How Effortful is Boredom? Studying Self-Control Demands Through Pupillometry

Registered Report Stage 1 (under review)

Vanessa C. Radtke[1,2], Wanja Wolff[2,3], and Corinna S. Martarelli[1]

[1] Faculty of Psychology, UniDistance Suisse, Brig, Switzerland

[2] Faculty of Sports Science, University of Konstanz, Konstanz, Germany

[3] Faculty of Educational Psychology, University of Bern, Bern, Switzerland

Author Note

We have no known conflict of interest to disclose. Study materials, data, and scripts will be shared on Open Science Framework (OSF) at https://osf.io/8rgbd/?view_only=09a4d990eaa24a13b3df3dc8a0f89065.

Correspondence concerning this article should be addressed to Vanessa C. Radtke, UniDistance Suisse, Schinerstrasse 18, 3900 Brig, Switzerland.

Email: vanessa.radtke@fernuni.ch

Abstract

Self-control is essential for managing our actions, yet its exertion is perceived as effortful. Performing a task may require effort not only because of its inherent difficulty but also due to its potential for inducing boredom, as boredom has been shown to be self-control demanding by itself. So far, the extent of self-control demands during boredom and its temporal dynamics remain elusive. We will employ a multimethod approach to address this knowledge gap. Ninety-five participants will take part in an easy and hard version of the Stroop task. During both tasks, they will indicate several times their current sensation of task difficulty, boredom, boredom-related cognitive effort, difficulty-related cognitive effort, overall effort, and fatigue. We will test if pupil size, as a physiological indicator for cognitive effort, is predicted more accurately by overall cognitive effort (difficulty- and boredom-related) than by task-difficulty-related cognitive effort alone. This research will uncover the level of cognitive effort in experiencing boredom which is pertinent not only for self-control research but also to any research area dealing with boredom or the performance of repetitive tasks.

Keywords: boredom, cognitive effort, self-control, pupillometry, Stroop

Public Significance Statement

This study seeks to uncover the relationship between boredom and cognitive effort during task completion. To disentangle the dynamic interplay between boredom and cognitive effort during both easy and difficult tasks, we will employ a combination of objective measures (pupil size) and subjective measures (self-reports). Understanding how boredom influences cognitive effort required during the fulfillment of different self-control demanding tasks will shed light on a crucial aspect of human behavior, specifically how enduring boredom impacts cognitive effort during task completion. Our findings will carry important implications for society and well-being, as they offer insights into strategies for enhancing task performance and maintaining motivation in various areas of life.

Registered Report: How Effortful is Boredom? Studying Self-Control Demands Through Pupillometry

To effectively navigate everyday life, people have to control their behavior on a regular basis (Hofmann et al., 2012). For example, in order to make progress towards the goal of alleviating their back pain, a person has to control the impulse to linger on their couch and has to engage in repeated stretching and muscle strengthening exercises instead. This example illustrates that self-control is essential for reaching our goals. Indeed, self-control is understood to be a fundamental aspect of human functioning (Ainslie, 2021; Bieleke & Wolff, 2021b) helping us to overcome impulses that offer short-term gratification but are not in line with our long-term goals (Hofmann et al., 2009). Better self-control is linked to a wide range of positive outcomes, such as success, health, and happiness (Hofmann et al., 2014; Moffitt et al., 2011). As self-control is defined as the "efforts people exert to stimulate desirable responses and inhibit undesirable responses" (de Ridder et al., 2012, p. 77), self-control is by definition linked to effort. Notably, self-control is also influenced by motivational aspects which would not only facilitate the regulation of behavior towards ones goal, but also reduce the sensation of effort (Wennerhold & Friese, 2023) and consequently its costs.

Cognitive effort which can be defined as "intensity of mental […] work that organisms apply towards some outcome" (Inzlicht et al., 2018), can be measured objectively (e.g., with pupillometry) or experienced subjectively (Bijleveld, 2018; Robinson & Morsella, 2014) which we will refer to as perceived effort[1]. A large body of research shows that the objective and the perceived investment of effort tends to feel unpleasant and aversive (David et al., 2022; Kool & Botvinick, 2018; Wolff et al., 2021). On the other hand, effort is considered to feel unpleasant due to the costs of continuing with the ongoing activity

[1] Please note that from now on, we will use the term "effort" to refer to "cognitive effort" for the sake of readability.

(Kurzban, 2016). Thus, while effort is instrumental for effective self-control, it appears to carry a momentary cost, and the prolonged exertion of effort creates cumulative costs, such as fatigue or tiredness (Ainslie, 2021; Hopstaken et al., 2015; Kurzban, 2016; Kurzban et al., 2013; Westbrook et al., 2013). In addition to efforts' intrinsic costs (Kool et al., 2013), mobilizing effort towards one goal creates opportunity costs: When we direct our effort towards one activity (e.g., exercising to tackle backpain), we have to forego other, potentially more rewarding activities (e.g., going for a walk with friends). In this case, the aversive sensation of effort is considered to index the costs of continuing with the ongoing activity (Kurzban, 2016). This can create added self-control demands that make pursuing one's goal and continuing with a task even more self-control demanding and by extension more effortful (Kurzban et al., 2013). In light of effort's costs (intrinsic and/or opportunity costs), people are selective about when to invest effort in the service of self-control and the sensation of effort has been conceived as an index for the momentary costs of self-control (Bieleke et al., 2023).

Momentary Self-Control Costs and Performance in Self-Control Demanding Tasks

The idea of selectively investing effort is consistent with influential theories like the motivational intensity theory (Brehm & Self, 1989) which states that individuals do only decide to invest effort into a task of known difficulty if the required amount of effort can be justified by the individuals' importance of success. Other theories like ego depletion or mental fatigue theories posit that prior engagement in demanding tasks or the exertion of self-control can result in impaired performance on subsequent tasks (Baumeister et al., 1998; Kurzban et al., 2013; Marcora et al., 2009). To test this, a plethora of studies has investigated the effects exerting self-control has on accumulated self-control costs (e.g., exertion, fatigue, tiredness; Wolff et al., 2021) as well as on performance in subsequent self-control demanding tasks (see Dang, 2018; Giboin & Wolff, 2019; Hagger et al., 2010). This has most frequently been studied with a sequential task design (Dang et al., 2021; Englert & Bertrams, 2021; Solomon et al., 1980; Vassena et al., 2014). Here, participants are asked to complete either a more difficult or less difficult version of a task before performance on a secondary task is measured. While a large body of research has found that prior self-control exertion has detrimental effects on performance in subsequent self-control demanding tasks (Giboin & Wolff, 2019; Hagger et al., 2016; Hagger et al., 2010), high-powered pre-registered replication failures and evidence for a large amount of grey literature (Wolff et al., 2018), as well as for publication bias (Carter & McCullough, 2014; Holgado & Mesquida, 2023) has called the evidential value of this proposition into question (Holgado & Mesquida, 2023). One potential reason for these mixed findings might be that the hard and easy tasks that are typically used in sequential task design studies, differ not only with respect to their structural self-control demands caused by task difficulty, but on other properties too. One such difference could be that easy variants might not only be less challenging but also more boring (Wolff & Martarelli, 2020). Take, for instance, the Stroop task (Stroop, 1935), which is frequently used in self-control research (Dang, 2018; Wolff et al., 2018). Here, participants are asked to categorize a stream of color words (e.g., red, blue) according to the font color they are presented in. In the easy version, the font color and the semantic meaning of the word are the same (i.e., congruent). In contrast, font color and semantic meaning differ (i.e., incongruent) in the hard version[2]. Consistent with the hypothesis that hard and easy Stroop versions might differ not only with respect to their structural self-control demands caused by task difficulty, participants tend to find the easy version of the Stroop to be more boring (Bieleke et al., 2021; Hunte et al., 2022; Mangin et al., 2021). This difference matters because continuing to work on a boring task is understood to be self-control demanding in its own right (Bieleke & Wolff, 2021a).

How Could Boredom Act as a Self-Control Demand?

Boredom occurs when one's resources feel not adequately utilized (Wolff et al.,

[2] In more demanding versions, participants even have to respond to instantaneous rule changes in the categorization task, thereby adding task switching as a further self-control demand.

2024). Put simply, we get bored when we feel we are wasting our time. Consistent with this conceptualization, boredom tends to occur when tasks are too easy (or too hard; Westgate & Wilson, 2018), feel meaningless (van Tilburg & Igou, 2017), and/or when a person feels they have no agency (Danckert & Eastwood, 2020). Crucially, boredom is consequential because it is understood to signal that one should do something else (Bench & Lench, 2013; Elpidorou, 2014). With respect to the mechanisms by which boredom operates, research shows that boredom leads to a devaluation of the current activity, and increased reward sensitivity (Milyavskaya et al., 2019), thereby prompting exploration behavior (Agrawal et al., 2022; Bench & Lench, 2013; Danckert, 2019; Geana et al., 2016). By biasing behavior towards alternative activities, boredom increases the opportunity costs of sticking with what one is currently doing, thereby increasing the self-control demands (Bieleke & Wolff, 2021a; Bieleke et al., 2023) and the perceived effort (Eastwood et al., 2012) needed to keep engaged with the boring task. Take for example a lengthy zoom call. During such a meeting, many readers have probably experienced moments of boredom and have felt the urge to look at their phone and browse social media. Not giving in to this boredom-induced urge has likely made it harder to stay focused on what is discussed in the call. Consequently, self-control and effort are not only required for performing challenging activities, but also for underchallenging ones, making boredom a possible confound in self-control research but also in other research fields (Meier et al., 2023) that has been largely overlooked for so far (Wolff & Martarelli, 2020).

