THE ROLES OF COGNITIVE LOAD AND APPRAISAL OF TASK DIFFICULTY IN PREDICTING SUBJECTIVE FATIGUE AND SUBSEQUENT TASK DISENGAGEMENT

by

Nicole E. Carmona

B.Sc.(Hons), Queen's University, 2016

A thesis presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Arts

in the program of

Clinical Psychology

Toronto, Ontario, Canada, 2019

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A THESIS

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my thesis may be made electronically available to the public.

Abstract

The Roles of Cognitive Load and Appraisal of Task Difficulty in Predicting Subjective Fatigue and Subsequent Task Disengagement

Master of Arts, 2019

Nicole E. Carmona

Psychology

Ryerson University

Central fatigue refers to an inability to sustain mental or physical performance in self-initiated tasks and an increased perception of effort (Chaudhuri & Behan, 2000), suggesting that fatigue results from a mismatch between the perceived resources needed to initiate a task and the availability of cognitive resources available to complete it. Consequently, fatigue may be considered a "stop-emotion" to preserve cognitive resources, resulting in task disengagement (Meijman, 2000). This study investigated: 1) the role of perceived cognitive resources in the development of mental fatigue by manipulating the task demands and appraisals of task difficulty, and 2) the subsequent effect of fatigue on task engagement. Fatigue increased and cognitive resources decreased with time on task, rather than as a result of the task demands × instruction of task difficulty interaction. Increases in fatigue did not predict measures of engagement in almost all cases. Implications and suggestions for future research are discussed.

Acknowledgements

Firstly, I would like to thank my supervisor, Dr. Colleen Carney, for trusting me and my creative vision and for bestowing upon me a large amount of independent responsibility in seeing this project through, from conception to completion. Her trust (and vote of confidence) has allowed me to realize my own potential as an independent scholar, and to feel confident following my own instincts. I am also extremely grateful for the valuable feedback that she has provided me with over the past two years that has not only shaped this thesis, but has also shaped me as a person. To my committee, thank you for your feedback and your thoughtful discussions, and for bringing such interesting perspectives to this project.

I would also like to thank Olya Bogouslavsky, my mentor and friend, for whom I am eternally grateful. I do not believe I would be in graduate school without her scaffolding and her unwavering confidence in my abilities as a student and a researcher. To say that working so closely with her as an undergraduate student was instrumental is an understatement, and I am thrilled that we had the opportunity to continue working together throughout my Master's degree. To the SAD Lab as a whole, thank you for supporting me over the past two years and for creating such a positive lab environment. It is such a pleasure to work on a team of passionate and likeminded people.

Thank you to my friends and family for keeping me grounded, keeping things fun, and for always stressing the importance of work-life balance. And finally, to Michael, thank you for your continuous love, support, and encouragement. This degree was not without trying times, and I am so thankful to have had you by my side to celebrate the highs and help me through the lows. To put it simply, this would not have been possible without you.

Table of Contents

List of Tables

List of Figures

Introduction

Fatigue, in the acute sense, is a common experience in the general population following periods of perceived exertion or stress (Chaudhuri & Behan, 2000), with potential detrimental effects on physical wellbeing, mood and motivation, as well as attention, memory, and decision-making processes (Marcora, Staiano, & Manning, 2009; Trejo et al., 2005; van der Linden, Frese, & Meijman, 2003). Notwithstanding its widespread experience, the mechanisms underlying the development of fatigue remain poorly understood. Fatigue has been widely studied in both clinical and non-clinical populations, although the term "fatigue" has been inconsistently used to refer to subjective "tiredness," as well as various physiological, cognitive, and behavioural states (Balkin & Wesensten, 2011). The conflation of these distinct states has led to some confusion regarding how to classify, research, and treat fatigue in different populations.

The neurobiological model of fatigue outlined by Chaudhuri and Behan (2000, 2004) suggests that fatigue can have either a peripheral or a central cause, with differential effects on physical and cognitive outcomes. Peripheral fatigue, also known as objective fatigue, refers to muscle fatigability (i.e., the inability to sustain a specified force or work rate during exercise or physical activity) due to failure of neuromuscular transmission, metabolic defects of the muscle, or a peripheral circulatory failure, with no effects on mental endurance (Chaudhuri & Behan, 2000). In contrast, central fatigue is subjective and refers to the failure to initiate or sustain mental or physical activities requiring self-motivation and is accompanied by an increased perception of effort (Chaudhuri & Behan, 2004). Because it is the result of perceptions of effort and motivational states, central fatigue is inherently subjective and cannot be measured objectively. It is thought that a deficit of the non-motor functions of the basal ganglia underlies the experience of chronic central fatigue (Chaudhuri & Behan, 2000, 2004).

The cognitive component of central fatigue, also known as mental fatigue, is of particular interest due to its deleterious effects on performance of day-to-day tasks in non-clinical populations (Gunzelmann, Moore, Gluck, Van Dongen, & Dinges, 2011). Although the experience of mental fatigue overlaps with excessive sleepiness, depression, and amotivation (Chaudhuri & Behan, 2000), fatigue is a distinct construct that can independently contribute to accidents, burnout, distress and impairment, especially when experienced chronically. Several models have been proposed to explain mental fatigue, but it is generally agreed upon that it is characterized by "a reluctance for further effort and changes in mood, motivation, and information processing" (Hopstaken, van der Linden, Bakker, & Kompier, 2015a, pp. 305). Additionally, central fatigue can be a significant symptom of various sleep (e.g., insomnia disorder), psychiatric (e.g., major depressive disorder, chronic fatigue syndrome), and neurological (e.g., multiple sclerosis, Parkinson's disease) disorders, resulting in significant distress and impairment. Therefore, an understanding of the determinants of mental fatigue in healthy populations has important implications for the development of fatigue treatments for use in clinical populations.

Mental Fatigue: Effort, Motivation, and Cognitive Resources

Mental fatigue induction typically rests on the assumption that cognitive resources will eventually be exhausted by task demands, though this can be achieved with different methods. The use of long periods of sustained cognitive activity ("time on task") is one of the most common paradigms for the induction of mental fatigue despite maintenance in vigilance over time (e.g., Bailey, Channon, & Beaumont, 2007; Boksem, Meijman, & Lorist, 2006; Gergelyfi et al., 2015; Gunzelmann et al., 2011; Hopstaken et al., 2015a; Hopstaken et al., 2015b; Sandry et al., 2014; Tanaka, Ishii, & Watanabe, 2014; van der Linden, Frese, & Meijman, 2003; van der Linden, Frese, & Sonnentag, 2003), although it has not consistently resulted in fatigue induction in the literature (Ackerman & Kanfer, 2009; Boksem & Tops, 2008). To account for this inconsistency, fatigue has been proposed to result from a cost-benefit analysis, in which sustained activity will only result in fatigue when energetical costs (i.e., effort) are high but potential rewards are low (Boksem & Tops, 2008). It has been hypothesized that a neural monitor of performance costs and physiological resources, elaborated below, underlies the experience of increased subjective effort associated with fatigue (Tops, Boksem, & Koole, 2013). As task engagement continues over time, increases in energy depletion outweigh potential rewards, leading to decreased motivation to continue working towards a goal that has not already come to fruition (Boksem & Tops, 2008). Consistent with this theory, several studies have found that manipulation of either extrinsic or intrinsic reward results in the reversal of mental fatigue and improved performance on cognitive tasks, as increased effort becomes more justified (Boksem, Meijman, & Lorist, 2006; Hopstaken, van der Linden, Bakker, & Kompier, 2015).

Building upon Chaudhuri and Behan's model of fatigue surrounding deficits in the basal ganglia, Boksem and Tops (2008) hypothesized that midbrain dopaminergic circuits underlie the effort-based decision making that contributes to mental fatigue, with the nucleus accumbens in the striatum (NAc; part of the basal ganglia) receiving information about rewarding and aversive qualities of potential actions from the orbitofrontal cortex and the basolateral amygdala, as well as integrated input from the anterior cingulate cortex guiding behaviour toward the course of action with the highest reward value. Information as to the costs of action comes from the insula, which relays information regarding available energetic resources to the NAc. Finally, dopaminergic and cholinergic projections to the prefrontal cortex contribute to cognitive control processes, with decreases in dopamine and acetylcholine contributing to increased distractibility and decreased cognitive control (Boksem & Tops, 2008). This proposed psychobiological mechanism has been supported by lesion studies using animal models and neuroimaging research among clinical and non-clinical samples (Boksem & Tops, 2008; Dobryakova, DeLuca, Genova, & Wylie, 2013).

In addition to manipulating time on task, an alternative approach to inducing mental fatigue has been to manipulate task demands (i.e., "cognitive load"), either by varying the complexity of the task (e.g., increasing working memory load using an *n*-back task) or by decreasing the available time to process ongoing task demands (Borragán, Slama, Bartolomei, & Peigneux, 2017). This paradigm has historically been based on the rationale that increasing cognitive load uses more cognitive resources and requires greater effort (Kahneman, 1973), resulting in greater levels of mental fatigue. Indeed, several studies have found that in non-clinical populations, increasing cognitive load by increasing working memory demands not only increases subjective fatigue, but is also associated with poorer performance on tasks requiring executive functions (e.g., cognitive flexibility, Shigihara et al., 2013a; working memory, Hopstaken et al., 2015), reduced P300 amplitude (Hopstaken et al., 2015; Käthner et al., 2014), and increases in baseline and stimulus-evoked pupil diameter (Hopstaken et al., 2015a, 2015b). Likewise, manipulation of cognitive load by identifying individuals' maximal processing speed capacity has also demonstrated subjective and objective fatiguing effects using a time load dual back task, in which cognitive load is manipulated by altering the time available to process ongoing task demands, regardless of task complexity (Borragán et al., 2017).

The assumption of a finite amount of resources that can be exhausted more readily by difficult tasks is controversial, as these “resources” have no discernible physiological substrates, there is no evidence that increases in effort consume large amounts of energy, and the experience of fatigue associated with low resources can be easily reversed by increasing motivation (e.g., by providing incentives; Hopstaken et al., 2015a, 2015b; Inzlicht & Marcora, 2016). It is perhaps more appropriate to consider that cognitive resources are not depleted by the execution of tasks but are differentially allocated to various tasks, with tasks associated with a higher cognitive load requiring the allocation of more attentional resources (Evans, Boggero, & Segerstrom, 2015; Kahneman, 1975). In many cases, individuals behave as if their cognitive resources are limited, even though this may not be the case (Evans, Boggero, & Segerstrom, 2015).

Two studies have illustrated this phenomenon clearly. In one study, participants completed an easy (low-load) or difficult (high-load) version of a letter-crossing task, after which participants either (1) received no feedback (i.e., controls), or (2) read a passage of text explaining that the colour of the page used for the previous task has been shown to either (a) replenish or (b) deplete mental resources (Clarkson, Hirt, Jia, & Alexander, 2010). They then completed an anagram task, and their performance on this task was analyzed. Providing participants with this situational feedback resulted in an interaction effect, in which participants in the low-load condition used feedback to *interpret* their amount of available mental resources for a second task, whereas participants in the high-load condition used feedback to *explain* their amount of available mental resources. Specifically, participants in the low-load condition who received feedback that they were depleted interpreted themselves as fatigued and performed similarly to those in the high-load control condition on the anagram task (Clarkson et al., 2010). Moreover, participants in the high-load condition who received feedback that they were replenished misattributed their current state of fatigue to the paper colour and were led to believe that they had more available mental resources, allowing them to perform similarly to those in the low-load control condition on the anagram task; participants in the high-load condition also reported significantly greater subjective "exhaustion" after replenished feedback compared to depleted feedback, underscoring the effect of appraisals and perceptions of resources in predicting performance (Clarkson et al., 2010). Additionally, Job, Dweck, and Walton (2010) demonstrated in a series of studies that manipulating participants' implicit theories about willpower can affect whether or not one experiences a performance decrement—specifically, after an initially fatiguing task, participants who were led to believe that willpower is a renewable resource maintained their performance on a second task, while participants who believed that willpower resources are finite performed worse on the second task.

In his Motivational Control Theory of cognitive fatigue, Hockey (2011) posited that effort is not just a product of task demands, but is an optional response to the "perception and appraisal of demands" that is under the control of the individual (pp. 169), citing evidence that high-effort instructions interact with greater task demands to produce greater levels of cognitive fatigue than if a normal amount of effort is expected (Earle, 2004 as cited in Hockey, 2011). In the same vein, the Job Demand-Resources model of burnout, which is characterized by high levels of fatigue and emotional exhaustion due to work stress, suggests that burnout results from an imbalance between work demands and the resources to cope with such demands (Bakker, Demerouti, & Sanz-Vergel, 2014; Fernández-Castro et al., 2017). This convergent evidence indicates that manipulation of one's appraisal of task demands may influence mental fatigue by influencing the expected amount of effort required. Surprisingly, despite the centrality of cognitive resources to our understanding of mental fatigue, no published studies have actually included measures for the subjective momentary assessment of cognitive resources in relation to mental fatigue to test this hypothesis.

The construct of mental fatigue, which has been the subject of research in the fields of cognitive psychology, psychophysiology, and neuroscience (to name a few), shares striking similarities with the social psychological construct of “ego depletion” (Baumeister, Bratslavsky, Muraven, & Tice, 1998). Ego depletion refers to a “state of diminished volitional resources” (Baumeister, Muraven, & Tice, 2000, pp. 131) characterized by decreased persistence on tasks after completing an initial task that requires self-control. Much of the research into these two similar constructs has occurred in parallel, with inconsistent generalization of the findings from mental fatigue to ego depletion and vice versa. Although an overview of the literature surrounding ego depletion is beyond the scope of the present thesis, the resource model of self-control on which ego depletion is based suggests that internal resources that enable self-control are limited; ego depletion results when engagement in “controlled, willful processing” (Inzlicht & Schmeichel, 2012, pp. 450) depletes this inner resource, limiting the individual’s ability to further engage in other subsequent tasks requiring self-control (Inzlicht & Schmeichel, 2012; Baumeister, Vohs, & Tice, 2007). This model bears resemblance to Kahneman’s (1973) model of attentional resources and, for the same reasons, appears to be insufficient in accounting for the inconsistencies in the ego depletion literature (e.g., increasing motivation by providing incentives has been shown to reverse ego depletion; Muraven & Slessareva, 2003). In recent years, other models have been proposed to overcome the pitfalls of the resource model. For example, the process model of ego depletion proposes that following an initial task requiring self-control, shifts in motivation and attention away from restraint and toward gratification of desires undermine self-control on a second task (Inzlicht & Schmeichel, 2012). This model is similar to the aforementioned cost-benefit hypothesis of mental fatigue and task engagement outlined by Boksem and Tops (2008).

Despite the apparent similarities between ego depletion and mental fatigue, there have been some inconsistencies across the two fields that require consideration. As mentioned above and reviewed in more detail below, working memory performance has been reliably shown to decrease with the experience of mental fatigue using both time-based and cognitive-load based paradigms (e.g., Hopstaken et al., 2015a), although the effect of ego depletion on working memory has been inconsistent. One recent study used a sophisticated design with four different ego-depleting paradigms to demonstrate that ego depletion does not result in working memory impairment relative to non-depleting conditions (Singh & Göritz, 2018). Interestingly, the lack of relationship between ego depletion and working memory performance was not moderated by trait self-control capacity or implicit theories about willpower (i.e., beliefs about self-control as a renewable versus non-renewable resource). A number of points may underlie this apparent discrepancy. The depleting tasks chosen by Singh and Göritz (2018) required attention and response inhibition (e.g., E-crossing task, colour Stroop task), but are not known to assess or tax working memory function. Therefore, this study answers the question of whether ego depletion (i.e., taxing self-control resources, in this case by inhibiting automatic responses) impairs working memory performance, but not the reverse question of whether engaging in working memory tasks induces ego depletion. In contrast, the mental fatigue literature indicates that working memory tasks (e.g., the *n*-back) are capable of inducing mental fatigue (and subsequent task disengagement, resulting in poorer performance; e.g., Hopstaken et al., 2015a). Additionally, ego depletion studies have used sequential-task paradigms to demonstrate that the self-control "depletion" is generalizable to various different tasks; for example, the initial task of ignoring words at the bottom of a screen while watching a video of a person's face results in poorer performance on a handgrip task (e.g., Muraven et al., 1998). This is true of the Singh and Göritz (2018) study as well, in which the initial depleting tasks were different from the second task. In contrast, studies investigating mental fatigue have generally not sought to reveal a common resource that can be depleted by disparate tasks; instead, they have demonstrated that adjusting the cognitive load of a certain task can lead to increased feelings of fatigue and impaired performance on that same task.

More broadly, several limitations of the ego depletion literature, outlined by Lurquin and Miyaki (2017) and briefly summarized here, underlie its "replication crisis," which may also contribute to the lack of clarity surrounding whether or not mental fatigue and ego depletion are one and the same. For instance, there is a lack of a clear operational definition of self-control across ego depletion studies, without tying the definition to any specific atomized cognitive functions. This issue is further compounded by the circular logic used to select self-control tasks, with tasks chosen because they have been previously found to have a depleting effect. The lack of independent validation of self-control tasks and their performance metrics results in ambiguity as to whether or not self-control is really being assessed. Furthermore, the assumption that self-control can be tapped by exhausting various cognitive, physical, or emotional domains also rests on the circular logic that generalization of self-control resources must exist because ego depletion is observed (Lurquin & Miyaki, 2017).

In sum, ego depletion and mental fatigue both purportedly result from a state of low motivation and perception of high effort following the completion of a task requiring executive functions and can both be reversed by manipulating motivational state or perceptions of cognitive resources and task demands. At this time, it seems appropriate to regard ego depletion and mental fatigue as related but distinct constructs – it is possible that mental fatigue is a subconstruct within the phenomenon of ego depletion in general. However, due to the limitations plaguing the ego depletion literature, it cannot be said with certainty whether these two phenomena are actually one (Lurquin & Miyaki, 2017). Future research should seek to clarify this possibility.

Fatigue and Cognition

It is well established that those experiencing fatigue do not feel like they are at their "mental best." Indeed, lapses in working memory and sustained attention have been found in studies of experimental fatigue manipulation, controlled sleep deprived states, shift workers, and clinical samples of individuals with various sleep, psychiatric, and neurological disorders (e.g., DeLuca, Johnson, Ellis, & Natelson, 1997; Durmer & Dinges, 2005; Geiger-Brown et al., 2012; Kahol et al., 2008; van der Linden, Frese, & Meijman, 2003; Van Dongen, Baynard, Maislin, & Dinges, 2004).

Specifically, higher-order cognitive functions are particularly vulnerable to the effects of mental fatigue due to their increased reliance on mental effort and controlled processing, while functions reliant on automatic processing are preserved (van der Linden, 2011). Recent behavioural and physiological research has supported the view that the psychomotor vigilance decrement seen with long times on task has been better explained by the *underload hypothesis*, demonstrating associations with boredom and sleepiness rather than cognitive fatigue (Pattyn, Neyt, Hendrickx, & Soetens, 2008; Shigihara et al., 2013). Additionally, decreases in psychomotor vigilance do not account for the deficits in higher-order cognitive functioning seen in cognitive fatigue (Borragán et al., 2017; Gergelyfi, Jacob, Olivier, & Zenon, 2015).

Van der Linden, Frese, and Meijman (2003) demonstrated that compromised executive functions when mentally fatigued suggest that fatigue decreases the likelihood of actions being guided by task goals or changes to task context, rather than causing fundamental alterations in cognitive functions. This study found that performance on the Tower of London and Wisconsin Card Sorting Task—specifically, executive functions such as planning and flexibility—suffered in participants who were mentally fatigued compared to non-fatigued controls, while non-executive aspects of these tasks such as short-term memory did not differ between fatigued and non-fatigued participants (van der Linden, Frese, & Meijman, 2003). Specifically, fatigued participants took longer to plan their first move on the Tower of London task and took less time to learn new rules on the Wisconsin Card Sorting Task, indicating perseveration and an increased reliance on automatic processing when fatigued (van der Linden, Frese, & Meijman, 2003).

However, it is important to note that mental fatigue does not always result in neurobehavioural performance deficits (Kanfer, 2011). Individual differences in the adaptive strategies chosen to maintain cognitive performance have been found, with some individuals choosing accuracy over speed and vice versa when motivated to improve performance on an action monitoring task (Boksem, Meijman, & Lorist, 2006). In addition, increases in subjective fatigue may sometimes be accompanied by compensatory increases in mental effort in order to achieve a specific goal, offering an explanation as to why subjective fatigue often precedes objective indicators of mental fatigue like performance deficits (Gergelyfi, Jacob, Olivier, & Zenon, 2015; Kanfer, 2011). Kanfer (2011) outlined a temporal relationship between subjective fatigue and performance decrements, wherein the increased perception of mental fatigue may lead to decreased motivation and willingness to sustain high levels of cognitive effort, resulting in a decrement in task performance. Consistent with the theory proposed by Boksem and Tops

(2008), this suggests that following the experience of subjective fatigue, the subjective appraisal of effort and the motivation to achieve specific goals is critical in determining if cognitive performance will suffer.

Fatigue and Disengagement

Task disengagement, whether intentional or not, is one of the most prominent characteristics of mental fatigue (Kanfer, 2011; van der Linden, 2011). Differing proposed mechanisms of mental fatigue have led to various hypotheses as to what contributes to task disengagement. For example, in line with the reward-cost theory of mental fatigue, disengagement has been speculated to occur when perceived effort becomes too great for a potential reward, resulting in the subjective feeling of mental fatigue (Tops, Lorist, Wijers, & Meijman, 2004). It has also been suggested that fatigue is an adaptive “stop-emotion” that urges the individual to reduce further effort or withdraw task engagement in order to prevent the over-investment of cognitive resources in a limited set of activities (Meijman, 2000, as cited in van der Linden, 2011). The notion that subjective mental fatigue signals the brain to reduce effort or terminate engagement entirely is quite dated; in 1900, Thorndike observed that “we can feel mentally fatigued without being so, so that the feelings serve as a sign to stop working long before our actual ability to work has suffered any important decrease” (pp. 481). Behaviourally, disengagement from a task requiring mental effort results in poor performance, suggesting that mental disengagement from cognitive tasks may underlie the cognitive impairments seen in fatigued individuals. Put differently, fatigued individuals may disengage from the task at hand in order to conserve cognitive resources, resulting in poor performance on cognitive tasks requiring executive function.

In support of this hypothesis, van der Linden, Frese, and Sonnentag (2003) investigated the effect of experimentally-manipulated mental fatigue on completion of an unfamiliar complex computer task requiring problem solving skills (e.g., goal setting, hypothesis formation, planning, and integration of feedback). The complex task required the completion of several consecutive subtasks. The authors proposed that systematic exploration of the task, associated with the use of problem-solving skills, requires a high level of engagement and would decrease as a consequence of mental fatigue. In contrast, they hypothesized that unsystematic trial and error and rigid behaviour (i.e., perseveration) would be associated with task disengagement as fewer cognitive resources are allocated to hypothesis formation, planning, and reflection, and that these unsystematic exploration methods would increase under conditions of mental fatigue. In this study, fatigued participants used significantly less systematic exploration and demonstrated significantly more rigid behaviour than the control group, although perseveration was lowest for fatigued participants with low experience with computers. Additionally, although fatigued participants made more errors while completing the task, no differences were found in the number of subtasks completed between the two groups, demonstrating that mental fatigue can lead to more subtle changes in behaviour prior to deterioration of performance (van der Linden, Frese, & Sonnentag, 2003). The results from this study indicate that consistent with prior research on fatigue and cognition, executive functions are particularly vulnerable to the effects of mental fatigue and contribute to reductions in task engagement, as cognitive resources are withdrawn from these higher-order functions.

