well as for efficacy, reward, and the efficacy x reward interaction.

Results

Behavioral

Accuracy

Estimated marginal means for each condition as a function of block half are presented in Fig. 2A. A generalized linear mixed model (GLMM) of accuracy (model specification was as follows: *accuracy* ~ 1 + efficacy × reward × half + (1 + efficacy + reward + half + efficacy: reward + efficacy: half | subject), with a binomial family and logit link function) revealed a significant main effect of efficacy, Wald $\chi^2(1) = 9.92$, $p = .002$. In line with our expectations, accuracy was higher in high-efficacy conditions compared to low-efficacy conditions, $b = 0.22$, $SE = 0.07$, $z = 3.15$, $p = .002$. Contrary to our expectations however, we found no significant main effects for reward or block half, and no significant interactions.

Response time

False responses and responses faster than 250 ms and slower than 900 ms were excluded from the analyses (including responses slower than the 900 ms cut-off did not meaningfully impact results). Overall, 10.40% of trials were excluded, comprising 6.87% incorrect responses, 3.53% trials with RTs above 900 ms, and no trials with RTs below 250 ms (the few trials that were faster than 250 ms were already filtered out for also containing incorrect responses). Exclusion rates were similar across conditions, ranging from 8.90% to 11.41%: 8.90% in efficacy/high-reward, 10.25% in efficacy/low-reward, 11.41% in no-efficacy/high-reward, and 11.05% in no-efficacy/low-reward trials. Exclusion rates were also virtually identical across block halves (first: 10.41%; second: 10.40%), indicating that later trials were not disproportionately excluded.

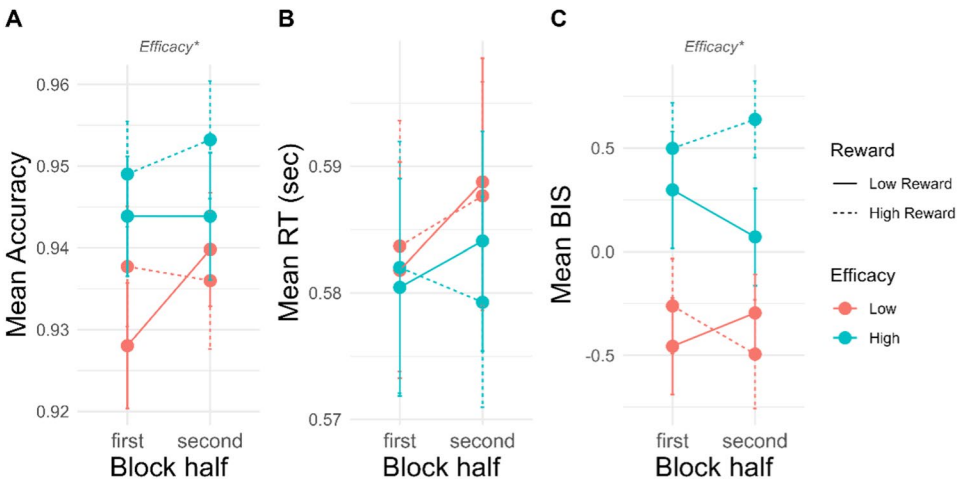

Fig. 2 Behavioral results as a function of efficacy, reward, and block half. (**A**) Estimated marginal means for accuracy in each condition. (**B**) Estimated marginal means for response time in each condition. (**C**) Mean BIS scores for each condition. Error bars represent standard error of the mean. Asterisked labels denote significant effects, $p < .05$

Estimated marginal means for each condition as a function of block half are presented in Fig. 2B. Contrary to our expectations, a GLMM of response time (model specification was as follows: *response time* ~ 1 + efficacy × reward × half + (1 + efficacy x reward x half | subject), with an inverse Gaussian family and log link function) showed no significant main effects, and no significant interactions.

Balanced integration score

Since individual effects for accuracy and RT can be the result of differential speed-accuracy trade-offs, which may partially camouflage as genuine performance effects, we furthermore used the BIS to assess overall task performance without overemphasizing accuracy or speed alone. For the RT component of the BIS, we applied the same exclusion criteria as in the RT analyses reported above. Mean BIS scores for each condition as a function of block half are presented in Fig. 2C. A repeated-measures ANOVA revealed a significant main effect of efficacy where scores were higher in the high-efficacy condition compared to the low-efficacy condition, $F(1, 29) = 12.67$, $p = .001$, $\eta^2_G = 0.084$. This thus confirms that participants achieved better overall performance under high-efficacy conditions, through higher accuracy without a cost in RT. None of the other main effects or interactions were significant.

Pupillometry

Tonic pupil size

The time course of mean tonic pupil size for each condition as a function of block half is presented in Fig. 3A. Analyses included all trials following our pre-processing procedures (see above). Estimated marginal means of tonic pupil size within our window of interest (-500 ms to stimulus presentation) are presented as a function of block half in Fig. 4A. An LMM (model specification was as follows: *pupil size* ~ 1 + efficacy × reward × half + (1 + efficacy + reward + half + efficacy: reward + efficacy: half + reward: half | subject) showed a significant main effect of reward, Wald $\chi^2(1) = 10.10$, $p = .001$, where pupils were larger in low-reward conditions compared to high-reward conditions, $b = -42.26$, $SE = 13.30$, $t = -3.18$, $p = .004$. However, it is important to note that this effect is in the opposite direction to what we expected, and will be expanded upon in the discussion section. In line with our predictions, the main effect of block half was also significant, Wald $\chi^2(1) = 13.74$, $p < .001$, with smaller mean pupil size in the second half of the blocks, compared to the first half of the blocks, $b = 20.24$, $SE = 5.46$, $t = 3.71$, $p < .001$. Contrary to our expectations, results did not show a significant main effect of efficacy, and no significant interactions.

Phasic pupil size

The time course of mean phasic pupil size for each condition as a function of block half is presented in Fig. 3B. Analyses included all trials following our pre-processing procedures (see above). Estimated marginal means of phasic pupil size within our window of interest (400 to 1500 ms post-stimulus) are presented for each condition as a function of block half in Fig. 4B. As predicted, an LMM (model specification was as follows: *pupil size* ~ 1 + efficacy × reward × half + (1 + efficacy + reward + half + efficacy: reward + efficacy: half + reward: half | subject) showed a significant main effect of efficacy, Wald $\chi^2(1) = 5.65$, $p = .017$, where pupils were larger across high-efficacy conditions compared to low-efficacy conditions, $b = 9.69$, $SE = 4.08$, $t = 2.38$, $p = .024$. Results did not show the predicted main effects of reward or block half, and no significant two-way interactions. The three-way interaction, however, was significant, Wald $\chi^2(1) = 10.06$, $p = .002$, but did not follow the expected overall pattern of high-efficacy and high-reward blocks attenuating a decrease in pupil size across block halves (see below). Because the analysis window includes the time of response, and because our behavioral results showed an effect of efficacy on accuracy, we also ran a model which included accuracy as a nuisance variable. Accuracy

