Title: Pupil response speed as a marker of cognitive fatigue in early Multiple Sclerosis.

Guillemin, C.[1,2], Hammad, G.[1], Read, J.[1], Requier, F.[1,2], Charonitis, M.[1,2], Delrue, G.[3], Vandeleene, N.[1], Lommers, E.[1,3], Maquet, P.[1,,3] & Collette, F.[1,2]

[1]*GIGA-CRC In Vivo Imaging, University of Liège, Liège, Belgium*
[2]*Psychology and Cognitive Neuroscience Research Unit, University of Liège, Liège, Belgium*
[3]*Department of Neurology, CHU of Liège Sart Tilman, Liège, Belgium*

Correspondance: Fabienne Collette: *f.collette@uliege.be*

AIM: To assess the effect of cognitive load on performance, subjective scales (fatigue and sleepiness) and eye metrics in early pwMS and HC

Abstract

Context: Cognitive fatigue (CF) is a disabling symptom frequently reported by patients with multiple sclerosis (pwMS). Whether pwMS in the early disease stages present an increased sensitivity to fatigue induction remains debated. Objective measures of CF have been validated neither for clinical nor research purposes. This study aimed at (i) assessing how fatigue induction by manipulation of cognitive load affects subjective fatigue and behavioral performance in newly diagnosed pwMS and matched healthy controls (HC); and (ii) exploring the relevance of eye metrics to describe CF in pwMS.

Methods: Nineteen pwMS with disease duration <5 years and 19 matched HC participated to this study. CF was induced with a dual-task in two separate sessions with varying cognitive load (High and Low cognitive load conditions, HCL and LCL). Accuracy, reaction times (RTs), subjective fatigue and sleepiness states were assessed. Bayesian Analyses of Variance for repeated measures (rmANOVA) explored the effects of time, group and load condition on the assessed variables. Eye metrics (number of long blinks, pupil size and pupil response speed: PRS) were obtained during the CF task for a sub-sample (16 pwMS and 15 HC) and analyzed with Generalized Linear Mixed Models (GLMM).

Results: Performance (accuracy and RTs) was lower in the HCL condition and accuracy decreased over time (BFs$_{incl}$ > 100) while RTs did not significantly vary. Performance over task and conditions followed the same pattern of evolution across groups (BFs$_{incl}$ <0.08) suggesting that pwMS did not show increased alteration of performance during fatigue induction. Regarding subjective state, both fatigue and sleepiness increased following the task (BFs$_{incl}$ >15), regardless of condition and group (BFs$_{incl}$ <3). CF in pwMS seems to be associated with PRS, as PRS decreased during the task among pwMS only and especially in the HCL condition (all p <.05). A significant Condition*Group interaction was observed regarding long blinks (p <.0001) as well as an expected effect of cognitive load condition on pupil diameter (p <.01).

Conclusion: These results suggest that newly diagnosed pwMS and HC behave similarly during fatigue induction, in terms of both performance decrement and accrued fatigue sensation. Eye metric data further reveal a susceptibility to CF in pwMS, which can be objectively measured.

Keywords: Early Multiple Sclerosis; Cognitive Fatigue; Cognitive Fatigability; Cognitive Load; Eye Metrics; Pupillometry

Introduction

Fatigue is highly prevalent in multiple sclerosis (MS)[1,2] and about two thirds of patients consider it as one of their worst symptoms[3]. Fatigue alters quality of life, and jeopardizes employment and occupational status[4,5]. Cognitive and motor fatigue are considered as distinct entities[6,7], with the former corresponding to a sensation of mental exhaustion, and pwMS can report one or the two. By comparison to healthy controls (HC), patients with MS (pwMS) report more frequent and severe fatigue sensation as measured subjectively (i.e. perceived feeling of exhaustion)[8]. Additionally, pwMS present faster and/or steeper decrease in performance during prolonged cognitive tasks (i.e. objective cognitive fatigability)[8], which is suggestive of an increased sensitivity to cognitive fatigue (CF). However, this might be due to a higher cognitive demand, rather than a specific sensitivity to fatigue, as the same task is usually proposed to both patients and controls[9,10]. Accordingly, Borragán and colleagues[11] showed no difference in performance decrement across time between pwMS and matched controls when cognitive load is adapted to participants' abilities.

Current objective measures of cognitive fatigability, such as decreased performance at the PASAT[12] (a challenging arithmetic cognitive test) rely on cognitive functions that are frequently impaired in pwMS (processing speed, attention, working memory…)[9], thus implying a higher cognitive demand which could explain fatigue vulnerability. Besides, the decrease in performance associated with a demanding or prolonged cognitive task is usually not related to the subjective feeling of fatigue[8]. It has been proposed that the latter arises as a bodily signal that encourages the interruption of an ongoing task when efficient goal pursuit is threatened[13], i.e. before any visible alteration of performance.

Intrinsic and extrinsic eye movements, as well as eyelid activity, are useful to measure fatigue[14,15]. The frequency and duration of eye blinks reliably signal accumulating fatigue and sleepiness[15]. Likewise, pupil photomotor reflex (i.e. pupil constriction to light and dilation to darkness) is blunted in pwMS in proportion to time on task, as expected for a fatigue biomarker[16–18], but its reliability as a proxy to CF remains debated[19].

In this context, this study aims to assess CF induced by a dual task (Time Load Dual Back: TLDB[11]), with individually adjusted cognitive load. In addition, in a subsample of participants, we explore the link between CF and eye metrics in the early stage of the disease, to assess if eye metrics are modulated by time-on-task and cognitive load in HC and pwMS. Due to individually adjusted cognitive demand, we expect a similar decrease of performance across time in the two groups. However, we expect an increased subjective fatigue level in pwMS compared to HC, and that pupil variables will be sensitive measures of fatigability.

Methods

Participants

Nineteen pwMS and 19 HC matched on sex, age and education participated in the study. Inclusion criteria for participants in the pwMS group were a disease duration below or equal to 5 years and a score at the Expanded Disability Status Scale[20] (EDSS) under 4. Patients presented with either a Relapsing-Remitting (RR) course or a Clinically Isolated Syndrome (CIS), according to the 2017 McDonald criteria[21] and were free from relapse for at least 6 months prior to the study. Exclusion criteria for both groups comprised the existence of other neurological or psychiatric diseases, a history of mild or severe traumatic brain injury, the use of medication impacting fatigue state and/or alertness (see supplemental material for details), substance abuse, colorblindness, native language other than French, incompatibility with Magnetic Resonance environment and age above 45 years old.

Study Design

This study was approved by the local ethics committees of the University Hospital of Liège (B707201835630). Participants received detailed information on the protocol and gave written consent prior to the study (this study is part of a larger research project, the details of which are available following this link:

https://osf.io/egr6d/?view_only=cdcc343c7d71406685b038e46e88145b). Participants were asked to observe a stable sleep-wake cycle during their participation, with a sleep duration of at least 7 hours per night, and to avoid stimulating beverage 24 hours prior an appointment. Participants were tested in a quiet environment, with temperature and light kept constant. To limit circadian confounds, appointments for a single participant were scheduled at the same time of the day, according to her/his preferences.

Data from three separated sessions are reported here. The first session consisted in participants filling-in questionnaires and performing a first training and calibration of the cognitive task (TLDB[22], see description below). Questionnaires assessed factors that could influence fatigue induction, namely trait fatigue (Fatigue Scale for Motor and Cognitive Functions: FSMC[23]), sleep quality (Pittsburgh Sleep Quality Index: PSQI[24]), daytime sleepiness (Epworth Sleepiness Scales: ESS[25]), anxiety (State-Trait Anxiety Inventory: STAI[26]) and depressive mood (Beck's Depression Inventory, second edition: BDI-II[27]). During sessions 2 and 3, the effects of mental load on CF were assessed by means of two conditions (High Cognitive Load: HCL; Low Cognitive Load: LCL) that have been counterbalanced across sessions.