Mechanistically, two psychobiological systems are proposed to work in parallel to contribute to task disengagement. Reductions in dopamine and acetylcholine influx to the prefrontal cortex from midbrain dopaminergic structures result in increased distractibility and decreased goal activation, promoting exploratory behaviour for potentially more rewarding outcomes (Boksem & Tops, 2008). In addition to the role of dopaminergic reward systems, the locus coeruleus-norepinephrine (LC-NE) system is also hypothesized to underlie task disengagement as a result of increasing levels of mental fatigue (van der Linden, 2011). These two systems likely work in concert, as the locus coeruleus receives input from the orbitofrontal and anterior cingulate cortices, which are involved in coding the reward value and costs of tasks or behaviour and have been implicated in the experience of mental fatigue, as described above (Aston-Jones et al., 2002; Boksem & Tops, 2008; Rajkowski, Lu, Zhu, Cohen, & Aston-Jones, 2000). Norepinephrine is associated with arousal, which is further associated with the preparedness of the brain for attention toward, and higher-order processing of, perceptual input (Aston-Jones & Cohen, 2005; van der Linden, 2011). Although activation of the LC-NE system cannot be measured directly, several physiological parameters have been identified as indirect measures of LC-NE activity, including pupil diameter and the P300 amplitude (Aston-Jones & Cohen, 2005). In their Adaptive Gain Theory, Aston-Jones and Cohen (2005) suggest that the LC-NE system's role in task (dis)engagement lies in its facilitation of the decision to either exploit the current task in anticipation of a potential rewarding outcome, or to explore the environment for other, potentially more rewarding activities. These two outcomes are associated with different modes of norepinephrine release; the phasic mode is associated with intermediate baseline (pre-stimulus) levels of norepinephrine and bursts of norepinephrine release after viewing a stimulus, indicating responding to the task at hand. The tonic mode corresponds with task disengagement and exploration of the environment for more rewarding outcomes, as it is characterized by high levels of both baseline and stimulus-evoked norepinephrine – the high level of baseline norepinephrine is indicative of attention to task-irrelevant stimuli. Indeed, evidence indicates that fatigued individuals explore the environment for more rewarding activities, suggesting minimal engagement in the task at hand under fatigued conditions (Aston-Jones & Cohen, 2005). In support of this theory, larger baseline pupil diameter has been found to correlate with exploratory behaviour and individual differences in baseline pupil diameter covary with individual differences in exploratory choice behaviour (Jepma & Nieuwenhuis, 2011). Interestingly, variability in baseline pupil diameter across trials tracks changes in exploratory versus exploitative behaviour consistent with changes in perceived task utility (Jepma & Nieuwenhuis, 2011). Additionally, Hopstaken and colleagues (2015b) suggest that the Adaptive Gain Theory leaves room for a third mode of norepinephrine release, aptly titled the "disengage mode," characterized by decreased levels of both baseline and stimulus-evoked norepinephrine and corresponding to low attention, low arousal, and disengagement from tasks.

Taking these factors into consideration, a recent study conducted by Hopstaken and colleagues (2015a) examined the relationships between mental fatigue, motivation, and cognition using subjective, behavioural (i.e., accuracy on the cognitive task), and physiological (i.e., pupil diameter and P300 amplitude) indicators of task engagement. In this study, fatigue was induced using two hours of continuous performance on an *n*-back task of varying levels of difficulty (i.e., 1-back, 2-back, and 3-back conditions). Subjective fatigue was found to significantly increase with time on task while subjective engagement was found to significantly decrease with time on task; unfortunately, the authors did not examine the main effect of cognitive load on subjective mental fatigue or subjective task engagement. Analyses of behavioural and physiological outcomes revealed that the effects of time on task were qualified by an interaction with cognitive load. Specifically, accuracy on the 1-back and 2-back conditions decreased over time while accuracy on the 3-back did not change, suggesting the possibility of a learning effect masking the effect of mental fatigue. Likewise, ERP results indicated that the P300b amplitude in response to presentation of the stimulus (i.e., letter) decreased with time on task for the 1-back and 2- back, but not for the 3-back. Finally, eye tracking results demonstrated a decrease in baseline pupil diameter with time on task; however, they also found that increasing task difficulty was associated with larger baseline pupil diameter. The authors interpreted this phenomenon as in accordance with "exploration" rather than "exploitation" of the task at hand; however, it may also simply be an index of greater mental effort and task demands (Kahneman & Beatty, 1966). Moreover, they found an interaction effect that revealed that greater decreases in pupil diameter were found over time for conditions with higher baseline diameter (i.e., more difficult conditions), supporting decreasing levels of arousal and decreased engagement as task difficulty increased. Multilevel modeling indicated that the four measures of task engagement were significantly correlated with one another, supporting the hypothesis that changes in one measure of disengagement would be accompanied by changes in others. Moreover, these correlations strengthened as task difficulty increased. After the two hours of continuous performance, a motivation manipulation reversed the effects of subjective mental fatigue, resulting in significant increases in subjective, behavioural, and physiological engagement. In support of both the rewards-costs theory of mental fatigue proposed by Boksem and Tops (2008) and the role of fatigue as a stop-emotion, the authors interpreted these findings to indicate that task disengagement may serve to preserve resources for the possibility of more rewarding outcomes in the future (Hopstaken et al., 2015a).

In a separate study, Hopstaken et al. (2015b) sought to explore the third mode of NE release, the "disengage mode," by analyzing changes in both baseline and stimulus-evoked pupil diameter while participants completed seven blocks of 183 trials of a *2*-back task. Consistent with their hypotheses, stimulus-evoked pupil diameter decreased as fatigue increased with time on task, indicating disengagement from responding to the task (i.e., stimulus) at hand. Indeed, several studies have supported the use of stimulus-evoked pupil diameter as an index of cognitive load and mental effort (for a review, see van der Wel & van Steenbergen, 2018). However, contrary to their hypotheses, baseline pupil diameter did not change with time on task, which the authors note may have been a floor effect due to a less arousing environment than that of the previous study (Hopstaken et al., 2015a, 2015b). Again, after a reward manipulation, pupil dynamics (i.e., both baseline and stimulus-evoked pupil diameter) were restored to baseline sizes. Interestingly, multilevel analyses demonstrated that stimulus-evoked pupil diameter was significantly correlated with subjective and behavioural engagement, but associations involving baseline pupil diameter were not significant (Hopstaken et al., 2015b). In an elegant series of studies, Tsukuhara, Harrison, and Engle (2016) demonstrated that baseline pupil diameter is more closely related to working memory capacity and fluid intelligence, demonstrating stability in spite of changes in mental effort and familiarity with the testing environment (as well as age, ethnicity, and drug use). Due to conflicting evidence, the significance of baseline pupil diameter as a proxy of engagement and LC-NE function is less clear.

Aims and Hypotheses

A significant body of research has accrued investigating the role of time on task in the development of mental fatigue, with variation of cognitive load receiving less attention. Few studies have examined the effect of subjective appraisal of task demands on fatigue induction. Moreover, in light of the fact that fatigue has been proposed to be a stop-emotion to conserve cognitive resources, it is surprising that no published studies have included momentary assessments of subjectively appraised available cognitive resources, nor evaluated the relationship between subjective fatigue and subjective levels of available cognitive resources. The purpose of this study was to investigate the relationship between fatigue, cognitive resources, and disengagement by manipulating cognitive load and perceptions of task difficulty. Based on the research reviewed herein, it is clear that the construct of mental fatigue is inherently subjective; that is, the experience of a discrepancy between task demands and available cognitive resources is reliant on subjective cognitive appraisals and is not rooted in any physiological or behavioural substrates that would allow for concomitant measurement. Therefore, the behavioural and physiological manifestations of such mental fatigue (i.e., in the form of task disengagement) must be consequences of this evaluation of discrepancy.

The primary aims of the following proposed study were threefold. First, this study assessed if subjective assessments of available cognitive resources are inversely related to subjective assessments of fatigue. Second, this study tested if the effect of cognitive load on subjective fatigue is influenced by subjective appraisal of the task demands. Third, this study investigated if increases in subjective fatigue predict subjective, behavioural, and physiological disengagement from a cognitive task, as participants reduce their mental effort to conserve cognitive resources. Significant outcomes on these various measures would not only validate the use of behavioural and physiological measures in future research as indices of disengagement, but would also increase confidence in a significant effect of fatigue on disengagement by mitigating the effects of shared method variance by replicating such a finding using various methodologies.

Participants underwent four task conditions in which both cognitive load and appraisal of the difficulty of the task (and, consequently, the effort required) were manipulated to identify how appraisals of task difficulty interact with cognitive load to produce varying levels of subjective fatigue. A within-groups design was employed to control for individual differences in cognitive performance and the development of fatigue (Ackerman & Kanfer, 2009; Borragán et al., 2017; Kanfer, 2011). Moreover, due to the preponderance of research supporting the role of increasing motivation (e.g., provision of incentives) in reversing the experience of mental fatigue (e.g., Boksem et al., 2006; Gergelyfi et al., 2015; Hopstaken et al., 2015; Muraven & Slessareva, 2003) and toward the aim of maintaining a parsimonious study design and maximizing power, the following experiment did not include a motivational manipulation, though motivation was examined at the end of each testing block to assess if changes in motivation mirrored those in fatigue (see Methods). The order in which all four conditions were experienced was counterbalanced to mitigate the effect of dwindling levels of motivation on interpretations of the relationship between cognitive appraisals, mental fatigue, and task disengagement with increased time on task.

Based on the literature reviewed above, we hypothesized the following: 1) that subjective fatigue and subjective cognitive resources would be negatively correlated at all time points, such that available cognitive resources decrease as subjective fatigue increases; 2) following the four fatigue manipulation conditions, a) increases in fatigue ratings from pre- to post-test would be greatest in the high load-hard instruction condition and smallest in the low load-easy instruction condition and b) decreases in ratings of available cognitive resources would be greatest in the high load-hard instruction condition and smallest in the low load-easy instruction condition; and 3) that changes in fatigue ratings from pre- to post-test (reflecting the cumulative effect of both the task demands and participants' expectations of task difficulty) would significantly predict task disengagement across all modalities of measurement (i.e., subjective, behavioural, and physiological).

Method

Participants

This study included a convenience sample of 72 males and females between the ages of 18 and 79 years (*M*=20.21, *SD*=3.71, Range=17-41). Two outliers were detected with respect to age; however, omission of these two cases did not significantly alter any of the results, therefore the original results are reported below. The demographics and clinical characteristics of the study sample can be found in Table 1. An a priori power analysis suggested that 72 participants would be needed to have sufficient power to detect a medium effect of fatigue on subjective disengagement and a small effect on pupil diameter among undergraduate students, using a two-tailed test with an alpha value of .05 and 1-β of .95. Participants were recruited from the Ryerson Psychology Research Pool (SONA) and were offered two course credits for their participation.

In order to participate, participants were required have self-reported normal or corrected-to-normal vision and to abstain from alcohol use in the 12 hours prior to study participation and caffeine use 6 hours prior to participation based on their half-lives. Adherence to these requirements were verified by self-report. One participant reported using caffeine prior to participating in the study and was therefore excluded from the analyses that follow. There were no other inclusion/exclusion criteria. Time of day of testing was also assessed due to circadian fluctuations in arousal (Wright, Lowry, & LeBourgeois, 2012).

All study-related materials, including informed consent and re-consent during the debriefing process, were de-identified. Each student enrolled in an introductory psychology course (and, therefore, SONA) was provided with a six-digit identification code for study participation. Consequently, no identifying information was collected in this study. Only lab personnel involved in the current study had access to study data and materials, which were kept in the Sleep and Depression Lab at Ryerson University and will be stored for 10 years after completion of the present study, upon which they will be destroyed.

Table 1

Demographic and clinical characteristics of study participants

Demographic Variables	*n (%)*
Gender	
Male	11 (14.7)
Female	63 (84)
Other	1 (1.3)
Ethnicity	
Aboriginal Canadian	1 (1.3)
African Canadian	3 (4)
Caribbean Canadian	6 (8)
East/Southeast Asian Canadian	13 (17.3)
European Canadian	26 (34.7)
Latin/Central South American Canadian	2 (2.7)
South Asian Canadian	14 (18.7)
West Asian/Arab Canadian	6 (8)
Pacific Islander Canadian	1 (1.3)
Other	3 (4)
Clinical Variables	*M (SD)*
Insomnia Severity Index	9.86 (4.59)
DASS-21	
Depression	11.94 (10.18)
Anxiety	8.62 (6.87)
Stress	13.18 (7.26)
Fatigue Severity Scale	4.40 (0.98)
Composite Morningness Questionnaire	31.08 (5.94)
Epworth Sleepiness Scale	9.07 (4.17)

Measures

Malingering. The Rey 15-item memory test for malingering (Rey, 1964), also known as the Memorization of Fifteen Items Task (MFIT), was included to assess participants' overall effort and engagement on a task. The MFIT is framed as a difficult memory task; however, in reality, the fifteen items are arranged in three rows of five logical and conceptually related sets, making them very easy to remember (Iverson & Franzen, 1996). Participants were given the following instructions: "I am going to show you a slide with 15 *(emphasized)* things on it to remember. When I take the card away, I want you to write down as many of the 15 things as you can remember" (Goldberg & Miller, 1986, pp. 794). Participants were then shown the slide for 10 seconds; after 5 seconds, participants were given the additional prompt of "be sure to look at all of them" (Goldberg & Miller, 1986, pp. 794). To score the MFIT, one point is given for every correctly drawn item, regardless of its location, resulting in a maximum score of 15. Though previous researchers have proposed a cut-off score of 8 (Bernard & Fowler, 1990) or 9 (Goldberg & Miller, 1986; Lezak, 1983; Schretlen, Brandt, Krafft, & Van Gorp, 1991) or lower to detect malingering, a meta-analysis of studies assessing the sensitivity and specificity of the MFIT suggested using a cut-off score of 7, which yields excellent specificity to detect malingering (Reznek, 2005). Reznek (2005) argued that "the [ethical] cost of misidentifying a non-malingerer as a malingerer is greater than that of missing a malingerer" (pp. 542), therefore high specificity should be prioritized over high sensitivity. Suspected malingering, as indicated by poor performance on the MFIT, would confound results on the engagement measures detailed below. All participants scored above the cutoff score of 7 on the MFIT (M=14.70, SD=1.02, Range=9-15) suggesting that none of the participants were malingering.

Fatigue. The Fatigue Severity Scale (FSS) is a widely used 9-item self-report questionnaire that measures the impact of fatigue symptoms on daily activities over the previous week (Krupp, LaRocca, Muir-Nash, & Steinberg, 1989). It was included in this study as a baseline measure of habitual fatigue. Using a 7-point Likert scale, the FSS assesses participants' level of agreement/disagreement with a number of statements. Total scores range from 9 to 63 and then are subsequently divided by the number of items to provide an average item score. The scale has strong psychometric properties including excellent internal consistency, test-retest reliability, and validity in both clinical and non-clinical samples (Krupp et al., 1989; Lerdal, Wahl, Rustoen, Hanestad, & Moum, 2005; Valko, Bassetti, Bloch, Held, & Baumann, 2008). In the current study, the FSS demonstrated high internal consistency (Cronbach's α=.83). The FSS has been shown to distinguish between clinical and non-clinical samples; the clinical cut-off for the FSS is an average item score of 4, indicating a significant impact of habitual fatigue symptoms on activities of daily living (Krupp et al., 1989). As seen in Table 1, the average FSS score for this sample exceeded the clinical cut-off (M=4.40, SD=0.98), indicating that this sample experienced a significant amount of impairment in activities of daily living related to fatigue symptoms.

A visual analogue scale of fatigue (VAS-F) was completed at baseline, and before and after each block of the cognitive task to examine temporal changes in fatigue after manipulation of cognitive load and expectation of task difficulty. Participants indicated their level of fatigue along a 100-point slider scale, anchored by "not at all fatigued" and "extremely fatigued;" participants were not able to see the numeric value of their selected level. In general, visual analogue scales have been shown to be reliable and valid measures of subjective phenomena (Gift, 1989). In particular, the visual analogue scale for fatigue has demonstrated reliability, validity and sensitivity to change, making it an ideal measure for momentary assessment (Tseng, Gajewski, & Kluding, 2010).

Cognitive resources. According to the neurobiological model of fatigue (Chaudhuri & Behan, 2004), fatigue is experienced when there is a mismatch between available cognitive resources and task demand. A 100-point VAS assessing availability of cognitive resources (VAS-C) was also included, anchored by the terms “no cognitive resources at all” and “a high level of cognitive resources.”

Cognitive task. Participants completed four blocks of a visual letter *n*-back task, a measure of working memory and attentional control that requires high levels of task engagement (Watter, Geffen, & Geffen, 2001). This task presents a sequence of letters and the participant must press the *m* key if the letter matches the letter presented *n* trials ago, and the *n* key if it does not match the letter presented *n* trials ago; in the case of the 0-back, participants were instructed to press the *m* key if they saw the letter “X.” Cognitive load was manipulated by increasing the *n* from 0-back to 2-back. The 0-back requires sustained attention for stimulus detection but does not require working memory capacity, while the 2-back involves a working memory load as the stimulus must constantly be maintained and updated in working memory for appropriate response (Bailey, Channon, & Beaumont, 2007).

Cognitive engagement. Assessment of task engagement was threefold, including both subjective and objective measures.

Subjective engagement. Following the methods of Hopstaken and colleagues (2015a), subjective task engagement was measured using a 100-point VAS (VAS-E). Participants responded to the question “How engaged did you feel during that task?” The VAS was anchored by the terms “Not at all engaged” and “Very much engaged.”

Behavioural engagement. Overall behavioural engagement was assessed using accuracy of performance on the *n*-back task, as disengagement from the task presumably results in decreased accuracy. Accuracy was operationalized as the *d*-prime score (*d*'), a measure of sensitivity based on the signal detection theory, taking into account appropriate detection of the "signal" against the "noise." *d'* was calculated by subtracting the *z*-score transformation of the false alarm rate (proportion of incorrect responses on non-match trials) from the *z*-score transformation of the hit rate (proportion of correct responses on match trials; Hopstaken et al., 2015a). Higher *d'* scores indicate higher sensitivity. Secondary analyses also investigated reaction time of responses to ensure that there are no accuracy/speed trade-offs, as per Hopstaken et al. (2015a). Finally, although not included in analyses of disengagement in previous studies, missed trials (i.e., no key presses when stimulus is presented) may also be an index of interest that is not accounted for by *d*', as it stands to reason that consistent with the proposed "disengage mode" of norepinephrine functioning, disengagement from the task results in increased lapses of attention and, consequently, missed stimulus detection.

Objective engagement (pupil diameter). Pupil diameter was measured continuously throughout the *n*-back task as an objective assessment of task engagement using the EyeLink 1000 Plus eye tracker apparatus (SR Research Ltd., Kanata, Canada) in the Institute for Stress and Wellbeing Research. Pupillometry, or the measurement of pupil diameter, is sensitive to cognitive load and attentional performance (Kahneman & Beatty, 1966). Moreover, pupil diameter indirectly assesses locus coeruleus activity, which is implicated in arousal and task engagement (Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010; Jepma & Nieuwenhuis, 2011). Recent evidence indicates that the relationship between pupil diameter and attentional performance is nonlinear after controlling for time on task; that is, impaired attention is seen with decreases and increases in pupil diameter, and optimal attentional performance is in the intermediate range (van den Brink, Murphy, & Nieuwenhuis, 2016). However, although we might expect disengagement to be associated with both increases and decreases in pupil diameter, decreases in both baseline and stimulus-evoked pupil diameter (corresponding to the "disengagement mode" of LC-NE activity) are of primary interest in this study for its association with low arousal, of which the behavioural consequences are very similar to those of fatigue (Hopstaken et al., 2015a, 2015b).

Pupillometric data were collected using remote, monocular eye tracking at a sampling rate of 500 Hz using an LCD arm mount and a 16 mm lens for the first 15 participants. The right eye was tracked for the majority of participants (n=70); the left eye was tracked for two participants due to technical difficulties with tracking their non-dominant right eye. A target sticker was placed on the participants' foreheads to locate their eye and they were seated with an eye-to-screen distance of approximately 500-700 mm. However, variable distances from the screen between and within participants can influence the diameter of the pupil, precluding accurate measurements of change in pupil diameter across and within n-back tasks. Consequently, for the remaining 57 participants, head position and height were held constant using a chin rest located 55 cm from the desktop monitor, consistent with recommendations from the manufacturer of the eye tracker (SR Research Ltd., Kanata, Canada). The height of the chin rest and the height of the chair were adjustable to fit the height of each participant comfortably. A desktop mounted eye tracker using a 35 mm lens collected continuous pupillometric data throughout the four n-back blocks at a sampling rate of 1000 Hz. Figure 1 illustrates the desktop-mounted eye tracking apparatus. Regardless of eye tracking method, the eye tracker was calibrated and validated using a 5-point fixation routine for all participants at the beginning of each block and again prior to each *n*-back task.

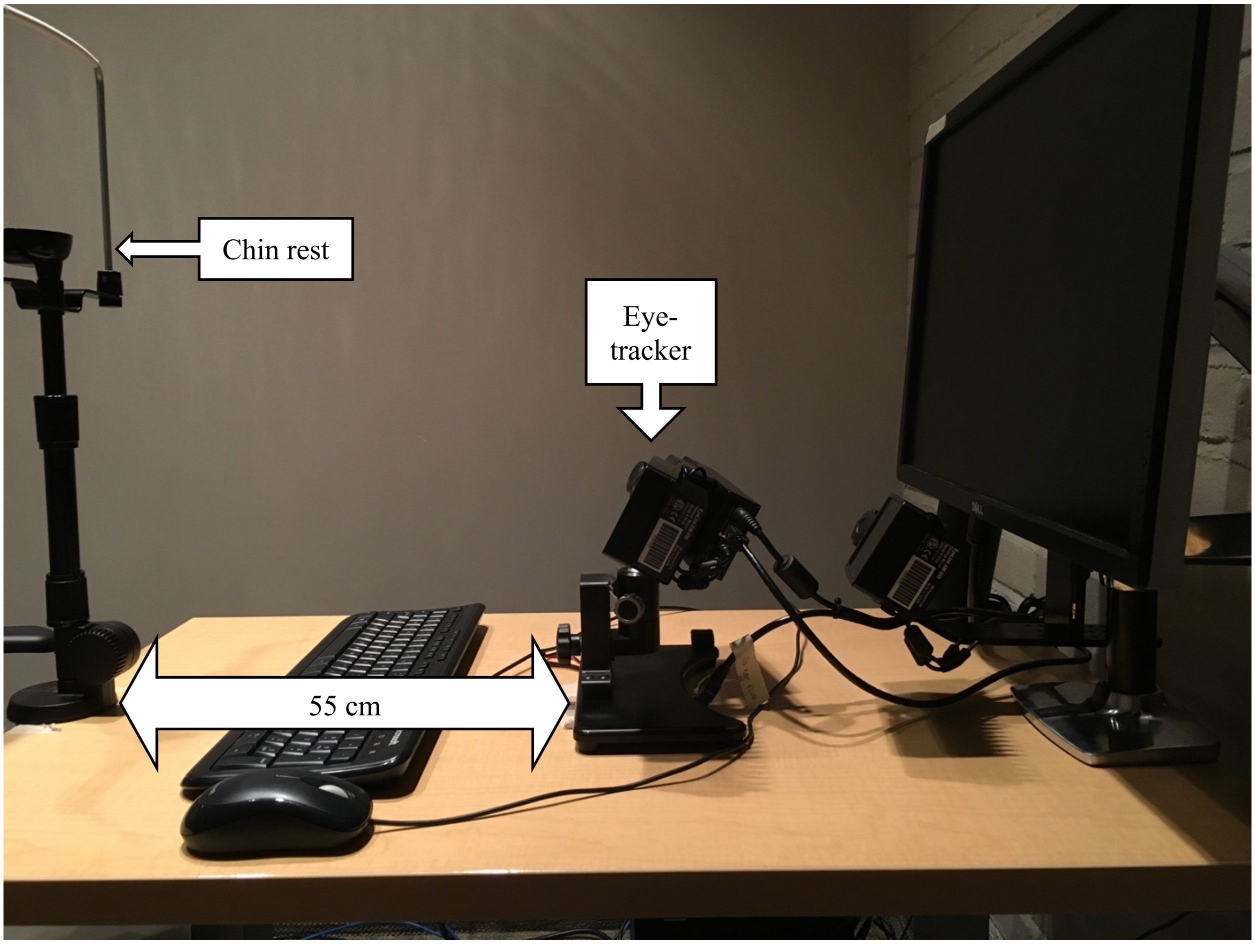

Figure 1. Desktop-mounted eye tracker.