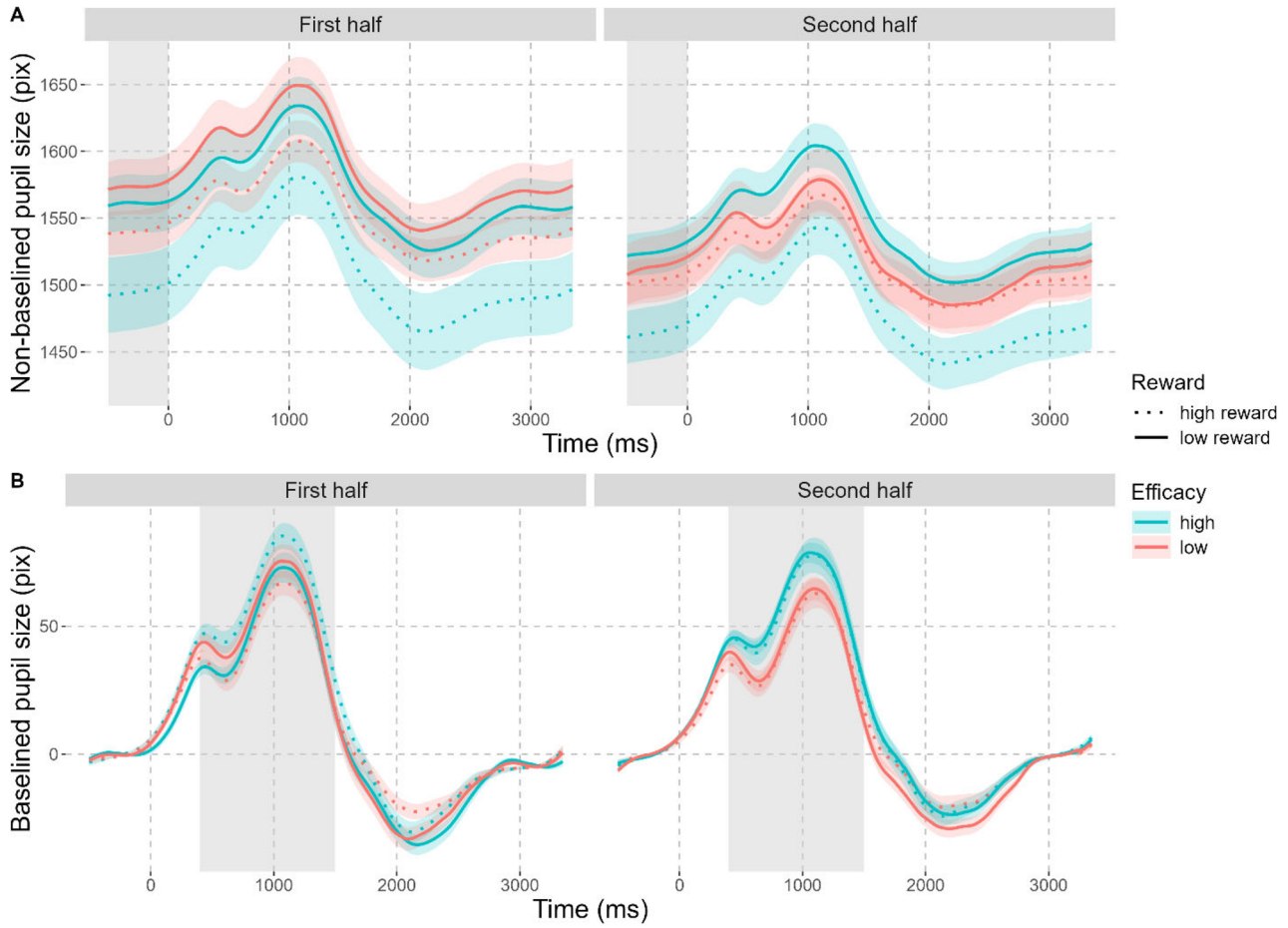

Fig. 3 Time course of mean pupil size for each condition as a function of block half. (**A**) Absolute tonic pupil size. (**B**) Baseline-corrected phasic pupil size. Shaded areas indicate windows of interest for the respective analyses. Ribbons represent Cousineau-Morey-corrected within-subject standard errors [64]

significantly predicted pupil size, with incorrect responses being associated with larger pupils, plausibly reflecting response monitoring. However, the inclusion of accuracy did not alter the pattern of results, in line with the fact that accuracy was very high throughout the task. Both the main effect of efficacy and the three-way interaction remained significant, and the pattern of estimated marginal means was essentially identical.

In line with our expectations, within the first block half, post-hoc contrasts with Tukey correction showed a significant effect of efficacy within high-reward conditions, where high efficacy was related to larger pupil size, b = -15.34, SE = 5.90, z = -2.60, p = .009, as well as a significant effect of reward within the high-efficacy conditions, where high reward was related to larger pupil size, b = -12.36, SE = 4.34, z = -2.85, p = .004. Within the second block half, however, Tukey corrected post-hoc contrasts showed a significant effect of efficacy only within the low-reward conditions, where high efficacy was related to larger pupils, b = -14.04, SE = 5.43, z = -2.59 p = .010. Finally, Tukey corrected post-hoc contrasts showed a significant effect of block half only in the high-efficacy/low-reward condition, where we did not observe attenuation of a decrease in pupil size as expected, but pupils were actually larger in the second block half than in the first, b = -8.72, SE = 3.95, z = -2.21, p = .027.

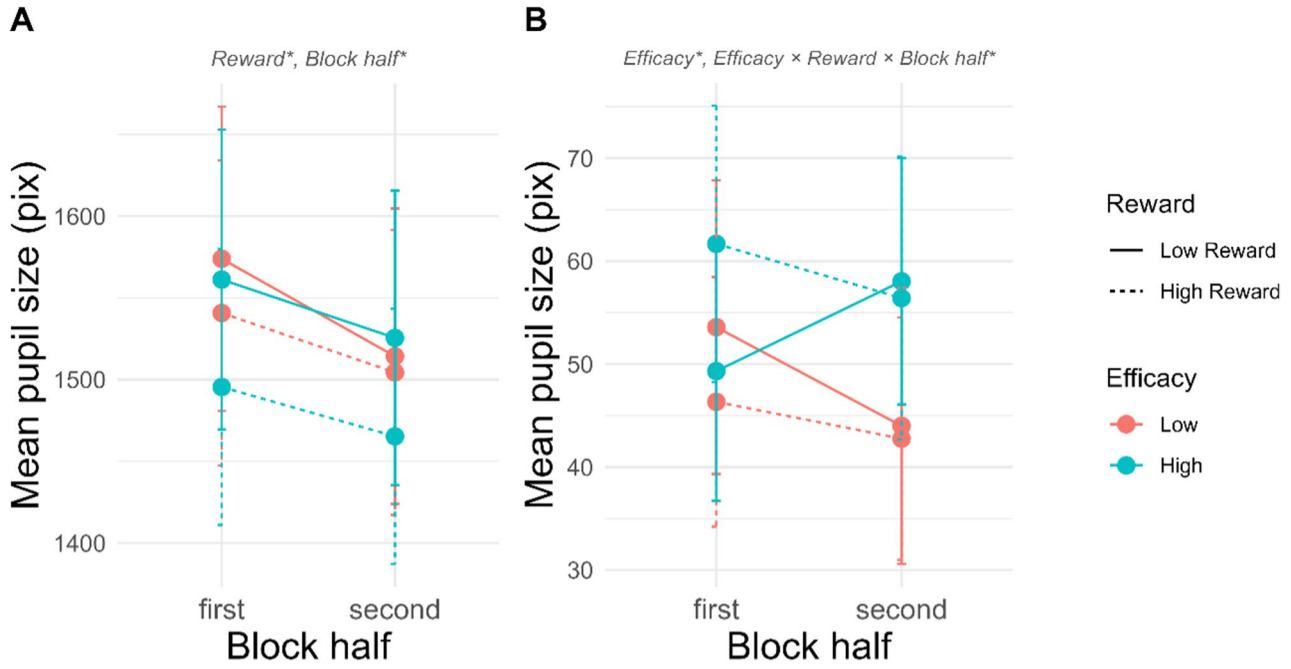

Fig. 4 Mean pupil size within our windows of interest for each condition as a function of block half. (**A**) Estimated marginal means for tonic pupil size from −500 ms to stimulus presentation. (**B**) Estimated marginal means for phasic pupil size from 400 to 1500 ms post-stimulus. Error bars represent standard error of the mean. Asterisked labels denote significant effects, $p < .05$