Fatigue Induction Protocol and Measures

The TLDB is a computerized dual task during which letters and digits are presented alternatively on a computer screen, combining a number parity judgment task and a classical N-Back working memory task[28]. Participants were instructed to determine if each digit was odd or even, and if each letter was the same as the previous one. By adjusting the time available to process stimuli, cognitive load induced by the task can be manipulated, while maintaining the same task complexity. Practically, a calibration of the task was performed in each session, during which the shortest Stimulus Time Duration (STD) at which participant's accuracy remained above 85% was assessed. This STD was used during the HCL condition of the task, and in the LCL, the STD was 50% slower (STD + ½ STD). Consequently, the pace of the task was adapted for each participant to induce the same level of cognitive load across participants, in both the high and low load conditions. This design is particularly relevant here, as pwMS frequently present deficits in information processing speed and working memory[9]. We used a 32-minutes extended version of the original 16-minute TLDB[22], as results from a previous study suggest that CF starts to arise at the very end of the original task in pwMS[11].

In order to assess the evolution of subjective states, a sleepiness scale (the Karolinska Sleepiness Scale, KSS[29]) and Visual Analog Scales (VAS[30]) were administered prior to and following the TLDB task.

Effects of Cognitive Load on Subjective State and Performance

Bayesian statistical inference[31] assessed the effect of fatigue on cognitive performance and subjective states, using JASP (Jeffreys's Amazing Statistics Program v.0.16; https://jasp-stats.org). Results are reported in terms of Bayes Factor (BF) which corresponds to the likelihood ratio of evidence provided by the data over two hypotheses (BF_{10} represents the marginal likelihood of the alternative hypothesis H1 over the one of the null hypothesis H0. BF_{incl} represents the likelihood of adding an effect in the statistical model over the null model). In short $BF_{10/incl}$ values <1 are in favor of the null hypothesis/model, and values >1 in favor of the alternative hypothesis/tested model. Significant results were determined using Jeffreys's grades of evidence[32] (Table 1).

Between-group differences on demographic data, self-reported questionnaires and STD were evaluated using Bayesian two sample t-tests. STD analysis was done to assess if pwMS needed a slower presentation rate than HC to maintain similar performance.

Performance at the TLDB was measured by means of accuracy and RTs, averaged across four blocks of 8 minutes approximately. The first 60 trials were considered as habituation trials and were removed from analysis. Weighted accuracy (correct answers/number of items) was extracted, with digits and letters accounting for 65 and 35% of the total score, respectively[22]. RTs were extracted for digits and

letters separately[1] and only successful trials were considered in analysis. Effects of cognitive load on performance were investigated using a Group (pwMS vs controls) x Condition (HCL vs LCL) x Block (4) Bayesian analysis of variance for repeated measures (rmANOVA). Similarly, effects of cognitive load on subjective fatigue and sleepiness were assessed by Group x Condition x Time (before vs after TLDB) Bayesian rmANOVAs. For all rmANOVAs, estimation of effects (BF$_{incl}$) was computed across matched models.

Effects of Fatigue Induction on Eye Metrics

During the TLDB task, images of the participant's right eye were recorded at a frequency of 120 Hz by a camera positioned on a portable glass-like device: the Drowsimeter R100 from Phasya© (Phasya s.a., Belgium https://www.phasya.com/en/drowsimeter-r100). Corrective lenses were available and could be integrated to the device for participants wearing glasses.

Objective measures (eyelid gap and pupil size in pixels) were extracted from each image with the Drowsilogic software from Phasya (v.4.2.5), generating time series with approximately 230k points for our 32-minutes long TLDB task. The pupil size signal was preprocessed before analysis by first suppressing blinks and spurious periods. For each block of approximately 8 minutes mean pupil size, number of long blinks (duration >300ms) and mean pupil variation speed during dilation and constriction were computed. Number of long blinks, rather than blink frequency, were considered to have a measure of fatigue and sleepiness states independent from the confounding effects of stimulus presentation rate (see supplemental material for a detailed presentation of preprocessing steps).

Frequentist generalized linear mixed models (GLMM) were used for eye metric analyses. This method minimizes the influence of removal of sessions with low data quality and enables a control of the inter-subject variability of pupil size.[2] GLMM analyses were implemented in SAS 14.2 (SAS Institute), with the subject (intercept) entered as random effect and a type III sum of squares hypothesis testing. The estimation of degrees of freedom was performed using Kenward-Roger's correction[33]. A statistical threshold for significance of *p* <.05 was used and results of post-hoc analysis conducted on significant effects were corrected for multiple testing with a Tukey adjustment. The tested models evaluated the effects of time (Blocks 1 to 4), condition (HCL vs LCL) and group (pwMS vs HC) on each dependent variable, along with interaction effects. Pupil response speed (PRS) was also analyzed as a dependent variable, in a single model testing the effects of time, condition, group and pupillary response (Dilation

---

[1] As the number of target letters is much lower than the number of prompted digits, RTs of the two kind of items were analyzed separately. Besides, working memory target detection and parity judgment tasks rely on distinct cognitive processes that might differ regarding processing speed/RTs.

[2] Pupil size was expressed in pixels and depended on the distance between the recorded eye and the camera. Hence, the value cannot be considered as metric and varies across participants due to subject's head morphology. It thus seems important to consider a random effect associated to each participant.

vs Constriction), as well as their interactions. Effect sizes of significant effects were estimated with semi-partial R^2 [34].

Material used for analyses is openly available at:
https://osf.io/egr6d/?view_only=cdcc343c7d71406685b038e46e88145b

Results

Effects of cognitive load on subjective state and performance

Demographics, Questionnaires and Task Calibration

No between-group difference was observed for sex, age and education or trait fatigue (Table 2). Eight (42%) patients reached the FSMC cutoff score for severe fatigue, 5 (26%) for moderate fatigue and the remaining 6 (32%) had normal to mild fatigue. In HC, 2 (11%) participants had severe fatigue, 7 (37%) had moderate fatigue, and the remaining 10 (53%) had normal to mild fatigue symptoms. Regarding the other questionnaires (supplemental Table S1), all t-Tests were inconclusive (0.33 <BF_{10} <3) except for trait anxiety (STAI-Y2) suggesting an absence of between-group difference.

Overall, STD in the pwMS group was longer than in HC, suggesting that patients needed more time to process stimuli. Yet these differences did not exceed anecdotic evidence.

Performance: Accuracy and Reaction Times

Figure 1 displays the effects of block, group and condition on accuracy and mean RTs at the TLDB task (see Table S2 for accuracy and RTs values). Bayesian rmANOVA on accuracy (Table 3) revealed block and condition effects, with a gradual decrease of accuracy over time and lower accuracy in the HCL condition (BFs_{incl} >100). Post-hoc analysis showed significant differences in accuracy between block 1 and blocks 2 to 4 (corrected BFs_{10} >100), as well as between block 2 and block 4 (corrected BF_{10} > 100) and block 3 and 4 (corrected BF_{10} >10). Strong evidence for an absence of effect for the interactions Block*Group and Block*Condition*Group was observed (BFs_{incl} <0.1), suggesting that decrease in accuracy over time and condition was similar in pwMS and HC. Analyses on the remaining effects on accuracy were inconclusive.

An effect of condition was observed on RTs (Table 3) with increased RTs in the LCL condition, as expected. RTs for letter items did not vary over time (BF_{incl} <0.033), regardless of condition (Block*Condition, BF_{incl} <0.33). Regarding interaction effects including group, we found evidence for an absence of effect (BFs_{incl} <0.33) on RTs for the Block*Group and the Block*Condition*Group (for letters only) interactions. Again, these results suggest that RTs vary over time in a similar way for the two groups. We also observed a significant Condition*Group effect on RTs for digits (BF_{incl} >100), with pwMS showing slower RTs in the LCL condition, which can be driven by the time available to provide an answer in that condition. Analyses on the remaining effects on RTs were not conclusive.