Two interest periods were defined to measure two distinct pupillometric indices of engagement (Hopstaken et al., 2015a; Hopstaken et al., 2015b): pre-trial baseline pupil diameter refers to the baseline pupil size during the interstimulus interval and was defined as the 500 ms before stimulus onset, and stimulus-evoked pupil diameter refers to the maximum pupil diameter in the 1500 ms following stimulus onset. Average values for each of these interest periods were calculated for each for each participant and each block. Blinks were automatically marked and removed by a blink detection algorithm using Eye Link's software. An interest area was defined in the centre of the screen where stimuli were presented. Pupil diameter was assessed based solely on fixations that fell within the interest area in an effort to minimize the pupil foreshortening error which can bias estimates of pupil diameter when the eye is being tracked from non-0° angles (Hayes & Petrov, 2016).

Manipulation check. Following the completion of each block of the cognitive task, participants completed three additional Likert scale "exit items" assessing perceptions of task difficulty and motivation as a manipulation check of the fatigue induction. In the first item, participants were asked "How difficult did you think the N-back task was going to be **before** you started it? With 0 meaning not at all difficult and 10 meaning extremely difficult." In the second item, participants were asked "How difficult did you find the N-back task **while** you did it? With 0 meaning not at all difficult and 10 meaning extremely difficult." In the third item, participants were asked "How motivated were you to do well on this task? With 0 meaning not at all motivated and 10 meaning extremely motivated." Due to previous research demonstrating the role of motivation in reversing the experience of fatigue, inclusion of an assessment of intrinsic motivation to uphold performance allowed for the study of how increasing fatigue covaries with decreasing motivation. Additionally, findings of null effects of the fatigue manipulation may relate to the maintenance of high levels of motivation throughout the study procedure.

Covariates. The sample was also characterized based on perceived sleep, insomnia symptom severity, psychological symptoms, chronotype, and daytime sleepiness, as all of these factors may influence the experience of fatigue and/or impair performance on the *n*-back.

Perceived Sleep. Due to the effects of sleep deprivation on both fatigue and cognitive performance (Baranski, 2011), participants were asked to indicate the number of hours they estimated they had slept the previous night, as well as if this amount of sleep was normal for them (i.e., was it more or less than usual). On average, participants received 6.63 hours of sleep the previous night (SD=1.67; range: 3-11 hours). Thirty-six participants (48%) reported receiving a normal amount of sleep the previous night, while 29 (38.7%) reported receiving less sleep and 10 (13.3%) reported receiving more sleep than normal.

Insomnia Severity Index. The Insomnia Severity Index (ISI; Morin, 1996) is a 7-item self-report scale used to assess the severity of insomnia symptoms over the past month, including sleep onset latency, sleep maintenance, and early morning awakening, as well as dissatisfaction with sleep and the impact of sleep symptoms on daytime functioning. Items are rated on a Likert scale from 0 ("no sleep difficulty") to 4 ("severe sleep difficulty"). Total scores are calculated by summing all of the items and range from 0 to 28. A total score of 14 is the recommended cut-off score for distinguishing between good sleepers and individuals with insomnia disorder (Buysse, Ancoli-Israel, Edinger, Lichstein, & Morin, 2006). The ISI has been found to be a reliable and valid measure for detecting insomnia disorder in the general population (Morin, Belleville, Bélanger, & Ivers, 2011). Using the cut-off score of 14, 18.7% of the present sample (n=14) reported moderate-to-severe insomnia symptom severity. In the current study, the ISI had good internal consistency (Cronbach's α=.80).

Composite Morningness Questionnaire. Individual differences in morningness/eveningness (i.e., chronotype) may have an impact on fatigue levels and, consequently, disengagement, as participants were tested at various times of day and it is possible that their selected timeslot was inconsistent with their preferred time of day for strenuous mental activity. The Composite Morningness Questionnaire (CMQ; Smith, Reilly, & Midkiff, 1989) is a 13-item self-report scale that includes items evaluating preferred rise and sleep times, ideal work schedule, levels of fatigue, alertness, and appetite after rising, reliance on an alarm clock, and so on. The items were selected based on factor analysis of the Morningness-Eveningness Questionnaire (Horne & Östberg, 1976) and a diurnal type scale (Torsvall & Akerstedt, 1980). Items are scored using a Likert scale from 1 to 4 or 5. Scores range from 13 to 55, with higher scores reflecting greater morningness preference. Cut-offs for morning type (44-55), intermediate type (23-43), and evening type (13-22) have been established and correspond to the 90th percentile and above, the 11th to 89th percentile, and the 10th percentile and below on total score distributions, respectively (Smith, Reilly, & Midkiff, 1989). The validation study demonstrated high internal consistency (α=.87) and convergent validity with several external criteria, including rise times and bedtimes on weekdays and weekends, time-of-day class schedule preference, perceived morning class performance, time of feeling most alert, and time of feeling best physically (Smith, Reilly, & Midkiff, 1989). Likewise, in the current study the CMQ demonstrated good internal consistency (α=.82). Using the recommended cut-off scores, the majority of the sample (*n*=65, 86.7%) reported an intermediate chronotype, with eight

(10.7%) participants reporting an evening chronotype and two (2.7%) participants reporting a morning chronotype.

Psychological symptoms. The 21-item short form of the Depression Anxiety Stress Scales (DASS-21) is an abridged version of the 42-item DASS, a self-report questionnaire that assesses symptoms of depression, anxiety, and stress over the past week (Lovibond & Lovibond, 1995). Participants indicate how much each item applies to them on a 4-point scale from 0 ("did not apply to me") to 3 ("applied to me very much, or most of the time"). The DASS-21 produces separate scale scores for the depression, anxiety, and stress scales by summing the relevant items. Scale scores are then doubled for comparison with the full-length DASS. Each scale of the DASS-21 has demonstrated adequate reliability and validity for distinguishing between symptoms of depression, anxiety, and stress in both clinical and community samples, as well as equivalent psychometric properties to the full length version (Antony, Bieling, Cox, Enns, & Swinson, 1998; Henry & Crawford, 2005). Reliability analyses using the present sample indicate excellent internal consistency for the Depression scale (α=.91) as well as the measure as a whole (α=.90); internal consistency was lower for the Anxiety (α=.67) and Stress (α=.73) scales. Responses on the DASS-21 in the current sample indicated that participants scored in the mild range for depression and anxiety symptoms, while stress symptoms were in the normal range.

Daytime Sleepiness. Daytime sleepiness was assessed using the Epworth Sleepiness Scale (ESS), a subjective measure that includes eight items for which participants indicate the likelihood of dozing in different situations on a scale of 0 ("would never doze off") to 3 ("high chance of dozing") (Johns, 1991). Item responses are summed to yield a total score, with scores $\geq$ 10 indicating excessive daytime sleepiness (Johns, 1991). Furthermore, the ESS is capable of reliably distinguishing healthy controls from individuals with a number of sleep-wake disorders

(Johns, 1991). In the current sample, the ESS demonstrated acceptable internal consistency (α=.74). Evidence indicates that excessive daytime sleepiness is associated with poorer cognitive function (Durmer & Dinges, 2009). Therefore, daytime sleepiness was assessed for potential inclusion as a covariate in statistical analyses. Participants in the current sample scored just below the clinical cut-off (M=9.07, SD=4.17), indicating that the sample was excessively sleepy.

Additionally, a 100-point VAS assessing sleepiness (VAS-S), anchored by "not at all sleepy" and "extremely sleepy," was also included at baseline and before and after each block of the *n*-back task to rule out a competing hypothesis that increases in sleepiness are responsible for detriments in performance (i.e., the underload hypothesis; Pattyn, Neyt, Hendrickx, & Soetens, 2008).

Boredom. Likewise, a 100-point VAS assessing boredom (VAS-B), anchored by "not at all bored" and "extremely bored," was included to rule out the underload hypothesis (Pattyn, Neyt, Hendrickx, & Soetens, 2008).

Procedure

Participants were recruited from SONA to participate in a study about mental fatigue and task engagement. Testing took place in the eye tracking suite located in the Institute for Stress and Wellbeing Research at Ryerson University. After providing their informed consent, participants completed a battery of questionnaires using Qualtrics, including demographic questions, the ISI, FSS, DASS-21, ESS, and CMQ. Using a slider, participants indicated their baseline levels of fatigue (VAS-F), cognitive resources (VAS-C), boredom (VAS-B) and sleepiness (VAS-S) along 100-point visual analogue scales. After completing the questionnaires, participants completed the MFIT.

Participants were then set up at the eye tracker, where they completed four testing blocks; a schematic of the procedure for each block can be found in Figure 2. Each block began with calibration of the eye tracker, followed by completion of the pre-task VAS-F, VAS-C, VAS-B and VAS-S.

Each block of the *n*-back included 15 practice trials using numbers as stimuli, followed by 60 trials using letters. The purpose of the 15-trial practice *n*-back was to familiarize participants with the task. Both practice and testing trials had a target rate of 25%. Four separate number and letter lists were generated such that each block included a unique list of numbers and letters. Numbers included 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 and letters included: A, B, C, D, E, H, I, K, L, M, O, P, R, S, T, and X (all capital). Stimuli were presented in the centre of the screen using 40-point Arial font. The font colour was randomized within low- and high-load conditions, respectively, such that one of the conditions within each set was presented in black and the other in burgundy. The luminance of the screen (in lux units), as measured using a photometer app on a mobile device, did not differ between the black and burgundy conditions; as such, it is assumed that the change in font colour did not have an effect on pupil diameter. The background colour was held constant across each block; it was set at a grey tone (RGB code: 155, 155, 155, 235) in an attempt to mitigate eye fatigue. Each stimulus was presented for 500 milliseconds in the centre of the screen, followed by a randomized interstimulus interval of 5 to 5.5 seconds. This interstimulus interval allows pupil diameter to return to baseline (Beatty, 1982). Ambient lighting of the eye tracker suite was consistent within and across all participants.

Following completion of the practice *n*-back task, participants received a verbal "instruction" as to the difficulty of the following task to manipulate the expectation of low or high required cognitive resources (i.e., task demand). In conditions where participants were to expect low effort demands, they were told: “most people find the task that you are about to do is *really* easy; they do not have to try very hard and they do not have to put much effort in at all, they just need to pay attention.” Likewise, when participants were to expect high effort demands, they were told: “most people find the task that you are about to do is *really* difficult. They find it really challenging and they have to put a lot of effort in to do well, so pay attention and just *try* to do your best” (italics added for emphasis). On half of the blocks, participants were told the task would be easy, and on the other half that it would be difficult. The eye tracker was then calibrated and validated again using a 5-point fixation routine. Participants then completed 60 trials of the target *n*-back. Half of the blocks were low cognitive load conditions (0-back) and half were high cognitive load conditions (2-back). This resulted in four within-subject conditions: low load-easy expectation (A; “real easy”), low load-hard expectation (B; “fake hard”), high load-easy expectation (C; “fake easy”), and high load-hard expectation (D; “real hard”).

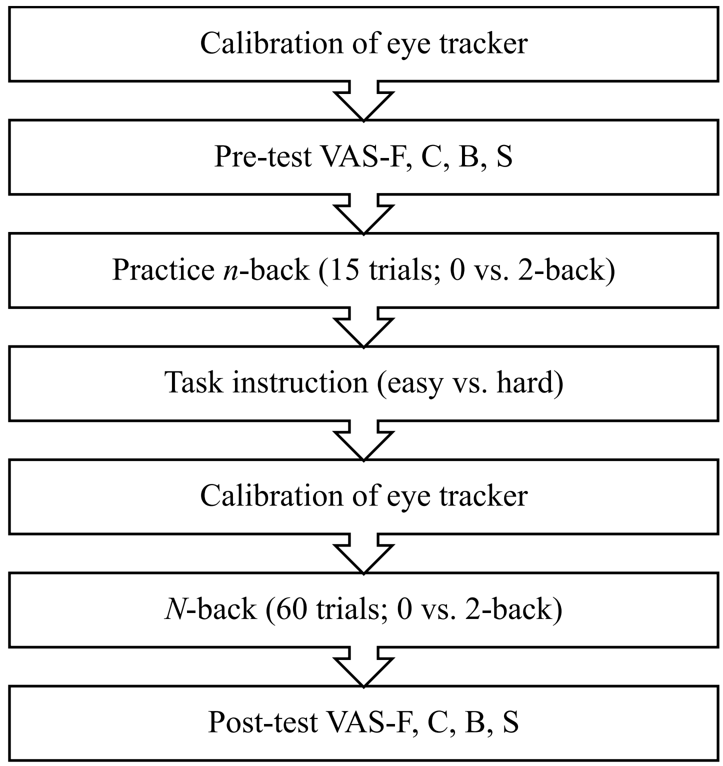

Figure 2. Schematic of testing block procedure. This procedure was repeated four times, corresponding to the four within-subject fatigue manipulation conditions. Cognitive load of the practice n-back was congruent with the cognitive load of that condition.

The order of the presentation of the blocks (load × instruction) was randomized using a reduced Latin square to minimize order effects and to control for time on task effects; 17 (23%) participants were randomized to receive the order ABCD, 17 (23%) received the order BCDA, 20 (27%) received the order CDAB and 20 (27%) received the order DABC. As previously mentioned, the font colour of the stimuli was randomized within low- and high-load conditions, such that one low-load and one high-load condition was presented in black and the others were presented in burgundy. The change in font colour was a distractor cue so participants would not be suspicious that they were doing an identical task to one they had already done but were being told it was easier or harder.

When participants finished the 60-trial *n*-back, they again completed the VAS-F, VAS-C, VAS-S and VAS-B, as well as the VAS-E to indicate their subjective engagement in the task. Participants then completed the post-test questions assessing perceptions of difficulty before and during each block, as well as motivation to do well. Between blocks, participants were given the opportunity to rest for up to 3 minutes to avoid carry-over effects of fatigue onto the next block.

Participation in this study protocol took between 105 and 120 minutes from beginning to completion. Upon completion of study protocol, participants were debriefed on the nature of the study, explaining that it is a study exploring the roles of task difficulty and perceptions of task difficulty in developing fatigue and the downstream effects of fatigue on task engagement. Debriefing also involved revealing to participants the use of deception in order to investigate these study questions (i.e., the use of incongruent verbal instructions to manipulate participants' beliefs about the difficulty of the cognitive tasks they were about to perform). Participants received a debriefing letter and were given the opportunity to ask any questions. Participants then indicated their written re-consent to the use of their data in light of the information that deception was used. Finally, participants received two course credits for their time.

Statistical Analyses

All analyses were conducted with IBM SPSS Statistics version 23 for Mac.

Data preparation. Several steps were taken to prepare the data for statistical analysis. To investigate the effects of both instruction and cognitive load on subjective fatigue, the change in fatigue (Δfatigue) over each block was calculated by subtracting pre-test VAS-F from post-test VAS-F. Thus, a higher Δfatigue score indicates a greater increase in fatigue over the block. Likewise, the change in available cognitive resources over each block was calculated by subtracting pre-VAS-C from post-VAS-C (Δcognitive resources). A higher Δcognitive resources score indicates a greater decrease in cognitive resources over the block.

As detailed in the Measures section, behavioural engagement in the *n*-back task was operationalized as the *d*' score, a measure of accuracy. The hit rate and false alarm rate on each block of the *n*-back were calculated and subsequently transformed into *z*-scores. To calculate the *d*' score, *z*-false alarm rate was subtracted from *z*-hit rate. The miss rate was calculated by dividing the number of missed *m*-key presses by the total number of appropriate *m*-key presses ("hits").

Primary analyses.

Association between subjective fatigue and cognitive resources. Pearson product moment correlations were conducted between pre- and post-test VAS-F and VAS-C for each block, resulting in eight correlations. *P*-values were adjusted using Bonferroni correction for multiple comparisons.

Fatigue manipulation. A 4 (condition) × 2 (time) × 4 (randomization order) mixed analysis of variance (ANOVA) was conducted with VAS-F scores as the dependent variable, including condition and time as within-subjects variables and randomization order as a between-subjects variable. A similar analysis including condition, time, and randomization order as independent variables was carried out with VAS-C scores as the dependent variable. Additionally, to examine if the manipulation of fatigue was effective, three factorial ANOVAs including condition (4 levels) and randomization order (4 levels) as independent variables compared scores on exit item 1 (perceived difficulty before the task), exit item 2 (perceived difficulty during the task) and exit item 3 (motivation to do well), respectively. Significant interactions were followed up with an analysis of multivariate and univariate simple effects, where applicable. Consistent with the recommendations of Barcikowski and Robey (1984), in cases where the assumption of sphericity was violated, degrees of freedom were corrected using Hyunh-Feldt estimates when $\varepsilon \geq .75$ and Greenhouse-Geisser estimates when $\varepsilon < .75$.

Association between fatigue and disengagement. Hierarchical linear regression analyses were used to investigate the effect of Δfatigue on subjective engagement (VAS-E), behavioural engagement (*d*'), and physiological engagement (baseline and stimulus-evoked pupil diameter), respectively, while controlling for the effect of baseline fatigue (VAS-F); baseline VAS-F was entered in the first block, followed by Δfatigue. Separate analyses were carried out for each condition, resulting in 16 regression analyses.

Pearson product moment correlations assessed if habitual fatigue and the proposed covariates (i.e., insomnia symptom severity, morningness-eveningness score, psychological symptoms, daytime sleepiness, momentary boredom, momentary sleepiness) were significantly related to VAS-E, *d*', and baseline and stimulus-evoked pupil diameter in any of the blocks. In the event of a significant correlation between one of these indices and a measure of engagement, the regression analysis was repeated to include the significant covariates in the third block, after baseline VAS-F in Δfatigue.

Finally, Pearson product moment correlations were conducted to determine the associations between VAS-E, *d*', baseline pupil diameter, and stimulus-evoked pupil diameter for each block in order to validate the use of multiple indices of disengagement.

Secondary analyses. To examine the possibility of accuracy/speed trade-offs on the *n*-back, the aforementioned regression analyses were repeated with reaction time and miss rate as the dependent variables.

Post hoc analyses. To examine patterns of engagement across the four conditions, 4 (condition) × 4 (randomization order) mixed model ANOVAs were conducted including each engagement variable as dependent variables (VAS-E, *d'*, baseline pupil diameter, and stimulus-evoked pupil diameter). Finally, to empirically test the underload hypothesis, two 4 (condition) × 2 (time) × 4 (randomization order) mixed model ANOVAs were conducted using VAS-B and VAS-S as dependent variables.

Results

Preliminary Analyses

All variables of interest were analysed graphically and using statistics for deviations from normality. The effect of time of day of testing on several dependent variables of interest (i.e., baseline VAS's for fatigue, cognitive resources, boredom, sleepiness; subjective, behavioural, and physiological indices of engagement across all conditions) was assessed using one-way ANOVAs, including time of day as a categorical independent variable with four levels (morning, early afternoon, afternoon, evening). Time of day was not significantly associated with any outcome variables, all ps>.05.

Primary Analyses

Association between fatigue and cognitive resources. Using a Bonferroni correction for the nine correlation analyses, type I error rate was adjusted to α=.005. As hypothesized, VAS-F and VAS-C were correlated at baseline, $r(74)=-.326$, $p=.004$, and at pre- and post-test for all four conditions. Table 2 displays the correlation coefficients for the relationship between VAS-F and VAS-C for each condition and each time point (i.e., pre or post-test).

Table 2

Pearson product moment correlations of VAS-F and VAS-C at pre-test and post-test

	Pearson Product Moment Correlation			
	Real Easy (A)	Fake Hard (B)	Fake Easy (C)	Real Hard (D)
Pre	-.56*	-.47*	-.50*	-.62*
Post	-.48*	-.49*	-.48*	-.62*

Note. *p<.001.

Fatigue manipulation.

Change in fatigue. To investigate the cumulative effect of the fatigue manipulation (including both the instruction of effort required and the cognitive load of the *n*-back) on momentary ratings of fatigue, a 4 (condition; within-subjects) × 2 (time; within-subjects) × 4 (randomization order; between-subjects) mixed ANOVA was conducted with VAS-F as the dependent variable. Mauchly's test indicated that the assumption of sphericity was met for the main effect of condition, $\chi^2(5)=8.41$, $p=.135$, but it was violated for the condition × time interaction, $\chi^2(5)=18.20$, $p=.003$, therefore degrees of freedom were corrected using Huynh-Feldt estimates ($\varepsilon=.93$). Mean ratings of VAS-F at pre- and post-test for each level combination of condition and randomization order can be found in Table 3 and statistics for the main and interaction effects can be found in Table 4. The main effect of time was significant while the main effects of condition and randomization order were not statistically significant. The interaction between condition and time was not statistically significant. The interaction between condition and randomization order was significant, indicating that the effect of the fatigue manipulation had differential effects on fatigue ratings (averaged across time) according to the order in which the conditions were experienced (i.e., order effects). To follow up this interaction, the multivariate simple effects of condition at each level of randomization order were examined, and significant multivariate effects were followed up with pairwise comparisons. At each level of randomization order, there was a significant effect of condition on ratings of fatigue, $p\text{s}\leq.001$.

Table 3.

Mean VAS-F ratings at pre- and post-test within each level combination of condition and randomization order.

	Mean VAS-F Ratings (SE)							
	ABCD		BCDA		CDAB		DABC	
Condition	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Real Easy (A)	53.19 (23.93)	70.38 (21.86)	75.18 (23.78)	82.76 (22.78)	64.32 (24.40)	76.63 (23.14)	50.11 (26.49)	66.47 (25.28)
Fake Hard (B)	60.06 (16.82)	73.94 (26.16)	47.82 (23.38)	62.12 (22.25)	66.79 (27.50)	73.47 (25.05)	56.26 (26.61)	74.42 (18.40)
Fake Easy (C)	71.44 (20.52)	79.87 (17.48)	58.24 (21.66)	72.24 (20.91)	50.42 (19.05)	68.58 (18.47)	63.32 (27.36)	69.74 (25.04)
Real Hard (D)	75.69 (19.36)	85.63 (15.14)	65.71 (21.225)	80.29 (24.15)	56.47 (22.56)	75.74 (23.82)	36.42 (24.52)	54.11 (24.79)

Table 4

Three-way mixed analysis of variance assessing the effects of condition, time, and randomization order on momentary ratings of fatigue

Variable	F	p	η_p^2
Main effects			
Condition	1.41	.240	.02
Time	94.44	<.001**	.59
Randomization order	1.27	.290	.05
Interactions			
Condition*Time	.74	.523	.01
Condition*Randomization order	20.53	<.001**	.48
Time*Randomization order	.16	.920	.01
Condition*Time*Randomization order	2.35	.018*	.10

Note. * p<.05, ** p<.01, *** p<.001.

Finally, these results were qualified by a significant three-way interaction between condition, time and randomization order. This interaction was followed up using multivariate simple effects of time at each level of condition, within each level of randomization order (see Table 5).

Table 5

Multivariate simple effects to explore the effect of the interaction between condition, time and randomization order on momentary ratings of fatigue

Variable	*V*	*F*	*p*
ABCD			
Real Easy (A)	.20	17.21	<.001
Fake Hard (B)	.11	8.52	.005
Fake Easy (C)	.06	4.56	.036
Real Hard (D)	.08	5.49	.022
BCDA			
Real Easy (A)	.05	3.56	.063
Fake Hard (B)	.13	9.61	.003
Fake Easy (C)	.17	13.35	.001
Real Hard (D)	.16	12.56	.001
CDAB			
Real Easy (A)	.14	10.49	.002
Fake Hard (B)	.03	2.35	.130
Fake Easy (C)	.27	25.10	<.001
Real Hard (D)	.27	24.48	<.001
DABC			
Real Easy (A)	.22	18.54	<.001
Fake Hard (B)	.21	17.33	<.001
Fake Easy (C)	.04	3.14	.081
Real Hard (D)	.24	20.63	<.001

Note. V=value of Pillai's trace. Multivariate simple effects examine the effect of time (change from pre- to post-test) on ratings of fatigue within each level combination of condition and randomization order.