Blinks

Blink rate

Mean blink rate for each condition is presented as a function of block half in Fig. 5A. In line with our expectations, a GLMM of blink rate (model specification was as follows: *blink rate* ~ 1 + efficacy × reward × half + (1 + efficacy + reward + efficacy: reward | subject), with a Gamma family and identity link function) showed a significant main effect of block half, Wald $\chi^2(1) = 6.99$, $p = .008$, where the number of blinks was higher in the second half of blocks compared to the first half, $b = -2.54$, $SE = 0.96$, $t = -2.64$, $p = .008$. None of the other main effects or two-way interactions were significant. Although the three-way interaction, however, was significant, $\chi^2(1) = 3.95$, $p = .047$, it did not follow the expected pattern.

Tukey corrected post-hoc contrasts showed that within the high-efficacy/low-reward condition, participants blinked significantly more often in the second block half, compared to the first block half, $b = -10.23$, $SE = 4.09$, $z = -2.50$, $p = .012$. None of the other contrasts showed significant differences.

Blink duration

Mean blink duration for each condition is presented as a function of block half in Fig. 5B. In line with our expectations, a GLMM of blink duration (model specification was as follows: *blink duration* ~ 1 + efficacy × reward × half + (1 + efficacy + reward + efficacy: reward | subject), with a Gamma family and identity link function) showed a significant main effect of block half, Wald $\chi^2(1) = 54.03$, $p < .001$, where blinks were longer in duration in the second block half compared to the first block half, $b = -4.32$, $SE = 0.59$, $t = -7.35$, $p < .001$. Results further showed a significant interaction between efficacy and reward, Wald $\chi2(1) = 4.71$, $p = .030$, as well as a significant three-way interaction, Wald $\chi2(1) = 15.89$, $p < .001$, with these interactions not fully following the pattern we expected.

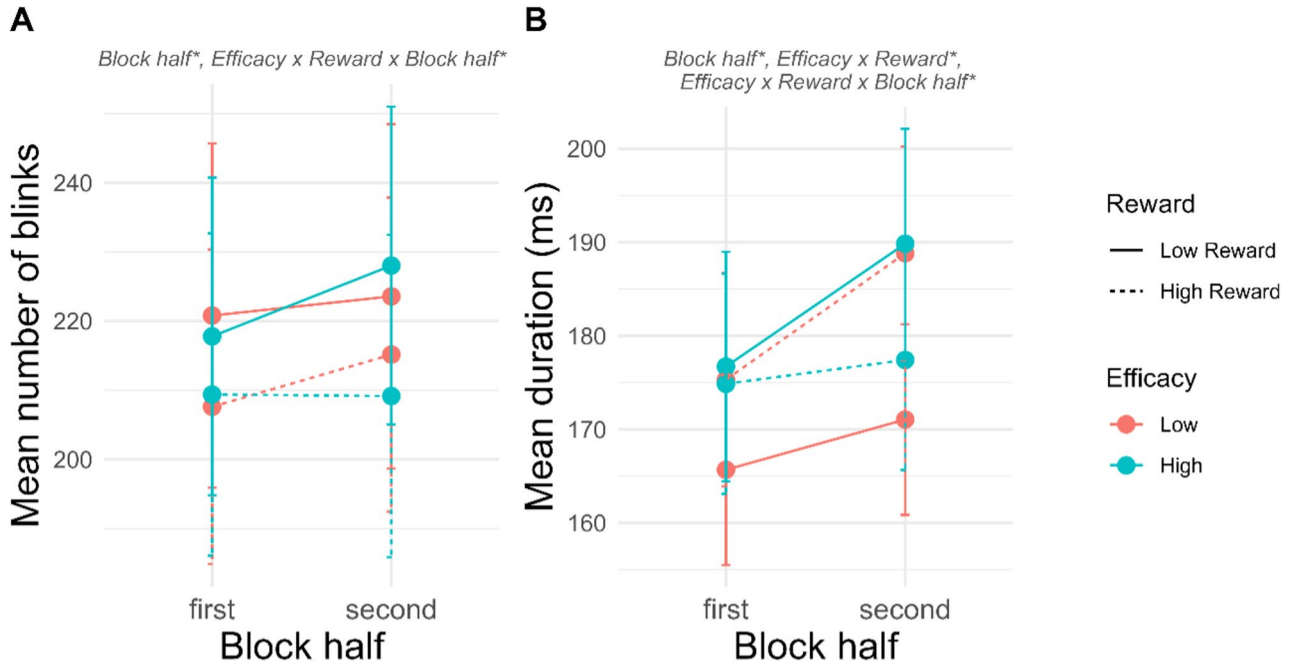

Fig. 5 Blink metrics for each condition as a function of block half. (**A**) Estimated marginal means for blink count. (**B**) Estimated marginal means for blink duration in ms. Error bars represent standard error of the mean. Asterisked labels denote significant effects, $p < .05$

Somewhat unexpectedly, Tukey corrected post-hoc contrasts showed that within the second block half of the low-reward conditions, blinks were significantly longer in the high-efficacy condition compared to the low-efficacy condition, $b = -18.79$, $SE = 7.39$, $z = -2.54$, $p = .011$. In addition, Tukey corrected post-hoc contrasts showed that within the second block half, there were significant effects of reward in both the low-efficacy and high-efficacy conditions, which were in opposite directions. Specifically, within the low-efficacy condition, blinks were longer in the high-reward condition compared to the low-reward condition, $b = -17.77$, $SE = 6.57$, $z = -2.70$, $p = .007$, whereas within the high-efficacy condition, blinks were shorter in the high-reward condition compared to the low-reward condition, $b = 12.42$, $SE = 5.86$, $z = 2.12$, $p = .034$. Lastly, and in line with our predictions, Tukey corrected post-hoc contrasts showed that blinks were significantly longer in the second block half compared to the first block half for the low-efficacy/low-reward, $b = -5.38$, $SE = 2.32$, $z = -2.31$, $p = .021$, low-efficacy/high-reward, $b = -13.52$, $SE = 2.44$, $z = -5.53$, $p < .001$, and high-efficacy/low-reward conditions, $b = -13.13$, $SE = 2.37$, $z = -5.55$, $p < .001$, but not for the high-efficacy/high-reward condition, where as expected, we found no difference in blink duration between block halves, $b = -2.55$, $SE = 2.27$, $z = -1.13$, $p = .261$.