Supplemental exploratory analyses were conducted on STD to determine if fatigue relates to individual load level. Bayesian Kendal's correlations were performed between STD and trait fatigue as measured by the FSMC, in each group separately (Table S6). We found moderate to very strong evidence for a correlation between STD and FSMC, in the pwMS group only (all $BFs_{10} > 3$). The strongest relationship was found between STD and the physical sub-scale of trait fatigue. In the HC group, all correlations were inconclusive. This result suggests that maximal load level is diminished in patients with a high fatigue complaint (see Figure S1, showing that pwMS with extreme RTs happen to experience severe trait fatigue according to the FSMC).

Evolution of Subjective States

As displayed in Table 4, subjective fatigue and sleepiness both increased after the task, regardless of condition (see Table S3 for fatigue and sleepiness values). We observed an absence of group effect on subjective state evolution (Time*Group for fatigue, and Time*Condition*Group for sleepiness, BFs_{incl} <0.33).

Effects of Fatigue Induction on Eye Metrics

A smaller sample of participants (16pwMS and 15HC) was included in the analyses due to data loss after quality control for eye metrics. Again, no between-group differences for demographics and trait fatigue were observed (Table 5).

Effects of fatigue induction on eye metrics are displayed in Figure 2. GLMM analysis on the number of long blinks revealed a significant Condition*Group interaction (p <.0001) as well as a trend for the simple effects of Block and Group (both ps =.06). Post-hoc analysis evidenced an effect of condition in both groups (p <.0001) with higher numbers of blinks during the HCL condition for HC, and in the LCL condition for pwMS. We also observed a trend for a between-group difference in the LCL condition only (p =.057) with patients showing more blinks than controls. Regarding pupil diameter, only a condition effect was observed, with larger pupil size in the HCL condition (p <.01). We observed a significant effect on pupillary response speed, with faster pupil changes during dilation compared to constriction (p <.05). The Block*Group and Block*Condition*Group interactions were also significant (p <.05). Post-hoc analysis showed a significant difference in PRS between block 1 and the last two blocks (p =.005 and .03, regardless of condition, p =.005 and .004 in the HCL condition), in the pwMS group only. This result demonstrates that PRS decreased over time for patients only, especially in the HCL condition, and regardless of response type (constriction vs. dilation).

Discussion

The main results are twofold: (1) pwMS and HC behaved in a similar way during fatigue induction by the TLDB task, with similar performance decrement and increased subjective fatigue; (2) PRS significantly decreased more in pwMS, as compared to HC, during the HCL task.

A majority of our pwMS presented moderate to severe fatigue, in line with the literature[35]. However, fatigue was also unexpectedly prevalent in HC. While fatigue in the healthy middle-aged population is not rare[36], the COVID-19 pandemic and its aftermaths probably explain this observation[37,38], as participants were recruited between 2019 and 2021. In any case, our sample is well suited to explore the effects of cognitive load on fatigue induction in MS, as the between-group differences observed cannot be attributed to mood nor sleep disorders despite their high prevalence in MS[39,40].

Patients with MS and HC showed a similar decrease in performance on the TLDB task (accuracy and RTs), suggesting a similar impact of fatigue induction on performance. These results confirm that, when cognitive demand is individually adapted, pwMS, at least those with limited disability only (mean EDSS score = 1.71 and 1.65, in the present study and in Borragán et al.[11], respectively), do not show increased performance fatigability, even for longer task duration[11]. Trait fatigue and cognition have often been linked in MS studies, with high fatigue scores repeatedly related to a worsening of processing speed[41–44]. Accordingly, STD was positively correlated to trait fatigue in pwMS only, meaning that patients with a fatigue complaint needed more time to process stimuli. As visually displayed (Figure S1), patients with high RTs values all presented severe fatigue according to the FSMC. Further studies should investigate whether performance at the TLDB decreases in pwMS according to their trait fatigue.

Subjective fatigue and sleepiness were not modulated by cognitive load condition. With a similar protocol, Borragán and colleagues[11] showed that subjective fatigue increased significantly following the HCL condition only, and no effect on sleepiness was observed. As our tasks were twice longer, increase in sleepiness and subjective fatigue during the LCL condition could be due to task length. Indeed, prolonged tasks induce fatigue, whatever their cognitive load, while short-duration tasks emphasize the differentiated effect of cognitive load[45].

Concerning eye blinks, an increased number of long blinks was observed at the end of the task, and pwMS tend to show more blinks than HC.[3] This might relate to the necessity for both groups to lubricate the eyes when cognitive demand is prolonged and fatigue arises[15], while the increased need for blinking in patients compared to HC could reflect a larger induction of fatigue and sleepiness. Finally, we observed more frequent long blinks in the HCL condition in HC and in the LCL condition for pwMS. We tentatively propose that the increased blink frequency in HC exclusively during HCL results from the high cognitive effort. By contrast, in pwMS, both conditions produced a high cognitive

---

[3] This result must be taken with caution as blinks rate shows very high inter-subject variability[15]

demand, yet the slow pace of stimulus presentation during the LCL condition offered the opportunity to produce more long blinks without missing items.

The pupil dilation induced by the cognitive effort and attention in high-load tasks [14,46] explain why pupil size was larger during HCL condition. By contrast, the effects of prolonged fatigue induction on pupil size variations are more controversial. While some studies report a link between mean pupil size and fatigue, others indicate that measures of pupil dynamics (as for example peak pupillary response) are more sensitive to detect fatigue[47]. Indeed, pupillary responses result from the sympathetic/parasympathetic balance, respectively inducing pupil dilation and constriction. We found a significant effect of response type (dilation vs. constriction) on PRS.[4] Moreover, we showed that PRS (regardless of response type) decreases as a function of time on task in pwMS only and especially in the HCL condition. Autonomic disturbance is frequent in MS[48], and corresponds to a blunted parasympathetic drive[16,17]. However, the absence of significant interaction between pupil response type (dilation vs. constriction) and time on task rather suggests that both autonomous systems are impaired[19], thought this result should be confirmed in further studies. Interestingly, the decrease in PRS could indicate an increased fatigue susceptibility in patients compared to controls, especially when cognitive load is high.

Conclusion

Despite the absence of behavioral between-group difference, pwMS remained sensitive to fatigue induction, especially when confronted to high mental load level. The better detection of CF by pupil responses possibly results from dependence of the latter on autonomic functions, which are disturbed early on during the course of disease. It is tantalizing to suggest that MS lesions originally disrupt the regulation of autonomic system, thereby causing excessive fatigue. By contrast, cognitive abilities are more easily maintained through compensation (including cognitive reserve)[18,49] Further research will quantify the ability of pupillometry to reliably estimate MS-related fatigue.

---

[4] In our study, dilation was found to be faster than constriction. The opposite is usually observed in study investing pupillary reflexes to light and darkness[55]. Yet it is to note that light remained constant during the task and fluctuation of pupillary response are thought to express activity of the Autonomic Nervous System (ANS). As preganglionic neurons of the sympathetic ANS have much shorter axons than in the parasympathetic ANS[56], it is not surprising to observe faster dilation than constriction.