Figure 3 demonstrates the estimated marginal means of VAS-F from pre- to post-test for each condition separately for each randomization order group. This figure demonstrates an increase in fatigue ratings at pre-test with each additional block of the *n*-back, and different magnitudes of the increase in fatigue from pre-to post-test across the conditions. Within ABCD (Figure 3.I), fatigue increased significantly from pre- to post-test in all four within-subjects conditions (all ps<.05). In the remaining randomization order groups, Figure 3 sections II, III and IV demonstrate that fatigue ratings did not increase significantly from pre- to post-test in the fourth block (condition A for BCDA, condition B for CDAB, and condition C for DABC). **Summary statement: contrary to the hypothesis that increases in fatigue would be highest in condition D ("real hard") and lowest in condition A ("real easy"), fatigue ratings increased with time-on-task, showing significant increases with each additional block until—in most cases—reaching a plateau in the final block.**

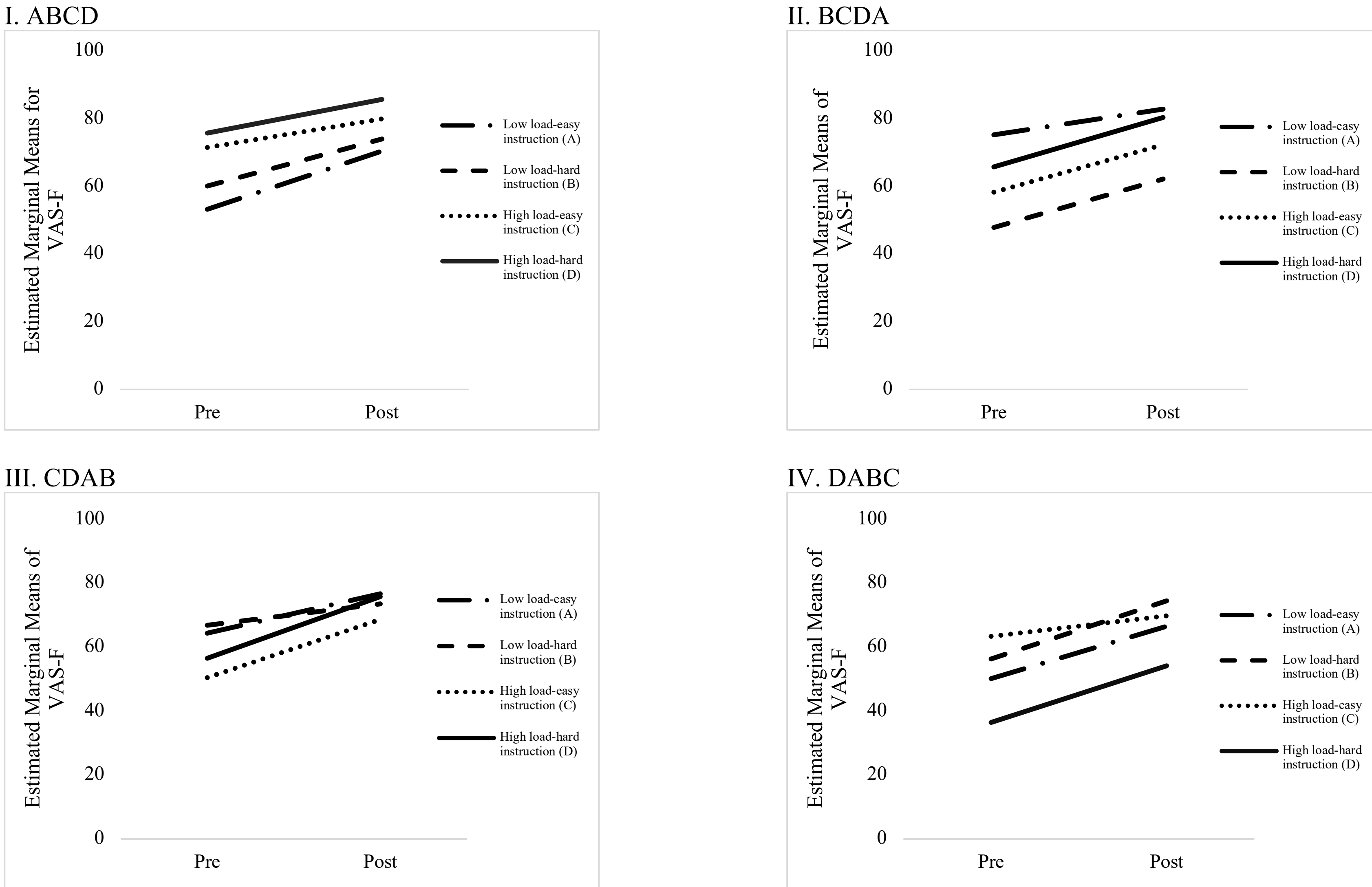

Figure 3. Estimated marginal means of VAS-F at pre- and post-test across the four conditions displayed separately for each randomization order. I) ABCD; II) BCDA; III) CDAB; IV) DABC.

Change in cognitive resources. To investigate the cumulative effect of the fatigue manipulation on momentary ratings of cognitive resources, a 4 (condition; within-subjects) × 2 (time; within-subjects) × 4 (randomization order; between-subjects) mixed ANOVA was run with VAS-C as the dependent variable. Mauchly's test indicated that the assumption of sphericity was violated for the main effect of condition, $\chi^2(5)=22.56$, $p<.001$, and the condition × time interaction, $\chi^2(5)=24.78$, $p<.001$, therefore degrees of freedom were corrected using Huynh-Feldt estimates (ε=.92 and ε=.85, respectively). Table 6 displays the means for VAS-C at pre- and post-test for each level combination of condition and randomization order. Once again, there was a main effect of time (see Table 7 for statistics). The main effects of condition and randomization order were not significant, *p*s>.05. The interaction between condition and randomization order was significant ($p<.001$). This interaction was followed up with multivariate simple effects of condition within each level of randomization order, and significant multivariate effects were followed up with pairwise comparisons. All four multivariate simple effects were significant (*p*s≤.005), indicating that there was a significant effect of condition on ratings of available cognitive resources at each level of randomization order.

Table 6

Mean VAS-C ratings at pre- and post-test within each level combination of condition and randomization order.

	M (SE)							
	ABCD		BCDA		CDAB		DABC	
Condition	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Real Easy (A)	53.06 (26.73)	41.94 (29.31)	39.06 (24.27)	31.12 (28.51)	37.26 (17.59)	34.89 (24.66)	60.58 (22.39)	49.32 (26.23)
Fake Hard (B)	46.63 (25.51)	39.25 (29.30)	54.00 (19.76)	46.88 (18.35)	35.89 (21.01)	34.47 (21.35)	56.47 (24.26)	44.58 (22.86)
Fake Easy (C)	39.63 (29.70)	31.31 (28.46)	53.47 (16.63)	41.41 (22.21)	51.63 (22.21)	43.68 (20.82)	46.05 (25.44)	40.68 (24.10)
Real Hard (D)	35.94 (29.63)	32.47 (28.55)	46.65 (17.91)	40.85 (21.70)	43.26 (22.11)	39.92 (17.69)	68.05 (17.15)	61.58 (18.47)

Table 7

Three-way mixed analysis of variance assessing the effects of condition, time, and randomization order on momentary ratings of available cognitive resources

Variable	F	p	η_p^2
Main effects			
Condition	1.16	.325	.02
Time	39.12	<.001	.37
Randomization order	1.71	.172	.07
Interactions			
Condition*Time	1.34	.265	.02
Condition*Randomization order	11.61	<.001	.34
Time*Randomization order	1.06	.373	.05
Condition*Time*Randomization order	1.06	.390	.05

Figure 4 displays the means of VAS-C in each condition within each level of randomization order.

Within the ABCD group, average ratings of cognitive resources were higher in condition A (M=47.50, SE=5.85) compared to conditions C (M=35.47, SE=5.69; p=.002) and D (M=34.20, SE=5.35; p<.001). Likewise, average ratings of cognitive resources were higher in condition B (M=42.94, SE=5.44) compared to conditions C (p=.037) and D (p=.007).

In the BCDA group, ratings of cognitive resources were higher in condition B (M=50.44, SE=5.28) compared to conditions D (M=43.75, SE=5.19; p=.010) and A (M=35.09, SE=5.68; p<.001), and ratings were higher in condition C (M=47.44, SE=5.52) compared to conditions D (p=.011) and A (p=.001).

In the CDAB group, cognitive resources were rated significantly higher in condition C (M=47.65, SE=5.22) compared to conditions D (M=41.59, SE=4.91; p=.001), A (M=36.08, SE=5.68; p=.001), and B (M=35.18, SE=5.00; p<.001).

Finally, in the DABC group, ratings of cognitive fatigue were higher in condition D (M=64.82, SE=4.91) compared to conditions A (M=54.95, SE=5.37; p=.027), B (M=50.53, SE=5.00; p=.002) and C (M=43.37, SE=5.22; p<.001). Moreover, ratings of cognitive resources in conditions A and B were also significantly higher than C (p=.001 and p=.029, respectively) but were not significantly different from each other (p>.05).

Summary statement: averaged across time, momentary ratings of available cognitive resources decreased incrementally from the first block to the fourth block in each randomization order group. This was contrary to the hypothesis that decreases in ratings of cognitive resources would be highest in condition D (“real hard”) and lowest in condition A (“real easy”).

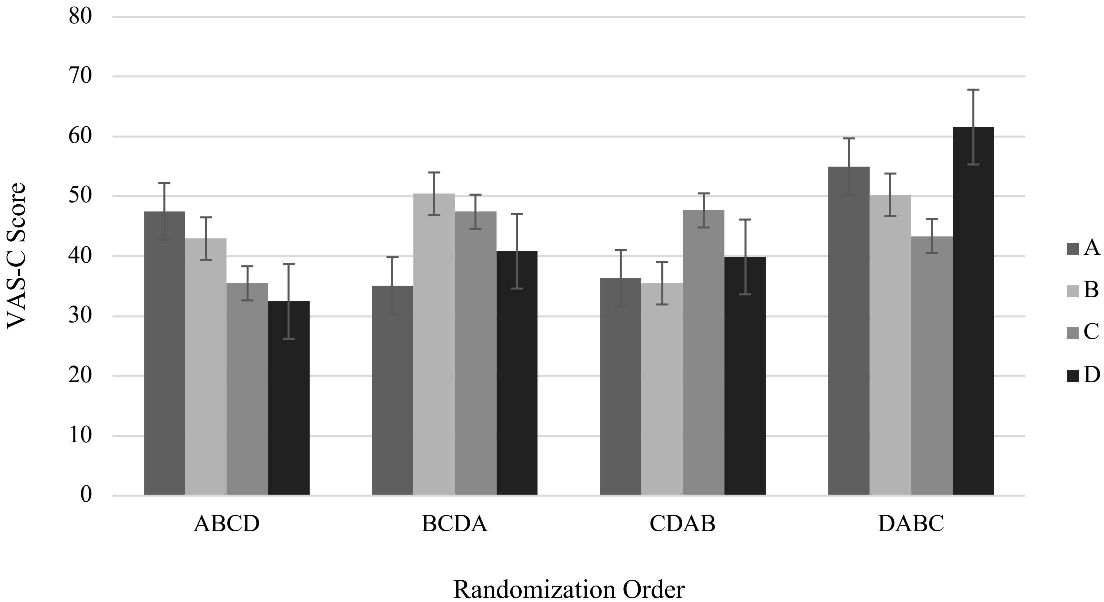

Figure 4. Examining the interaction between condition and randomization order on VAS-C. Mean VAS-C scores are displayed for each condition within each randomization order group.

Appraised difficulty before the n-back. A two-way (condition × randomization order) mixed ANOVA was carried out to investigate if the verbal instruction was successful in manipulating participants' perceptions of task difficulty and the effort required to meet the demands of the task, using the first exit item ("How difficult did you think the N-back task was going to be **before** you started it?") as the dependent variable. Mauchly's test was not significant, $\chi^2(5)=10.33$, $p=.067$, indicating that the assumption of sphericity was met. The main effect of condition was significant, $F(3,186)=20.99$, $p<.001$, $\eta_p^2=.25$ (Figure 5). Follow-up tests using pairwise comparisons demonstrated that condition A ($M=2.50$, $SE=.25$) was perceived to be significantly less difficult than conditions B ($M=3.41$, $SE=.32$; $p=.035$), C ($M=4.21$, $SE=.31$; $p<.001$), and D ($M=5.12$, $SE=.32$; $p<.001$); and condition B was perceived to be significantly less difficult than condition D ($p<.001$). The differences between conditions B and C, and C and D were not significant ($p=.102$ and $p=.904$, respectively). The main effect of randomization order was not significant, $F(3,62)=.80$, $p=.501$, $\eta_p^2=.04$, nor was the condition × randomization order interaction, $F(9,186)=.40$, $p=.932$, $\eta_p^2=.02$.

Summary statement: conditions receiving the same task instruction (i.e., "real easy" and "fake easy," "real hard and "fake hard") were not rated as equally difficult *before* starting the task.

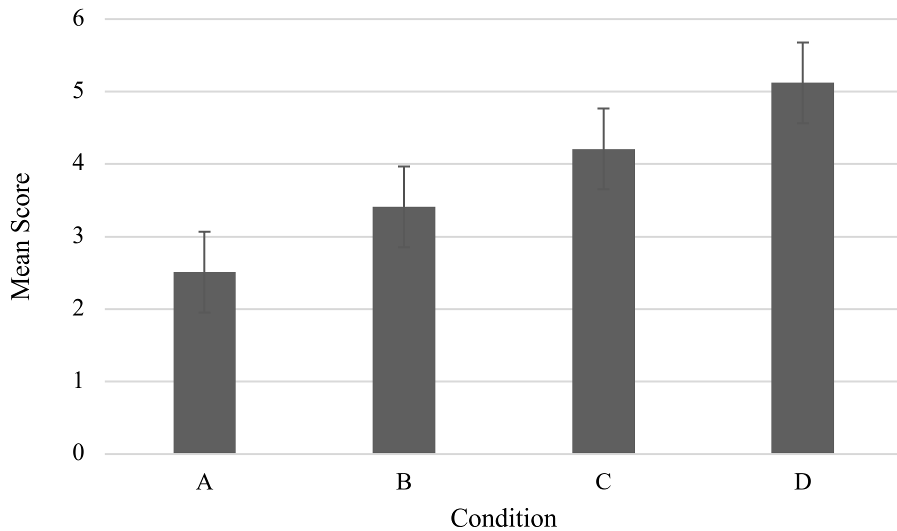

Figure 5. Mean item scores for exit item 1 (appraised difficulty before doing the *n*-back) across each condition.

Appraised difficulty during the n-back. Similarly, a two-way (condition × randomization order) mixed model ANOVA was carried out to investigate if the verbal instruction was successful in manipulating participants' perceptions of task difficulty and the effort required to meet the demands of the task, using the second exit item ("How difficult did you think the N-back task was going to be **while** you did it?") as the dependent variable. Mauchly's test was significant indicating that the assumption of sphericity was not met, $\chi^2(5)=33.64$, $p<.001$, therefore Greenhouse-Geisser corrected tests are reported ($\varepsilon=.73$). There was a significant main effect of condition, $F(2.20,138.39)=33.85$, $p<.001$, $\eta_p^2=.35$, which was followed up using pairwise comparisons (Figure 6). Ratings of task difficulty were not significantly different between conditions with the same cognitive load (A *vs.* B: $p=1.00$; C *vs.* D: $p=.859$), but *0*-back conditions (i.e., A and B) were both rated as significantly less difficult relative to *2*-back conditions (i.e., C and D), all ps<.001.

Summary statement: despite receiving different task instructions, tasks of the same cognitive load were rated as equally difficult *during* the task (i.e., "real easy" and "fake hard," "fake easy" and "real hard").

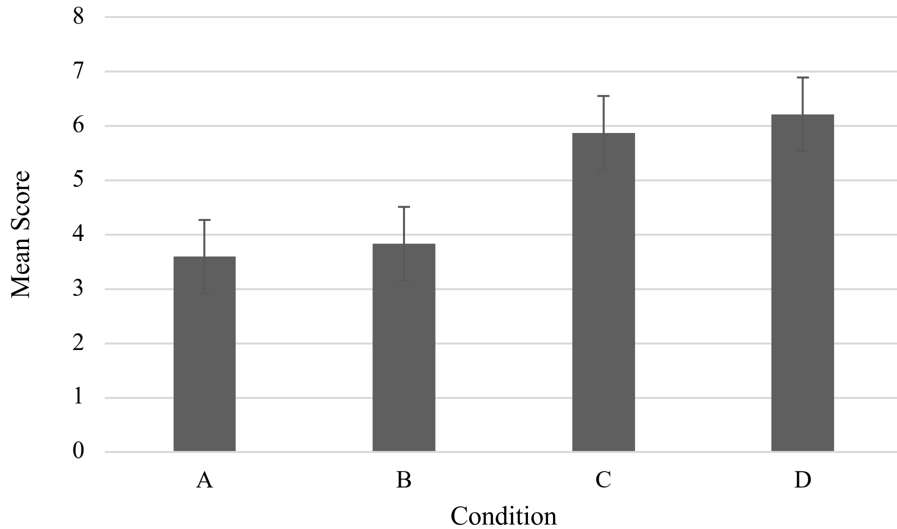

Figure 6. Mean scores for exit item 2 (appraised difficulty during the n-back) across each condition.

Motivation. Another two-way (condition × randomization order) mixed model ANOVA was carried out to investigate if motivation decreased along with fatigue, using the third exit item ("How motivated were you to do well on this task?") as the dependent variable. Mauchly's test was significant, therefore the assumption of sphericity was not met, $\chi^2(5)=18.41$, $p=.002$; degrees of freedom were corrected using Huynh-Feldt estimates ($\varepsilon=.90$). The main effect of condition was not significant, $F(2.71,170.51)=.70$, $p=.552$, $\eta_p^2=.01$, but there was a main effect of randomization order, $F(3,63)=2.83$, $p=.046$, $\eta_p^2=.12$. There was a significant condition × randomization order interaction, $F(8.12,170.51)= 12.21$, $p<.001$, $\eta_p^2=.36$, suggesting that the average ratings of motivation in each condition differed based on the order in which the conditions were experienced (Figure 7). To understand this interaction, univariate tests of the effect of randomization order within each level of condition were conducted (Table 8), and significant omnibus tests were followed up with pairwise comparisons.

Table 8

Univariate simple effects to break down the effect of the interaction between condition and randomization order on ratings of motivation to do well

Variable	F	p	η_p^2
Real Easy (A)	5.34	.002	.20
Fake Hard (B)	7.09	<.001	.25
Fake Easy (C)	1.30	.283	.06
Real Hard (D)	7.35	<.001	.26

Note. Significant univariate effects indicate significant differences in ratings of motivation between the four randomization order groups, within each level of condition.

Within condition A, ratings of motivation to do well in the CDAB group (M=3.05, SE=.62) significantly differed from ABCD (M=6.31, SE=.67; p<.001), BCDA (M=4.87, SE=.70; p=.041), and DABC (M=6.00, SE=.65; p=.001). Put differently, ratings of motivation to do well in the low load-easy instruction condition were lower when this condition was presented in the third block compared to when it was in the first, second, and fourth blocks.

Within condition B, ratings of motivation to do well in the CDAB group (M=3.00, SE=.59) were significantly lower than the ABCD group (M=5.25, SE=.64; p=.007), the BCDA group (M=7.07, SE=.67; p<.001) and the DABC group (M=4.88, SE=.63; p=.020). These results indicate that ratings of motivation to do well in the low load-hard instruction condition were lower when this condition was presented in the fourth block compared to when it was in the first, second, and third blocks. Additionally, ratings of motivation to do well were higher when this condition was presented first (BCDA) compared to when it was third (DABC; p=.020).

There were no significant differences in ratings of motivation to do well in the high load-easy instruction condition (C) depending on when it was presented.

Lastly, ratings of motivation to do well were significantly higher in the DABC group (M=7.71, SE=.62) compared to ABCD (M=3.88, SE=.64; p<.001), BCDA (M=5.20, SE=.66; p=.009) and CDAB (M=4.42, SE=.59; p<.001). This indicates that motivation to do well in the high load-hard instruction condition were higher when this condition was presented first compared to when it was presented second, third, and fourth. Figure 7 reveals order effects for ratings of motivation to do well, such that ratings of motivation in conditions A, B, and D were dependent on the order in which those conditions were presented.

Summary statement: motivation decreased with time-on-task, but not in the "fake easy" condition (condition C).

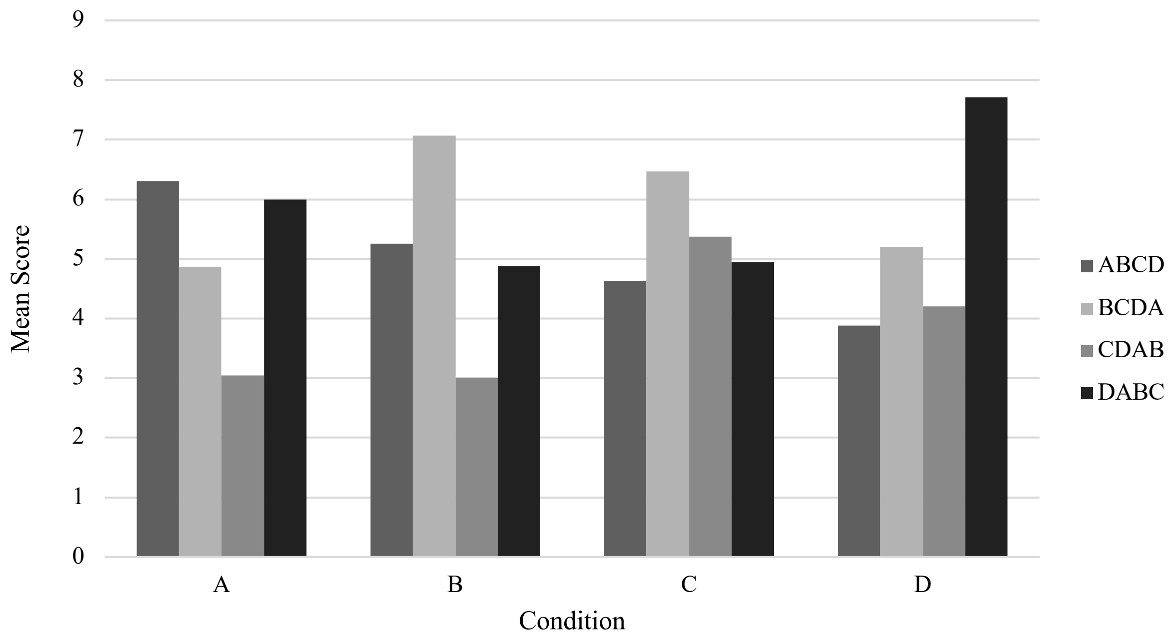

Figure 7. Mean scores for exit item 3 (motivation to do well) across each condition within each level of randomization order.

Association between fatigue and disengagement.

Regressions. Hierarchical regression analyses were conducted for each index of engagement in each condition, including baseline VAS-F in block 1 and Δfatigue in block 2 to examine if the increase in fatigue over the course of each condition would significantly predict disengagement over and above the effect of baseline fatigue ratings. Model 1, including only baseline VAS-F, was significant in predicting VAS-E scores in condition A, condition B, and condition C (Table 9); in all cases, baseline fatigue ratings only accounted for a small amount of the variance in subjective engagement. Model 1 was not significant in predicting any other dependent variable. Moreover, ΔR^2 was not significant for any of the 4 regression analyses including VAS-E as a dependent variable; statistics are reported for Model 2 (including both baseline fatigue ratings and Δfatigue as predictors) in Table 10. Model 2 was significant in predicting VAS-E in all conditions, with baseline VAS-F remaining a significant predictor of VAS-E scores.