The Stroop task can provide an intuitive example of this. At the very beginning of the task, a period to become accustomed to it is required. During this phase, participants need to familiarize themselves with several key aspects, such as the colors presented, the location of the corresponding keys, and the task instructions. These additional challenges can result in a higher task difficulty, potentially leading to a greater demand for cognitive effort to successfully complete the task. In addition, the task would still feel novel, thereby likely being less boring for the participant. However, as the task progresses, participants become increasingly familiar with its demands. They encountered all the words and colors in different combinations, repeatedly pressed the keys, and have a clear understanding of the instructions. Consequently, the task can get easier, leading to a reduction in the self-control demands of the task. However, as the same stimuli appear and the same keys are pressed over a longer period, the task might also get more and more boring. Thus, while potentially less effort is required to deal with the structural difficulties of the task, more effort might be needed to ward of the boredom-induced urge to get the task over with (Bieleke et al., 2021). Notably, recent research indicates that both difficult and boring tasks can contribute to fatigue (Pickering et al., 2023). Fatigue is related to a decreased perception of value in exerting effort and less willingness to continue investing effort (Dora et al., 2022; Müller & Apps, 2019). Thus, maintaining focus on a task can become more effortful over time due to fatigue resulting from both task difficulty and boredom. While perceived task difficulty and task-induced boredom are likely to dynamically vary in hard and easy tasks, it is likely that these dynamics are not identical. More specifically, hard and easy tasks likely differ in their perceived difficulty and boringness from the start, and the temporal dynamics of perceived difficulty and boringness are likely to follow different trajectories too (Bieleke et al., 2021; Wolff & Martarelli, 2020). As a result, it is possible for both easy and hard tasks to exhibit various levels of effort throughout their execution. By exclusively assessing perceived effort in relation to task difficulty (as is traditionally done in self-control research), the additional effort requirements that can arise from task-induced boredom might be overlooked. Furthermore, assessing perceived effort only after task completion neglects the temporal fluctuations in effort demands during a self-control-demanding task.

To uncouple the temporal dynamics of effort (due to task difficulty and boredom) and boredom during different self-control tasks, triangulation of methods holds promise. First, self-reports can provide a valuable reading of people's state (Cooper-Martin, 1994; Johnson et al., 1995) and researchers have highlighted the need to track the dynamics of people's feelings with higher resolution (Mills & Christoff, 2018; Waugh et al., 2015). Complementing self-reports, pupil size can serve as a physiological indicator of effort with, generally speaking, greater pupil size indicating increased effort: Phasic (stimulus-evoked) changes in pupil diameter have been found during the completion of various tasks that require effort with a greater pupil size being related to a greater extend of effort (e.g., in inhibition, updating, working memory tasks; van der Wel & van Steenbergen, 2018). Moreover, pupil size measurements, as an objective physiological measure of effort, show consistency and correlate with the self-reported perception of effort (e.g., Koelewijn et al., 2015; Wals & Wichary, 2023; Zénon et al., 2014).

While phasic (stimulus-evoked) changes in pupil size tend to occur as reactions to the immediate demands of a task, tonic (baseline) pupil changes tend to reflect the state of the individual more generally (Cohen Hoffing et al., 2020). Considering changes of pupil size from a general psychophysiological perspective, the activation of the sympathetic pathway of the autonomic nervous system leads to the dilation of the pupil whereas the activation of the parasympathetic pathway induces its constriction (Kardon, 2005; Mathôt, 2018; McDougal & Gamlin, 2008). While our understanding of the exact neurological processes of pupil dilation during the exertion of effort is somewhat restricted (Mathôt, 2018), we do know that the locus coeruleus (LC), a brain area suggested to play an important role in behavioral regulation and, consequently, self-control (Aston-Jones & Cohen, 2005), affects changes in pupil size. Research indicates that when the LC is more active, the pupils tend to enlarge (Joshi et al., 2016). Connections from the orbitofrontal cortex (OFC) and the anterior cingulate cortex (ACC) to the LC were found (Aston-Jones et al., 2002; Rajkowski et al., 2000). As those brain regions are linked to the evaluation of rewards and costs, a responsiveness of the LC to ongoing cost-reward evaluations is suggested, consequently shaping the resulting behavior (Aston-Jones & Cohen, 2005). Given the association of the LC activation to cost-reward evaluations and self-control, the relation to pupil size, and the link between greater effort and larger pupils, these findings imply that the LC likely contributes to the dilation of pupils when self-control and effort (either difficulty- or boredom-related) are exerted during the engagement in cognitive tasks. Research investigating the dynamics of pupil dilation during the performance of cognitive tasks has demonstrated a gradual reduction in phasic (stimulus-evoked) pupil diameter during the execution of both high and low cognitive-demanding tasks (Hopstaken et al., 2015; Timme et al., 2022). This decline in pupil dilation aligns with the reduced demands and required effort of cognitive tasks over time. However, pupil size has not only been found to decrease over time but also to increase (Bijleveld, 2018; Timme et al., 2022). Bijleveld (2018) reported that both the perceived feeling of effort and physiological effort, as measured by phasic peak pupil dilation, increased over time in easy and hard trials of a cognitively demanding task. This increase aligns well with the idea of a rise in boredom over time and the proposal that staying engaged with a boring task might enhance the effort that has to be invested to complete the task (Wolff & Martarelli, 2020). Although Hopstaken et al. (2015) did not explicitly mention this observation, Figure 5 in their paper suggests a tendency of an increase in phasic peak pupil dilation in the first block of the task (which had a comparable duration to the whole task in Bijleveld, 2018) before the peak pupil dilation started to decrease over time. This could indicate a progressive increase in boredom, starting early in the task and intensifying over time and related to this, an increase in effort needed to keep a good performance while the individual's willingness to perform is still present. Over time other mechanisms might become more relevant in explaining total pupil dilation, such as the individual's decision to stop investing effort into the task all together, which could result from too high levels of boredom or fatigue. This idea aligns well with motivational theories like the motivational intensity theory (Brehm & Self, 1989) which suggests that effort is only mobilized to the degree that is justified by a tasks potential reward value. Within this framework, it is conceivable that fluctuations in boredom alter how much effort should be mobilized toward the task because boredom has been theorized to reduce the value people ascribe to a boredom-inducing activity (Wolff & Martarelli, 2020). However, it should be noted that Timme et al. (2022) found the averaged pupil dilation over a period of ten minutes to first decrease before showing a tendency to increase which is opposite to the findings of other two studies described above. Although pupil dilation was calculated differently in this study, these differences demonstrate the persistent uncertainty regarding how and why the perception of effort and physiological indicators of effort change during the performance of cognitive tasks over longer periods of time. Moreover, none of these studies included the assessment of boredom. Although it is highly speculative why these differences in results emerge between studies, it highlights the importance of employing a combination of self-report and pupillometry to allow for a deeper comprehension of the temporal dynamics of task difficulty and boredom dependent effort during self-controlled behavior.

The Present Study

In the present study we will investigate as how effortful task-induced boredom is subjectively perceived and physiologically measurable. To achieve this, we will access the temporal dynamics of task difficulty and boredom-related effort during the performance of an easy and hard version of a cognitive task (Stroop task). We integrate a promising research protocol that complements the subjective assessment of perceived effort with the implementation of pupillometry (tonic and phasic pupil size) as objective measure of effort, resulting in a high temporal resolution of physiological and subjective data. Subjective experiences will be assessed several times during the experiment with thought probes asking for the perception of boredom, task difficulty, task difficulty related effort, boredom related effort, overall effort, and fatigue. Thereby, we understand task difficulty related effort as the perception of effort that individuals feel like having to expend into the task in response to the experienced task difficulty, and boredom related effort as the perception of effort that individuals have to expend into the task in response to the experience of boredom. A secondary task (flanker task) will be included in our study to further access how the processing of the first task influences performance in the second task. However, note that our primary focus lies on examining the temporal dynamics of task difficulty and boredom-related effort in the easy and hard version of the first task (Stroop task). To the best of our knowledge, this is the first study directly assessing the perception of effort due to boredom and disentangling task difficulty and boredom related effort. Moreover, implementing a within-subjects design, this study contributes valuable insights into how boredom and effort interact across varying difficulty levels of cognitive tasks within the same participants. By uncoupling effort from its direction (i.e., effort to deal with an intrinsically hard task vs. effort to keep working on a boring task), this study addresses the research gap on how enduring boredom impacts effort during the performance of cognitive tasks. As a result, this study will enhance our understanding of boredom and its role in self-control. We predict boredom and boredom related effort to increase over time while we assume task difficulty and task difficulty related effort to decrease. Furthermore, we predict that pupil size (tonic and phasic) will be predicted more accurately by overall cognitive effort (difficulty- and boredom-related) than by effort related solely to task difficulty.

Proposed Method

Proposed Sample Characteristics

To ensure a significant difference in the difficulty level and the potential for boredom induction by our tasks, we conducted an online study that included four Stroop task versions (easy and hard versions of color Stroop and numerical Stroop, see Appendix A). Based on the results, we concluded that the color Stroop task was the most suitable variant for our study. To determine the required sample size for the present study, we conducted a G*Power Analysis (Faul et al., 2007). Given that the effect size observed in the calculated t-tests was larger for task difficulty (d = 0.92) than for boredom (d = 0.34), our power analysis focuses on the effect size of the difference in boredom between the easy and the hard color Stroop. Thus, we are adopting a conservative approach in calculating the required sample size. The power analysis for a one-tailed paired t-test was calculated based on this effect size. The analysis indicated that 95 participants are necessary to detect a difference in boredom between the tasks with a power of 95% at an alpha level of 0.05. Replacement participants will be recruited to achieve the calculated sample size, if participants drop out, z-transformed error rate in the Stroop task is more than three standard deviations away from the mean of all participants or errors in the recording of oculomotor data occur (i.e., participants for whom the calibration procedure did not work correctly, indicated by an average error in measuring horizontal and vertical eye positions exceeding an angular accuracy of 0.8°).

Participants (age range of 16 to 45 years, all genders) without self-reported color blindness will be recruited from the general population and from our institute.