Tables

Table 1. Jeffreys's Bayes factor evidence category[32]

BF_{10}	Interpretation
>100****	Extreme evidence for H1
30–100***	Very Strong evidence for H1
10–30**	Strong evidence for H1
3–10*	Moderate evidence for H1
1–3	Anecdotal evidence for H1
1	No evidence
0.333–1	Anecdotal evidence for H0
0.1–0.333*	Moderate evidence for H0
0.033–0.1**	Strong evidence for H0
0.01–0.033***	Very Strong evidence for H0
<0.01****	Extreme evidence for H0

Level of evidence for the alternative (H1) and the null (H0) hypothesis depending on Bayes Factor (BF_{10} here). For instance, a BF_{10} of 5 indicates that the data observed are 5 times more likely to happen under the alternative hypothesis rather than the null, corresponding to a moderate evidence for H1. The same logic applies to BF_{incl}, with $BF_{incl} > 3$ indicating evidence of an effect of the variable in the studied model, and $BF_{inc} < 0.333$ indicating an absence of effect.

Table 2. Participants' demographics and disease characteristics, trait fatigue and task calibration measures

	pwMS (*n* = 19)	HC (*n* = 19)	BF_{10} T-Test
Age, y, mean (SD)	31.58 (5.23)	31.42 (5.75)	**0.316***
Women, *n* (%)	14 (73.68)	14 (73.68)	0.341[a]
Education, y, mean (SD)	14.26 (2.05)	14.63 (1.36)	0.368
Disease duration, y, mean (SD)	2.04 (0.95)	n.a.	/
Time since diagnosis, y, mean (SD)	1.8 (1.22)	n.a.	/
EDSS, median (range)	1.5 (1–3)	n.a.	/
Trait Fatigue			
FSMC cog, mean (SD)	29.32 (8.23)	25.74 (8.15)	0.548
FSMC phys, mean (SD)	28.95 (10.09)	23.32 (8.23)	1.237
FSMC Total, mean (SD)	58.26 (19.61)	49.05 (15.75)	0.847
Task Calibration			
HCL STD, s, mean (SD)	0.889 (0.314)	0.742 (0.154)	1.154
LCL STD, s, mean (SD)	1.326 (0.508)	1.082 (0.204)	1.356

pwMS: People with Multiple Sclerosis; **HC**: Healthy Controls; **EDSS**: Expanded Disability Status Scale; **FSMC**: Fatigue Scale for Motor and Cognitive function (cog: cognitive sub-scale; phys: physical sub-scale); **HCL**: High Cognitive Load condition, **LCL**: Low Cognitive Load condition, **STD**: Stimulus Time Duration (expressed in seconds). **Disease duration** corresponds to years since first presumed symptoms. [a] Contingency table analysis for independent multinomial sampling[50]

Table 3. Analysis of Effects for rmANOVA on TLDB performance

	Accuracy $BF_{inclusion}$	Digits RT $BF_{inclusion}$	Letters RT $BF_{inclusion}$
Block	**>100****	2.120	**0.018****
Condition	**>100****	**>100****	**>100****
Group	0.459	1.688	1.230
Block * Condition	0.960	0.405	**0.106***
Block * Group	**0.068****	**0.077****	**0.048****
Condition * Group	1.044	**>100****	1.465
Block * Condition * Group	**0.073****	0.341	**0.072****

Evidence for the effects of Block (1 to 4), Condition (HCL vs LCL), Group (pwMS vs HC) and their interaction effects on mean accuracy[22], and reaction times for digits and letters items at the Time Load Dual Back task.

Table 4. Analysis of Effects for rmANOVA on subjective states

	VAS Fatigue $BF_{inclusion}$	Sleepiness (KSS) $BF_{inclusion}$
Time	**15.596****	**>100******
Condition	**0.191***	**0.171***
Group	0.348	0.343
Time * Condition	**0.265***	**0.239***
Time * Group	**0.274***	0.377
Condition * Group	**0.303***	**0.296***
Time * Condition * Group	0.423	**0.326***

Evidence for the effects of Time (pre-vs post-task), Condition (HCL vs LCL), Group (pwMS vs HC) and their interaction effects on subjective states. **VAS**: Visual Analogue Scale; **KSS**: Karolinska Sleepiness Scale.

Table5. Participants' characteristics (subsample for analyses on eye metrics)

	pwMS (*n* = 16)	HC (*n* = 15)	BF_{10} T-Test
Age, y, mean (SD)	31.13 (5.52)	31.47 (6.01)	0.344
Women, *n* (%)	12 (75)	12 (80)	0.374[a]
Education, y, mean (SD)	14.50 (2.07)	14.67 (1.68)	0.348
FSMC cog, mean (SD)	28.29 (10.51)	24.33 (8.03)	0.721
FSMC phys, mean (SD)	26.71 (10.08)	22.67 (8.66)	0.711
FSMC Total, mean (SD)	55.00 (20.04)	47.00 (16.25)	0.743

pwMS: People with Multiple Sclerosis; **HC**: Healthy Controls; **FSMC**: Fatigue Scale for Motor and Cognitive function (cog: cognitive sub-scale; phys: physical sub-scale). [a] Contingency table analysis for independent multinomial sampling[50]

Table 6. GLMM results for the number of long blinks, pupil diameter and response speed

	Long Blinks	Pupil Diameter	PRS
Block	$F_{3, 1} = 165.10$ $p = .06$	$F_{3, 163} = 1.02$ $p = .38$	$F_{3, 354.8} = 2.08$ $p = .11$
Condition	$F_{1, 192} = 2.24$ $p = .14$	$F_{1, 164.4} = 7.52$ $p = .007**$ $R^2_{\beta*} = 0.04$	$F_{1, 362.2} < 0.09$ $p = .99$
Group	$F_{1, 27.53} = 3.92$ $p = .06$	$F_{1, 28.98} = 0.56$ $p = .46$	$F_{1, 28.75} = 0.15$ $p = .70$
Block * Condition	$F_{3, 1} = 27.40$ $p = .14$	$F_{3, 163} = 0.56$ $p = .64$	$F_{3, 354.8} = 1.21$ $p = .30$
Block * Group	$F_{3, 1} = 41.19$ $p = .11$	$F_{3, 163} = 1.91$ $p = .13$	$F_{3, 354.8} = 3.96$ $p = .009**$ $R^2_{\beta*} = 0.03$
Condition * Group	$F_{1, 192} = 47.61$ $p <.0001***$ $R^2_{\beta*} = 0.20$	$F_{1, 164.4} = 0.01$ $p = .91$	$F_{1, 362.2} = 1.81$ $p = .18$
Block * Condition * Group	$F_{3, 1} = 25.16$ $p = .15$	$F_{3, 163} = 1.71$ $p = .17$	$F_{3, 354.8} = 3.55$ $p = .01*$ $R^2_{\beta*} = 0.03$
Pupil Response	/	/	$F_{1, 354.8} = 4.22$ $p = .04*$ $R^2_{\beta*} = 0.01$
Block * Pupil Response	/	/	$F_{3, 354.8} = 0.08$ $p = .97$
Condition * Pupil Response	/	/	$F_{1, 354.8} = 0.14$ $p = .71$
Group * Pupil Response	/	/	$F_{1, 354.8} = 1.53$ $p = .22$
Block * Condition * Pupil Response	/	/	$F_{3, 354.8} = 0.05$ $p = .99$
Block * Group * Pupil Response	/	/	$F_{3, 354.8} = 0.05$ $p = .98$
Condition * Group * Pupil Response	/	/	$F_{1, 354.8} = 0.05$ $p = .82$
Block * Condition * Group * Pupil Response	/	/	$F_{3, 354.8} = 0.08$ $p = .97$

Models with fixed effects of Block (1 to 4), Condition (HCL vs LCL), Group (pwMS vs HC), Pupil Response (Dilation vs Constriction) and their interaction effects on dependent variables. Analysis of long blinks was performed with a Poisson GLMM, pupil diameter and response speed with a Gamma GLMM. *** $p <.001$, **$p <.01$, *$p <.05$