Additionally, Model 2 was significantly improved in predicting both baseline and stimulus-evoked pupil diameter in condition A only compared to Model 1; in both cases, the addition of Δfatigue to the regression model explained 15% more variance in pupil diameter, and this change was significant (baseline: p=.002; stimulus-evoked: p=.003). Only Δfatigue was a significant predictor of pupil diameter (baseline: B=-9.45, SE=2.94, p=.002; stimulus-evoked: B=-9.92, SE=3.19, p=.003); specifically, as Δfatigue increased by one unit, baseline pupil diameter decreased by 9.45 units, and stimulus-evoked pupil diameter decreased by 9.92 units when baseline fatigue was equal to 0. Model 2 was not significant in predicting any of the other indices of engagement.

Summary statement: contrary to our hypothesis, increases in fatigue from pre- to post-test did not significantly predict indices of subjective, behavioural, and physiological engagement across all conditions. Baseline fatigue ratings predicted subjective engagement, and increases in fatigue only predicted pupillometric variables in condition A.

Table 9

Model statistics and unstandardized coefficients for Model 1 predicting engagement variables

		B (*SE*)	Model statistics		
Dependent Variable		Baseline VAS-F	*F*	Adjusted R^2	*p*-value
Real Easy	VAS-E	-.26 (.12)*	4.57	.05	.036
	d'	.00 (.01)	.00	-.01	.955
	BL pupil	1.53 (2.37)	.42	-.01	.522
	SE pupil	1.51 (2.55)	.35	-.01	.558
Fake Hard	VAS-E	-.39 (.11)**	12.02	.13	.001
	d'	.01 (.01)	.50	-.01	.482
	BL pupil	2.07 (2.40)	.74	-.01	.393
	SE pupil	2.17 (2.62)	.68	-.01	.413
Fake Easy	VAS-E	-.29 (.11)*	6.86	.08	.011
	d'	.00 (.01)	.06	-.01	.815
	BL pupil	2.09 (2.73)	.59	-.01	.447
	SE pupil	2.24 (2.89)	.60	-.01	.441
Real Hard	VAS-E	-.22 (.13)	2.89	.03	.094
	d'	.00 (.01)	.07	-.01	.794
	BL pupil	.94 (2.43)	.15	-.02	.699
	SE pupil	1.02 (2.60)	.15	-.02	.698

Note. **p*<.05, ***p*<.01, ****p*<.001. P-values above refer to significance testing for variance explained by Model 1.

Table 10

Model statistics and unstandardized coefficients for Model 2 predicting engagement variables

		B (SE)		Model statistics		
Dependent Variable		Baseline VAS-F	Δfatigue	*F*	Adjusted R^2	*p*-value
Real Easy	VAS-E	-.29 (.12)*	-.27 (.17)	3.60	.07	.032
	d'	.00 (.01)	.01 (.01)	.22	-.02	.807
	BL pupil	.61 (2.21)	-9.45 (2.94)**	5.40	.13	.007
	SE pupil	.54 (2.40)	-9.92 (3.19)**	5.04	.12	.010
Fake Hard	VAS-E	-.41 (1.12)**	-.10 (.14)	6.25	.13	.003
	d'	.01 (.01)	.00 (.01)	.25	-.02	.781
	BL pupil	2.63 (2.51)	2.50 (3.01)	.70	-.01	.500
	SE pupil	2.71 (2.74)	2.41 (3.35)	.60	-.01	.553
Fake Easy	VAS-E	-.33 (.11)**	-.25 (.16)	4.70	.10	.012
	d'	.00 (.01)	-.01 (.01)	.54	-.01	.584
	BL pupil	1.72 (2.92)	-1.62 (4.21)	.36	-.02	.698
	SE pupil	1.90 (3.08)	-1.46 (4.45)	.35	-.01	.706
Real Hard	VAS-E	-.27 (.13)*	-.34 (.17)	3.48	.07	.036
	d'	.00 (.01)	-.01 (.01)	.67	-.01	.516
	BL pupil	.87 (2.45)	-1.32 (2.27)	.16	-.03	.856
	SE pupil	.93 (2.63)	-1.53 (3.51)	.17	-.03	.844

Note. *p<.05, **p<.01, ***p<.001. P-values above refer to significance testing for variance explained by Model 2.

Correlations between engagement variables and covariates. To investigate whether any of the other potential covariates should be included in the regression analyses, Pearson product-moment correlations were conducted including all of the indices of engagement in each condition as dependent variables. Correlation coefficients are reported in Tables 11-14.

In condition A, VAS-E was significantly related to baseline VAS-B ($r(59)=-.29$, $p=.027$), ESS total score, DASS-21 Stress score, and VAS-F, VAS-C, VAS-B and VAS-S at pre- and post-test. Accuracy and pupil diameter variables were not significantly correlated with any covariates in condition A.

In condition B, VAS-E was significantly correlated with ESS total score, and VAS-F, VAS-C, VAS-B and VAS-S at pre- and post-test. Accuracy in condition B was significantly associated with DASS-21 Depression score and VAS-S at post-test.

In condition C, VAS-E was related to baseline VAS-C ($r(58)=.27$, $p=.041$) and baseline VAS-B ($r(58)=-.39$, $p=.002$), FSS total score, and VAS-F, VAS-C, VAS-B and VAS-S at pre- and post-test. In addition, *d'* was significantly correlated with VAS-B at pre- and post-test, as well as VAS-F and VAS-S at post-test.

In condition D, VAS-E was related to baseline VAS-C ($r(58)=.35$, $p=.007$) and baseline VAS-B ($r(58)=-.29$, $p=.028$), ISI total score, and VAS-F, VAS-C, VAS-B and VAS-S at pre- and post-test. Accuracy was significantly related to VAS-C and VAS-B at pre- and post-test, as well as VAS-F only at post-test.

Table 11

Pearson product-moment correlations between study variables of interest for low load-easy instruction condition (A; "real easy")

Variable	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*
1. VAS-E	1																		
2. d'	-.12	1																	
3. BL pupil	-.08	.04	1																
4. SE pupil	-.07	.03	1.00**	1															
5. ISI	-.07	-.21	-.07	-.08	1														
6. DASS-D	-.13	.12	-.08	-.10	.51**	1													
7. DASS-A	-.09	.21	-.13	-.15	.37**	.52**	1												
8. DASS-S	-.24*	.11	-.18	-.19	.36*	.64**	.61**	1											
9. FSS	-.01	.07	-.06	-.06	.28*	.39**	.38**	.35**	1										
10. CMQ	-.06	.09	.10	.09	-.13	-.07	.09	.14	.08	1									
11. ESS	-.28*	.05	-.01	-.01	.00	.05	.20	.19	.12	-.07	1								
12. VAS-F pre	-.35**	-.02	.30*	.28*	.33*	.20	.13	.06	.11	-.30*	.14	1							
13. VAS-C pre	.37**	.02	-.14	-.13	-.31*	-.09	-.11	-.01	-.12	.16	-.12	-.54**	1						
14. VAS-B pre	-.50**	-.10	.14	.14	.09	.07	-.16	-.17	-.10	-.14	.23	.54**	-.33*	1					
15. VAS-S pre	-.38**	-.00	.21	.20	.36**	.21	.20	.13	.07	-.36**	.18	.92**	-.51**	.55**	1				
16. VAS-F post	-.48**	.02	.04	-.04	.39**	.27*	.19	.24	.07	-.25	.28*	.79**	-.38**	.48**	.82**	1			
17. VAS-C post	.50**	-.14	-.01	-.01	-.23	-.12	-.16	-.14	-.12	-.09	-.12	-.36**	-.77**	.23	-.35**	-.46**	1		
18. VAS-B post	-.52**	.21	.13	.13	-.03	.24	-.12	-.02	.02	.12	.14	.16	-.13	.58**	.17	.19	-.15	1	
19. VAS-S post	-.47**	.07	.06	.06	.34**	.14	.10	.25	-.01	-.16*	.33*	.66**	-.32*	.44**	.77**	.87**	-.39**	.20	1

Table 12

Pearson product-moment correlations between study variables of interest for low load-hard instruction condition (B; "fake hard")

Variable	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*
1. VAS-E	1																		
2. d'	-.16	1																	
3. BL pupil	-.10	-.14	1																
4. SE pupil	-.11	-.16	1.00**	1															
5. ISI	-.18	.23	.00	-.04	1														
6. DASS-D	-.28*	-.06	-.08	-.10	.51**	1													
7. DASS-A	-.12	-.16	-.12	-.13	.37**	.52**	1												
8. DASS-S	-.20	-.01	-.18	-.19	.36*	.64**	.61**	1											
9. FSS	-.05	.03	-.07	-.08	.28*	.39**	.38**	.35**	1										
10. CMQ	-.01	-.19	.09	.06	-.13	-.07	.09	.14	.08	1									
11. ESS	-.27*	-.03	.00	.03	.00	.05	.20	.19	.12	-.07	1								
12. VAS-F pre	-.28*	.13	-.00	.01	.37**	.27*	.17	.23	.17	-.37*	.27*	1							
13. VAS-C pre	.40**	-.11	-.08	-.09	-.33*	-.23	-.12	-.10	-.21	.14	-.11	-.49**	1						
14. VAS-B pre	-.43**	.08	-.10	-.10	.11	.21	-.24	.06	-.02	-.10	.18	.47**	-.43*	1					
15. VAS-S pre	-.31**	.12	-.04	-.03	.37**	.24	.07	.31*	.14	-.37**	.31*	.88**	-.40**	.53**	1				
16. VAS-F post	-.30**	.12	.06	.06	.27*	.25	.05	.16	.23	-.15	.29*	.67**	-.34**	.30**	.67**	1			
17. VAS-C post	.50**	-.11	-.05	-.06	-.25	-.21	-.10	-.15	-.22	-.09	-.09	-.33**	.82**	-.35**	-.30**	-.52**	1		
18. VAS-B post	-.62**	.25	.17	.17	.11	.34*	-.12	.04	.15	.07	.17	.16	-.26*	.60**	.23	.37**	-.39**	1	
19. VAS-S post	-.27**	.30*	.07	.08	.31**	.14	.00	.19	.15	-.24*	.27*	.59**	-.27*	.26*	.75**	.74**	-.36**	.34**	1

Table 13

Pearson product-moment correlations between study variables of interest for high load-easy instruction condition (C; "fake easy")

Variable	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*
1. VAS-E	1																		
2. d'	.33*	1																	
3. BL pupil	.08	.20	1																
4. SE pupil	.09	.20	1.00**	1															
5. ISI	-.23	-.07	.04	.02	1														
6. DASS-D	-.19	-.07	-.12	-.14	.51**	1													
7. DASS-A	-.04	-.12	-.19	-.20	.37**	.52**	1												
8. DASS-S	.00	.05	-.22	-.23	.36**	.64**	.61**	1											
9. FSS	-.32*	-.04	-.01	-.01	.28*	.39**	.38**	.35**	1										
10. CMQ	.07	.07	.11	.11	-.13	-.07	.09	.14	.08	1									
11. ESS	-.05	.16	.01	.02	.00	.05	.20	.19	.12	-.07	1								
12. VAS-F pre	-.40**	-.15	.11	.11	.29*	.23	.12	.24	.20	-.26*	.14	1							
13. VAS-C pre	.48**	.31	.03	.03	-.21	-.23	-.07	-.15	-.36**	-.03	.02	-.52**	1						
14. VAS-B pre	-.49**	-.34**	.05	.04	-.01	.12	-.20	-.08	.01	-.05	.00	.35**	-.29*	1					
15. VAS-S pre	-.37**	-.16	.08	.09	.22	.14	.12	.24	.16	-.24	.15	.88**	-.52**	.39**	1				
16. VAS-F post	-.44**	-.28*	.06	.07	.25	.15	.08	.13	.13	-.25	.21	.79**	-.45**	.37**	.71**	1			
17. VAS-C post	.60**	.35*	-.01	-.00	-.18	-.11	-.10	-.14	-.26	-.05	.01	-.45**	.83**	.39**	-.39**	-.46**	1		
18. VAS-B post	-.51**	-.34**	-.04	-.03	-.05	.13	-.25	-.09	-.05	-.01	.09	.21	-.20	.66**	.20	.45**	-.30*	1	
19. VAS-S post	-.48**	-.24	.01	.01	.29**	.00	.06	.10	.08	-.33*	.24	.71**	-.36**	.43**	.78**	.78**	-.44**	.36**	1

Table 14

Pearson product-moment correlations between study variables of interest for high load-hard instruction condition (D; "real hard")

Variable	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*
1. VAS-E	1																		
2. d'	.36*	1																	
3. BL pupil	-.05	.08	1																
4. SE pupil	-.04	.09	1.00**	1															
5. ISI	-.42**	-.07	-.03	-.05	1														
6. DASS-D	-.25	-.14	-.15	-.17	.53**	1													
7. DASS-A	-.04	-.10	-.15	-.17	.32*	.47**	1												
8. DASS-S	-.07	-.02	-.22	-.24	.34*	.63**	.62**	1											
9. FSS	-.19	.06	-.12	-.13	.28*	.39**	.38**	.35**	1										
10. CMQ	.03	-.04	.14	.13	-.13	-.07	.09	.14	.08	1									
11. ESS	-.17	.12	-.04	-.04	.00	.05	.20	.19	.12	-.07	1								
12. VAS-F pre	-.50**	-.24	.10	.11	.45*	.28	.27	.17	.22	-.15	.11	1							
13. VAS-C pre	.51**	.29*	.05	.06	-.39*	-.24	-.21	-.16	-.34**	.00	.00	-.61**	1						
14. VAS-B pre	-.57**	-.32*	-.04	-.03	.17	.23	-.05	-.02	-.07	-.03	.10	.59**	-.40*	1					
15. VAS-S pre	-.46**	-.04	.09	.10	.42**	.17	.20	.17	.21	-.23	.17	.90**	-.57**	.55**	1				
16. VAS-F post	-.64**	.30*	.07	.07	.46**	.30*	.20	.19	.31*	-.24	.23	.81**	-.53**	.61**	.77**	1			
17. VAS-C post	.60**	.35**	.02	.03	-.36**	-.21	-.23	-.16	-.33*	-.01	-.07	-.63**	.96**	-.45**	-.59**	-.61**	1		
18. VAS-B post	-.66**	-.26*	-.13	-.03	.17	.31*	-.03	.06	.14	-.03	.26	.39**	-.34**	.73**	.42	.56**	-.42**	1	
19. VAS-S post	-.47**	-.10	.11	-.13	.39**	.25	.22	.17	.26	-.27*	.35	.76**	-.45*	.44**	.80**	.83**	-.50**	.43**	1

Regressions including covariates. Unstandardized regression coefficients for the following analyses can be found in Table 15. In cases wherein both pre- and post-test VAS values were correlated with dependent variables of interest, change scores were used as predictor variables for consistency with Δfatigue. Using hierarchical regression analysis, VAS-E in condition A was regressed on baseline fatigue (step 1); Δfatigue (step 2); and baseline VAS-B, ESS total score, Δcognitive resources, Δboredom and Δsleepiness (step 3). Addition of the covariates significantly improved the model, ΔR^2=.19, p=.024. Only the ESS total score remained a significant predictor of VAS-E, B=-1.57, SE=.68, p=.025.

Next, VAS-E in condition B was regressed on baseline fatigue (step 1), Δfatigue (step 2), and Epworth Sleepiness Scale (ESS) total score, Δcognitive resources, Δboredom and Δsleepiness (step 3). Addition of the covariates did not significantly improve the model, ΔR^2=.10, p=.076, and only baseline fatigue remained a significant predictor of VAS-E, B=-.38, SE=.12, p=.002.

Accuracy (d') in condition B was regressed on baseline fatigue (step 1), Δfatigue (step 2), and DASS-21 Depression total score and VAS-S at post-test (step 3). The addition of the covariates did not significantly improve the model, ΔR^2=.01, p=.767, with none of the predictors demonstrating a significant association with d', all ps>.05.

VAS-E in condition C was regressed on baseline fatigue (step 1), Δfatigue (step 2), and baseline VAS-C, baseline VAS-B, FSS total score, Δcognitive resources, Δboredom and Δsleepiness (step 3). Addition of the covariates significantly improved the model, ΔR^2=.20, p=.009. Only FSS total score, B=-6.09, SE=2.37, p=.013, and baseline VAS-B, B=-.22, SE=.09, p=.019, significantly predicted VAS-E in condition C.

D' in condition C was regressed on baseline fatigue (step 1), Δfatigue (step 2), and Δboredom and VAS-S at post-test (step 3). Addition of the covariates did not significantly improve the model, ΔR^2=.08, p=.059, although VAS-S at post-test emerged as a significant predictor of accuracy in this condition, B=-.02, SE=.01, p=.028.

In condition D, VAS-E was regressed on baseline fatigue (step 1), Δfatigue (step 2), and baseline VAS-C, baseline VAS-B, ISI total score, Δcognitive resources, Δboredom and Δsleepiness (step 3). Addition of the covariates significantly improved the model, ΔR^2=.33, p<.001. ISI total score emerged as a significant predictor, with increasing insomnia severity predicting decreased engagement, B=-1.41, SE=.60, p=.022. Additionally, baseline VAS-C B=.44, SE=.13, p<.001, and Δcognitive resources significantly predicted engagement in this condition, B=1.32, SE=.35, p<.001.

Finally, *d'* in condition D was regressed on baseline fatigue (step 1), Δfatigue (step 2), and Δcognitive resources and Δboredom (step 3). Addition of these covariates did not significantly improve the model, ΔR^2=.03, p=.386, and none of the predictors emerged as significant, *p*s>.05.

Summary statement: subjective engagement was predicted by different variables in each condition. demonstrating associations with underload (i.e., sleepiness) in the "real easy" condition (condition A), baseline fatigue in the "fake hard" condition (condition B), impairment related to fatigue symptoms in the "fake easy" condition (condition C), and cognitive resources and insomnia symptom severity in the "real hard" condition (condition D).

Table 15

Regression coefficients for Model 3 regressions.

	Dependent Variable						
	Real Easy (A)	Fake Hard (B)		Fake Easy (C)		Real Hard (D)	
	VAS-E	VAS-E	d'	VAS-E	d'	VAS-E	d'
Predictor	*B (SE)*						
Baseline VAS-F	-.08 (.14)	-.38 (.12)**	.01 (.01)	-.23 (.11)	.01 (.01)	-.01 (.12)	.00 (.01)
Δfatigue	-.14 (.25)	.05 (.21)	.00 (.01)	-.07 (.19)	-.02 (.01)	-.19 (.19)	-.01 (.01)
ISI	-	-	-	-	-	-1.41 (.60)*	-
FSS	-	-	-	-6.09 (2.37)*	-	-	-
ESS	-1.57 (.68)*	-1.18 (.64)	-	-	-	-	-
DASS-21 Depression	-	-	-.02 (.02)	-	-	-	-
DASS-21 Stress	-.51 (.44)	-	-	-	-	-	-
Baseline VAS-C	-	-	-	.10 (.12)	-	.43(.14)*	-
Baseline VAS-B	-.15 (.11)	-	-	-.22 (.09)*	-	-.11(.10)	-
Δcognitive resources	.36 (.19)	.27 (.25)	-	.23 (.20)	-	1.32 (.35)*	.02 (.03)
Δboredom	-.08 (.11)	-.26 (.14)	-	.01 (.11)	.01 (.01)	.04 (.12)	.01 (.01)
Δsleepiness	.32 (.21)	.30 (.19)	-	-.17 (.20)	-	.13 (.16)	-
VAS-S post-test	-	-	-.00 (.01)	-	-.02 (.01)*	-	-

Correlations between engagement variables. Lastly, Pearson product-moment correlations were calculated for the measures of subjective, behavioural, and physiological engagement in each condition (Table 16). The results indicated that VAS-E scores in each of the conditions were significantly correlated with each other. Similarly, all baseline pupil diameter variables were significantly correlated, and all stimulus-evoked pupil diameter variables were correlated with each other; additionally, baseline and stimulus-evoked pupil diameter were correlated with each other at each time point, with very large effect sizes (rs $\geq$.90). In contrast, accuracy (*d'*) was significantly correlated in the low load conditions (A and B) and the high load conditions (C and D), but correlations between accuracy in low load and high load conditions were not significant. VAS-E was related to *d'*, but only in the high load conditions; pupil diameter variables were not significantly correlated with other measures of engagement.

Summary statement: contrary to our hypothesis, subjective, behavioual and physiological engagement variables were not significantly correlated in all conditions. Relationships between subjective and behavioural engagement became stronger as the cognitive load (i.e., task difficulty) increased, but the pupillometric variables were not significantly correlated with the other indices of engagement.

Table 16

Pearson product-moment correlations between subjective, behavioural, and physiological engagement variables.

Variable	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*
1. A: VAS-E	1.00															
2. A: *d'*	-.12	1.00														
3. A: BL pupil	-.08	.04	1.00													
4. A: SE pupil	-.07	.03	1.00**	1.00												
5. B: VAS-E	.69**	-.14	-.06	-.06	1.00											
6. B: *d'*	.05	.27*	-.10	-.12	-.16	1.00										
7. B: BL pupil	-.07	.03	.96**	.95**	-.10	-.14	1.00									
8. B: SE pupil	-.08	.02	.95**	.96**	-.11	-.16	.99**	1.00								
9. C: VAS-E	.45**	-.19	.06	.06	.47**	-.00	.04	.03	1.00							
10. C: *d'*	-.01	.06	.18	.16	-.08	.05	.20	.19	.33*	1.00						
11. C: BL pupil	-.05	.04	.91**	.91**	-.07	-.09	.94**	.94**	.08	.20	1.00					
12. C: SE pupil	-.05	.03	.92**	.92**	-.06	-.11	.94**	.95**	.09	.20	1.00**	1.00				
13. D: VAS-E	.52**	-.02	-.09	-.08	.45**	.09	-.12	-.13	.45**	.33*	-.14	-.15	1.00			
14. D: *d'*	.18	-.05	.08	.08	.17	.01	.09	.09	.20	.51**	.10	.11	.36**	1.00		
15. D: BL pupil	-.06	.04	.92**	.92**	-.06	-.12	.92**	.91**	.15	.19	.94**	.94**	-.05	.08	1.00	
16. D: SE pupil	-.05	.02	.93**	.93**	-.06	-.14	.93**	.92**	.15	.18	.94**	.94**	-.04	.09	1.00**	1.00

Note. A = real easy (low load-easy instruction) condition. B = fake hard (low load-difficult instruction) condition. C = fake easy (high load-easy instruction) condition. D = real hard (high load-difficult instruction) condition. VAS-E = Visual Analogue Scale for Engagement. *d'* = accuracy on *n*-back. BL pupil = baseline pupil diameter. SE pupil = stimulus-evoked pupil diameter.

Secondary Analyses

Reaction time. The regression analyses were repeated for each condition, in which baseline fatigue ratings and Δfatigue were regressed on reaction time. Unstandardized coefficients and model statistics for these analyses are reported in Table 17. In all four analyses, neither baseline fatigue nor Δfatigue were significant predictors of reaction time, with the regression models accounting for little-to-none of the variance.

Miss rate. The regression analyses were again repeated for each condition, in which baseline fatigue ratings and Δfatigue were regressed on miss rate. Unstandardized coefficients and model statistics for these analyses are also reported in Table 17. In all instances, neither baseline fatigue nor Δfatigue were significant predictors of reaction time, with the regression models accounting for little-to-none of the variance.

Summary statement: increases in fatigue did not predict secondary measures of behavioural engagement.