Discussion

The expected value of control (EVC) model [6] posits that decisions pertaining to cognitive effort allocation rely on a cost-benefit calculation based on the expected value derived from engaging in different control processes. Important empirical evidence in support of the EVC account and related models has come from research focusing on the interaction between efficacy and reward [7–9], and less directly also from studies investigating the interaction between reward and time-on-task [10, 12, 27–29]. The present study aimed to further integrate these concepts by examining whether and how efficacy, reward, and time-on-task interact to influence cognitive effort over relatively long periods within a block design. While the behavioral data showed no effect of time-on-task or interactions with it, and no effect of reward, they generally displayed strong effects of efficacy, a pattern that was

also largely mirrored in phasic pupil size. This suggests that participants consistently allocated more effort when their performance had an influence on reward outcomes, leading to better overall performance in high-efficacy blocks. This was accompanied by a complex pattern of results in our ocular measures. Namely, tonic pupil size displayed a main effect of reward which ran counter to our expectation (with larger pupils in low-reward blocks) but nevertheless showed the expected general decrease in pupil size over time. Blink rate, on the other hand, displayed an expected general increase over time, with a three-way interaction which reflected the same general time-related increase reaching statistical significance only within the high-efficacy/low-reward condition. Finally, blink duration also increased over time as expected, but this effect depended on reward and efficacy in a three-way interaction. Together, the present data hence show time-on-task effects only on a physiological level, which may indicate compensatory processes keeping performance relatively stable over time. However, due to the absence of convergence with our behavioral data, we interpret these time-on-task effects with caution.

At the behavioral level, the separate results for accuracy and response speed indicated higher accuracy for high-efficacy blocks, whereas RT was largely unaffected by efficacy. This general pattern of results was further supported by an additional analysis using the balanced integration score (BIS), which integrates response speed and accuracy. These results showed that participants indeed performed better in the high-efficacy conditions compared to the low-efficacy conditions (i.e., the respective effect on accuracy was not brought about by a differential speed-accuracy trade-off). These results were consistent with our expectations, and provide corroborating evidence for the finding that efficacy information is integrated into the EVC calculation assumed to underlie cognitive effort allocation [7–9], extending this pattern to a context in which efficacy information was provided only once at the start of an experimental block. However, contrary to our expectations, we found no clear effects of reward or time-on-task. Nevertheless, it is interesting to note that in general (i.e., across accuracy, RT, and consequently also the BIS), numerically, performance seemed to be best and most stable for the high-efficacy and high-reward condition, which is in line with our expectation. While this could point to a lack of statistical power, it is important to emphasize that the task set-up was such that we expected behavioral time-on-task decrements, based on earlier work such as for example [10, 65], and for those to interact with our experimental factors. On the one hand, the absence thereof could mean that the task was not tiring (which can happen both for very easy and hard tasks, but is less likely in a task of medium difficulty), and/or that participants managed to compensate, and keep performance stable by applying extra effort (and/or profited from task improvements related to training; see below), which should be visible in the ocular measures we obtained.

Although our findings of better overall performance in the high-efficacy conditions are in line with results from earlier studies [7–9], our results also show some important deviations. While Frömer et al. [7] also found better accuracy in the high-efficacy conditions compared to the low-efficacy conditions, and no effect of reward (presented in supplementary Table 3 of their paper), Eayrs et al. [9] and Toobaei et al. [8] both did not find any effects of efficacy and reward on accuracy. More strikingly, all three studies report effects of efficacy and reward on response time [7–9], while we did not find any effects on accurate response times. One possible explanation for this difference is that the present study employed a block design, whereas the other three studies cued reward and efficacy at the trial level. The possibility that this difference in cueing format may have affected how effort was employed will be further elaborated below.

Results from our analysis of phasic pupil size were largely consistent with our behavioral results, with larger pupils under high-efficacy conditions than under low-efficacy conditions, but with an additional complex three-way interaction where in the first block half we found effects of efficacy and reward that were in the expected directions, but in the second block half the pattern reflected only an effect of efficacy in the low-reward conditions. This suggests that the increased accuracy within the high-efficacy blocks resulted from increased effort allocation, given that previous research has found phasic pupil size to be a valid marker for cognitive effort [18–20]. The effect of efficacy was more pronounced in the second block halves, possibly reflecting a learning effect where, with time-on-task, the efficacy manipulation became more salient or clear to participants. Somewhat unexpectedly, the more pronounced effect of efficacy in the second block halves seemed to come from the high-efficacy/low-reward condition seeing an increase in pupil size, whereas the other conditions each showed

a decrease in pupil size. Over the whole blocks however, similarly to our behavioral results, numerically we observed the largest mean pupil size within the high-efficacy/high-reward condition, so that the direction of these effects was as expected. Note that the above result pattern stayed the same when accounting for accuracy on a trial level, which had a main effect (larger pupils for incorrect trials, as one would expect), but otherwise did not interact with our results.

In contrast to our findings for phasic pupil size, tonic pupil size did reveal an effect of reward, which, however, ran counter to our expectations, as pupils were larger during low-reward blocks compared to high-reward blocks. This result is also not in accordance with earlier work showing that tonic pupil size increases with increased rewards [36–38]. In addition, there was an effect of time-on-task where tonic pupil size decreased with time, resulting in smaller average pupil size in the second half of blocks when compared to the first halves. This effect aligned with our expectations, as studies have shown that tonic pupil size decreases with time-on-task [28, 33, 34]. However, we found no evidence for any effect of efficacy on tonic pupil size, or of any interaction effects.

Furthermore, our tonic pupil size data do not appear to correspond, at first glance, either to our behavioral findings or to our results for phasic pupil size. One explanation for this could be that although (larger) tonic pupil size has been found to be an indicator of (increased) cognitive effort [31, 32], tonic pupil size has also been implicated in other cognitive processes. In particular, many studies have implicated tonic pupil size as a broad measure of arousal, which is affected by numerous factors [35]. As such, the present tonic pupil size results may be hard to interpret in a very specific fashion. Specifically, the significant decrease in tonic pupil size might reflect a drop in arousal levels in a fashion that relates to fatigue, whereas changes in this pattern may stem from motivational factors, varying the level of engagement with the task. A fitting alternative, and somewhat broader account, can be found in adaptive gain theory [66], which implicates the locus coeruleus (LC) in the regulation of behavior, which can be inferred via pupillometry [18]; see also [67]. Adaptive gain theory assumes two behavioral modes, exploitation and exploration, the former being focused on obtaining rewards in a given context, and the latter being more distractable to allow for changes in foraging strategy. Because exploration is associated with high (tonic) LC activity, it is similarly also associated with larger tonic pupil size [18]. Crucially, experimental studies have found evidence in favor of this account [68, 69]. Within the context of the adaptive gain theory, our findings may indicate that participants showed less exploration and task disengagement (which may furthermore be related to mind wandering; e.g. [70, 71]), in the high-reward blocks, compared to the low-reward blocks, though this effect seems to be primarily driven through the much smaller tonic pupil size within the high-efficacy/high-reward block. However, although this speculative alternative account provides a post-hoc explanation for why we may have found a reward effect in the opposite direction to what we expected, it does not provide an explanation for the lack of convergence with our behavioral findings and our phasic pupil size results.