Figures

*Figure 1. Accuracy and RTs performance at the TLDB task*
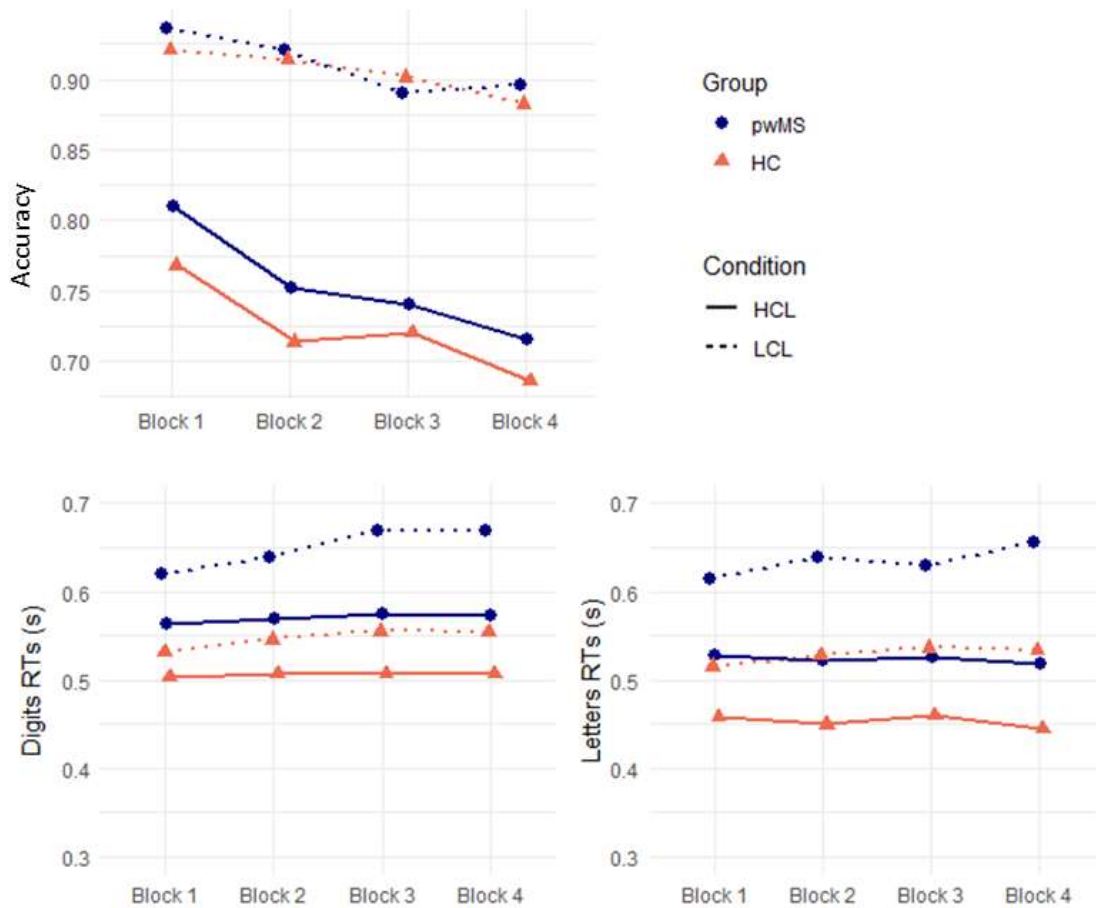
Evolution of accuracy (upper panel) and reaction times (digits: left lower panel; letters: right lower panel) at Time Load Dual Back task during the High Cognitive Load (HCL, solid line) and the Low Cognitive Load conditions (LCL, dashed lines) in people with Multiple Sclerosis (pwMS, navy-blue circles) and Healthy Controls (HC, orange triangles).

Figure 2. Eye metrics variables during the TLDB task
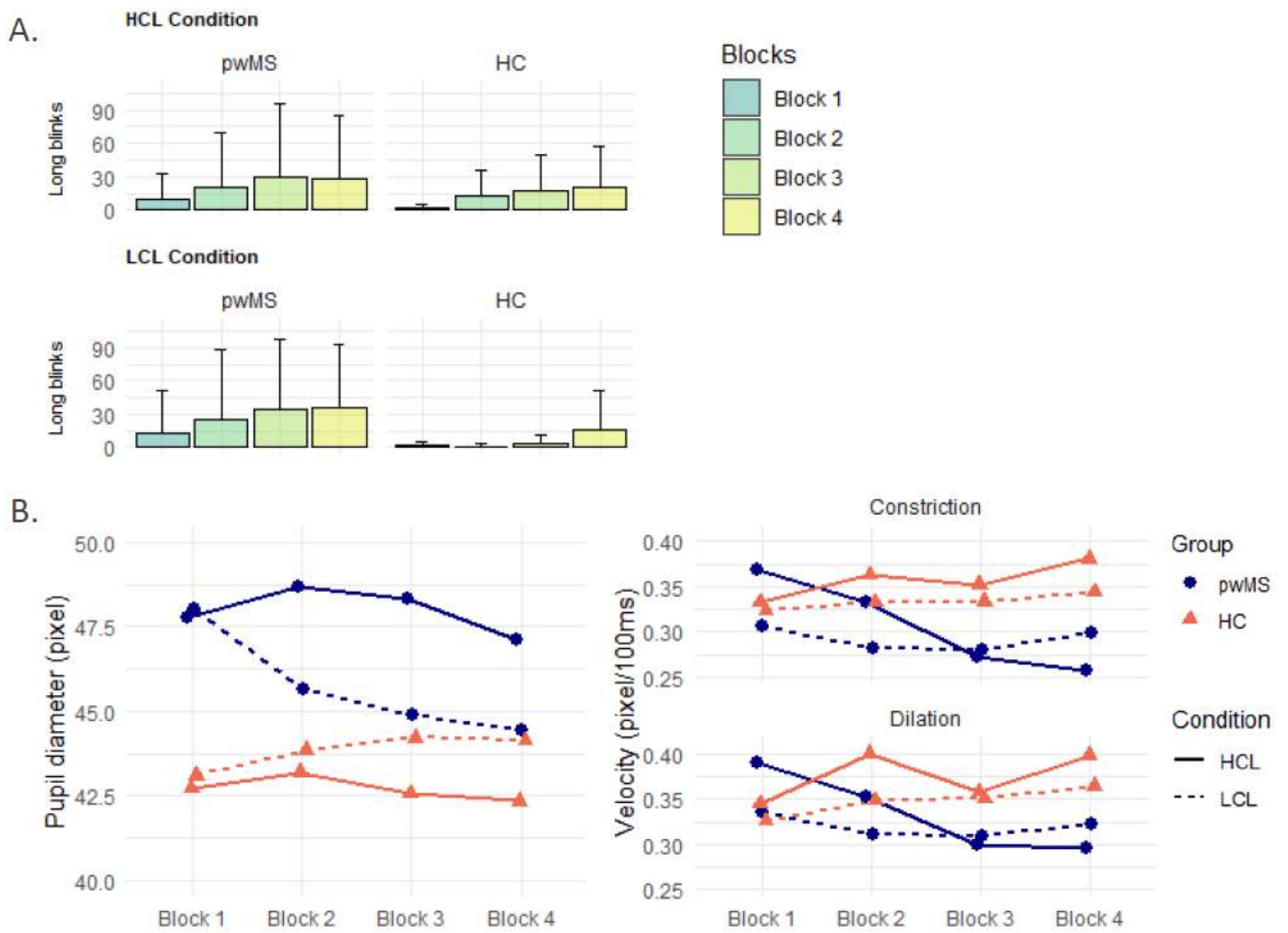
Number of long blinks (A.) and evolution of pupil diameter and response speed (B.) during the four blocks of the TLDB task in the High Cognitive Load (HCL, solid line) and the Low Cognitive Load conditions (LCL, dashed lines) in people with Multiple Sclerosis (pwMS, navy-blue circles) and Healthy Controls (HC, orange triangles).