Table 17

Model statistics and unstandardized coefficients for Model 2, with reaction time and miss rate as the dependent variables

Dependent Variable		Unstandardized Coefficients (*SE*)		Model statistics		
		Baseline VAS-F	Δfatigue	*F*	Adjusted R^2	*p*-value
Real Easy (A)	RT	-.09 (.92)	1.95 (1.29)	1.22	.01	.302
	Miss rate	.00 (.01)	.00 (.01)	1.02	.03	.366
Fake Hard (B)	RT	-.31 (.83)	-.40 (1.00)	.20	-.02	.899
	Miss rate	.00 (.00)	.00 (.00)	.04	-.03	.958
Fake Easy (C)	RT	-.13 (1.40)	.33 (1.99)	.02	-.03	.977
	Miss rate	.00 (.00)	.00 (.00)	1.94	.03	.151
Real Hard (D)	RT	-.64 (1.37)	.73 (1.84)	.23	-.02	.793
	Miss rate	.00 (.00)	.00 (.00)	1.22	.01	.302

Post-Hoc Analyses

Engagement. Exploratory two-way (condition × randomization order) mixed model ANOVAs were conducted for each measure of engagement (VAS-E, *d'*, baseline pupil diameter, and stimulus-evoked pupil diameter) to determine if the fatigue manipulation differentially affected engagement across the four conditions.

Subjective engagement. Mauchly's test was not significant, $\chi^2(5)=8.70$, $p=.122$, therefore the assumption of sphericity was met. Mean VAS-E ratings can be found in Table 18. The main effect of condition was significant, $F(3,201)=4.10$, $p=.008$, $\eta_p^2=.06$. The main effect of randomization order was not significant, $F(3,67)=2.07$, $p=.113$, $\eta_p^2=.09$. However, the main effect of condition was qualified by a significant condition × randomization interaction, $F(3,67)=293$, $p=.040$, $\eta_p^2=.12$.

To understand this interaction, simple effects of randomization order on VAS-E ratings at each level of condition were conducted. Only conditions B ($F(3,67)=2.82$, $p=.045$, $\eta_p^2=.11$) and D ($F(3,67)=5.58$, $p=.002$, $\eta_p^2=.20$) yielded significant effects of randomization order on ratings of subjective engagement. Within condition B, pairwise comparisons revealed that VAS-E ratings were significantly different between the BCDA group and the CDAB group ($p=.005$), indicating that engagement in the low load-hard instruction condition was significantly higher when this condition was presented first compared to when it was presented last. Within condition D, pairwise comparisons revealed that VAS-E ratings in the DABC group were significantly different from VAS-E ratings in the ABCD group ($p=.008$), the BCDA group ($p=.010$), and the CDAB group ($p<.001$). This demonstrates that engagement was highest in condition D (high load-hard instruction) when it was presented as the first block compared to when it was presented second, third, or fourth. Of note, engagement during condition D was lowest when it was presented in the second block (i.e., CDAB) after already completing one block of the *2*-back.

Summary statement: time-on-task effects were seen for subjective engagement, but only when participants were provided with a hard task instruction (i.e., conditions B and D), with the highest engagement occurring when these conditions were presented first. Subjective engagement was lowest for condition D ("real hard") when it was presented second rather than fourth (i.e., last).

Table 18

Mean VAS-E ratings for after each condition for each randomization order group.

	M (SD)			
	ABCD	BCDA	CDAB	DABC
Real Easy (A)	35.44 (23.56)	31.29 (22.13)	17.90 (17.10)	34.00 (29.58)
Fake Hard (B)	31.69 (19.56)	44.35 (18.15)	22.10 (20.94)	30.00 (29.70)
Fake Easy (C)	33.44 (25.12)	42.71 (18.14)	35.50 (20.74)	36.53 (24.10)
Real Hard (D)	33.56 (24.55)	34.71 (23.41)	25.75 (21.07)	55.32 (23.82)

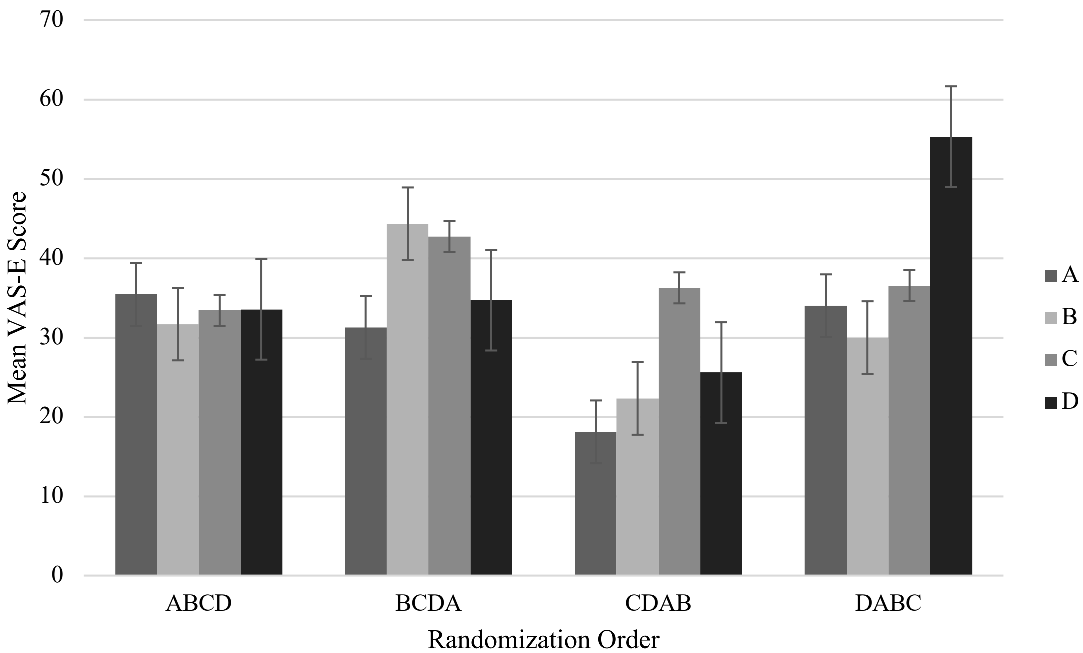

Figure 8. Mean VAS-E scores (subjective engagement) across each condition within each level of randomization order.

Behavioural engagement.

Accuracy. Mauchly's test was significant for the main effect of condition, $\chi^2(5)=17.76$, $p=.003$, therefore Huynh-Feldt estimates are reported ($\varepsilon=.93$). The main effects of condition, $F(2.79,186.65)=.03$, $p=.992$, $\eta_p^2=.00$, and randomization order, $F(3,67)=.25$, $p=.974$, $\eta_p^2=.00$, were not significant. Additionally, the interaction term was also non-significant $F(8.36,186.65)=1.67$, $p=.104$, $\eta_p^2=.07$. Null effects indicate that there were no significant differences in accuracy on the *n*-back across the four conditions.

Reaction time. To investigate a trade-off between accuracy and reaction time, this analysis was repeated including reaction time as the dependent variable. Mauchly's test was significant, $\chi^2(5)=39.16$, $p<.001$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates ($\varepsilon=.70$). The main effect of condition was significant, $F(2.11,141.32)=79.09$, $p<.001$, $\eta_p^2=.54$. Reaction times were significantly longer in the high load conditions, with condition D ($M=749.70$, $SE=29.97$) and condition C ($M=751.05$, $SE=30.39$) being significantly different from both conditions A ($M=521.85$, $SE=20.52$) and B ($M=497.75$, $SE=17.87$), all *p*s<.001. These results indicate that there was no trade-off between accuracy and reaction time, as reaction times were increased with increased cognitive demands. The main effect of randomization order ($F(3,67)=1.48$, $p=.227$, $\eta_p^2=.06$) and the interaction between condition and randomization order ($F(6.33,141.32)=1.08$, $p=.376$, $\eta_p^2=.05$) were not statistically significant.

Miss rate. To investigate potential differences in the rate of misses across the conditions due to disengagement from the *n*-back, this analysis was repeated including miss rate as the dependent variable. Average miss rates for each condition within each level of randomization order can be found in Table 19. Mauchly's test was significant for the main effect of condition, $\chi^2(5)=40.99$, $p<.001$, therefore Greenhouse-Geisser corrected degrees of freedom are reported (ε=.71). The main effect of condition was significant, $F(2.13, 144.85)=32.29$, $p=.002$, $\eta_p^2=.32$. Follow-up pairwise comparisons indicated that both low load conditions (A and B) significantly differed from both high load conditions (C and D), all *p*s<.001, but conditions of the same cognitive load did not significantly differ from each other. The main effect of randomization order was not significant, $F(3,68)=.51$, $p=.677$. However, the interaction between condition and randomization order was significant, $F(6.39,144.85)=2.14$, $p=.048$. Pairwise comparisons revealed that for groups ABCD, BCDA, and CDAB, the low load conditions were associated with significantly fewer misses than the high load conditions, *p*s<.05; the two *0*-back tasks did not differ from each other, nor did the two *2*-back tasks. However, in the DABC group, condition C (the fourth block) was associated with a significantly higher miss rate compared to all three of the other conditions.

Summary statement: accuracy was maintained throughout all conditions, although longer reaction times were necessary to maintain accuracy in conditions with higher cognitive load (i.e., conditions C and D). Additionally, tasks of higher cognitive load were associated with a higher miss rate than low cognitive load tasks in most cases, but the miss rate was highest when the "fake easy" condition was presented last.

Table 19

Average miss rates for each condition within each level of randomization order.

	M (SD)			
	ABCD	BCDA	CDAB	DABC
Real Easy (A)	.06 (.08)	.09 (.11)	.10 (.11)	.10 (.10)
Fake Hard (B)	.05 (.07)	.11 (.07)	.08 (.08)	.12 (.13)
Fake Easy (C)	.30 (.16)	.21 (.08)	.28 (.21)	.27 (.16)
Real Hard (D)	.27 (.14)	.20 (.17)	.28 (.24)	.13 (.12)

Physiological engagement. A third between-subjects factor, pupillometry method (remote *vs*. desktop mounted), was added to the analyses including pupillometric dependent variables to investigate potential differences in pupil size depending on how pupil size was measured. Therefore, two three-way mixed model ANOVAs were conducted including condition, randomization order, and pupillometry method as independent factors and baseline pupil diameter and stimulus-evoked pupil diameter as the dependent variables, respectively.

Baseline pupil diameter. Mauchly's test was significant indicating that the assumption of sphericity was not met for the main effect of condition, $\chi^2(5)=15.20$, $p=.010$, therefore Huynh-Feldt corrected tests are reported ($\varepsilon=1.00$). The main effects of condition, $F(3,168)=1.63$, $p=.184$, $\eta_p^2=.03$, and randomization order, $F(3,56)=.337$, $p=.799$, $\eta_p^2=.02$, were not significant. However, the main effect of pupillometry method was significant, $F(1,56)=38.38$, $p<.001$, $\eta_p^2=.40$, such that pupil diameter was significantly larger for participants who were tested using the desktop-mounted apparatus than those who were tested using remote tracking. None of the interaction terms were significant (condition × randomization order: $F(3,56)=.337$, $p=.799$, $\eta_p^2=.02$; condition × pupillometry method: $F(3,168)=.63$, $p=.596$, $\eta_p^2=.01$, condition × pupillometry method × randomization order: $F(6,168)=.32$, $p=.927$, $\eta_p^2=.01$).

As previously noted, the distance between the eye being tracked and the eye tracker can influence the size of the pupil, and participants' distance was not controlled in the remote tracking condition. Due to this methodological limitation, a two-way (condition × randomization) mixed model ANOVA was conducted selecting only participants who were tested using the desktop-mounted apparatus. Once again, Mauchly's test was significant, $\chi^2(5)=15.20$, $p=.010$ and Huynh-Feldt corrections for degrees of freedom are reported ($\varepsilon=.97$). There was no main effect of randomization order, $F(3,51)=1.19$, $p=.324$, $\eta_p^2=.07$, but the main effect of condition was significant, $F(2.90, 148.02)=6.72$, $p<.001$, $\eta_p^2=.12$. These effects were qualified by a significant interaction between condition and randomization, $F(8.71,148.02)=2.92$, $p=.004$, $\eta_p^2=.15$. This interaction was followed up with multivariate simple effects of the association between condition and baseline pupil diameter within each level of randomization order. Multivariate effects were significant for the CDAB, $V=.30$, $F(3,49)=7.13$, $p<.001$, and DABC groups, $V=.17$, $F(3,49)=3.43$, $p=.024$, but not significant for the ABCD, $V=.07$, $F(3,49)=1.13$, $p=.346$, or BCDA groups, $V=.06$, $F(3,49)=1.04$, $p=.382$.

Within the CDAB group, average baseline pupil diameter was significantly larger in condition C ($M=1238.91$, $SE=118.28$) than all other conditions (A: $M=1030.49$, $SE=102.97$, $p<.001$; B: $M=1071.54$, $SE=101.84$, $p<.001$; D: $M=1139.08$, $SE=104.04$, $p=.011$). Additionally, pupil diameter was significantly larger in condition D compared to condition A ($p=.007$). Within the DABC group, average baseline pupil diameter was significantly larger in condition D ($M=1197.44$, $SE=100.73$) than all of the other conditions (A: $M=1087.53$, SE=99.70, $p=.005$; B: $M=1082.60$, $SE=98.60$, $p=.007$; C: $M=1119.37$, $SE=114.52$, $p=.039$).

Stimulus-evoked pupil diameter. Mauchly's test was significant for the main effect of condition indicating that the assumption of sphericity was violated, $\chi^2(5)=12.12$, $p=.033$, therefore Huynh-Feldt corrected tests are reported ($\varepsilon=1.00$). Once again, there was a main effect of pupillometry method, $F(1,56)=39.50$, $p<.001$, $\eta_p^2=.41$, such that participants tested with the desktop-mounted apparatus yielded larger pupil diameters than those tested using remote tracking. There were no main effects of condition, $F(3,168)=1.96$, $p=.122$, $\eta_p^2=.03$, or randomization order, $F(3,56)=.37$, p=.777, $\eta_p^2=.02$. However, the condition × randomization interaction was significant, $F(9,168)=2.20$, $p=.024$, $\eta_p^2=.11$, and was followed up with multivariate simple effects of condition on stimulus-evoked pupil diameter within each level of randomization order. The multivariate test was only significant for the DABC group, $V=.21$, $F(3,54)=4.65$, $p=.006$. Pairwise comparisons yielded significant differences in stimulus-evoked pupil diameter between condition D ($M=1296.33$, $SE=103.35$) and all other conditions (A: $M=1172.44$, $SE=102.91$, $p=.001$, B: $M=1168.36$, $SE=102.25$, $p=.002$, C: $M=1208.74$, $SE=115.19$, $p=.015$). All of the other interaction terms were not significant (condition × pupillometry method: $F(3,168)=.76$, $p=.519$, $\eta_p^2=.01$; randomization order × pupillometry method: $F(2,56)=.32$, $p=.729$, $\eta_p^2=.01$; condition × pupillometry method × randomization order: $F(6,168)=.43$, $p=.857$, $\eta_p^2=.02$).

To remove the effect of the pupillometry method, another two-way (condition × randomization order) mixed model ANOVA was conducted selecting only participants who were tested using the desktop-mounted apparatus. Mauchly's test was significant for the main effect of condition, $\chi^2(5)=11.06$, $p=.050$, therefore Huynh-Feldt corrected tests are reported ($\varepsilon=1.00$). Similar to the results of baseline pupil diameter, the main effect of condition was significant, $F(2.99,152.41)=8.03$, $p<.001$, $\eta_p^2=.14$. The main effect of randomization order was not significant, $F(3,51)=1.22$, $p=.312$, $\eta_p^2=.07$.

There was a significant interaction between condition and randomization order, $F(8.97, 152.41)=3.82$, $p<.001$, $\eta_p^2=.18$. Multivariate simple effects of condition within each level of randomization order revealed significant effects of condition within the CDAB group, $V=.36$, $F(3,49)=9.02$, $p<.001$, and the DABC group, $V=.21$, $F(3,49)=4.24$, $p=.010$. Within the CDAB group, stimulus-evoked pupil diameter was significantly larger in condition C ($M=1345.32$, $SE=124.56$) compared to conditions A ($M=1112.71$, $SE=111.26$; $p<.001$), B ($M=1168.28$, $SE=110.55$; $p<.001$) and D ($M=1236.58$, $SE=111.76$; $p=.006$), and significantly larger in condition D compared to condition A ($p=.003$). Within the DABC group, pupil diameter was significantly larger in condition D (M=1296.33, SE=108.21) compared to conditions A (M=1172.44, SE=107.73; p=.002), B (M=1168.36, SE=107.04; p=.003), and C (M=1208.74, SE=120.61; p=.020).

Summary statement: pupillometric variables demonstrated significant order effects when the high cognitive load conditions were presented first (i.e., CDAB and DABC), with the high cognitive load conditions demonstrating larger pupil diameter than the low load conditions.

Boredom. To rule out the underload hypothesis as an explanation for null findings, a three way (condition × time × randomization order) mixed model ANOVA was conducted with VAS-B scores as the dependent variable. Mauchly's test was not significant for the main effect of condition, $\chi^2(5)=10.55$, p=.061, or the condition × time interaction term, $\chi^2(5)=9.35$, p=.096, therefore the assumption of sphericity was met. The main effects of condition, $F(3,201)=.83$, p=479, $\eta_p^2=.01$, and randomization order, $F(3,67)=1.90$, p=.138, $\eta_p^2=.08$, were not significant. However, the main effect of time was significant, indicating an increase in boredom from pre- to post-test averaged across the four conditions, $F(9,201)=18.22$, $p<.001$, $\eta_p^2=.45$. Analysis of two-way interaction terms revealed the interaction between time and randomization order, $F(3,67)=1.80$, p=.156, $\eta_p^2=.07$ was not statistically significant. However, there was a significant interaction between condition and randomization order, $F(9,201)=18.22$, $p<.001$, $\eta_p^2=.45$, and condition and time, $F(3,201)=4.95$, p=.002, $\eta_p^2=.07$.

These interaction effects were further qualified by a significant three-way interaction between condition, time and randomization, $F(9,201)=8.05$, $p<.001$, $\eta_p^2=.27$. To explore this interaction, multivariate simple effects of time were conducted within each level combination of condition and randomization order. Within the ABCD group, only conditions A, V=.28,

$F(1,67)=25.38$, $p<.001$, and B, $V=.20$, $F(1,67)=17.06$, $p<.001$, resulted in significant increases in boredom from pre- to post-test, while the increase was not significant in conditions C and D, *p*s>.05. Within the BCDA group, only conditions B, $V=.36$, $F(1,67)=37.37$, $p<.001$, and D, $V=.09$, $F(1,67)=6.58$, $p=.013$ resulted in significant increases in boredom from pre- to post-test, while the increase was not significant in conditions A and C, *p*s>.05. Within the CDAB group, increases in boredom over time were only found in conditions C, $V=.35$, $F(1,67)=35.55$, $p<.001$, and D, $V=.14$, $F(1,67)=10.71$, $p=.002$, while the change was not significant in conditions A and B, *p*s>.05. Lastly, in the DABC group, significant increases in boredom over time were found for conditions D, $V=.45$, $F(1,67)=54.15$, $p<.001$, A, $V=.38$, $F(1,67)=40.36$, $p<.001$, and B, $V=.22$, $F(1,67)=19.27$, $p<.001$, while the change was not significant in condition C, $p=.524$.

Summary statement: boredom increased with time-on-task and reached a plateau by the third or fourth block in all randomization orders.

Sleepiness. Another three way (condition × time × randomization order) mixed model ANOVA was conducted with VAS-S scores as the dependent variable to rule out an underload hypothesis that the fatigue manipulation resulted in sleepiness, which was in turn responsible for decreased engagement. Mauchly's test was significant for the main effect of condition, $\chi^2(5)=16.10$, $p=.007$, therefore the assumption of sphericity was violated and corrected degrees of freedom are reported using Huynh-Feldt estimates ($\varepsilon=.96$). Mauchly's test was not significant for the condition × time interaction, $\chi^2(5)=2.77$, $p=.735$, therefore the assumption of sphericity was met for this interaction term. The main effects of condition, $F(2.88,192.85)=1.62$, $p=.185$, $\eta_p^2=.02$, and randomization order, $F(3,67)=.40$, $p=.751$, $\eta_p^2=.02$, were not significant. Consistent with the mixed model ANOVA results of the other VAS measures, there was a significant main effect of time, $F(1,67)=85.99$, $p<.001$, $\eta_p^2=.56$, with increased sleepiness scores at post-test ($M=74.85$, $SE=2.64$) relative to pre-test ($M=62.87$, $SE=2.68$).

The interaction between condition and randomization order was significant, $F(8.64,192.85)=20.09$, $p<.001$, $\eta_p^2=.47$, suggesting that order effects influenced the magnitude of the differences in sleepiness ratings on VAS-S across the four conditions. The two-way interactions between condition and time, $F(3,201)=2.12$, $p=.099$, $\eta_p^2=.03$, and time and randomization order, $F(3,67)=.98$, $p=.408$, $\eta_p^2=.04$, were not significant. These interactions were qualified by a significant three-way interaction between condition, time, and randomization order, $F(9, 201)=3.50$, $p<.001$, $\eta_p^2=.14$. This interaction was followed up with multivariate simple effects of time within each level combination of condition and randomization order. Within the ABCD group, the increase in sleepiness from pre- to post-test was only significant for conditions A, $V=.20$, $F(1,67)=16.32$, $p<.001$, and B, $V=.10$, $F(1,67)=7.44$, $p<.001$, while the differences in sleepiness over time were not significant for conditions C and D, ps>.05. Within the BCDA group, sleepiness significantly increased over time in conditions B, $V=.16$, $F(1,67)=12.91$, $p=.001$, C, $V=.10$, $F(1,67)=7.63$, $p=.007$, and D, $V=.11$, $F(1,67)=7.85$, $p=.007$, while the increase was not significant in condition A, $V=.03$, $F(1,67)=2.17$, $p=.145$. Likewise, within the CDAB group, significant differences were found for the first three blocks (C: $V=.24$, $F(1,67)=20.56$, $p<.001$; D: $V=.15$, $F(1,67)=12.00$, $p=.001$; A: $V=.10$, $F(1,67)=7.56$, $p=.008$) but not in the fourth block (B: $V=.00$, $F(1,67)=.03$, $p=.875$). This pattern was mirrored again in the DABC group, with significant increases in sleepiness reported for the first three testing blocks (D: $V=.35$, $F(1,67)=23.23$, $p<.001$; A: $V=.31$, $F(1,67)=30.39$, $p<.001$; B: $V=.12$, $F(1,67)=9.34$, $p=.003$) but not the fourth block (C: $V=.03$, $F(1,67)=2.36$, $p=.130$).

Summary statement: sleepiness increased with time-on-task and reached a plateau by the third or fourth block in all randomization orders.

Discussion

The aims of this study were to 1) assess the magnitude and direction of the relationship between subjectively rated fatigue and appraised cognitive resources, 2) investigate how the interaction between actual task demands and appraisals of task demands (i.e., effort required) affects changes in ratings of a) fatigue and b) cognitive resources, and 3) evaluate how changes in fatigue relate to subjective, behavioural, and physiological measures of engagement. The following discussion will address each of these aims in turn.

Aim 1: Fatigue and cognitive resources

Consistent with the first hypothesis, ratings of fatigue and cognitive resources were significantly negatively correlated at baseline and at pre- and post-test in each condition of the fatigue manipulation, suggesting that appraisals of available cognitive resources to meet task demands are central to the appraisal of the self as fatigued, lending further credibility to the understanding of fatigue as a stop-emotion in order to conserve cognitive resources (Meijman, 2000). This finding also allows for a richer understanding of previous research from the self-control literature that indicates that we tend to act in accordance with how "replenished" we *think* our resources for self-control are, irrespective of how depleted they actually are by clearly illustrating that our perceptions of ourselves as fatigued (or depleted) is closely related with appraisals of the availability of cognitive resources (Clarkson et al., 2010; Job et al., 2010).