Participants will give written-informed consent and be free to end the experiment at any time. They will receive 100 CHF or course credits as compensation if they are students at UniDistance Suisse. The local Ethics Committee approved the study, which will be conducted according to the principles of the Declaration of Helsinki.

Proposed Materials and Design

In this within-subjects design study participants will take part in two sessions separated by at least one day, each consisting of a Stroop task and a subsequent flanker task. In one session, participants will complete the Stroop task in its easy congruent-only version (easy Stroop), while in the other session, they will complete the hard version with modified instructions (hard Stroop). Pupil size and thought probe data will be recorded during the Stroop task only, while behavioral data (response time, accuracy) will be recorded during both tasks.

Tasks

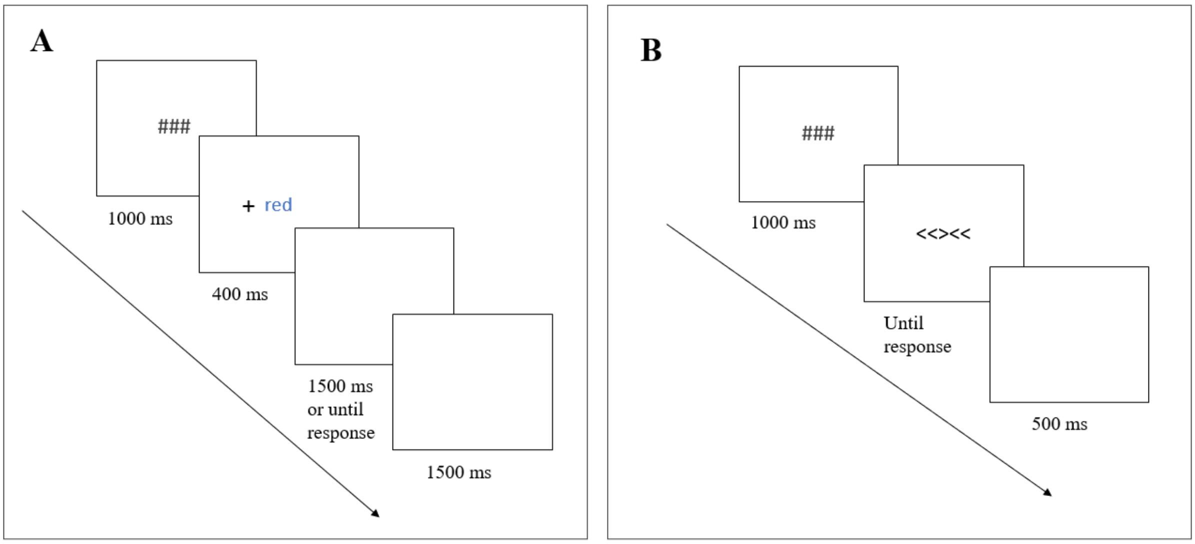
Stroop task. As primary task participants will perform a color Stroop task (original

Stroop task version first implemented by Stroop, (1935) comprising 34 practice trials and 360 experimental trials. Each trial consists of the presentation of fixation stimuli (###, including pixels of all four colors of the color words to ensure a more similar luminance, displayed for 1000 ms on screen), a color word (green, red, blue, yellow) that will be presented for 400 ms either in green (RGB: 0,85,0,255), blue (RGB: 0,0,85,255), red (RGB: 85,0,0,255) or yellow (RGB: 85,85,0,255), a blank screen (displayed until reaction or for 1500 ms), and an intertrial interval (displayed for 1500 ms). Participants are instructed to press one of four keys on the keyboard (“1” for green, “2” for blue, “9” for red, “0” for yellow) to indicate their response. The background of the screen will be displayed in grey (150,150,150,255). Self-control demands and the induction of boredom will be manipulated using an easy version of the Stroop task and a hard version with modified instructions. While in the easy condition participants will get the instruction to indicate the color of the presented color word, the task in the hard condition switches according to another stimulus presented on screen which either demands to indicate the color if a “+” is displayed (in 80 % of trials) or the word if a “x” is displayed (in 20% of trials). The instructions will inform participants that in the easy condition all trials will be congruent (color and color word matching) whereas in the hard condition all trials will be incongruent (color and word not matching each other, see Figure 1 A). Response times and accuracy will serve as performance indicators.

Flanker task. As secondary task, participants will perform a flanker task (original task version first implemented by Eriksen & Eriksen, 1974), comprising 24 practice trials and 140 experimental trials. Each trial consists of fixation stimuli (###, displayed for 1000 ms on screen), five arrows presented on the middle of the screen, and an intertrial interval (displayed for 500 ms). The arrows will remain on screen until participants press the keys “1” for right or “0” for left, indicating the direction of which the central arrow is pointing. The other four arrows will either be congruent and thus pointing to the same direction as the central arrow (50 % of all trials) or incongruent and thus pointing to the contrary direction (see Figure 1 B). The task performance (accuracy, response times) will be recorded. The flanker task will be implemented to investigate potential effects of the primary tasks on secondary task performance.

Figure 1

Example of an Incongruent Trial of the Stroop Task (A) and of the Flanker Task (B)



Note. Naming the font color ("+" presented next to the word) will be the task in 80 % of the trials of the hard Stroop task (A). In 20 % of the trials a "x" will be presented and thus, the task will be to indicate the word. All trials of the hard Stroop task will be incongruent whereas all trials of the easy Stroop task will be congruent. Flanker task trials (B) will be either incongruent or congruent (50 % of all trials).

Thought Probes

During the Stroop task participants will be prompted eleven times to report how bored they are ("How much boredom do you feel?"), as how difficult they are experiencing the task ("How difficult is the task?"), how much effort they are investing due to task difficulty ("Due to the difficulty of the task, how much effort do you have to invest into the task?"), due to boredom ("Due to boredom, how much effort do you have to invest into the task?") and overall ("Overall, how much effort do you have to invest into the task?"), and how much fatigue they are experiencing ("How much fatigue do you feel?"). They will indicate their answer by pressing a key between 1 (not at all) and 9 (very much) on their keyboard. The probes will be displayed randomly after blocks of 30 to 35 trials. The order of the probes will be randomized for each presentation.

Pupil Size

Tonic (baseline) pupil size and phasic (stimulus-evoked) pupil size during the tasks will be recorded at a sampling rate of 1000 Hz, a spatial resolution of 0.01°, and a gaze position accuracy of 0.05° using an EyeLink 1000 Plus eye-tracker. A chinrest will be placed at a distance of 45 cm to the monitor (1920 x 1080 pixels) in order to prevent head movements. At the start of the experiment the eye-tracker will be calibrated and validated using a 9-point grid. Participants where the average horizontal and vertical error exceeds an angular accuracy of 0.8° will be excluded. The testing room will be cut off from external light sources, and it will not be illuminated throughout all testing sessions.

The data will be extracted using EyeLink DataViewer version 3.4.1. Missing data points due to off-screen fixations or eye blinks will be removed before analyzing pupil diameter. Pupil size will be reported for the right eye and reported in pixels.

Tonic (baseline) pupil size will be averaged separately for each last 500 ms of the fixation stimuli (###) presentation. Phasic (stimulus-evoked) peak pupil dilation for each trial will be obtained by correcting peak pupil dilation during the presentation of the blank screen (as this time period typically corresponds to the occurring of peak pupillary response in the Stroop task, see for example Hershman & Henik, 2020) for baseline pupil size (averaged pupil size during a period of 500 ms before stimulus onset). Both tonic (baseline) pupil size and phasic (stimulus-evoked) pupil size will be averaged in blocks with each block consisting of the last five trials before each probe. A larger phasic (stimulus-evoked) peak pupil dilation will serve as indicator for higher stimulus-induced cognitive effort. Conversely, a lager tonic (baseline) pupil size will serve as indicator for higher cognitive effort, irrespective of the effort induced by the task stimulus. Analyzing tonic (baseline) pupil size is essential for our research question as effort related to boredom persists throughout the whole trial, i.e., during the presentation of the fixation stimuli and during the presentation of the task stimulus. Relying solely on tonic pupil size as a means of baseline correcting stimulus-evoked pupil size and not analyzing it independently would neglect the possible effect of the endurance of boredom on cognitive effort and could lead to an overlook of this effect.

Supplementary Measures

In addition to the measures needed to investigate the core research question of this paper we will implement supplementary measures to provide a more comprehensive description of the sample and to enhance our understanding of experiencing boredom and related characteristics. For this purpose, participants will answer three questionnaires before they take part in the experiment, the Short Boredom Proneness Scale (SBPS; Struk et al., 2017); German version by (Martarelli et al., 2021), the Beck Depression Inventory II (BDI-II, Beck et al., 1996); German version by (Kühner et al., 2007), and the Brief Self-control Scale (BSCS; Tangney et al., 2018); German version by (Bertrams & Dickhäuser, 2009). They will also indicate whether they have been diagnosed with ADHD, if they have consumed any stimulants on the day of testing (caffeine, nicotine, amphetamine, others), and whether they feel sleep deprived (“Do you feel sleep deprived*?”* Answered with “yes” or “no”). We will include responses in the sample description. After the completion of the experiment, participants will get offered to take with them some sweets. The amount chosen by each participant will be measured.