Supplemental Material

Supplemental comments on methods:

This study is part of a larger research project investigating the effects of cognitive fatigue in early MS (the full protocol is available on OSF, following this link: https://osf.io/egr6d/?view_only=cdcc343c7d71406685b038e46e88145b). Overall, participation to this research protocol implied in 5 appointments. Participants completed the full protocol within 7 weeks, with each session from 2 to 14 days apart.

Treatments excluded:

The following medications were considered as an exclusion criteria if they were not stopped at least 5 half-life prior participation to the study: benzodiazepines, neuroleptics, tricyclic antidepressants, beta blockers, anticholinergics and anticonvulsant drugs.

KSS and VAS:

In order to assess the evolution of subjective states, the Karolinska Sleepiness Scale (KSS[29]) and Visual Analog Scales (VAS[30]) were administered at different time-points. The KSS Likert scale ranged from 1 (very alert) to 9 (very sleepy) and the VAS scales comprised assessment of seven states (Motivation, Happiness, **Fatigue**, Openness, Stress, Anxiety, Effort), on a scale from 0 to 100. For the fatigue scale, a score of 0 represents low fatigue, and a score of 100 corresponds to extreme fatigue sensation.

Eye metrics quality assessment, pre-processing & analyses:

Because several factors can alter the quality of the recorded eye metrics, this experiment was conducted on a subset of participants. For instance, an accurate measure of the pupil size could not be acquired when participants had long and curved eyelashes. Some participants felt uncomfortable with the corrective lenses and the measure of eye metrics was left behind to let them wear their personal glasses and preserve the quality of behavioral data. Another reason for bad quality data was the observed signal loss if eyelids gap was too small, that was more likely to occur at the end of the fatiguing task, when participants started to feel fatigued.

Regarding preprocessing of eye data, blinks were first eliminated using a cubic-spline interpolation as described by Mathôt and colleagues[51,52]. The reconstructed signal was filtered using a low-pass 4-th order Butterworth filter with a cutoff frequency of 4 Hz[53,54]. Finally, spurious periods were removed from the signal where the eyelid gap was smaller than the reconstructed (but unfiltered) pupil size. The preprocessed signal was next split into four blocks of approximately 8 minutes. For each block, mean pupil size, number of long blinks (duration > 300ms) and mean pupil variation speed during dilation and constriction were computed. Number of long blinks, rather than blink frequency, were considered to provide a measure of fatigue and sleepiness states independent from the confounding effects of stimulus presentation rate. Mean speed of pupil size variation, referred to pupil response speed (PRS), was defined as the 1st order difference in mean pupil size across consecutive time bins of 100 ms.

The quality of the recording was assessed visually in order to exclude sessions of poor quality. Considering the high number of sessions excluded from analysis and the distribution of the variables of interest, frequentist generalized linear mixed models (GLMM) were used for these analyses. The number of long blinks was analyzed as dependent variable following a Poisson distribution, mean pupil size and PRS following a gamma distribution.

Table S1. Self-assessed questionnaires in the whole sample (TLDB task analyses)

	pwMS (*n* = 19)	HC (*n* = 19)	BF_{10}
STAI-Y1, mean (SD)	34.54 (10.17)	32.74 (8.16)	0.383
STAI-Y2, mean (SD)	44.68 (11.45)	43.53 (12.43)	**0.326***
BDI-II, mean (SD)	12.16 (11.17)	10.74 (10.28)	0.336
PSQI, mean (SD)	6.32 (2.85)	5.00 (1.83)	0.958
ESS, mean (SD)	10.37 (5.23)	8.37 (4.37)	0.597

pwMS: People with Multiple Sclerosis; **HC**: Healthy Controls; **STAI**: State (Y1) Trait (Y2) Anxiety Inventory; **BDI-II**: Beck Depression Inventory second edition; **PSQI**: Pittsburg Sleep Quality Index; **ESS**: Epworth Sleepiness Scale.

Table S2. Mean and standard deviation of accuracy rate and reaction times at the Time Load Dual Back

	pwMS (*n* = 19)				HC (*n* = 19)			
	Block 1	Block 2	Block 3	Block 4	Block 1	Block 2	Block 3	Block 4
Accuracy								
HCL	0.81 (0.10)	0.75 (0.10)	0.74 (0.12)	0.72 (0.12)	0.77 (0.07)	0.71 (0.09)	0.72 (0.08)	0.69 (0.10)
LCL	0.94 (0.05)	0.92 (0.07)	0.89 (0.10)	0.90 (0.09)	0.92 (0.06)	0.91 (0.06)	0.90 (0.07)	0.88 (0.09)
Digits Reaction Times (seconds)								
HCL	0.564 (0.12)	0.569 (0.11)	0.574 (0.13)	0.572 (0.11)	0.503 (0.09)	0.507 (0.09)	0.507 (0.09)	0.507 (0.10)
LCL	0.620 (0.17)	0.639 (0.17)	0.669 (0.20)	0.668 (0.19)	0.531 (0.09)	0.546 (0.09)	0.555 (0.10)	0.553 (0.09)
Letters Reaction Times (seconds)								
HCL	0.528 (0.18)	0.523 (0.18)	0.525 (0.17)	0.519 (0.21)	0.458 (0.08)	0.450 (0.09)	0.460 (0.10)	0.445 (0.13)
LCL	0.615 (0.26)	0.638 (0.30)	0.629 (0.20)	0.655 (0.26)	0.515 (0.08)	0.527 (0.09)	0.536 (0.09)	0.534 (0.08)

pwMS: People with Multiple Sclerosis; HC: Healthy Controls; HCL: High Cognitive Load; LCL: Low Cognitive Load.

Table S3. Mean and standard deviation of subjective states

		pwMS (*n* = 19)			HC (*n* = 19)		
		Fatigue	Motivation	Sleepiness	Fatigue	Motivation	Sleepiness
HCL	Pre-task	47.55 (24.87)	31.42 (23.61)	4.63 (2.48)	46.19 (18.59)	25.99 (19.63)	4.37 (1.57)
	Post-task	57.35 (20.05)	41.38 (26.86)	5.90 (2.03)	50.27 (23.89)	43.62 (25.63)	6.11 (2.54)
LCL	Pre-task	43.82 (24.96)	30.06 (21.54)	4.53 (2.46)	44.33 (24.49)	26.90 (19.33	4.53 (2.07)
	Post-task	54.28 (28.40)	40.72 (24.91)	5.68 (2.50)	54.28 (28.40)	37.77 (19.25)	6.32 (2.16)

Fatigue/motivation scores represent values on Visual Analogue Scale (0–100; high score = high fatigue/low motivation), sleepiness was assessed with the Karolinska Sleepiness Scale (high score = high sleepiness). pwMS: People with Multiple Sclerosis; HC: Healthy Controls; HCL: High Cognitive Load; LCL: Low Cognitive Load.

Table S4. Kendal's correlation between STD at the TLDB and trait fatigue

| **Correlation with STD**	pwMS (*n* = 19)	HC (*n* = 19)
FSMC cog	$BF_{10} = 4.14$* $\tau = 0.40$	$BF_{10} = 0.50$ $\tau = 0.18$
FSMC phys	$BF_{10} = 30.58$*** $\tau = 0.53$	$BF_{10} = 1.62$ $\tau = 0.32$
FSMC Total	$BF_{10} = 12.31$** $\tau = 0.47$	$BF_{10} = 1.05$ $\tau = 0.28$

STD: Stimulus Time Duration; FSMC: fatigue scale for motor and cognitive functions (cog: cognitive sub-scale, phys: physical sub-scale); pwMS: people with multiple sclerosis; HC: healthy controls. *Moderate evidence, **Strong evidence, ***Very strong evidence for a correlation.