Aim 2: Influence of fatigue manipulation on ratings of fatigue and cognitive resources

We hypothesized that the fatigue manipulation (i.e., task difficulty × instruction) would result in the greatest increases in fatigue for the high load-difficult instruction condition (D) and the lowest increases in the easy load-easy instruction condition (A). Contrary to this hypothesis, the fatigue manipulation was not successful in producing differing increases in fatigue from pre-to post-test across the four conditions, indicated by a non-significant interaction between condition and time. Rather, the significant three-way interaction between condition, time, and randomization order revealed that the pattern of increases in fatigue was associated with time on task (i.e., stepwise increases in fatigue following each consecutive cognitive task) rather than the fatigue manipulation, with significant increases in VAS-F scores from pre- to post-test for the first three blocks of the *n*-back and a non-significant increase in the fourth block due to ceiling effects, regardless of which condition was presented. The one exception to this finding was in the ABCD randomization group, in which significant increases were found from pre- to post-test in all four blocks of the *n*-back. In this regard, it is possible that the fatigue manipulation was influential – specifically, in the other four randomization orders, participants may have believed that they used up their available cognitive resources because they had already completed the high load-difficult instruction condition, whereas in the ABCD group they had thus far completed two low load conditions and a single high load condition that they were told was going to be easy. Put differently, participants in the ABCD may have perceived themselves as having more room to become more fatigued on the last two blocks.

Similarly, we hypothesized that the fatigue manipulation would result in the largest decreases in ratings of cognitive resources for the high load-difficult instruction condition (D) and the smallest decreases in the easy load-easy instruction condition (A); in other words, that participants would perceive having spent more cognitive resources in completing condition D compared to condition A, with conditions B and C having intermediate results. Unlike the results of the fatigue manipulation on VAS-F, the analysis including VAS-C as the dependent variable yielded a significant two-way interaction between condition and randomization order only. The significant two-way interaction indicates that averaged from pre- to post-test, the differences in average ratings of VAS-C across the conditions were dependent on the order in which the conditions were presented. Across all randomization orders, ratings of VAS-C decreased incrementally with each additional block of the *n*-back; however, cognitive resources decreased the most in the first two blocks, suggesting a possible floor effect.

To better understand these results, we investigated if the fatigue manipulation was successful in affecting participants' appraisals of actual task demands and expectations of effort/difficulty by analyzing the three exit items on the post-test questionnaire: appraised difficulty of the *n*-back before, appraised difficulty of the *n*-back during, and motivation to do well. Regarding the first exit item (appraised difficulty before), condition A was considered the least difficult, the incongruent conditions (B and C) were considered equally difficult, and condition D was considered the most difficult. However, a non-significant difference between conditions C and D indicates that the instruction as to how difficult the upcoming *2*-back would be was not sufficiently powerful in influencing participants' beliefs. A limitation of this exit item is that it may have been influenced by retrospective recall and/or biased by the actual task demands (i.e., cognitive load) because it was completed at the end of the block after having completed the *n*-back, rather than right after they were given the verbal instruction. An additional limitation is that participants completed a practice *n*-back with the same cognitive load as the task they were going to complete in that block *before* receiving the verbal instruction. As a result, participants were already aware of the actual task demands by the time they received the verbal instruction. This may be why C and D were not significantly different while B and D were significantly different; in conditions B and D participants were told the task they were about to do would be difficult, but the cognitive load was increased from condition B to D. Future studies could have participants rate their expectations of task difficulty immediately after receiving such an instruction.

Results from the second exit item (appraised difficulty during) indicated that the regardless of the instruction, participants perceived both low load conditions and both high load conditions to be equally difficult while completing the *n*-back task. These results corroborate those of the first exit item by suggesting that the instruction as to the task's difficulty did not influence how difficult participants found the task while they were doing it. This may be attributed to the within-subjects design; in order to test participants under all four conditions of the fatigue manipulation, deception was used to have participants believe that the tasks were qualitatively different (i.e., by a change in font colour) which led to one version of the task being more difficult than the other. Participants' evaluations of the two *0*-back and two *2*-back tasks as equally difficult provide further evidence that the manipulation was not strong or convincing enough to test the interaction between cognitive load and appraisals of task demands. Had the instruction been effective, we would have anticipated that participants' evaluations of the tasks' difficulties would have been influenced by this instruction, i.e., the perceived difficulty of the *0*-back would have been greater in the low load-hard instruction condition relative to the low load-easy condition, and the *2*-back would have been perceived as less difficult in the high load-easy instruction condition relative to the high load-hard instruction condition. Relatedly, it is likely that participants' familiarity with one version of the *0*- or *2*-back tasks ended up influencing their perception of how difficult the task is during the second *0*- or *2*-back, even if they received a different verbal instruction.

Consistent with the VAS-F and VAS-C results, motivation to do well was also associated with a significant interaction between condition and randomization order, with motivation decreasing with time on task. Interestingly, ratings of motivation to do well in conditions A, B, and D depended on when they were presented (e.g., motivation was rated highest for each condition when it was the first block), but there were no significant order effects for motivation in condition C. It is possible that in this case, motivation to do well reflects participants' appraisals of their own performance. In this condition, participants were told that the task would be easy when it was actually difficult and this may have prompted them to view their performance as poor relative to that of most people, which may have been perceived as demotivating. This is further compounded by the fact that participants had already had exposure to the practice *2*-back, wherein they may have already appraised the task demands as difficult before receiving information that most other people find it easy. According to the Social Cognitive Theory, cognitive and affective states provide information about self-efficacy and influence motivated behaviour (Bandura, 1986); applied to this context, *feeling* negatively (e.g., anxious, sad, embarrassed) about anticipated performance relative to that of others may have promoted decreased self-efficacy to do well on the task and, in turn, decreased motivation to apply effort.

Collectively, these results provide further evidence that mental fatigue increases significantly using a time on task paradigm (e.g., Bailey, Channon, & Beaumont, 2007; Boksem, Meijman, & Lorist, 2006; Gergelyfi et al., 2015; Gunzelmann et al., 2011; Hopstaken et al., 2015a; Hopstaken et al., 2015b; Sandry et al., 2014; Tanaka, Ishii, & Watanabe, 2014; van der Linden, Frese, & Meijman, 2003; van der Linden, Frese, & Sonnentag, 2003). Additionally, the results from this study provide evidence that increases in fatigue coincide with decreases in ratings of cognitive resources, but cognitive resources do not decrease over time to the same degree that fatigue increases over time (i.e., from pre- to post-test). Although substrates for a cognitive resource have not yet been identified, results from this study support the theory that *perceptions* of cognitive resources as limited are important to the experience of fatigue (Clarkson et al., 2010; Job et al., 2010). Moreover, although not a central aim of this study, we observed that motivation decreased concurrently with increasing fatigue and decreasing cognitive resources, which is consistent with past findings (Boksem & Tops, 2008; Hopstaken et al., 2015a). Because of order effects, this study was unable to test if increasing the cognitive load of the task required more cognitive resources than the low load tasks. Moreover, due to the insufficiently convincing instruction of effort, this study was unable to adequately test Hockey's (2011) Motivational Control Theory of fatigue that effort is a product of both task demands and appraisals of task demands, nor could this study replicate previous findings that high-effort instructions interact with greater task demands to produce increased mental fatigue (Earle, 2004). As such, the results herein can neither support nor refute a mechanistic role for situational appraisals of task demands (and, implicitly, of one's own available cognitive resources to meet those demands) in the development of mental fatigue. It would be interesting for future research to examine this hypothesis further, particularly in incongruent conditions (i.e., conditions B and C in the present study), which would clarify the respective roles of task demands versus appraisals of task difficulty in the development of mental fatigue.

Aim 3: Association between fatigue and disengagement

We wanted to understand if fatigue increases predicted task disengagement and found that subjective engagement was the only variable that was consistently predicted by the model, but it was predicted by baseline fatigue, not increases in fatigue. Such a finding is not surprising, given that the sample was highly fatigued at baseline and reported high rates of habitual fatigue as well. In addition, baseline ratings of fatigue were not subject to order effects, while the increases in fatigue from pre- to post-test for each condition were dependent on the order in which the conditions were presented. Consequently, the results of the regression analyses must be interpreted with caution, as they do not account for the order effects on ratings of fatigue at pre- and post-test. Similarly, order effects also affected engagement as measured by VAS-E, miss rate, baseline pupil diameter, and stimulus-evoked pupil diameter. This is likely a contributing factor to the null associations between fatigue and measures of engagement.

To provide greater context to these null findings in the regression analyses between increases in fatigue and disengagement, several post-hoc analyses were conducted to examine the effect of the fatigue manipulation on engagement. Interestingly, subjective engagement was lower on average in the *0*-back conditions compared to the *2*-back conditions. Specifically, follow-up tests revealed that subjective engagement was only significantly different between conditions A and C; in both of these conditions, participants received instructions that the task they were about to do was easy, but the higher task demands of condition C was associated with more engagement. This demonstrates that increased task demands resulted in increased effort and engagement. This finding is inconsistent with Meijman's (2000) theory of fatigue as an adaptive "stop-emotion," which predicts that greater levels of fatigue should result in decreased engagement as efforts to conserve dwindling cognitive resources are increased. Furthermore, a significant interaction between condition and randomization order revealed that the association between condition and subjective engagement was largely dependent on order effects, with engagement being rated highest after the first block of the *n*-back and decreasing with additional blocks. Interestingly, simple effects were not significant for the ABCD group only, indicating that the four conditions did not significantly differ from each other in terms of subjective engagement when presented in this order. One explanation for this is that the ABCD order is the only order in which the tasks become progressively more difficult (in terms of *either* perceptions of task difficulty or actual task difficulty), suggesting that this order may have been associated with the least amount of burnout in the first couple of blocks and allowing for engagement levels to be maintained across all four testing blocks. From a statistical standpoint, in the other randomization orders, the easier and more difficult tasks are interspersed, allowing for easier detection of significant effects.

The primary measure of behavioural engagement, accuracy (*d'*), was surprisingly unaffected by the fatigue manipulation or by order effects, with participants performing similarly on all four *n*-back tasks. This is in contrast with previous findings in which performance on the *n*-back, as indicated by *d'*, significantly worsened with increased time on task for both *1*- and *2*-back tasks (Hopstaken et al., 2015a, 2015b). *D'* is a ratio of signal to noise, or the ratio of hits to false alarms. Although this is a well-validated metric of accuracy that has previously been used as an index of engagement in the fatigue literature, it appears that in this study participants did not make a significant number of false alarms while completing the task to yield significant results. This suggests that participants may not have completed a sufficient number of trials in each block to yield an increase in false alarms; whereas each block in the present study contained 15 practice trials and 60 test trials, each time-on-task block in Hopstaken and colleagues' studies (2015a, 2015b) consisted of 183 trials. It is also possible that this finding is due to practice effects, as participants practiced the *n*-back at the beginning of each testing block before initiating the actual task, resulting in a ceiling effect of performance. There is some evidence in the literature that fatigue does not always result in neurobehavioural performance deficits (Kanfer, 2011), and this is supported by findings that individuals with insomnia report performing poorly on cognitive tasks because of fatigue but generally perform well on tasks

(Goldman-Mellor et al., 2015). For example, van der Linden, Frese & Sonnentag found that fatigued participants completed the same number of subtasks but demonstrated impairments in the process in a novel computer task (i.e., less systematic exploration and more errors). As discussed by Gergelyfi and colleagues (2015), consistent with the temporal relationship between fatigue and performance deficits proposed by Kanfer (2011), it is possible that participants increased their effort in order to maintain their performance, which might explain significant increases in fatigue in the absence of objective performance deficits. Furthermore, participants did not rate any of the *n*-back tasks as very difficult either before or while they were completing them (all average ratings were below 5.5/10 for difficulty before and all ratings were below 6.5/10 for difficulty during), indicating that they did not appraise the tasks as requiring a significant amount of mental effort. To frame this within the cost-benefit analysis terms proposed by Boksem and Tops (2008), although motivation decreases were coincident with increases in fatigue, perceptions of mental effort may not have been sufficiently high to promote behavioural disengagement from the task.

The second behavioural measure, reaction time, was also analyzed and was found to be longer during the more difficult tasks, thereby ruling out the possibility of a trade-off between speed and accuracy. Additionally, miss rate was included as a third behavioural measure, which has heretofore not been explored in the literature. It is reasonable to expect that a behavioural biproduct of the tonic mode of norepinephrine functioning put forth by the Adaptive Gain Theory (Aston-Jones & Cohen, 2005) would be increased misses as exploration of the environment for more rewarding tasks is promoted. The results indicated that the results were relatively robust to order effects; namely, that the miss rate was higher for more cognitively demanding conditions compared to the low demand conditions in three out of the four randomization order groups. In the DABC group, condition C was associated with significantly more misses than conditions A, B, and D. Examination of the marginal means revealed that across all of the randomization orders, condition C was consistently associated with the highest miss rate, followed by condition D; in the DABC group, the miss rate for condition D was reduced because this condition was completed first (i.e., no cumulative time on task effects) and participants were motivated to do well, contributing to a significant difference from condition C, which was completed last. Because miss rate has not yet been examined as a measure of engagement, further research is needed to validate this index against other measures of mental effort (e.g., heart rate variability) and engagement.

The first index of physiological engagement, baseline pupil diameter, found a main effect of condition in which the tasks with high load (C and D) demonstrated larger pupil diameter than the low load tasks. These findings are consistent with those of Hopstaken and colleagues (2015a), in which higher task demands (increased *n* on the *n*-back) were associated with larger baseline pupil sizes. As suggested by Hopstaken et al. (2015a), this increase in baseline pupil size may reflect a tendency toward disengagement and exploration behaviour (i.e., in search of more rewarding tasks). The interaction with randomization order was also significant, in which there were no significant differences in baseline pupil diameter for the ABCD and BCDA groups, but significant pairwise comparisons between the conditions were found in the CDAB and DABC groups. The lack of significant differences in the ABCD and BCDA groups reflects a floor effect; the low load conditions were already associated with small pupil size and as time on task increased, increases in pupil size associated with higher load conditions were blunted due to disengagement, which is in line with the results of baseline pupil diameter found in Hopstaken et al. (2015b). In contrast, in the CDAB group, condition C was associated with a larger pupil diameter than all of the other conditions (which reflects the fact that it was a high load condition *and* it was presented in the first block), and pupil size in condition D was larger than condition A. These results can be interpreted in two ways: 1) pupil size was significantly lower in condition D because the instruction was effective in convincing participants that this task was more difficult than condition C, resulting in decreased engagement, or 2) these results indicate disengagement with time on task, underscored by the significant decreases in pupil size between conditions C and D and again between D and A, finally leveling off in conditions A and B. Because task difficulty *during* the task was rated as equivalent for conditions C and D, the second hypothesis is more plausible. Furthermore, in the DABC group, pupil size in condition D was significantly larger than all of the other conditions – again, pupil size was equivalent in conditions A and B, and increased again (non-significantly) in condition C due to the increased load of the task. Additionally, these results mirror those of the VAS-E (above), which found that the low load conditions were *less* engaging than the high load conditions.

Similarly, the second index of physiological engagement, stimulus-evoked pupil diameter, also demonstrated an effect of condition in which pupil diameter was larger for the high load conditions compared to the low load conditions, consistent with increased mental effort to meet the higher task demands (Kahneman & Beatty, 1966). Again, a significant interaction demonstrated order effects for randomization groups BCDA, CDAB, and DABC, while there were no significant pairwise comparisons for the ABCD group. In the BCDA group, condition C was associated with larger stimulus-evoked pupil size than condition A, reflecting an effect of decreased cognitive load and increased time on task in condition A, while in the CDAB and DABC groups, the first condition presented (C and D, respectively) was associated with significantly larger pupil size than all of the other conditions. The significant order effects for the majority of the indices of engagement provide a likely explanation for the null regression findings, which were carried out based on condition and not on randomization order.

It is interesting that both baseline and stimulus-evoked pupil diameter were significantly predicted by increases in fatigue, but only in the low load-easy instruction condition. The results are only significant for the low load-easy instruction condition because pupil diameter was consistently small for this condition due to the low load of the task, regardless of the order in which it was presented. The consistency of baseline and stimulus-evoked pupil diameter in this condition and their significant associations with increases in fatigue lend further support to the hypothesis that if order effects were controlled, there is a greater likelihood of finding an association between increases in fatigue and the indices of engagement across the various conditions. For the remainder of the conditions, baseline and stimulus-evoked pupil size were affected by order effects and demonstrated negligible relationships with increases fatigue.

Put together, the engagement results tell an interesting story. Subjective engagement decreased with each additional testing block, though engagement was rated higher on average for the more cognitively demanding tasks than the easier tasks. This makes sense within the context of the underload literature (Pattyn et al., 2008; Smallwood et al., 2004), as participants found the more difficult tasks more challenging and less monotonous, thereby increasing engagement. Consistent with this finding, pupillometric results revealed that pupil diameter was larger for the more cognitively demanding tasks, reflecting greater effort, and this effect was also more pronounced when these tasks were completed in the first half of the experiment compared to the second half of the experiment, when participants' effort and engagement decreased and fatigue increased. Although not significant, it is an interesting observation that in the DABC group, pupil diameter decreased in the low load conditions (A and B) and increased again in the high load-easy instruction condition (C), despite it being in the last block. Notwithstanding this physiological increase in effort, the miss rate was also highest for this condition when presented last, and reaction times were longer. It is helpful to conceptualize the high load-easy instruction condition as an “unpleasant surprise,” wherein the expectation of an easy task is violated by being presented with a challenging task. The convergence of increased effort and decreased performance in this condition indicates that having this expectation violated is taxing.

Covariate regression models. Consistent with the underload hypothesis (Pattyn, Neyt, Hendrickx, & Soetens, 2008), addition of daytime sleepiness to the regression model significantly predicted subjective engagement in condition A, leading to a non-significant relationship between baseline fatigue ratings and subjective engagement. Interestingly, fatigue severity emerged as a significant predictor of subjective engagement in condition C and insomnia severity emerged as a significant predictor of subjective engagement in condition D. These results indicate that engagement was decreased on more cognitively demanding tasks as participants’ habitual experience of impairment due to their fatigue symptoms increased. Moreover, these findings were robust enough to remain significant despite order effects in the dependent variable (see results of the mixed model ANOVA for VAS-E). These results are consistent with findings that clinical populations experiencing significant mental fatigue (e.g., insomnia disorder, multiple sclerosis) often hold maladaptive or mistaken beliefs about fatigue and consequently attempt to conserve their energy by minimizing strenuous mental or physical activity to prevent worsening of their fatigue symptoms (Harris & Carney, 2012; Moss, Carney, Haynes, & Harris, 2015). Additionally, boredom at baseline remained a significant predictor of engagement in conditions C. Participants’ ratings of the high load conditions as only moderately difficult suggest that the task in condition C may not have been sufficiently cognitively demanding to overcome boredom. Another significant finding is that the decrease in cognitive resources from pre- to post-test significantly predicted decreased engagement in the high load-hard instruction condition, indicating the centrality of appraisals of available cognitive resources in our ability to remain subjectively engaged in difficult tasks. Lastly, despite significant correlations between *d'* and other covariates in conditions B, C, and D, none of the covariates emerged as significant predictors of performance on the *n*-back task. This may be explained by the fact that performance on all four cognitive tasks was upheld, resulting in little variability in participants' performance on the four tasks using this index; additionally, the maintenance in accuracy occurred independent of increases in fatigue and decreases in subjective engagement.

Correlations between measures of engagement. Consistent with the results of Hopstaken et al. (2015a), correlations between the different indices of engagement strengthened with task difficulty. Specifically, VAS-E and *d'* were not significantly related for the low load conditions but were significantly related for the high load conditions. Unfortunately, this did not extend to correlations involving pupil dynamics, which remained uncorrelated with the other engagement measures. This is in contrast with previous studies, in which baseline pupil diameter and stimulus-evoked pupil diameter were both correlated with subjective and behavioural indices of engagement, albeit not simultaneously (Hopstaken et al., 2015a, 2015b; van der Wel & van Steenbergen, 2018). Therefore, it is surprising that neither measure of pupil dynamics was associated with other indices of engagement. It is possible that significant relationships between the measures are masked by the variability in pupil diameter within each condition due to order effects.

Underload Hypothesis

We also wanted to eliminate the underload hypothesis as a competing explanation for the results, in which changes in performance (i.e., the psychomotor vigilance decrement) due to long times on task are associated with increases in sleepiness and boredom, rather than mental fatigue (Pattyn, Neyt, Hendrickx, & Soetens, 2008). Mixed model ANOVAs indicated that increases in boredom were related to order effects, with boredom increasing significantly for the first two to three blocks and reaching a ceiling effect by the fourth block. Similarly, participants' ratings of sleepiness increased from pre- to post-test in nearly all conditions within each level of randomization order, with the exception of conditions in which a ceiling effect was observed. Physiologically, sleepiness reflects an increase in pressure on the body's homeostatic sleep system to produce slow-wave sleep and is related to a build-up of adenosine in the basal forebrain as a by-product of cellular activity (i.e., using up energy) as time awake increases, and would also be expected to increase with mental activity as glycogen becomes depleted (Bjorness & Greene, 2009). However, participants should not be able to detect marked changes in sleepiness from pre- to post-test on each block given the short time span of the cognitive tasks because in parallel with the accrual of adenosine, there would also be minute increases in alerting chemical activity (e.g., histamines, glutamate, orexin) to compete with sleepiness. Additionally, participants were already quite sleepy at baseline, scoring just below the clinical cut-off score on the ESS, and it is possible that participants' excessive sleepiness simply became more apparent due to boredom during the tasks. Unfortunately, layperson definitions for fatigue, sleepiness and boredom often overlap and there is significant variability in how people define these constructs. This was supported in the present study by significant relationships between subjective measures of fatigue, cognitive resources, boredom, and sleepiness, indicating insufficient delineation (and measurement) of these distinct constructs. Therefore, a significant limitation of this research is that standardized definitions of fatigue and sleepiness, particularly with respect to their differences, were not provided to participants at the beginning of the study. Due to the concomitant changes in fatigue, cognitive resources, boredom, and sleepiness, the underload hypothesis cannot be definitively supported nor refuted.

It is plausible that the decreased engagement seen in the low load conditions may have been attributable to boredom as these tasks required psychomotor vigilance but did not include any working memory demands. Such a hypothesis is corroborated by the physiological findings that pupil size was smallest in the low load conditions, indicative of low arousal. This is consistent with the literature on the psychomotor vigilance decrement seen with tasks of sustained attention, which is associated with mindlessness and boredom due to the monotony of the task and the subjective experience of the task demands as insufficiently challenging, meaningless, repetitive and/or monotonous, resulting in task disengagement (Pattyn et al., 2008; Smallwood et al., 2004). However, in spite of decreased subjective and physiological engagement during the low load tasks, behavioural performance was upheld according to *d'*, reaction time, and miss rate. This suggests that the low load tasks were associated with enough "underload" to result in decreased effort and subjective disengagement but were not sufficiently mindless to translate into impaired performance. Collectively, these results indicate that boredom and sleepiness do not account for the primary analysis findings.

Furthermore, notwithstanding the fact that the background colour of the experiment was made grey in an effort to reduce eye strain, it is possible that participants experienced visual fatigue nonetheless as a result of high screen luminance (i.e., the brightness of the screen; Benedetto, Carbone, Drai-Zerbib, Pedrotti, & Baccino, 2014) and long times on task (Park et al., 2017). Visual fatigue can contribute to visual discomfort, dry eyes and fewer blinks and less efficient vision, which can impair visual task performance (Sheedy, 1992). In this vein, it may be that participants were interpreting symptoms of visual fatigue (e.g., dry eyes requiring longer blinks) as symptoms of sleepiness.

Limitations

Several limitations of the present study warrant consideration. An undergraduate sample was selected to understand the interactions between task demands and appraisals of task difficulty in healthy individuals. Previous research has criticized the use of undergraduate student samples in psychology research as lacking generalizability to adult populations (e.g., Hanel & Vione, 2016). However, others have made the case for the use of undergraduate analogue samples for the purpose of testing theoretical models on the basis that 1) psychological phenomena of interest exist on a continuum, wherein clinical samples exhibit the same phenomena to a higher degree, and 2) the feasibility of recruiting analogue samples allows for the use of more complex experimental study designs (Flett, Vrendenburg & Krames, 1997; Ree, Pollitt, & Harvey, 2006; Stopa & Clark, 2001). Ironically, although intended to serve as a healthy sample, this sample was actually a highly fatigued and sleepy group at baseline, in addition to reporting mild levels of depression and anxiety, which also happen to be associated with clinical levels of fatigue. The effect of these baseline characteristics suggest that ceiling effects may be a significant limitation in understanding the relationships between cognitive tasks, fatigue, sleepiness and engagement in healthy adults. In addition, the present sample was predominantly White and female which may preclude generalization of the results to non-White, non-female populations.