Given its relationship to both fatigue [30, 40–43] and mind wandering [44–47], we also analysed blink data. First, as we expected, the blink rate results were also suggestive of a time-on-task effect, with participants blinking more frequently in the second half of blocks compared to the first half of blocks. While we also observed a significant three-way interaction, efficacy and reward did not attenuate this time-on-task increase in blink rate as we expected, as we only found a significant increase in blink rate within the high-efficacy/low-reward condition. Though not significant, blink rates were numerically lower in the high-reward conditions than in the low-reward conditions, which would at least seem consistent with our findings regarding tonic pupil size when these are viewed in the context of adaptive gain theory, although this interpretation remains speculative.

Blink durations similarly suggested the expected time-on-task effects, with longer blinks in the second half of blocks compared to the first. In contrast to blink rates, however, blink durations additionally showed interactions between reward and efficacy, as well as a three-way interaction. The blink duration results were the only measure that showed a pattern broadly consistent with attenuation of time-on-task effects in the high-efficacy/high-reward condition, such that blinks were longer in the second half of blocks for each condition except for the combined high-efficacy/high-reward block, in which the mean blink duration remained stable over time. Given that blink duration has been shown to be a reliable marker for cognitive fatigue and mind wandering [30, 42–44, 47, 48], this could be taken to suggest that time-on-task effects were less pronounced when both efficacy and reward were

high, although this interpretation should be treated cautiously given the mixed pattern across measures. In contrast to the previously discussed results however, where we observed (at least numerically) the best performance, the largest phasic pupil size, and the smallest tonic pupil size within the high-efficacy/high-reward condition, blinks were numerically shortest in the low-efficacy/low-reward condition. One explanation might be that in this condition, participants possibly were differently engaged with the task from the start, given that the mean blink duration in the first block half is essentially equal for the other three conditions, but considerably shorter in the low-efficacy/low-reward condition.

Jointly, our findings pertaining to blink data as well as tonic pupil size seem to suggest that there were some time-on-task effects. It is, however, remarkable that results at the behavioral level do not reflect these time-on-task effects. For this reason, as previously stated, we interpret these time-on-task effects with caution. Though block designs similar to the one employed in the present study have reported a decrease in performance over time (e.g [27, 28]), , our experimental design might nevertheless have been a factor in obscuring time-on-task effects on behavioral measures. For example, although it is well-established that cognitively effortful tasks are generally avoided [2, 3], which furthermore should arguably lead to fatigue if they cannot be avoided, there are circumstances where this does not hold. Embrey et al. [72] found that when participants are given a choice between performing a cognitive task and doing nothing at all for an extended period of time, they are likely to opt for the task in an attempt to avoid boredom, in particular in a lab context (as compared to online participation). Furthermore, while their study was not conducted in a time-on-task context, Wu et al. [73] found that within just 100 trials, participants chose to perform a Stroop task rather than do nothing in 40% of the trials. It hence seems possible that although our task was probably fatiguing to a certain degree, participants may have been willing to compensate for such effects in a fashion that was not strictly tied to rewards.

Another possible limiting factor could have been the Stroop task employed here in a block design. While Stroop effects are known to be very robust [74], and the task clearly requires the engagement of cognitive effort, it is important to note that we followed Eayrs et al. [9] and included only incongruent trials in the present study, while Frömer et al. [7] did include neutral and congruent trials as well. We made this choice in an attempt to maintain a consistently challenging task, and to avoid a possible congruency x performance confound in our results (more successful/rewarded congruent than incongruent trials). However, including only incongruent trials does mean that the present task does not actually test the Stroop effect, which presents as the difference in response times between congruent and incongruent trials. Although Eayrs et al. [9] broadly replicated Frömer et al.'s [7] behavioral findings, exclusively using incongruent trials eliminates a traditional Stroop baseline, representing a fundamental context change that may have reduced our sensitivity to reward and time-on-task effects.

The present block design may furthermore have affected how effort was employed. Specifically, the present experiment was based on tasks from Eayrs et al. [9], and Frömer et al. [7], both of which used trial-based cues. This is expected to make the motivational factors of reward and efficacy highly salient, in turn leading to more responsive allocation of cognitive control (but see [75]). In contrast, within our block design, the information regarding efficacy and reward was only presented once at the start of each block (though the feedback for each trial also communicated information about efficacy implicitly and information about reward explicitly). Consequently, the modified paradigm may have encouraged more sustained use of efficacy and reward information rather than transient control adjustments. Within this stable environmental structure, the robust effects of efficacy on accuracy, BIS, and phasic pupil size suggest that efficacy information continued to guide effort allocation when maintained across an experimental block, whereas reward effects were less consistently observed than in previous trial-wise implementations. It is possible that the stable, block-level format attenuated the immediate behavioral salience of reward incentives, although this explanation remains speculative because the present study did not directly compare blocked and trial-wise implementations. In addition, previous research has indicated that the Stroop task can be sensitive to practice [76, 77] (but see [65]). Although the nature of practice effects in a purely incongruent Stroop task might in some ways differ from a standard Stroop task, this could, in theory, still have led to the task getting easier over time, counteracting time-on-task effects as correct task performance would require less cognitive effort. Although our phasic pupil size (which did not decrease over time) and behavioral

results (which did not improve over time) do not immediately align with this, it is conceivable that training contributed in subtle ways to our results to some extent.

Furthermore, it is conceivable that participants got habituated to the trial-based feedback they received [78], causing them to disregard it and simply carry out the task. This could have especially impacted our results pertaining to reward, as the size of the reward remained constant for the entirety of each block, and could be an explanation for why we did not find effects of reward on behavior or phasic pupil size, which was in contrast to earlier studies [7–9]. Theoretically, the absence of a reward effect on performance could in principle also reflect range-adaptation or familiarity differences, given that participants practiced only under the high-efficacy/high-reward condition in the present study, unlike prior studies that included all conditions during practice [7, 9]. Including counterbalancing condition as a nuisance predictor did not change any behavioral or pupil results however, suggesting that practice-related familiarity is unlikely to account for the lack of reward effects in our behavioral data, or the surprising effect of reward in our tonic pupil results. Including a fixed response deadline of 900 ms may also have obscured reward effects. Although this ensured that the reward criterion was identical across participants, it may not have been equally challenging or attainable for everyone. However, the studies by Frömer [7] and Eayrs [9] also used fixed response deadlines (of 750 ms for both congruent and incongruent trials, and 1000 ms incongruent-trials-only respectively), and did observe an effect of reward on their behavioral measures. Furthermore, given that accuracy was generally very high, and we found no evidence of a speed-accuracy trade-off, the absence of a reward effect is unlikely to be explained by participants adopting a speed-focused strategy at the expense of accuracy. Nevertheless, the fixed deadline may have limited the sensitivity of the reward manipulation by being insufficiently tailored to individual differences in baseline response speed or perceived reward attainability.

Despite these possible explanations for the absence of reward effects, our behavioral and phasic pupil size results together suggest that increased effort allocation in the high-efficacy conditions nevertheless had a significant impact on performance. The fact that efficacy effects were more consistent across our measures than reward effects, and that interactions were often absent, could also be taken to suggest that participants may not have perceived the contrast between high and low reward as sufficiently strong. This may have been compounded by the fact that in our design they were not directly pitted against each other in a trial-based fashion. A recommendation for future studies would therefore be to employ a stronger and more salient reward manipulation when investigating the efficacy x reward interaction in a block design, such as using points to exaggerate the difference (e.g., 1 point versus 1000 points), and/or randomly selecting a larger portion of trials to contribute to the final reward. An additional recommendation for future studies would be to investigate the use of tasks which have been subjectively rated as taking more mental effort to successfully complete, such as for example arithmetic tasks or mental rotation tasks [79], to maximize the effects of experimental manipulations of both efficacy and reward.