Table S5. Mean and standard deviation for number of long blinks, pupil size and PRS.

	pwMS				HC			
	Block 1	Block 2	Block 3	Block 4	Block 1	Block 2	Block 3	Block 4
Long Blinks								
HCL	9.15 (22.49)	20.42 (48.96)	29.75 (65.77)	28.67 (56.60)	2.08 (3.50)	12.83 (22.86)	17.25 (31.98)	19.83 (36.96)
LCL	12.27 (39.47)	25.53 (62.63)	35.00 (61.92)	35.93 (57.17)	1.77 (3.19)	1.23 (2.42)	3.92 (8.03)	16.46 (35.13)
Pupil size (pixels)								
HCL	47.78 (10.46)	48.68 (11.93)	48.31 (11.79)	47.11 (11.17)	42.70 (8.19)	43.19 (8.59)	42.55 (8.12)	42.34 (7.91)
LCL	48.00 (9.54)	45.65 (9.43)	44.91 (9.53)	44.44 (8.77)	43.10 (9.67)	43.85 (10.63)	44.21 (10.29)	44.14 (10.10)
Constriction Response Speed (pixels/100ms)								
HCL	0.37 (0.16)	0.33 (0.13)	0.27 (0.13$6)	0.26 (0.13)	0.33 (0.11)	0.36 (0.16)	0.35 (0.13)	0.38 (0.13)
LCL	0.31 (0.14)	0.28 (0.15)	0.28 (0.16)	0.30 (0.17)	0.32 (0.15)	0.33 (0.17)	0.33 (0.18)	0.34 (0.17)
Dilation Response Speed (pixels/100ms)								
HCL	0.39 (0.20)	0.35 (0.14)	0.30 (0.16)	0.30 (0.13)	0.34 (0.14)	0.40 (0.23)	0.36 (0.14)	0.40 (0.15)
LCL	0.34 (0.17)	0.31 (0.18)	0.31 (0.18)	0.32 (0.17)	0.33 (0.16)	0.35 (0.22)	0.35 (0.21)	0.36 (0.22)

HCL: High Cognitive Load condition; LCL: Low Cognitive Load condition; pwMS: people with multiple sclerosis (n = 12 in HCL & 15 in LCL); HC: healthy controls (n=12 in HCL & 13 in LCL).

Figure S1. Distribution of reaction times during the TLDB according to severity of trait fatigue
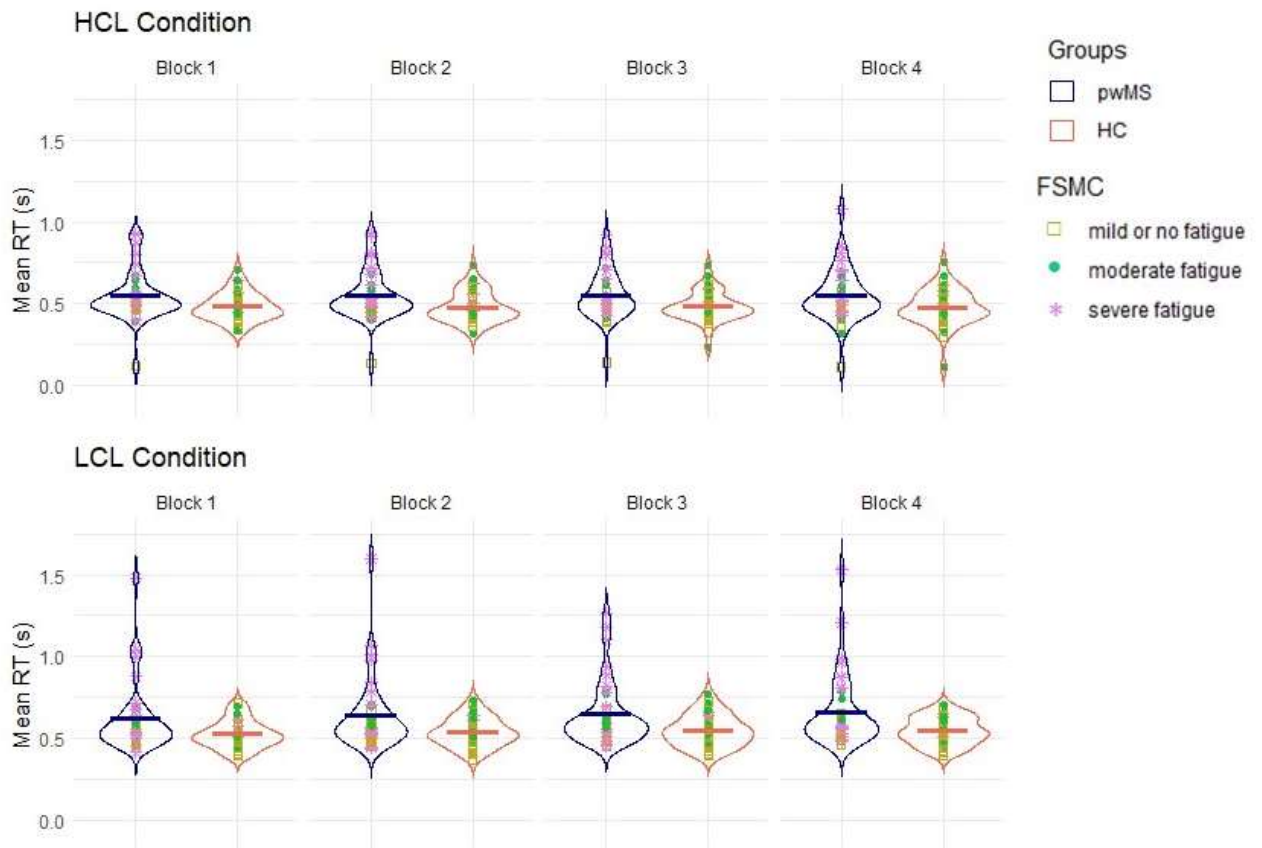
Mean reaction times at the TLDB task (digit and letter items together) for the High Cognitive Load (HCL, upper panel) and the Low Cognitive Load (LCL, lower panel) conditions. Participant's score at the Fatigue Scale for Motor and Cognitive functions (FSMC, total score) were converted according to available cutoffs[23] into 3 fatigue classes (no to mild fatigue: yellow squares; moderate fatigue: green dots; severe fatigue: pink stars). People with multiple sclerosis (pwMS) are depicted in navy-blue violin, Healthy Controls (HC) in orange. Horizontal bars represent group means.

References

1. Weiland, T. J. *et al.* Clinically Significant Fatigue: Prevalence and Associated Factors in an International Sample of Adults with Multiple Sclerosis Recruited via the Internet. *PLoS One* **10**, (2015).

2. Bakshi, R. *et al.* Fatigue in multiple sclerosis and its relationship to depression and neurologic disability. *Mult. Scler.* **6**, 181–185 (2000).

3. Fisk, J. D., Pontefract, A., Ritvo, P. G., Archibald, C. J. & Murray, T. J. The Impact of Fatigue on Patients with Multiple Sclerosis. *Can. J. Neurol. Sci. / J. Can. des Sci. Neurol.* **21**, 9–14 (1994).

4. Sainz de la Maza, S. *et al.* Measuring productivity loss in early relapsing-remitting multiple sclerosis. *Mult. Scler. Relat. Disord.* **58**, (2022).

5. Chen, J. *et al.* Estimating MS-related work productivity loss and factors associated with work productivity loss in a representative Australian sample of people with multiple sclerosis. *Mult. Scler. J.* 1–11 (2018). doi:10.1177/1352458518781971

6. Penner, I. K. & Paul, F. Fatigue as a symptom or comorbidity of neurological diseases. *Nat. Rev. Neurol.* **13**, 662–675 (2017).