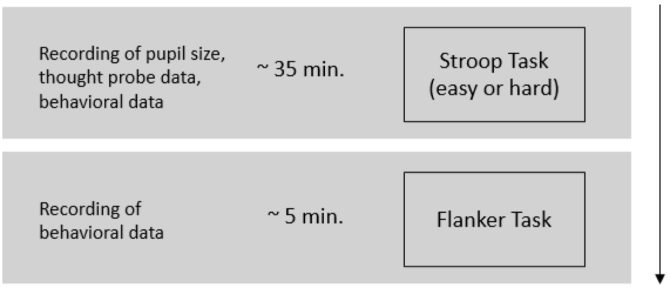
Proposed Procedure

Before starting the first experimental session in the laboratory, participants will answer the Short Boredom Proneness Scale (Struk et al., 2017), the Brief Self-Control Questionnaire (Tangney et al., 2018) and the BDI-II (Beck et al., 1996) as well as the other supplementary measures online. Afterwards they will receive an invitation to the laboratory together with the request not to drink any coffee at least for two hours before the start of the experiment. In each study part, online and offline, participants will be informed about the experimental procedure and asked to consent into the study conditions. The experimental part of the study consists of two sessions. Every participant takes part in both sessions (within-subjects design). Before each session participants will report if they have consumed any stimulants on the day of testing. The procedure of both experimental sessions (see Figure 2) will be similar with the difference that the first sequential paradigm task (Stroop task) will be an easy version in one session and a hard in the other one. Both sessions will involve the flanker task as secondary task. The order of the sessions will be counterbalanced. Each session will take approximately 40 to 50 minutes. During the task participants will answer to probes indicating their perception of boredom, task difficulty, effort they have to invest due to boredom and due to task difficulty, overall effort and fatigue. There will be a break of at least one day in between the sessions.

In the experimental sessions participants place their head on a chinrest and look at the screen (at 45 cm distance to head) where the tasks will be presented while oculomotor data (pupil size) and behavioral data (task performance, probe answers) will be recorded. Participants will be instructed to stay as still as possible, to direct their gaze to the center of the screen and to answer the questions (probes) presented on screen by pressing the corresponding number on the keyboard without removing their head from the chinrest.

Figure 2

Proposed Procedure for Each Session



Note. Behavioral data (reaction time, accuracy) will be recorded during the Stroop task and during the Flanker task. Pupil and thought probe data will be recorded during the Stroop task only.

Transparency and Openness

We share the dataset of our online study as well as the R code to reproduce our analyses on Open Science Framework (OSF) available at https://osf.io/8rgbd/?view_only=09a4d990eaa24a13b3df3dc8a0f89065. We will further incorporate the research materials, the full non-aggregated dataset as well as the analysis scripts for the present study after completing data collection. Data will be analyzed using R, version 4.3.1 (R Core Team, 2023).

Proposed Analysis and Predicted Outcomes

Manipulation Check

We will confirm the effectiveness of our experimental manipulations by assessing their impact on participants' subjective perception and performance. This manipulation check will allow us to ensure that the intended differences between the conditions concerning the manipulated difficulty level are indeed present.

Task Difficulty

We hypothesize that the hard condition will exhibit significantly greater overall perceived task difficulty compared to the easy condition. To test this hypothesis, we will perform a paired t-test comparing the overall task difficulty (dependent variable) in the easy and hard Stroop (difficulty level as independent variable). The overall perceived task difficulty will be calculated as the mean of the perceived difficulty ratings for the answered probes.

Boredom

In the context of boredom levels, we expect the easy condition to induce significantly greater overall boredom compared to the hard condition. To evaluate this hypothesis, we will conduct a paired t-test comparing overall boredom (dependent variable) in the easy and hard condition (difficulty level as independent variable). The overall boredom will be obtained by calculating the mean of the boredom ratings for the answered probes.

Performance

Performance differences between the two conditions will be assessed by examining error rates and reaction times. We expect that participants will demonstrate a greater error rate and longer reaction time in the hard condition compared to the easy condition. We will conduct two paired t-tests, one comparing the error rate and one comparing the reaction time between the easy and hard condition, implementing the error rate and the reaction time as dependent variables each in one analysis and the difficulty level (easy, hard) as independent variable.

Thought Probes

We expect changes over time in participants' self-reports for both the easy and in the hard variant of the task. Changes over time concerning those variables will be analyzed conducting six Linear Mixed Models including the perception of boredom, boredom related effort, task difficulty and task difficulty related effort, overall effort and fatigue as outcome variables, task (easy, hard) and time (probe one to eleven) as fixed effects accounting for differences among participants by including random intercepts and different effects of time on the outcome variables by including random slopes for time. We expect a main effect of time. More precisely, we expect an increase in boredom, boredom related effort, fatigue, and overall effort and a decrease of task difficulty and task difficulty related effort over time for both Stroop versions (easy and hard). Further, we predict a main effect of difficulty level (easy, hard). We anticipate greater levels of boredom and greater boredom related effort, but less task difficulty and task difficulty related effort for the easy Stroop version. We expect an interaction for perceived task difficulty and task difficulty related effort with a greater decrease over time in the hard Stroop. We further expect an interaction for perceived overall effort with a greater increase in overall effort in the easy Stroop version. Six plots showing the time (probe one to eleven) on the x-axis, the scores on the y-axis, and two different lines for the two tasks, will illustrate how the perception of boredom, boredom related effort, task difficulty related effort, task difficulty, fatigue and effort change over time.

Effort and Pupil Size

To test our prediction that pupil size will show a stronger association with perceived boredom related effort and task difficulty related effort together than with task difficulty related effort only, we will analyze tonic (baseline) pupil size and phasic (stimulus-evoked) pupil size separately. We hypothesize that tonic (baseline) and phasic (stimulus-evoked) pupil sizes would be more accurately predicted when considering both perceived effort due to boredom and effort due to task difficulty, as opposed to considering perceived effort due to task difficulty alone.

To investigate this hypothesis, we will employ eight Linear Mixed Models (LMM) to assess the influence of these factors on tonic pupil size and phasic pupil size separately. Tonic pupil size for each probe will be determined by calculating the average tonic (baseline) pupil size (pupil size during the last 500 ms of the presentation of the fixation stimuli) across the last five trials preceding the probe presentation. Phasic pupil size for each probe will be determined by calculating the peak stimulus-evoked pupil size. After baseline correcting phasic pupil size it will be averaged across the last five trials preceding the probe presentation. We will construct four different models for each outcome variable (tonic pupil size, phasic pupil size). Our first model will predict pupil size with the task’s difficulty level (easy, hard), time (thought probe one to eleven), and perceived effort due to task difficulty as fixed predictors, accounting for general differences between participants in tonic and phasic pupil size by including random intercepts and different effects of time on pupil size among participants by including random slopes for the variable time. Our second model will differ from model one by implementing effort due to boredom as fixed predictor instead of effort due to task difficulty. Our third model will integrate effort due to task difficulty and effort due to boredom as fixed predictors, while remaining with the other predictors. Our fourth model will expand upon model three by including additional covariates, namely participants’ score of the Short Boredom Proneness Scale (SBPS), the Beck Depression Inventory II (BDI-

II), and the Brief Self-Control Scale (BSCS). To evaluate the relative performance of these models in predicting pupil size, the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and WAIC (Widely Applicable Information Criterion) of the models will be reported. Our decision regarding the optimal balance between model fit and complexity will be based on the BIC.

As an illustration three plots will be implemented for each outcome variable (tonic pupil size, phasic pupil size) showing time and the scores of perceived boredom related effort, task difficulty related effort on the x-axis, the outcome variable on the y-axis and two different lines indicating the manipulated difficulty level (easy, hard).

Task Performance in Secondary Task

For assessing task performance in the secondary task, we will focus on the error rate and reaction time separately. We aim to analyze if participants' performance in the secondary task (flanker) is influenced by the difficulty level of the preceding task (Stroop task; easy, hard). We further intend to test whether perceived task difficulty related effort exerted in the Stroop task predicts task performance in the flanker task, and whether boredom and task difficulty related effort together predict task performance more accurately.

Difficulty Manipulation's Influence on Performance

To access whether the level of difficulty in the first task (Stroop task; easy, hard) influences participants' performance in the secondary task (flanker), we will conduct two 2 (previous task type: easy, hard) x 2 (flanker type: congruent, incongruent) ANOVAs. One ANOVA will analyze participants' error rate as dependent variable, while the other will focus on the reaction time. Results will be visualized with two bar charts (one showing the error rate, the other one showing the reaction time) providing the difficulty level on the x-axis, the error rate or the reaction time on the y-axis, and four differing bars (one for the easy Stroop version and congruent flanker trials, one for the easy Stroop version and incongruent flanker trials, one for the hard Stroop version and congruent flanker trials and one for the easy Stroop version and incongruent flanker trials).

Perceived Effort's Influence on Performance

We will explore the extent to which the perceived effort, resulting from perceived task difficulty in the Stroop task, can account for variations in error rates and reaction time during the flanker task. Additionally, we will investigate whether perceived effort due to task difficulty in the Stroop task predicts performance (error rate, reaction time) in the flanker task, and whether considering perceived boredom related and task difficulty related effort in the Stroop task together, provides a more accurate prediction of the error rate and reaction time in the flanker task than task difficulty related effort alone. To investigate this, we will employ several Linear Mixed Models (LMMs), using either the error rates or the reaction time in the flanker task as the outcome variables. Perceived effort due to boredom and effort due to task difficulty will be obtained by calculating the mean of this variables across the probes. Three distinct models will be tested for each performance variable (error rate, reaction time). The first model will assess the impact of the difficulty level (easy, hard; fixed effect) and perceived effort due to task difficulty (fixed effect) in the Stroop task as well as the type of flanker trial (congruent, incongruent) on the reaction time in the flanker task while accounting for random effects among participants with random intercepts and random slopes. The second model will differ from the first by incorporating boredom related effort as a fixed effect instead of task difficulty related effort. The third model will include both perceived task difficulty and boredom related effort as fixed effects. AIC, BIC, and WAIC will be reported. The model with the lowest BIC will be considered as the most optimal one.

Two plots for each performance variable (error rate, reaction time) will illustrate the relationship between performance in the flanker task on the y-axis and perceived effort scores (effort due to boredom and effort due to task difficulty) on the x-axis. Two distinct lines will represent the difficulty levels of the task (easy, hard).