Conclusion

In sum, our behavioral and psychophysiological results provide corroborative evidence in favor of the account that information regarding efficacy is integrated into the cost-benefit analysis underlying the allocation of cognitive control resources [7–9]. Extending previous findings obtained using trial-wise cueing paradigms, these effects were also observed during relatively long stretches of time within the blocked task structure employed in the present study. Specifically, our results pertaining to phasic pupil size and the balanced integration score (BIS) suggest that participants allocated more effort in conditions where their performance had an influence on reward outcomes, leading to better overall performance in high-efficacy blocks. In contrast, we did not observe the hypothesized effects of reward in behavioral performance or phasic pupil size, and reward-related effects observed in other ocular measures did not follow the predicted pattern. Moreover, the hypothesized interactions among efficacy, reward, and time-on-task were not supported across the behavioral and psychophysiological measures, with significant interactions generally not following the predicted pattern. Although tonic pupil

size and blink metrics suggested time-on-task changes, these patterns diverged from our behavioral and phasic pupillary findings and lacked a clear theoretical explanation. Consequently, they warrant cautious interpretation. Nevertheless, the data point to a possible split across measures: phasic pupil size and behavior showed effects consistent with efficacy-based effort allocation, whereas tonic pupil size and blink measures showed progressive time-on-task shifts, potentially reflecting emerging fatigue or disengagement, although the specific interpretation of these latter effects remains uncertain. Ultimately, this study contributes to the cognitive effort literature by jointly examining efficacy, reward, and time-on-task within a single paradigm, while underscoring the nuanced multidimensionality of ocular measures in tracking cognitive control over time.

Abbreviations

EVC	Expected value of control
CNV	Contingent negative variation
BIS	Balanced integration score
(G)LMM	(Generalized) linear mixed model
LC	Locus coeruleus
RT	Response time

Acknowledgements Not applicable.

Author contributions S.T.S., N.P., J.O.E., N.K., W.N., R.M.K., J.R.W., and C.N.B. conceived and designed the study. S.T.S. collected the data. W.J.L. and S.T.S. analysed and interpreted the data. W.J.L., S.T.S., and C.N.B. drafted the manuscript. All authors provided reviews and critical revisions, and approved the submitted version of the manuscript.

Funding This work was supported by a concerted research project awarded to R.M.K, C.N.B., W.N., and J.R.W. (Special Research Fund Ghent University, Grant number BOF20-. GOA-004).

Data availability The datasets generated and analysed during the current study are available in the Open Science Framework repository, [https://osf.io/fht7c].

Declarations

Ethics approval and consent to participate The present study was conducted in accordance with the ethical rules and regulations presented in the General Ethics Protocol of the Faculty of Psychology and Educational Sciences of Ghent University (https://www.ugent.be/pp/en/research/ec/general_ethics_protocol_fppw.pdf), and was conducted in accordance with the Declaration of Helsinki. In line with the General Ethics Protocol, specific approval from the faculty's Ethics Committee was not required, as the study fell within the scope of research for which formal review is waived (see Sect. 3.1). All participants provided written informed consent prior to participation.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

1. Shenhav A, Musslick S, Lieder F, Kool W, Griffiths TL, Cohen JD, et al. Toward a rational and mechanistic account of mental effort. Annu Rev Neurosci. 2017;40:99–124. https://doi.org/10.1146/annurev-neuro-072116-031526.
2. Kool W, McGuire JT, Rosen ZB, Botvinick MM. Decision making and the avoidance of cognitive demand. J Exp Psychol Gen. 2010;139:665–82. https://doi.org/10.1037/a0020198.
3. Westbrook A, Braver TS. Cognitive effort: A neuroeconomic approach. Cogn Affect Behav Neurosci. 2015;15:395–415. https://doi.org/10.3758/s13415-015-0334-y.
4. Inzlicht M, Shenhav A, Olivola CY. The Effort Paradox: Effort is both costly and valued. Trends Cogn Sci. 2018;22:337–49. https://doi.org/10.1016/j.tics.2018.01.007.
5. Massin O. Towards a definition of efforts. Motiv Sci. 2017;3:230–59. https://doi.org/10.1037/mot0000066.
6. Shenhav A, Botvinick MM, Cohen JD. The expected value of control: An integrative theory of anterior cingulate cortex function. Neuron. 2013;79:217–40. https://doi.org/10.1016/j.neuron.2013.07.007.
7. Frömer R, Lin H, Dean Wolf CK, Inzlicht M, Shenhav A. Expectations of reward and efficacy guide cognitive control allocation. Nat Commun. 2021;12:1030. https://doi.org/10.1038/s41467-021-21315-z.
8. Toobaei M, Taghavi M, Goodarzi MA, Sarafraz M, Jobson L. Exploring expected reward and efficacy in enhancing cognitive control in patients with depression. J Clin Exp Neuropsychol. 2023;45:636–46. https://doi.org/10.1080/13803395.2023.2287782.
9. Eayrs JO, Tobing HS, Steendam ST, Prutean N, Notebaert W, Wiersema JR, et al. Reward and efficacy modulate the rate of anticipatory pupil dilation. Psychophysiology. 2025;62:e14761. https://doi.org/10.1111/psyp.14761.
10. Esterman M, Grosso M, Liu G, Mitko A, Morris R, DeGutis J. Anticipation of monetary reward can attenuate the vigilance decrement. PLoS ONE. 2016;11:e0159741. https://doi.org/10.1371/journal.pone.0159741.
11. Garner LD, Shuman J, Robison MK. Piece-rate time-based incentives improve sustained attention. J Appl Res Mem Cogn. 2024;13:342–53. https://doi.org/10.1037/mac0000141.
12. Herlambang MB, Taatgen NA, Cnossen F. The role of motivation as a factor in mental fatigue. Hum Factors. 2019;61:1171–85. https://doi.org/10.1177/0018720819828569.
13. Bafna T, Hansen JP. Mental fatigue measurement using eye metrics: A systematic literature review. Psychophysiology. 2021;58:e13828. https://doi.org/10.1111/psyp.13828.
14. Bandura A. Self-efficacy: Toward a unifying theory of behavioral change. Psychol Rev. 1977;84:191–215. https://doi.org/10.1037/0033-295X.84.2.191.
15. Bandura A. The explanatory and predictive scope of self-efficacy theory. J Soc Clin Psychol. 1986;4:359–73. https://doi.org/10.1521/jscp.1986.4.3.359.
16. Li X, Xia B, Shen G, Dong R, Xu S, Yang L. The interplay of depressive symptoms and self-efficacy in adolescents: a network analysis approach. Front Psychol. 2024;15:1419920. https://doi.org/10.3389/fpsyg.2024.1419920.
17. Zhang B-F, Zhang X-Y. Correlation between self-efficacy, parental parenting patterns, and severe depression in adolescents. World J Psychiatry. 2024;14:1827–35. https://doi.org/10.5498/wjp.v14.i12.1827.
18. Mathôt S, Pupillometry. Psychology, physiology, and function. J Cogn. 2018;1:16. https://doi.org/10.5334/joc.18.
19. Alnaes D, Sneve MH, Espeseth T, Endestad T, Van De Pavert SHP, Laeng B. Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. J Vis. 2014;14:1–20. https://doi.org/10.1167/14.4.1.
20. Van Der Wel P, Van Steenbergen H. Pupil dilation as an index of effort in cognitive control tasks: A review. Psychon Bull Rev. 2018;25:2005–15. https://doi.org/10.3758/s13423-018-1432-y.
21. Unsworth N, Miller AL. Individual differences in the intensity and consistency of attention. Curr Dir Psychol Sci. 2021;30:391–400. https://doi.org/10.1177/09637214211030266.
22. Hockey GRJ. A motivational control theory of cognitive fatigue. In: Ackerman PL, editor. Cognitive fatigue: Multidisciplinary perspectives on current research and future applications. Washington: American Psychological Association; 2011. pp. 167–87. https://doi.org/10.1037/12343-008.
23. Van Tilburg WAP, Igou ER. On boredom: Lack of challenge and meaning as distinct boredom experiences. Motiv Emot. 2012;36:181–94. https://doi.org/10.1007/s11031-011-9234-9.
24. Kunasegaran K, Ismail AMH, Ramasamy S, Gnanou JV, Caszo BA, Chen PL. Understanding mental fatigue and its detection: a comparative analysis of assessments and tools. PeerJ. 2023;11:e15744. https://doi.org/10.7717/peerj.15744.
25. Tanaka M, Ishii A, Watanabe Y. Neural effects of mental fatigue caused by continuous attention load: A magnetoencephalography study. Brain Res. 2014;1561:60–6. https://doi.org/10.1016/j.brainres.2014.03.009.
26. Kurzban R, Duckworth A, Kable JW, Myers J. An opportunity cost model of subjective effort and task performance. Behav Brain Sci. 2013;36:661–79. https://doi.org/10.1017/S0140525X12003196.
27. Hopstaken JF, Van Der Linden D, Bakker AB, Kompier MAJ. The window of my eyes: Task disengagement and mental fatigue covary with pupil dynamics. Biol Psychol. 2015;110:100–6. https://doi.org/10.1016/j.biopsycho.2015.06.013.
28. Hopstaken JF, Van Der Linden D, Bakker AB, Kompier MAJ. A multifaceted investigation of the link between mental fatigue and task disengagement. Psychophysiology. 2015;52:305–15. https://doi.org/10.1111/psyp.12339.