The fatigue manipulation was not sufficiently deceiving to examine the interaction between cognitive load and appraisals of task demands/effort, precluding a test of the hypothesis that such an interaction would produce differing levels of fatigue, and that the differences in fatigue would subsequently predict varying levels of task engagement; that is, although the font colour was changed for one of each *0*- and *2*-back conditions, participants were not convinced that this minor qualitative difference was responsible for changing others' perceptions of the task from very easy and requiring little effort to very difficult and requiring lots of effort. Future research should consider explicitly stating that font colour changes people's perceptions of the task being easy versus difficult to make the manipulation more believable (see Clarkson et al., 2010 for an example).

Further, a within-subjects design was chosen to maximize power and control for individual differences in the propensity to become fatigued and cognitive functioning. However, this study design proved detrimental to detecting a significant effect of the cognitive load × task instruction manipulation on ratings of fatigue and cognitive resources from pre- to post-test. Completion of the four *n*-back blocks in a single testing session resulted in order effects and ceiling effects for the fourth block, and it is clear that participants' optional 3-minute break between blocks was insufficient in restoring mental fatigue to normal levels. Lim and Kwok (2016) evaluated the effects of varying break lengths (1, 5, and 10 minutes) on attention performance using a time on task paradigm and found that performance on the subsequent block initially improved with longer break durations. However, in spite of these initial improvements in performance, the longest break was associated with a steeper decline in performance over the course of the following block as participants perceived their cognitive resources to be replenished and therefore expended greater effort, when they may have still been fatigued (Lim

& Kwok, 2016). Given the convergent evidence that breaks of both short and long duration are insufficient in restoring performance, future research should consider testing over the course of several days to mitigate order and ceiling effects, while controlling for baseline fatigue levels on each day. Employing such a study design would clarify whether the cognitive load × task instruction manipulation is effective in producing varying levels of fatigue without being confounded by other non-study related factors and would allow for the prediction of levels of engagement from changes in fatigue.

Finally, it is likely that the analyses including pupil diameter as a dependent variable were underpowered to detect an effect of condition within subjects in a single testing session. Unfortunately, 15 participants had to be excluded from these analyses due to noisy pupil data. Consistent with recommendations in the literature (Hayes & Petrov, 2016), a chin rest was purchased after 15 participants were tested using remote pupil tracking. The desktop-mounted apparatus was superior to remote tracking due to unreliability of the pupil data as distance was not held constant when tested remotely, which changes the pupil foreshortening error and the size (area) of the pupil (Hayes & Petrov, 2016). The significant relationship between an increase in fatigue and a decrease in pupil diameter in the low load-easy instruction condition suggests that this may be a fruitful avenue for future research, and removal of the effect of condition presentation order may provide further clarity on the effect of appraisals of task difficulty on pupil dynamics. Future research seeking to predict physiological engagement from changes in fatigue should seek to recruit larger sample sizes.

Implications

Convergent evidence indicates that time on task, particularly doing effortful tasks, increases fatigue and decreases task engagement. These results have implications for the workplace and the classroom with respect to fatigue interventions, including taking frequent breaks and managing the cognitive demands of workplace activities such that they are not so low that boredom and disengagement ensues, and not so high that they are exceedingly difficult. For example, the physiological results herein indicate that the challenge of the high load tasks increased physiological engagement, but when asked, participants rated these tasks as only moderately difficult.

Furthermore, these results contribute to the clinical literature that clinically fatigued individuals disengage from cognitively demanding tasks. These results underscore the need for more effective fatigue interventions, particularly education as to the nature of fatigue, its various causes, and the importance of activity in reducing fatigue. Previous research has found that such educational interventions have been successful (Harris & Carney, 2012).

Conclusion

In conclusion, this study demonstrated an inverse relationship between ratings of fatigue and ratings of available cognitive resources, and concurrent increases in fatigue and decreases in ratings of cognitive resources with increasing time on task. These results provide preliminary evidence that appraisals of task demands are central to the experience of mental fatigue. However, the methodology used in this study did not directly test hypothesis that following an appraisal of demands, individuals immediately appraise their own resources because VAS-C was completed before the *n*-back but not post-instruction. Consequently, future research should evaluate the mediating role of appraisals of cognitive resources in the phenomenon of mental fatigue, with additional ratings of VAS-F and VAS-C completed immediately after receiving the instruction and prior to completing the cognitive task. In addition to providing the opportunity to evaluate the mechanistic role of appraisals of cognitive resources, such methodology would allow researchers to directly assess the impact of verbal instructions on perceptions of fatigue and cognitive resources and to evaluate the relative contributions of subjective momentary appraisals and cognitive load in the development of mental fatigue.

With respect to engagement, contrary to what was hypothesized, subjective and physiological engagement were lower for low load tasks and higher for high load tasks. Future studies should investigate how differing task difficulties correspond with task engagement to identify optimal task difficulties to maintain engagement. These results contribute to the vast literature that fatigue increases with time on task and makes a significant contribution that ratings of cognitive resources covary with this increase. Additionally, this study revealed that engagement is highest when participants receive an "unpleasant surprise," but find this expectation violation taxing; this interesting finding merits future investigation. Although this study makes several unique contributions to the literature on fatigue and engagement, understanding the mechanistic role of appraisals of cognitive resources in the development of fatigue and subsequent task engagement is an important next step.

References

Ackerman, P. L., Kanfer, R. (2009). Test length and cognitive fatigue: An empirical examination of effects on performance and test-taker reactions. *Journal of Experimental Psychology: Applied, 15,* 163-181.

Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M. W., & Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychological Assessment*, *10*(2), 176–181. https://doi.org/10.1037/1040-3590.10.2.176

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.

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

Bailey, A., Channon, S., & Beaumont, J. G. (2007). The relationship between subjective fatigue and cognitive fatigue in advanced multiple sclerosis. *Multiple Sclerosis, 13,* 73-80.

Bakker, A. B., Demerouti, E., & Sanz-Vergel, A. I. (2014). Burnout and work engagement: The JD-R approach. *Annual Review of Organizational Psychology and Organizational Behavior, 1,* 389-411.

Balkin, T. J., & Wesensten, N. J. (2011). Differentiation of sleepiness and mental fatigue effects. In P.L. Ackerman (Ed.), *Cognitive fatigue: multidisciplinary perspectives on current research and future applications* (pp. 47-66). Washington, DC: American Psychological Association.

Bandura, A. (1986). *Social foundations of thought and action.* New York, NY: Prentice-Hall.

Baranski, J. V. (2011). Sleep loss and the ability to self-monitor cognitive performance. In P. L. Ackerman (Ed.), *Cognitive fatigue: Multidisciplinary perspectives on current research and future applications* (pp. 67-82). Washington, DC: American Psychological Association.

Barcikowski, R. S., & Robey, R. R. (1984). Decisions in single group repeated measures analysis: Statistical tests and three computer packages. *The American Statistician, 38,* 148-150.

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,* 1252-1265.

Baumeister, R. F., Muraven, M., & Tice, D. M. (2000). Ego depletion: A resource model of volition, self-regulation, and controlled processing. *Social Cognition, 18,* 130-150.

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science, 16,* 351-356.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276–292.

Benedetto, S., Carbone, A., Drai-Zerbib, V., Pedrotti, M., & Baccino, T. (2014). Effects of luminance and illuminance on visual fatigue and arousal during digital reading. *Computers in Human Behavior, 41,* 112-119.

Bernard, L. C., & Fowler, W. (1990). Assessing the validity of memory complaints: Performance of brain-damaged and normal individuals on Rey's task to detect malingering. *Journal of Clinical Psychology, 46,* 434-436.

Bjorness, T. E., & Greene, R. W. (2009). Adenosine and sleep. *Current Neuropharmacology, 7,* 238-245.

Boksem, M. A. S., Meijman, T. F., & Lorist, M. M. (2006). Mental fatigue, motivation and action monitoring. *Biological Psychology, 72,* 123-132.

Boksem, M. A. S., & Tops, M. (2008). Mental fatigue: Costs and benefits. *Brain Research Reviews, 59,* 125-139.

Buysse, D. J., Ancoli-Israel, S., Edinger, J. D., Lichstein, K. L., & Morin, C. M. (2006). Recommendations for a standard research assessment of insomnia. *Sleep*, *29*(9), 1155–1173.

Borragán, G., Slama, H., Bartolomei, M., & Peigneux, P. (2017). Cognitive fatigue: A time-based resource-sharing account. *Cortex, 89,* 71-84.

Chaudhuri, A., & Behan, P. O. (2004). Fatigue in neurological disorders. *The Lancet, 363,* 978-988.

Chaudhuri, A., & Behan, P. O. (2000). Fatigue and basal ganglia. *Journal of the Neurological Sciences, 179,* 34-42.

Clarkson, J. J., Hirt, E. R., Jia, I., & Alexander, M. B. (2010). When perception is more than reality: The effects of perceived versus actual resource depletion on self-regulatory behavior. *Journal of Personality and Social Psychology, 98,* 29-46.

DeLuca, J., Johnson, S. K., Ellis, S. P., & Natelson, B. H. (1997). Cognitive functioning is impaired in patients with chronic fatigue syndrome devoid of psychiatric disease. *Journal of Neurology, Neurosurgery, and Psychiatry, 62,* 151-155.

Dobryakova, E., DeLuca, J., Genova, H. M., & Wylie, G. R. (2013). Neural correlates of cognitive fatigue: Cortico-striatal circuitry and effort-reward imbalance. *Journal of the*

International Neuropsychological Society, 19, 1-5.

Durmer, J. S., & Dinges, D. F. (2009). Neurocognitive Consequences of Sleep Deprivation. *Seminars in Neurology*, *25*(1), 117–129.

Earle, F. (2004). *The construct of psychological fatigue: A psychometric and experimental analysis* (Unpublished doctoral dissertation). University of Hull, Hull, England.

Evans, D. R., Boggero, I. A., & Segerstrom, S. C. (2015). The nature of self-regulatory fatigue and "ego depletion": Lessons from physical fatigue. *Personality and Social Psychology Review.*

Fernández-Castro, J., Martínez-Zaragoza, F., Rovira, T., Ego, S., Solanes-Puchol, Á., Martín-del-Río, B., … Doval, E. (2017). How does emotional exhaustion influence work stress? Relationships between stressor appraisals, hedonic tone, and fatigue in nurses' daily tasks: A longitudinal cohort study. *International Journal of Nursing Studies, 75,* 43-50.

Flett, G. L., Vrendenburg, K., & Krames, L. (1997). The continuity of depression in clinical and nonclinical samples. *Psychological Bulletin, 121,* 395-416.

Geiger-Brown, J., Rogers, V. E., Trinkoff, A. M., Kane, R. L., Bausell, R. B., & Scharf, S. M. (2012). Sleep, sleepiness, fatigue, and performance of 12-hour-shift nurses. *Chronobiology International, 29,* 211-219.

Gift, A. G. (1989). Visual analogue scales: Measurement of subjective phenomena. *Nursing Research*, *38*(5), 286–288.

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective and Behavioral Neuroscience*, *10*(2), 252–269.

Goldberg, J. O., & Miller, H. R. (1986). Performance of psychiatric inpatients and intellectually deficient individuals on a task that assesses the validity of memory complaints. *Journal of Clinical Psychology, 42,* 792-795.

Goldman-Mellor, S., Caspi, A., Gregory, A. M., Harrington, H. L., Poulton, R., Moffitt, T. E., et al. (2015). Is insomnia associated with deficits in neuropsychological functioning? Evidence from a population-based study. *Sleep, 38*(4), 623-631.

Gunzelmann, G., Moore, L. R., Gluck, K. A., Van Dongen, H. P. A., & Dinges, D. F. (2011). Fatigue in sustained attention: Generalizing mechanisms for time awake to time on task. In P.L. Ackerman (Ed.), *Cognitive fatigue: multidisciplinary perspectives on current research and future applications* (pp. 83-101). Washington, DC: American Psychological Association.

Hanel, P. H. P., & Vione, K. C. (2016). Do student samples provide an accurate estimate of the general public? *PLoS One, 11,* e0168354.

Harris, A. L., & Carney, C. E. (2012). Can we modify maladaptive attributions for fatigue? *Cognitive Behaviour Therapy, 41*(1), 40-50.

Hayes, T. R., & Petrov, A. A. (2016). Mapping and correcting the influence of gaze position on pupil size measurements. *Behavior Research Methods, 48*(2), 510-527.

Henry, J. D., & Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, *44*(2), 227–239.

Hockey, G. R. (2011). A motivational control theory of cognitive fatigue. In P.L. Ackerman (Ed.), *Cognitive fatigue: multidisciplinary perspectives on current research and future applications* (pp. 167-188). Washington, DC: American Psychological Association.

Hopstaken, J. F., van der Linden, D., Bakker, A. B., & Kompier, M. A. J. (2015a). A multifaceted investigation of the link between mental fatigue and task disengagement. *Psychophysiology*, *52*(3), 305–315.

Hopstaken, J. F., van der Linden, D., Bakker, A. B., & Kompier, M. A. J. (2015b). The window of my eyes: Task engagement and mental fatigue covary with pupil dynamics. *Biological Psychology, 110,* 100-106.

Horne, J. A., & Östberg, O. (1976). A self-assessment questionnaire to determine morningness and eveningness in human circadian rhythms. *International Journal of Chronobiology*, *4*, 97–110.

Inzlicht, M., & Marcora, S. M. (2016). The central governor model of exercise regulation teaches us precious little about the nature of mental fatigue and self-control failure. *Frontiers in Psychology, 7,* 656.

Inzlicht, M., & Schmeichel, B. J. (2012). What is ego depletion? Toward a mechanistic revision of the resource model of self-control. *Perspectives on Psychological Science, 7,* 450-463.

Iverson, G. L., & Franzen, M. D. (1996). Using multiple objective memory procedures to detect simulated malingering. *Journal of Clinical and Experimental Neuropsychology*, *18*(1), 38–51.

Jepma, M., & Nieuwenhuis, S. (2011). Pupil diameter predicts changes in the exploration–exploitation trade-off: Evidence for the Adaptive Gain Theory. *Journal of Cognitive Neuroscience*, *23*(7), 1587–1596.

Job, V., Dweck, C. S., & Walton, G. M. (2010). Ego depletion—is it all in your head? Implicit theories about willpower affect self-regulation. *Psychological Science, 21,* 1686-1693.

Johns, M. W. (1991). A New Method for Measuring Daytime Sleepiness: The Epworth Sleepiness Scale. *Sleep*, *14*(6), 540–545.

Kahneman, D. (1975). Attention and effort. *The American Journal of Psychology, 88,* 339.

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

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*(3756), 1583–1585.

Kahol, K., Leyba, M. J., Deka, M., Deka, V., Mayes, S., Smith, M., … Panchanathan, S. (2008). Effect of fatigue on psychomotor and cognitive skills. *The American Journal of Surgery, 195,* 195-204.

Kanfer, R. (2011). Determinants and consequences of subjective cognitive fatigue. In P. L. Ackerman (Ed.), *Cognitive fatigue: Multidisciplinary perspectives on current research and future applications* (pp. 189-207). Washington, DC: American Psychological Association.

Käthner, I., Wriessnegger, S. C., Müller-Putz, G. R., Kübler, A., & Halder, S. (2014). Effects of mental workload and fatigue on the P300, alpha and theta band power during operation of an ERP (P300) brain-computer interface. *Biological Psychology, 102*, 118e129.

Krupp, L. B., LaRocca, N. G., Muir-Nash, J., & Steinberg, A. D. (1989). The Fatigue Severity Scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. *Archives of Neurology*, *46*(10), 1121–1123.

Lerdal, A., Wahl, A. K., Rustoen, T., Hanestad, B. R., & Moum, T. (2005). Fatigue in the general population: A translation and test of the psychometric properties of the Norwegian version of the fatigue severity scale. *Scandinavian Journal of Public Health*, *33*, 123–130.

Lezak, M. D. (1983). *Neuropsychological assessment* (2nd ed.). New York, NY: Oxford University Press.

Lim, J., & Kwok, K. (2016). The effects of varying break length on attention and time on task. *Human Factors, 58*(3)*,* 462-481.

Lurquin, J. H., & Miyaki, A. (2017). Challenges to ego-depletion research go beyond the replication crisis: A need for tackling the conceptual crisis. *Frontiers in Psychology, 8,* 568.

Marcora, S. M., Staiano, W., & Manning, V. (2009). Cognitive fatigue impairs physical performance in humans. *Journal of Applied Physiology, 106,* 857-864.

Meijman, T. F. (2000, July). *The theory of the stop-emotion: On the functionality of fatigue.* Paper presented at the conference for Ergonomics and Safety for Global Business Quality and Production, Warsaw, Poland.

Morin, C. M. (1996). *Insomnia: psychological assessment and management*. New York, NY: Guilford Press.

Morin, C. M., Belleville, G., Bélanger, L., & Ivers, H. (2011). The Insomnia Severity Index: Psychometric indicators to detect insomnia cases and evaluate treatment response. *Sleep*, *34*(5), 601–608.

Moss, T. G., Carney, C. E., Haynes, P., & Harris, A. L. (2015). Is daily routine important for sleep? An investigation of social rhythms in a clinical insomnia population. *Chronobiology International, 32*(1), 92-102.

Muraven, M., & Slessareva, E. (2003). Mechanisms of self-control failure: Motivation and limited resources. *Personality and Social Psychology Bulletin, 29,* 894-906.

Muraven, M., Tice, D. M., & Baumeister, R. F. (1998). Self-control as a limited resource: Regulatory repletion patterns. *Journal of Personality and Social Psychology, 74,* 774-789.

Park, S., Choi, D., Yi, J., Lee, S., Lee, J. E., Choi, B., … Kyung, G. (2017). Effects of display curvature, display zone, and task duration on legibility and visual fatigue during visual search task. *Applied Ergonomics, 60,* 183-193.

Pattyn, N., Neyt, X., Henderickx, D., & Soetens, E. (2008). Psychophysiological investigation of vigilance decrement: Boredom or cognitive fatigue? *Physiology & Behavior, 93,* 369-378.

Rajkowski, J., Lu, W., Zhu, Y., Cohen, J., & Aston-Jones, G. (2000). Prominent projections from the anterior cingulate cortex to the locus coeruleus in Rhesus monkey. *Society for Neuroscience Abstracts, 26,* 838.15.

Ree, M. J., Pollitt, A., & Harvey, A. G. (2006). An investigation of interpretive bias in insomnia: An analog study comparing normal and poor sleepers. *Sleep, 29,* 1359-1362.

Rey, A. (1964). *L'examen clinique en psychologie [The clinical examination in psychology].* Paris: Presses Universitaires de France.

Reznek, L. (2005). The Rey 15-item memory test for malingering: A meta-analysis. *Brain Injury*, *19*(7), 539–543.

Sandry, J., Genova, H. M., Dobryakova, E., DeLuca, J., & Wylie, G. (2014). Subjective cognitive fatigue in multiple sclerosis depends on task length. *Frontiers in Neurology, 5,* 214.

Schretlen, D., Brandt, J., Krafft, L., & Van Gorp, W. (1991). Some caveats in using the Rey 15-item memory test to detect malingered amnesia. *Psychological Assessment, 3,* 667-672.

Sheedy, J. E. (1992). Reading performance and visual comfort on a high-resolution large monitor compared to a VGA monitor. *Journal of Electronic Imaging, 1,* 405-410.

Shigihara, Y., Tanaka, M., Ishii, A., Tajima, S., Kanai, E., Funakura, M., & Watanabe, Y. (2013). Two different types of mental fatigue produce different styles of task performance. *Neurology, Psychiatry, and Brain Research, 19,* 5-11.

Singh, R. K., & Göritz, A. S. (2018). Ego depletion does not interfere with working memory performance. *Frontiers in Psychology, 9,* 538.

Smallwood, J., Davies, J. B., Heim, D., Finnigan, F., Sudberry, M., O'Connor, R., & Obonsawin,

M. (2004). Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness and Cognition, 13,* 657-690.

Smith, C. S., Reilly, C., & Midkiff, K. (1989). Evaluation of three circadian rhythm questionnaires with suggestions for an improved measure of morningness. *Journal of Applied Psychology, 74,* 728-738.

Stopa, L., & Clark, D. M. (2001). Social phobia: Comments on the viability and validity of an analogue research strategy and British norms for the Fear of Negative Evaluation Questionnaire. *Behavioural and Cognitive Psychotherapy, 29,* 423-430.

Tanaka, M., Ishii, A., & Watanabe, Y. (2014). Neural effects of mental fatigue caused by continuous attention load: A magnetoencephalography study. *Brain Research, 1561,* 60-66.

Thorndike, E. L. (1900). Mental fatigue: I. *Psychological Review, 7,* 466-482.

Tops, M., Boksem, M. A. S., & Koole, S. L. (2013). Subjective effort derives from a neurological monitor of performance costs and physiological resources: A commentary on Kurzban et al.'s "An opportunity cost model of subjective effort and task performance". *Behavioral and Brain Sciences, 36,* 703-704.

Tops, M., Lorist, M. M., Wijers, A. A., & Meijman, T. F. (2004). To stress or to relax: Neurochemical aspects of activity and rest. *Gedrag en Organisatie, 17,* 32-42.

Torsvall, L., & Akerstedt, T. (1980). A diurnal type scale. Construction, consistency and validation in shift work. *Scandinavian Journal of Work Environment Health, 6,* 283-290.

Trejo, L. J., Kochavi, R., Kubitz, K., Montgomery, L. D., Rosipal, R., & Matthews, B. (2005). Measures and models for predicting cognitive fatigue. *Proceedings of the SPIE, 5797,* 105-115.

Tseng, B. Y., Gajewski, B. J., & Kluding, P. M. (2010). Reliability, responsiveness, and validity of the Visual Analog Fatigue Scale to measure exertion fatigue in people with chronic stroke: A preliminary study. *Stroke Research and Treatment*, *2010*, 1–7.

Tsukahara, J. S., Harrison, T. L., & Engle, R. W. (2016). The relationship between baseline pupil size and intelligence. *Cognitive Psychology, 91,* 109-123.

Valko, P. O., Bassetti, C. L., Bloch, K. E., Held, U., & Baumann, C. R. (2008). Validation of the Fatigue Severity Scale in a Swiss cohort. *Sleep*, *31*(11), 1601–1607.

van den Brink, R. L., Murphy, P. R., & Nieuwenhuis, S. (2016). Pupil diameter tracks lapses of attention. *PLoS One*, *11*(10), e0165274.

van der Linden, D. (2011). The urge to stop: The cognitive and biological nature of acute mental fatigue. In P. L. Ackerman (Ed.), *Cognitive fatigue: Multidisciplinary perspectives on current research and future applications* (pp. 149-164). Washington, DC: American Psychological Association.

van der Linden, D., Frese, M., & Meijman, T. F. (2003). Mental fatigue and the control of cognitive processes: Effects on perseveration and planning. *Acta Psychologica, 113,* 45-65.

van der Linden, D., Frese, M., & Sonnentag, S. (2003). The impact of mental fatigue on exploration in a complex computer task: Rigidity and loss of systematic strategies. *Human Factors, 45,* 483-494.

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,* 2005-2015.

Van Dongen, H. P. A., Baynard, M. D., Maislin, G., & Dinges, D. F. (2004). Systematic interindividual differences in neurobehavioral impairment from sleep loss: Evidence of trait-like differential vulnerability. *Sleep, 27,* 423-433.

Watter, S., Geffen, G. M., & Geffen, L. B. (2001). The n-back as a dual-task: P300 morphology under divided attention. *Psychophysiology*, *38*(6), 998–1003.

Wright, K. P., Lowry, C. A., & LeBourgeois, M. K. (2012). Circadian and wakefulness-sleep modulation of cognition in humans. *Frontiers in Molecular Neuroscience, 5,* 50.