Data Summary

Descriptive Statistics

To comprehensively understand our variables, we will conduct an analysis of descriptive statistics and present them in a table. We will calculate the means and standard deviations for various measures across eleven different time points. These statistics will be computed separately for our two difficulty levels (easy, hard). Tonic pupil size, phasic peak pupil size, ratings of boredom, fatigue, task difficulty, boredom related effort, task difficulty related effort, overall effort as well as the reaction time and error rates during the Stroop task and the flanker task will be reported in eleven blocks corresponding to the thought probes. Reaction times and error rates in the flanker task will be reported separately for congruent and incongruent trials, as well as combined. Phasic peak pupil dilation during the Stroop task will be reported as an average across all trials for both tasks. Additionally, it will be reported separately for the hard task distinguishing between trials involving the color naming task and trials involving the word naming task. Each block includes the five preceding trials before the presentation of the probe.

Correlation Analysis

To elucidate the relationships between the key variables in our study, we will conduct correlation analyses and present them in a correlation table. Boredom, perceived task difficulty, effort due to boredom, effort due to difficulty, overall effort, and fatigue which will be assessed with the thought probes will be included in the analysis. Concerning the tasks, the error rate and the reaction time in the easy version and in the hard version of the Stroop task as well as the error rate and reaction time in the flanker task will be integrated. Regarding the pupil size, the average tonic pupil size during the easy and during the hard version of the Stroop as well as the average phasic pupil size during the easy and hard version of the Stroop will be encompassed.

Explorative Analysis

We will explore associations between our key variables and the quantity of sweets participants took with them.

Pre-Test

To test the feasibility of our study design we tested two participants and recorded pupil size (tonic and phasic pupil size) and the answers to the thought probes. See Appendix B for the corresponding descriptive data.

Interpretation Given Different Outcomes

Concerning our primary research question, several result patterns are possible. Should our results, as hypothesized, show that pupil size as physiological indicator of effort is predicted more accurately when perceived effort due to boredom is considered as predictor, we would characterize the endurance of boredom as effortful. This would emphasize the importance of recognizing boredom as a confounding variable within self-control research and, more general, within the broader scope of study designs. Conversely, if pupil size is not predicted by perceived effort due to boredom, it could be argued that being bored during task performance might not trigger changes in this physiological correlate of effort. However, this outcome could even indicate a connection between boredom and low arousal levels, a correlation that remains ambiguous. In this case, the perceived effort elicited by boredom might be less influential on the pupillary activity than the possible decrease in arousal. But even when perceived effort induced by boredom does not manifest physiologically, the evaluation of the thought probes could still indicate a psychological perception of effort as consequence of the sensation of boredom. In this case, boredom might not serve as a confounding variable when cognitive effort is measured physiologically, but it would impact psychological assessments of cognitive effort. In case our results demonstrate that boredom does induce neither physiological nor the psychological perception of effort, boredom might not be a confounding factor in experiments that induce boredom during task performance.

However, further investigations would be warranted to explore whether boredom could still evoke cognitive effort in cases involving entirely undemanding tasks or no tasks at all.

Limitations

It is important to acknowledge limitations regarding the measurement of certain constructs in this study. Specifically, the validity of items used to assess perceptions of task difficulty, boredom and related effort need further validation.

Moreover, a constraint lies in the task switching component of the hard Stroop version, which may result in varying levels of required cognitive control throughout the task. Similar designs incorporating switching components (e.g., changes in the proportion of congruent and incongruent trials) have been demonstrated to impact Stroop interference (Rothermund et al., 2022), potentially affecting not only performance, but also pupillary reaction to the trial’s demands as well as other psychological states measured in this study.

Future research will need to test whether the results generalize to other tasks, including Stroop tasks of intermediate difficulty without task-switching components and with various levels of congruency.

References

Ainslie, G. (2021). Willpower with and without effort. *Behavioral and Brain Sciences*, *44*, e30. https://doi.org/10.1017/S0140525X20000357

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*, 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

Aston-Jones, G., Rajkowski, J., Lu, W., Zhu, Y., Cohen, J. D., & Morecraft, R. J. (2002). Prominent projections from the orbital prefrontal cortex to the locus coeruleus in monkey. *Society for Neuroscience Abstract*, *28*, 86–89.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, *74*(5), 1252–1265. https://doi.org/10.1037/0022-3514.74.5.1252

Beck, A. T., Steer, R. A., & Brown, G. (1996). *Depression Inventory-II (BDI-II)*. American Psychological Association (APA) PsycTests. https://doi.org/10.1037/t00742-000

Bertrams, A., & Dickhäuser, O. (2009). Messung dispositioneller Selbstkontroll-Kapazität. *Diagnostica*, *55*(1), 2–10. https://doi.org/10.1026/0012-1924.55.1.2

Bieleke, M., Barton, L., & Wolff, W. (2021). Trajectories of boredom in self-control demanding tasks. *Cognition & Emotion*, *35*(5), 1018–1028. https://doi.org/10.1080/02699931.2021.1901656

Bieleke, M., & Wolff, W. (2021). The self-regulation of human performance. *Performance Enhancement & Health*, *9*(2).

Bijleveld, E. (2018). The feeling of effort during mental activity. *Consciousness and Cognition*, *63*, 218–227. https://doi.org/10.1016/j.concog.2018.05.013

Brehm, J. W., & Self, E. A. (1989). The intensity of motivation. *Annual Review of Psychology*, *40*(1), 109–131. https://doi.org/10.1146/annurev.ps.40.020189.000545

Cohen Hoffing, R. A., Lauharatanahirun, N., Forster, D. E., Garcia, J. O., Vettel, J. M., & Thurman, S. M. (2020). Dissociable mappings of tonic and phasic pupillary features onto cognitive processes involved in mental arithmetic. *PLoS ONE*, *15*(3), e0230517. https://doi.org/10.1371/journal.pone.0230517

Cooper-Martin, E. (1994). Measures of cognitive effort. *Marketing Letters*, *5*(1), 43–56. https://doi.org/10.1007/BF00993957

Danckert, J., & Eastwood, J. D. (2020). *Out of My Skull: The Psychology of Boredom*. Harvard University Press.

David, L., Vassena, E., & Bijleveld, E. (2022). *The aversiveness of mental effort: A meta-analytic review of the association between mental effort and negative affect*. https://doi.org/10.31234/osf.io/m8zf6

de Ridder, D. T. D., Lensvelt-Mulders, G., Finkenauer, C., Stok, F. M., & Baumeister, R. F. (2012). Taking stock of self-control: A meta-analysis of how trait self-control relates to a wide range of behaviors. *Personality and Social Psychology Review*, *16*(1), 76–99. https://doi.org/10.1177/1088868311418749

Dora, J., van Hooff, M. L. M., Geurts, S. A. E., Kompier, M. A. J., & Bijleveld, E. (2022). The effect of opportunity costs on mental fatigue in labor/leisure trade-offs. *Journal of Experimental Psychology: General*, *151*(3), 695–710. https://doi.org/10.1037/xge0001095

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Hershman, R., & Henik, A. (2020). Pupillometric contributions to deciphering Stroop conflicts. *Memory & Cognition*, *48*(2), 325–333. https://doi.org/10.3758/s13421-019-00971-z

Hofmann, W., Friese, M., & Strack, F. (2009). Impulse and Self-Control From a Dual-Systems Perspective. *Perspectives on Psychological Science*, *4*(2), 162–176. https://doi.org/10.1111/j.1745-6924.2009.01116.x

Hofmann, W., Luhmann, M., Fisher, R. R., Vohs, K. D., & Baumeister, R. F. (2014). Yes, but are they happy? Effects of trait self-control on affective well-being and life satisfaction. *Journal of Personality*, *82*(4), 265–277. https://doi.org/10.1111/jopy.12050

Hopstaken, J. F., van der Linden, D., Bakker, A. B., & Kompier, M. A. J. (2015). The window of my eyes: Task disengagement and mental fatigue covary with pupil dynamics. *Biological Psychology*, *110*, 100–106. https://doi.org/10.1016/j.biopsycho.2015.06.013

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. https://doi.org/10.1016/j.tics.2018.01.007

Johnson, N. E., Saccuzzo, D. P., & Larson, G. E. (1995). Self-Reported Effort versus Actual Performance in Information Processing Paradigms. *The Journal of General Psychology*, *122*(2), 195–210. https://doi.org/10.1080/00221309.1995.9921232

Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex. *Neuron*, *89*(1), 221–234. https://doi.org/10.1016/j.neuron.2015.11.028

Kardon, R. H. Anatomy and physiology of the autonomic nervous system. In Miller, N. R., Newman, N. J., Biousse, V., & Kerrison, J. B. (Ed.), *Wash and Hoyt's Clinical Neuro-Ophthalmology* (6th ed., 649-714).