29. Boksem MAS, Meijman TF, Lorist MM. Mental fatigue, motivation and action monitoring. Biol Psychol. 2006;72:123–32. https://doi.org/10.1016/j.biopsycho.2005.08.007.
30. Yamada Y, Kobayashi M. Detecting mental fatigue from eye-tracking data gathered while watching video: Evaluation in younger and older adults. Artif Intell Med. 2018;91:39–48. https://doi.org/10.1016/j.artmed.2018.06.005.
31. Zénon A. Eye pupil signals information gain. Proc R Soc B Biol Sci. 2019;286:20191593. https://doi.org/10.1098/rspb.2019.1593.
32. Peysakhovich V, Vachon F, Dehais F. The impact of luminance on tonic and phasic pupillary responses to sustained cognitive load. Int J Psychophysiol. 2017;112:40–5. https://doi.org/10.1016/j.ijpsycho.2016.12.003.
33. Burke SE, Babu Henry Samuel I, Zhao Q, Cagle J, Cohen RA, Kluger B, et al. Task-based cognitive fatigability for older adults and validation of mental fatigability subscore of Pittsburgh fatigability scale. Front Aging Neurosci. 2018;10:327. https://doi.org/10.3389/fnagi.2018.00327.
34. Hopstaken JF, Van Der Linden D, Bakker AB, Kompier MAJ, Leung YK. Shifts in attention during mental fatigue: Evidence from subjective, behavioral, physiological, and eye-tracking data. J Exp Psychol Hum Percept Perform. 2016;42:878–89. https://doi.org/10.1037/xhp0000189.
35. Laeng B, Sirois S, Gredebäck G, Pupillometry. A window to the preconscious? Perspect Psychol Sci. 2012;7:18–27. https://doi.org/10.1177/1745691611427305.
36. Da Silva Castanheira K, LoParco S, Otto AR. Task-evoked pupillary responses track effort exertion: Evidence from task-switching. Cogn Affect Behav Neurosci. 2021;21:592–606. https://doi.org/10.3758/s13415-020-00843-z.
37. Cole L, Lightman S, Clark R, Gilchrist ID. Tonic and phasic effects of reward on the pupil: implications for locus coeruleus function. Proc R Soc B Biol Sci. 2022;289:20221545. https://doi.org/10.1098/rspb.2022.1545.
38. Kostandyan M, Bombeke K, Carsten T, Krebs RM, Notebaert W, Boehler CN. Differential effects of sustained and transient effort triggered by reward – A combined EEG and pupillometry study. Neuropsychologia. 2019;123:116–30. https://doi.org/10.1016/j.neuropsychologia.2018.04.032.
39. Stern JA, Boyer D, Schroeder D. Blink rate: A possible measure of fatigue. Hum Factors. 1994;36:285–97. https://doi.org/10.1177/001872089403600209.
40. Groen Y, Börger NA, Koerts J, Thome J, Tucha O. Blink rate and blink timing in children with ADHD and the influence of stimulant medication. J Neural Transm. 2017;124:27–38. https://doi.org/10.1007/s00702-015-1457-6.
41. Maffei A, Angrilli A. Spontaneous eye blink rate: An index of dopaminergic component of sustained attention and fatigue. Int J Psychophysiol. 2018;123:58–63. https://doi.org/10.1016/j.ijpsycho.2017.11.009.
42. McIntire LK, McKinley RA, Goodyear C, McIntire JP. Detection of vigilance performance using eye blinks. Appl Ergon. 2014;45:354–62. https://doi.org/10.1016/j.apergo.2013.04.020.
43. Zargari Marandi R, Madeleine P, Omland Ø, Vuillerme N, Samani A. Eye movement characteristics reflected fatigue development in both young and elderly individuals. Sci Rep. 2018;8:13148. https://doi.org/10.1038/s41598-018-31577-1.
44. Grandchamp R, Braboszcz C, Delorme A. Oculometric variations during mind wandering. Front Psychol. 2014;5. https://doi.org/10.3389/fpsyg.2014.00031.
45. Riby LM, Marr L, Barron-Millar L, Greer J, Hamilton CJ, McGann D, et al. Elevated blink rates predict mind wandering: Dopaminergic insights into attention and task focus. J Integr Neurosci. 2025;24:26508. https://doi.org/10.31083/JIN26508.
46. Smilek D, Carriere JSA, Cheyne JA. Out of mind, out of sight: Eye blinking as indicator and embodiment of mind wandering. Psychol Sci. 2010;21:786–9. https://doi.org/10.1177/0956797610368063.
47. Hollander J, Huette S. Extracting blinks from continuous eye-tracking data in a mind wandering paradigm. Conscious Cogn. 2022;100:103303. https://doi.org/10.1016/j.concog.2022.103303.
48. Huette S, Mathis A, Graesser A. Blink durations reflect mind wandering during reading. Proc Annu Meet Cogn Sci Soc. 2016;38.
49. Cori JM, Anderson C, Shekari Soleimanloo S, Jackson ML, Howard ME. Narrative review: Do spontaneous eye blink parameters provide a useful assessment of state drowsiness? Sleep Med Rev. 2019;45:95–104. https://doi.org/10.1016/j.smrv.2019.03.004.
50. Brysbaert M, Stevens M. Power analysis and effect size in mixed effects models: A tutorial. J Cogn. 2018;1:9. https://doi.org/10.5334/joc.10.
51. Mathôt S, Schreij D, Theeuwes J, OpenSesame. An open-source, graphical experiment builder for the social sciences. Behav Res Methods. 2012;44:314–24. https://doi.org/10.3758/s13428-011-0168-7.
52. Mittner M. pypillometry: A Python package for pupillometric analyses. J Open Source Softw. 2020;5:2348. https://doi.org/10.21105/joss.02348.
53. Mathôt S. A simple way to reconstruct pupil size during eye blinks. Figshare. 2013. https://doi.org/10.6084/M9.FIGSHARE.688001.
54. Posit Team. RStudio: Integrated development environment for R. 2024. Available: http://www.posit.co/
55. R Core Team. R: A language and environment for statistical computing. 2021. Available: https://www.R-project.org/

56. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67. https://doi.org/10.18637/jss.v067.i01.
57. Lenth Remmeans. Estimated Marginal Means, aka Least-Squares Means. 2025. Available: https://rvlenth.github.io/emmeans/
58. Voeten CC. buildmer: Stepwise elimination and term reordering for mixed-effects regression. 2025. Available: https://CRAN.R-project.org/package=buildmer
59. Singmann H, Bolker B, Westfall J, Aust F, Ben-Schachar M. afex: Analysis of factorial experiments. 2024. Available: https://CRAN.Rproject.org/package=afex
60. Liesefeld HR, Fu X, Zimmer HD. Fast and careless or careful and slow? Apparent holistic processing in mental rotation is explained by speed-accuracy trade-offs. J Exp Psychol Learn Mem Cogn. 2015;41:1140–51. https://doi.org/10.1037/xlm0000081.
61. Townsend JT, Ashby FG. The stochastic modeling of elementary psychological processes. Cambridge [Cambridgeshire]; New York: Cambridge University Press; 1983.
62. Liesefeld HR, Janczyk M. Combining speed and accuracy to control for speed-accuracy trade-offs(?). Behav Res Methods. 2019;51:40–60. https://doi.org/10.3758/s13428-018-1076-x.
63. Willett SM, Maenner SK, Mayo JP. The perceptual consequences and neurophysiology of eye blinks. Front Syst Neurosci. 2023;17:1242654. https://doi.org/10.3389/fnsys.2023.1242654.
64. Morey RD. Confidence intervals from normalized data: A correction to Cousineau (2005). Tutor Quant Methods Psychol. 2008;4:61–4. https://doi.org/10.20982/tqmp.04.2.p061.
65. Mangin T, Audiffren M, Lorcery A, Mirabelli F, Benraiss A, André N. A plausible link between the time-on-task effect and the sequential task effect. Front Psychol. 2022;13:998393. https://doi.org/10.3389/fpsyg.2022.998393.
66. Aston-Jones G, Cohen JD. An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. Annu Rev Neurosci. 2005;28:403–50. https://doi.org/10.1146/annurev.neuro.28.061604.135709.
67. Wienke C, Grueschow M, Haghikia A, Zaehle T. Phasic, event-related transcutaneous auricular vagus nerve stimulation modifies behavioral, pupillary, and low-frequency oscillatory power responses. J Neurosci. 2023;43:6306–19. https://doi.org/10.1523/JNEUROSCI.0452-23.2023.
68. Gilzenrat MS, Nieuwenhuis S, Jepma M, Cohen JD. Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. Cogn Affect Behav Neurosci. 2010;10:252–69. https://doi.org/10.3758/CABN.10.2.252.
69. Jepma M, Nieuwenhuis S. Pupil diameter predicts changes in the exploration–exploitation trade-off: Evidence for the adaptive gain theory. J Cogn Neurosci. 2011;23:1587–96. https://doi.org/10.1162/jocn.2010.21548.
70. Franklin MS, Broadway JM, Mrazek MD, Smallwood J, Schooler JW. Window to the wandering mind: Pupillometry of spontaneous thought while reading. Q J Exp Psychol. 2013;66:2289–94. https://doi.org/10.1080/17470218.2013.858170.
71. Oyarzo P, Preiss D, Cosmelli D. Attentional and meta-cognitive processes underlying mind wandering episodes during continuous naturalistic reading are associated with specific changes in eye behavior. Psychophysiology. 2022;59:e13994. https://doi.org/10.1111/psyp.13994.
72. Embrey JR, Mason A, Newell BR. Too hard, too easy, or just right? The effects of context on effort and boredom aversion. Psychon Bull Rev. 2024;31:2801–10. https://doi.org/10.3758/s13423-024-02528-x.
73. Wu R, Ferguson AM, Inzlicht M. Do humans prefer cognitive effort over doing nothing? J Exp Psychol Gen. 2023;152:1069–79. https://doi.org/10.1037/xge0001320.
74. Müller VI, Cieslik EC, Ficco L, Tyralla S, Sepehry AA, Aziz-Safaie T, et al. Not all Stroop-type tasks are alike: Assessing the impact of stimulus material, task design, and cognitive demand via meta-analyses across neuroimaging studies. Neuropsychol Rev. 2024. https://doi.org/10.1007/s11065-024-09647-1. [cited 26 Aug 2025].
75. Kukkonen N, Braem S, Allaert J, Eayrs JO, Prutean N, Steendam ST, et al. The cost of regulating effort: reward and difficulty cues with longer prediction horizons have a stronger impact on performance. J Cogn. 2025;8:9. https://doi.org/10.5334/joc.415.
76. Chen A, Tang D, Chen X. Training reveals the sources of Stroop and flanker interference effects. Costantini M, editor. PLoS ONE. 2013;8: e76580. https://doi.org/10.1371/journal.pone.0076580
77. Macleod CM. Training on integrated versus separated Stroop tasks: The progression of interference and facilitation. Mem Cognit. 1998;26:201–11. https://doi.org/10.3758/BF03201133.
78. Luo Y, Zhang X, Jiang H, Chen X. The neural habituation to hedonic and eudaimonic rewards: Evidence from reward positivity. Psychophysiology. 2022;59:e13977. https://doi.org/10.1111/psyp.13977.
79. So WKY, Wong SWH, Mak JN, Chan RHM. An evaluation of mental workload with frontal EEG. PLoS ONE. 2017;12:e0174949. https://doi.org/10.1371/journal.pone.0174949.

Authors and Affiliations

Willem Jacob Louter[1] · S. Tabitha Steendam[1] · Nicoleta Prutean[1] · Joshua O. Eayrs[2] · Nanne Kukkonen[1] · Wim Notebaert[1] · Ruth M. Krebs[1] · Jan R. Wiersema[3] · C. Nico Boehler[1]

✉ Willem Jacob Louter
willem.louter@ugent.be

[1] Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, Ghent 9000, Belgium

[2] School of Psychology, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK

[3] Department of Experimental Clinical and Health Psychology, Ghent University, Henri Dunantlaan 2, Ghent 9000, Belgium