Koelewijn, T., Kluiver, H. de, Shinn-Cunningham, B. G., Zekveld, A. A., & Kramer, S. E. (2015). The pupil response reveals increased listening effort when it is difficult to focus attention. *Hearing Research*, *323*, 81–90. https://doi.org/10.1016/j.heares.2015.02.004

Kool, W., & Botvinick, M. (2018). Mental labour. *Nature Human Behaviour*, *2*(12), 899–908. https://doi.org/10.1038/s41562-018-0401-9

Kühner, C., Bürger, C., Keller, F., & Hautzinger, M. (2007). Reliabilität und Validität des revidierten Beck-Depressionsinventars (BDI-II). Befunde aus deutschsprachigen Stichproben [Reliability and validity of the Revised Beck Depression Inventory (BDI-II). Results from German samples]. *Der Nervenarzt*, *78*(6), 651–656. https://doi.org/10.1007/s00115-006-2098-7

Kurzban, R. (2016). The sense of effort. *Current Opinion in Psychology*, *7*, 67–70. https://doi.org/10.1016/j.copsyc.2015.08.003

Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *The Behavioral and Brain Sciences*, *36*(6), 661–679. https://doi.org/10.1017/S0140525X12003196

Marcora, S. M., Staiano, W., & Manning, V. (2009). Mental fatigue impairs physical performance in humans. *Journal of Applied Physiology*, *106*(3), 857–864. https://doi.org/10.1152/japplphysiol.91324.2008

Martarelli, C. S., Bertrams, A., & Wolff, W. (2021). *Short Boredom Proneness Scale - German Version (SBPS) [Database record].* APA PsycTests. https://doi.org/10.1037/t82956-000

Mathôt, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of Cognition*, *1*(1), 1–23. https://doi.org/10.5334/joc.18

McDougal, D. H., & Gamlin, P. Pupillary Control Pathways. In *The Senses: A Comprehensive Reference* (Vol. 1, pp. 521–536). https://doi.org/10.1016/B978-012370880-9.00282-6

Mills, C., & Christoff, K. (2018). Finding Consistency in Boredom by Appreciating its

Instability. *Trends in Cognitive Sciences*, *22*(9), 744–747. https://doi.org/10.1016/j.tics.2018.07.001

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thomson, W. M., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(7), 2693–2698. https://doi.org/10.1073/pnas.1010076108

Müller, T., & Apps, M. A. J. (2019). Motivational fatigue: A neurocognitive framework for the impact of effortful exertion on subsequent motivation. *Neuropsychologia*, *123*, 141–151. https://doi.org/10.1016/j.neuropsychologia.2018.04.030

Pickering, T., Wright, B., Schücker, L., & MacMahon, C. (2023). Active or passive? Investigating different types of cognitive fatigue. *Canadian Journal of Experimental Psychology = Revue Canadienne De Psychologie Experimentale*, *78*(1), 50–65. https://doi.org/10.1037/cep0000312

Rajkowski, J., Lu, W., Zhu, Y., Cohen, J., & Aston-Jones, G [G.] (2000). Prominent projections from the anterior cingulate cortex to the locus coeruleus in Rhesus monkey. *Society for Neuroscience Abstract*, *26*, 838–15.

Robinson, M. M., & Morsella, E. (2014). The subjective effort of everyday mental tasks: Attending, assessing, and choosing. *Motivation and Emotion*, *38*(6), 832–843. https://doi.org/10.1007/s11031-014-9441-2

Rothermund, K., Gollnick, N., & Giesen, C. G. (2022). Accounting for Proportion Congruency Effects in the Stroop Task in a Confounded Setup: Retrieval of Stimulus-Response Episodes Explains it All. *Journal of Cognition*, *5*(1), 39. https://doi.org/10.5334/joc.232

Struk, A. A., Carriere, J. S. A., Cheyne, J. A., & Danckert, J. (2017). A Short Boredom Proneness Scale. *Assessment*, *24*(3), 346–359. https://doi.org/10.1177/1073191115609996

Tangney, J. P., Boone, A. L., & Baumeister, R. F. (2018). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. In R. F. Baumeister (Ed.), *World library of psychologists. Self-regulation and self-control: Selected works of Roy Baumeister* (pp. 173–212). Routledge. https://doi.org/10.4324/9781315175775-5

Timme, S., Wolff, W., Englert, C., & Brand, R. (2022). Tracking Self-Control - Task Performance and Pupil Size in a Go/No-Go Inhibition Task. *Frontiers in Psychology*, *13*, 915016. https://doi.org/10.3389/fpsyg.2022.915016

Tyler, J. M., & Burns, K. C. (2008). After Depletion: The Replenishment of the Self's Regulatory Resources, *7*(3), 305–321. https://doi.org/10.1080/15298860701799997

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), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y

van Tilburg, W. A. P., & Igou, E. R. (2017). Boredom begs to differ: Differentiation from other negative emotions. *Emotion*, *17*(2), 309–322. https://doi.org/10.1037/emo0000233

Wals, S. F., & Wichary, S. (2023). Under Pressure: Cognitive Effort During Website-Based Task Performance is Associated with Pupil Size, Visual Exploration, and Users' Intention to Recommend. *International Journal of Human-Computer Interaction*, *39*(18), 3504–3515. https://doi.org/10.1080/10447318.2022.2098576

Waugh, C. E., Shing, E. Z., & Avery, B. M. (2015). Temporal Dynamics of Emotional Processing in the Brain. *Emotion Review*, *7*(4), 323–329. https://doi.org/10.1177/1754073915590615

Wennerhold, L., & Friese, M. (2023). Challenges in the conceptualization of trait self-control as a psychological construct. *Social and Personality Psychology Compass*, *17*(3), Article e12726. https://doi.org/10.1111/spc3.12726

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

Westgate, E. C., & Wilson, T. D. (2018). Boring Thoughts and Bored Minds: The MAC Model of Boredom and Cognitive Engagement. *Psychological Review*, *125*(5), 689–713. https://doi.org/10.31234/osf.io/9j86p

Wolff, W., & Martarelli, C. S. (2020). Bored Into Depletion? Toward a Tentative Integration of Perceived Self-Control Exertion and Boredom as Guiding Signals for Goal-Directed Behavior. *Perspectives on Psychological Science*, *15*(5), 1272–1283. https://doi.org/10.1177/1745691620921394

Wolff, W., Radtke, V. C., & Martarelli, C. S. (2024). Same same but different. In M. Bieleke, W. Wolff, & C. Martarelli (Eds.), *The Routledge International Handbook of Boredom* (pp. 5–29). Routledge. https://doi.org/10.4324/9781003271536-3

Wolff, W., Sieber, V., Bieleke, M., & Englert, C. (2021). Task duration and task order do not matter: No effect on self-control performance. *Psychological Research*, *85*(1), 397–407. https://doi.org/10.1007/s00426-019-01230-1

Zénon, A., Sidibé, M., & Olivier, E. (2014). Pupil size variations correlate with physical effort perception. *Frontiers in Behavioral Neuroscience*, *8*, 286. https://doi.org/10.3389/fnbeh.2014.00286

Appendix A

Online Study

To ensure a significant divergence in the difficulty level and the potential for boredom induction by our tasks, we conducted an online study consisting of four Stroop task versions. These versions included both easy and hard versions of the color Stroop and the numerical Stroop. Our primary objective was to assess the participant's perception of task difficulty and boredom across the four tasks, thereby enabling us to select the most suitable Stroop task (color or numerical Stroop) for our present study and determining the appropriate sample size based on the effect size observed in our online study. The R code to reproduce analyses as well as the data set is available on OSF https://osf.io/8rgbd/?view_only=09a4d990eaa24a13b3df3dc8a0f89065.

Methods

Data collection for all four sessions of our within-subjects design study was done online via CloudResearch's Connect platform. Participants were assigned to one out of two participation orders (order one: easy color, easy numerical, hard color, hard numerical; order two: hard color, hard numerical, easy color, easy numerical). The interval between sessions varied from one to seven days. After completing the tasks for ten minutes each, participants were asked to indicate how much they agree or disagree with statements about their experience during the task (boredom and perceived task difficulty).

Participants

Starting with an initial cohort of 328 participants in session one, the study involved 210 participants (42.38 % female, 54.28 % male, 3.33 % other genders or no answer) who completed all four sessions of the study seriously (z-transformed error rate in the tasks less than three standard deviations away from the mean of all participants, attention checks answered correctly) and answered the relevant questions. The mean age of participants was 32.20 (*SD* = 5.31). Respondents were compensated with a fixed amount of $3.50 for their participation in each session. They had the opportunity to earn an additional variable compensation of up to $1 in two of the sessions. All participants gave informed consent before starting the experiment. The study was approved by the local Ethics Committee and carried out in accordance with the Helsinki Declaration.

Tasks

Stroop tasks. The color Stroop versions closely resemble those suggested in the present study. However, there are some distinctions: in the online study, a task reminder remained on screen, and the task was indicated not only by the "+" and "x" on screen during stimulus presentation but also with a fixation cross before stimulus onset which was either a "+" or an "x". To maintain participant's gaze at the center of the screen, we will omit the task reminder for the present study. Further, we will show the same fixation stimuli (###) in both tasks to keep the physiological influence on baseline (tonic) pupil size consistent across both task versions. To improve the quality of the pupillary data, we will present the stimuli for a fixed duration of 400 ms, instead of presenting them until participants react as we did in the online study. Lastly, while the hard Stroop task in the online study consisted of 50% incongruent trials, the hard Stroop task in the present study will consist of 100% incongruent trials. This change is implemented with the aim to (1) increase the overall difficulty of the hard task and (2) enhance the clarity of our analysis, particularly regarding pupillary reactions, as pupil size is greater in incongruent than in congruent trials. Otherwise, by calculating the mean of the five trials preceding the probe, congruent and incongruent trials would get mixed up.

In the numerical Stroop version (Schwarz & Ischebeck, 2003; Tzelgov et al., 1992) each trial consisted of a fixation cross (displayed for 1000 on screen), two presented numbers (one with a larger numerical value and one in a larger font size), and an intertrial interval lasting 500 ms. The numbers remained on screen until participants pressed one of two keys on the keyboard ("1" for left, "0" for right) to indicate their response. In the easy congruent-only version one number was both larger in value and in font size. Participants were instructed to identify the number with the larger font size. In the hard version, the numbers were presented either congruent or incongruent (50 % of the trials). In incongruent trials one number was larger in font size while the other was larger in value. The instructions switched during the task. In 80 % of the trials, participants were instructed to indicate the number with the lager font size, and in 20 % of the trials, they were instructed to indicate the number with the larger numerical value. A task reminder was displayed on screen. The task was further indicated by an “x” (indicate which number has the larger numerical value) or an “+” (indicate which number has a larger font size).

Subjective experience. After completing each task participants were asked to indicate how much they agree or disagree with statements about their experience during the task on a 9-point Likert scale (*1* = strongly disagree, *9* = strongly agree). Among these statements were one on boredom (“I was bored while performing the task”) and one on task difficulty (“the task was difficult for me”).

Supplementary measures. In addition to the tasks and questions described, we collected additional variables that are not pertinent to the present study.

Results

Results of a two-way repeated measures ANOVA which was performed to analyze the effect of task variant (color Stroop, numerical Stroop) and difficulty level (easy, hard) on boredom revealed a main effect of task variant ($F(1, 204) = 4.90$, $p = .028$, partial η^2 = 0.002), indicating that participants perceived the numerical Stroop ($M = 5.67$, $SD = 2.78$) as more boring than the color Stroop ($M = 5.42$, $SD = 2.78$). Additionally, a main effect of difficulty level was observed ($F(1, 204) = 35.51$, $p < .001$, partial $\eta^2 = 0.017$), highlighting that the easy tasks ($M = 5.90$, $SD = 2.72$) were more boring than the difficult tasks ($M = 5.19$, $SD = 2.74$). There was no significant interaction between the effect of task variant and level of difficulty ($F(1, 204) = 2.98$, $p = .086$, partial $\eta^2 = 0.001$).

Another two-way repeated measures ANOVA analyzing the effect of task variant (color Stroop, numerical Stroop) and difficulty level (easy, hard) on perceived task difficulty indicated a main effect of task version ($F(1, 204) = 96.46$, $p < .001$, partial $\eta^2 = 0.060$), with a higher perceived task difficulty in the color Stroop ($M = 3.75$, $SD = 2.58$) than in the numerical Stroop ($M = 2.68$, $SD = 2.03$). The ANOVA further revealed a main effect of difficulty level ($F(1, 204) = 170.31$, $p < .001$, partial $\eta^2 = 0.116$), with a higher perceived task difficulty in the hard ($M = 3.99$, $SD = 2.52$) than in the easy ($M = 2.44$, $SD = 1.94$) tasks. A significant interaction effect was found between task variant and difficulty level ($F(1, 204) = 76.80$, $p < .001$, partial $\eta^2 = 0.043$). This interaction suggests that the impact of difficulty level (easy, hard) on perceived task difficulty depends on the task variant (color, numerical). Specifically, in the color Stroop task, participants perceived the task to be more difficult in the hard ($M = 4.98$, $SD = 2.55$) compared to the easy condition ($M = 2.53$, $SD = 1.95$). In contrast, in the numerical Stroop task, the difference in perceived difficulty between the hard ($M = 3.00$, $SD = 2.07$) and easy ($M = 2.36$, $SD = 1.93$) conditions was less pronounced.

As the focus of the present study lies in the difference between the easy and hard versions of the color Stroop (see *Conclusion* for the rationale), we further conducted two t-tests comparing the perceived task difficulty and boredom between these two tasks. Results show that perceived task difficulty is greater in the hard version ($M = 4.98$, $SD = 2.54$) compared to the easy version of color Stroop task ($M = 2.54$, $SD = 1.97$), $t(209) = 13.34$, $p < .01$, $d = 0.92$. Boredom on the other hand is less pronounced in the hard ($M = 5.01$, $SD = 2.79$) than in the easy color Stroop ($M = 5.90$, $SD = 2.67$), $t(209) = 4.91$, $p < .01$, $d = 0.34$.

Conclusion

In both versions of the Stroop task (color, numerical) participants found the hard versions to be more difficult and less boring than the easy counterparts. Overall, the color Stroop was perceived as more difficult and less boring than the numerical Stroop. The difference in perceived task difficulty between the two difficulty levels was more pronounced in the color than in the numerical Stroop variant. Consequently, the color Stroop emerges as the more suitable choice for the present study. Our power analysis should focus on the difference in boredom between the hard and the easy color Stroop (d = 0.34) to adopt a conservative approach in calculating the required sample size.

Appendix B

Pre-test

Using the described experimental procedure, we tested two participants. Descriptive analyses revealed for the easy Stroop version a mean in boredom of 5.14 (*SD* = 1.04), in boredom related effort 5.59 (*SD* = 1.06), in task difficulty 2.77 (*SD* = 1.31), in task difficulty related effort 1.86 (*SD* = 0.37), in fatigue 6.41 (*SD* = 0.48), and in perceived effort 5.64 (*SD* = 0.60). In the hard Stroop version, we calculated a mean in boredom of 4.02 (*SD* = 0.59), in boredom related effort 3.68 (*SD* = 1.06), in task difficulty 4.76 (*SD* = 1.16), in task difficulty related effort 4.82 (*SD* = 1.20), in fatigue 5.27 (*SD* = 1.41), and in perceived effort 5.95 (*SD* = 0.57). Across all participants and probes, the mean baseline pupil size during the easy Stroop task was 918.38 (*SD* = 47.78), while the peak stimulus-evoked baseline corrected pupil size was 85.92 (*SD* = 12.30). In the case of the hard Stroop task, the average baseline pupil size was 788.07 (*SD* = 108.32), and the peak stimulus-evoked baseline corrected pupil size was 124.84 (*SD* = 23.96). The averages of thought probes are presented in table 1, the average tonic and phasic pupil size (averaged across the five trials preceding the thought probe) are shown in table 2.

Table 1

Means and Standard Deviations of Variables Corresponding to Each Thought Probe in the Easy und Hard Stroop Version

Variable	Difficulty Level		1	2	3	4	5	6	7	8	9	10	11
BrE	Easy	*M*	4.50	5.50	5.00	5.50	5.50	5.50	6.00	6.00	6.00	6.00	6.00
		SD	(0.71)	(0.71)	(1.41)	(0.71)	(0.71)	(0.71)	(0.00)	(1.41)	(1.41)	(1.41)	(1.41)
	Hard	*M*	2.00	2.50	2.50	2.50	3.50	4.00	4.00	5.00	5.00	4.50	5.00
		SD	(0.00)	(0.71)	(0.71)	(0.71)	(0.71)	(1.41)	(1.41)	(1.41)	(1.41)	(0.71)	(1.41)
DrE	Easy	*M*	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.50	2.00	1.50	1.50
		SD	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.71)	(0.00)	(0.71)	(0.71)
	Hard	*M*	5.00	5.00	5.00	5.00	4.50	4.00	6.00	5.00	4.50	4.50	4.50
		SD	(1.41)	(0.00)	(0.00)	(1.41)	(2.12)	(1.41)	(0.00)	(0.00)	(0.71)	(0.71)	(0.71)
Effort	Easy	*M*	5.00	5.50	5.00	5.50	5.50	5.50	6.00	6.00	6.00	6.00	6.00
		SD	(0.00)	(0.71)	(0.00)	(0.71)	(0.71)	(0.71)	(0.00)	(1.41)	(0.00)	(0.00)	(0.00)
	Hard	*M*	5.50	5.00	5.00	4.50	5.50	6.00	6.00	6.50	6.50	7.50	7.50
		SD	(0.71)	(0.00)	(0.00)	(0.71)	(0.71)	(0.00)	(0.00)	(0.71)	(0.71)	(0.71)	(0.71)
Fatigue	Easy	*M*	5.50	6.50	6.00	6.00	6.00	6.50	6.50	7.00	7.00	6.50	7.00
		SD	(0.71)	(0.71)	(0.00)	(0.00)	(0.00)	(0.71)	(0.71)	(0.00)	(0.00)	(0.71)	(0.00)
	Hard	*M*	4.0	4.00	4.00	4.00	2.50	5.00	5.50	6.50	7.50	7.50	7.50
		SD	(1.41)	(1.41)	(1.41)	(1.41)	(2.12)	(1.41)	(2.12)	(0.71)	(0.71)	(0.71)	(0.71)
Difficulty	Easy	*M*	3.50	2.5	2.00	2.50	3.00	2.50	3.50	3.00	2.50	3.00	2.50
		SD	(2.12)	(0.71)	(0.00)	(0.71)	(1.41)	(0.71)	(2.12)	(1.41)	(0.71)	(2.83)	(0.71)
	Hard	*M*	5.00	5.00	5.00	4.50	5.00	4.50	5.50	5.00	4.33	4.50	4.50
		SD	(1.41)	(0.00)	(0.00)	(0.71)	(0.00)	(0.71)	(0.71)	(0.00)	(3.21)	(0.71)	(0.71)
Boredom	Easy	*M*	4.00	5.00	5.50	5.50	6.00	6.00	6.00	6.50	6.50	6.00	6.00
		SD	(0.00)	(0.00)	(0.71)	(0.71)	(1.41)	(1.41)	(1.41)	(0.71)	(0.71)	(1.41)	(1.41)
	Hard	*M*	2.5	2.00	2.50	2.00	4.00	4.00	5.0	5.50	5.67	5.50	5.50
		SD	(0.71)	(0.00)	(0.71)	(0.00)	(0.00)	(0.00)	(0.00)	(0.71)	(1.15)	(0.71)	(0.71)

Note. BrE = Boredom related Effort, DrE = Difficulty related Effort.

Table 2

Means and Standard Deviations of Tonic and Phasic Pupil Size Corresponding to Each Thought Probe in the Easy und Hard Stroop Version

Variable	Difficulty Level		1	2	3	4	5	6	7	8	9	10	11
Tonic PS	Easy	*M*	976.4	162.2	972.6	958.9	918.5	921.5	884.3	923.9	833.5	891.7	858.7
		SD	(174.7)	(168.7)	(161.8)	(132.8)	(106.1)	(45.3)	(17.9)	(82.0)	(75.0)	(45.5)	(113.5)
	Hard	*M*	970.3	915.1	811.5	865.1	828.8	784.8	795.2	747.1	673.5	644.8	632.0
		SD	(738.7)	(666.1)	(565.7)	(596.7)	(567.6)	(489.9)	(453.2)	(378.0)	(265.2)	(211.8)	(218.3)
Phasic PS	Easy	*M*	93.9	91.2	85.5	87.3	97.4	54.4	83.0	84.2	101.7	86.3	80.2
		SD	(16.7)	(11.4)	(6.4)	(2.3)	(4.6)	(20.1)	(0.2)	(1.7)	(2.7)	(6.6)	(1.9)
	Hard	*M*	156.1	160.6	160.1	124.2	105.1	118.8	107.2	113.5	123.2	115.6	88.9
		SD	(97.1)	(135.4)	(106.2)	(80.8)	(59.8)	(52.1)	(45.5)	(90.0)	(77.3)	(53.7)	(37.5)

Note. Tonic PS = Tonic (Baseline) Pupil Size, Phasic PS = Phasic (Stimulus-evoked) Peak Pupil Size, baseline corrected. Tonic Pupil Size and Phasic Pupil Size are averaged across the five trials preceding the thought probe and reported in pixels.

Appendix C

Study-Design Template

Research Question	Hypothesis	Sampling Plan	Analysis Plan	Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis	Interpretation given different outcomes	Theory that could be shown wrong by the outcomes
Manipulation Check: Do hard and easy Stroop task versions differ in terms of their perceived overall task difficulty, the overall level of boredom they induce, and in terms of how good someone performs in the tasks?	Participants report greater overall levels of boredom and less overall task difficulty and show a better performance in the easy compared to the hard.	Ninety-five non-color-blind participants between 16 and 45 years from the general population will be recruited[3].	To test this hypothesis, we will perform four paired t-tests comparing either task difficulty, boredom, error rates or reaction times (as dependent variables) in the easy and hard Stroop version (as independent variable).		Differences: Stroop Tasks classified as difficult (and high self-control demanding tasks) are perceived as more difficult, they induce less boredom and lead to a worse performance in comparison to easy Stroop Tasks (often classified as low self-control tasks). No differences: The manipulation of the difficulty level of the Stroop Tasks does not impact perceived difficulty levels or performance. This would indicate that these variables are not affected by the type of Stroop task that is used.	The premise that structurally more demanding tasks are also perceived as more difficult and yield lower performance scores. In addition: that self-reported boredom levels are higher in structurally easy tasks.

[3] To determine the required sample size for the present study, we conducted a G*Power Analysis Faul et al. (2007). Given that the effect size observed in the calculated t-tests in our online study was larger for task difficulty (d = 0.92) than for boredom (d = 0.34), our power analysis focuses on the effect size of the difference in boredom between the easy and the hard color Stroop. The power analysis for a one-tailed paired t-test was calculated based on this effect size. The analysis indicated that 95 participants are necessary to detect a difference in boredom between the tasks with a power of 95% at an alpha level of 0.05.

					We will conduct our main analyses even if no differences between the tasks emerge, as we would still gain valuable insights into the dynamics of boredom and cognitive effort during task performance.	
Main Research Question 1: Do boredom, effort demands due to boredom, perceived task-difficulty, effort demands due to perceived task difficulty, overall perceived effort and fatigue change over time during both tasks (easy and hard Stroop)?	We expect an increase in boredom, effort due to boredom, fatigue, and overall effort and a decrease of task difficulty and effort due to task difficulty over time for both Stroop versions (easy and hard). We expect a greater decrease of task difficulty and task-difficulty-related effort over time in the hard than in the easy version. We further expect an interaction for overall effort with a greater increase in overall effort in the easy Stroop version.		Six Linear Mixed Models including boredom, boredom related effort, task difficulty and task difficulty related effort, overall effort and fatigue as outcome variables, task (easy, difficult) and time (probe one to eleven) as fixed effects accounting for differences among participants by including random intercepts and different effects of time on the outcome variables by including random slopes for time.		If hypothesis is supported: Over time tasks become more boring which increases effort demands to continue with a boring task. When tasks become easier over time, this reduces effort demands due to task difficulty. At the same time overall effort and fatigue increase. This holds true for tasks with different difficulty levels. Nevertheless, task difficulty and the related effort decrease stronger in hard tasks (due to the initially higher task difficulty). Further, as task difficulty decreases to a greater extent in hard tasks than in easy tasks, and boredom increases in both tasks, overall effort (which is expected to include boredom and difficulty related boredom) increases less in harder tasks than in easy tasks.	N/A

					If our hypothesis is not supported: When results show that it depends on the tasks' difficulty level (easy, hard) if boredom, perceived task difficulty, the related effort, perceived overall effort, and fatigue decrease or increase over time, we would conclude that task characteristics are relevant for predicting how these variables change over time.	
Main Research Question 2: Are physiological measures of effort solely influenced by task-difficulty-related effort or are they also affected by boredom-related effort?	Pupil size (tonic and phasic), as physiological indicator of effort, is predicted more accurately by task-difficulty-related effort and boredom-related effort than by task-difficulty related effort alone.		We will construct four Linear Mixed Models (LMM) for each outcome variable (tonic pupil size, phasic pupil size). Difficulty level of the task (easy, hard), time (thought probe one to eleven), effort due to task difficulty and effort due to boredom will serve as fixed predictors and participants' score of the Short Boredom Proneness Scale (SBPS), the Beck Depression Inventory II (BDI-II), and the Brief Self-Control Scale (BSCS) as covariates in different models. We will account for general differences between participants in tonic and phasic pupil		Should our results, as hypothesized, show that pupil size as physiological indicator of effort is predicted more accurately when effort due to boredom is considered as predictor, we would characterize boredom as effortful. This would emphasize the importance of recognizing boredom as a confounding variable within self-control research and, more general, within the broader scope of study designs. Conversely, if pupil size is not predicted by cognitive effort due to boredom, it could be argued that being bored during task performance might not	Performing easy and hard tasks is not effortful on a physiological level. Enduring boredom is not effortful on a physiological level.

			size by including random intercepts and different effects of time on pupil size among participants by including random slopes for the variable time. To evaluate the relative performance of these models in predicting pupil size, the AIC, BIC, and WAIC will be reported. We will base our decision on the most optimal model regarding model fit and complexity on the BIC.		trigger changes in this physiological correlate of effort. However, this outcome could even indicate a connection between boredom and low arousal levels, a correlation that remains ambiguous. In this case, the perceived effort elicited by boredom might be less influential on the pupillary activity than the possible decrease in arousal.	
Research Question 3A: Is task performance in a secondary task influenced by the level of difficulty (easy, hard) in the first task?	Given the inconsistent results in ego depletion research (Giboin & Wolff, 2019; Hagger et al., 2016; Hagger et al., 2010), this research question will be tested exploratively.		We will conduct two paired t-tests. One t-test will analyze participants' error rate as dependent variable, while the other will focus on the reaction time. The difficulty level of the first task (easy, hard) will serve as the independent variable in both tests.		If secondary task performance is worse after performing the hard task: Working on a structurally more demanding task leads to reduced performance on a subsequent task. If secondary task performance is worse after performing the easy task: Working on a structurally less demanding task leads to reduced performance on a subsequent task. If secondary task performance does not differ: Structural demand	Ego Depletion: Self-Control needs in a preceding task leads to less remaining self-control for a secondary task which manifests in a worsened task performance. (Please note that even if a reduction in performance is observed, this would not constitute evidence for a depletion of a specific resource, as we do not measure such a resource in this research.) Mental Fatigue Theories: Mental fatigue occurs after a period of prolonged

					differences in a prior mental task do not differentially carry over to performance on a secondary mental task.	mental activity which leads to decrements in task performance.
Research Question 3B: Can task performance in a secondary task be explained by the perceived amount of effort due to task difficulty invested in the first task? And does considering perceived boredom related effort in the first task provide a more accurate prediction of performance in the second task than considering task difficulty related effort alone?	This research question will be tested exploratively.		Three distinct LMMs will be tested for each performance variable in the flanker task (error rate, reaction time). The first model will assess the impact of the difficulty level (easy, hard; fixed effect) and the perception of effort due to task difficulty (fixed effect) in the Stroop task as well as the type of flanker trial (congruent, incongruent) on the reaction time in the flanker task while accounting for random effects among participants with random intercepts and random slopes. The second model will differ from the first by incorporating perceived boredom related effort as a fixed effect instead of perceived task difficulty related effort. The third model will include both perceived task difficulty and boredom related effort as fixed effects.		If secondary task performance can be explained by task difficulty related effort in the preceding task: Subjectively investing effort due to the difficulty of a task reduces performance on a subsequent task. If secondary task performance can be explained more accurately by boredom related effort and task difficulty related effort together: Both the perception of a higher task difficulty related effort and boredom related effort in a first task, reduce performance in a subsequent task. Thus, performing boring as well as difficult tasks can lead to a worsened performance in a secondary task if effort (related to boredom or task difficulty) is perceived to be invested in the preceding task.	Ego Depletion: Self-Control needed in a preceding task leads to less remaining self-control for a secondary task which manifests in a worsened task performance. (Please note that even if a reduction in performance is observed, this would not constitute evidence for a depletion of a specific resource, as we do not measure such a resource in this research.) Mental Fatigue Theories: Mental fatigue occurs after a period of prolonged mental activity which leads to decrements in task performance.

			The model with the lowest BIC will be considered as the most optimal one.		If performance in the secondary task cannot be explained by task difficulty or boredom related effort in a first task: Performance in a secondary is not impacted by the perceived amount of effort (either task difficulty or boredom related) invested in a preceding task.	
Exploratory: Does boredom proneness affect the quantity of sweets participants take with them after completing the experiment?	1) Controlling for trait self-control, boredom proneness effects the amount of sweets taken. 2) No other hypothesis. Other analysis with key variables fully exploratory.		1) We will perform a multiple regression analysis with the amount of sweets taken as dependent variable, boredom proneness as predictor and trait self-control as control variable. 2) We will conduct correlation analyses between the amount of sweets taken and our key variables.		If boredom proneness affects the amount of sweets taken: Reward sensitivity after the performance of a prolonged task differs with higher levels of boredom proneness. General alterations in reward sensitivity are likely. This might impact behavior of individuals with higher boredom proneness in various ways. If boredom proneness does affect the amount of sweets taken: Reward sensitivity after the performance of a prolonged task is not affected by boredom proneness. There are either no alterations in the reward system of individuals with	Individuals with higher boredom proneness levels show alterations in their reward sensitivity (which impacts need for rewards after experiencing boredom).

					higher boredom proneness or the alterations do not manifest in the need for sweets after performing prolonged tasks.	