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

8. Linnhoff, S., Fiene, M., Heinze, H. J. & Zaehle, T. Cognitive fatigue in multiple sclerosis: An objective approach to diagnosis and treatment by transcranial electrical stimulation. *Brain Sci.* **9**, 1–23 (2019).

9. Chiaravalloti, N. D. & DeLuca, J. Cognitive impairment in multiple sclerosis. *Lancet Neurol.* **7**, 1139–1151 (2008).

10. Brochet, B. & Ruet, A. Cognitive Impairment in Multiple Sclerosis With Regards to Disease Duration and Clinical Phenotypes. *Front. Neurol.* **10**, 1–7 (2019).

11. Borragán, G. *et al.* Cognitive Fatigue, Sleep and Cortical Activity in Multiple Sclerosis Disease. A Behavioral, Polysomnographic and Functional Near-Infrared Spectroscopy Investigation. *Front. Hum. Neurosci.* **12**, 378 (2018).

12. Agyemang, C., Berard, J. A. & Walker, L. A. S. Cognitive fatigability in multiple sclerosis: How does performance decline over time on the Paced Auditory Serial Addition Test? *Mult. Scler. Relat. Disord.* **54**, 103130 (2021).

13. Boksem, M. A. S. & Tops, M. Mental fatigue: Costs and benefits. *Brain Res. Rev.* **59**, 125–139 (2008).

14. Joshi, S. & Gold, J. I. Pupil Size as a Window on Neural Substrates of Cognition. *Trends Cogn. Sci.* **24**, 466–480 (2020).

15. Rodriguez, J. D. *et al.* Blink: Characteristics, Controls, and Relation to Dry Eyes. *Curr. Eye Res.* **43**, 52–66 (2018).

16. Pozzessere, G. *et al.* Autonomic involvement in multiple sclerosis: A pupillometric study. *Clin. Auton. Res.* **7**, 315–319 (1997).

17. de Seze, J. *et al.* Pupillary disturbances in multiple sclerosis: Correlation with MRI findings. *J. Neurol. Sci.* **188**, 37–41 (2001).

18. de Rodez Benavent, S. A. *et al.* Fatigue and cognition: Pupillary responses to problem-solving in early multiple sclerosis patients. *Brain Behav.* **7**, e00717 (2017).

19. Surakka, J. *et al.* Pupillary function in early multiple sclerosis. *Clin. Auton. Res.* **18**, 150–154 (2008).

20. Kurtzke, J. F. Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology* **33**, 1444–1452 (1983).

21. Thompson, A. J. *et al.* Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol.* **17**, 162–173 (2018).

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

23. Penner, I. K. *et al.* The Fatigue Scale for Motor and Cognitive Functions (FSMC): Validation of a new instrument to assess multiple sclerosis-related fatigue. *Mult. Scler.* **15**, 1509–1517 (2009).

24. Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R. & Kupfer, D. J. The Pittsburg Sleep Quality Index: A New Instrument for Psychiatric Practice and Research. *Psychiatry Research* **28**, 193–213 (1998).

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

26. Spielberger, C. D., Gorsuch, R. L., Ushene, R., Vagg, P. R. & Jacobs, G. A. *Manual for the State-Trait Anxiety Inventory.* (Consulting Psychologists Press, 1983).

27. Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. & Erbaugh, J. An Inventory for Measuring Depression. *Arch. Gen. Psychiatry* **4**, 561 (1961).

28. Kirchner, W. K. Age differences in short-term retention of rapidly changing information. *J. Exp. Psychol.* **55**, 352–358 (1958).

29. Åkerstedt, T. & Gillberg, M. Subjective and Objective Sleepiness in the Active Individual. *Int. J. Neurosci.* **52**, 29–37 (1990).

30. Lee, K. A., Hicks, G. & Nino-Murcia, G. Validity and reliability of a scale to assess fatigue. *Psychiatry Res.* **36**, 291–298 (1991).

31. Keysers, C., Gazzola, V. & Wagenmakers, E.-J. Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nat. Neurosci.* **23**, 788–799 (2020).

32. Jeffreys, H. *Theory of Probability. 3rd Edition*. (Clarendon Press, 1961).

33. Kenward, M. G. & Roger, J. H. An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Comput. Stat. Data Anal.* **53**, 2583–2595 (2009).

34. Jaeger, B. C., Edwards, L. J., Das, K. & Sen, P. K. An R^2 statistic for fixed effects in the generalized linear mixed model. *J. Appl. Stat.* **44**, 1086–1105 (2017).

35. Rooney, S., Wood, L., Moffat, F. & Paul, L. Prevalence of fatigue and its association with clinical features in progressive and non-progressive forms of Multiple Sclerosis. *Mult. Scler. Relat. Disord.* **28**, 276–282 (2019).

36. Galland-Decker, C., Marques-Vidal, P. & Vollenweider, P. Prevalence and factors associated with fatigue in the Lausanne middle-aged population: A population-based, cross-sectional survey. *BMJ Open* **9**, 1–10 (2019).

37. Cellini, N. *et al.* Changes in sleep timing and subjective sleep quality during the COVID-19 lockdown in Italy and Belgium: age, gender and working status as modulating factors. *Sleep Med.* **77**, 112–119 (2021).

38.   Torrente, F. *et al.* Psychological symptoms, mental fatigue and behavioural adherence after 72 continuous days of strict lockdown during the COVID-19 pandemic in Argentina. *BJPsych Open* **8**, (2022).

39.   Boeschoten, R. E. *et al.* Prevalence of depression and anxiety in Multiple Sclerosis: A systematic review and meta-analysis. *J. Neurol. Sci.* **372**, 331–341 (2017).

40.   Sakkas, G. K., Giannaki, C. D., Karatzaferi, C. & Manconi, M. Sleep Abnormalities in Multiple Sclerosis. *Curr. Treat. Options Neurol.* **21**, 4 (2019).

41.   Guillemin, C. *et al.* The Complex Interplay Between Trait Fatigue and Cognition in Multiple Sclerosis. *Psychol. Belg.* **62**, 108 (2022).

42.   Diamond, B. J., Johnson, S. K., Kaufman, M. & Graves, L. Relationships between information processing, depression, fatigue and cognition in multiple sclerosis. *Arch. Clin. Neuropsychol.* **23**, 189–199 (2008).

43.   Pokryszko-Dragan, A. *et al.* Cognitive performance, fatigue and event-related potentials in patients with clinically isolated syndrome. *Clin. Neurol. Neurosurg.* **149**, 68–74 (2016).

44.   Andreasen, A. K. *et al.* Structural and cognitive correlates of fatigue in progressive multiple sclerosis. *Neurol. Res.* **41**, 168–176 (2019).

45.   Sandry, J., Genova, H. M., Dobryakova, E., DeLuca, J. & Wylie, G. Subjective cognitive fatigue in multiple sclerosis depends on task length. *Front. Neurol.* **5**, 1–7 (2014).

46.   Kahneman, D. & Beatty, J. Pupil diameter and load on memory. *Science (80-.).* **154**, 1583–1585 (1966).

47.   Bafna, T. & Hansen, J. P. Mental fatigue measurement using eye metrics: A systematic literature review. *Psychophysiology* **58**, 1–23 (2021).

48.   Merkelbach, S. *et al.* Multiple sclerosis and the autonomic nervous system. *J. Neurol.* **253**, 21–25 (2006).

49.   Schoonheim, M. M., Meijer, K. A. & Geurts, J. J. G. Network Collapse and Cognitive Impairment in Multiple Sclerosis. *Front. Neurol.* **6**, 82 (2015).

50.   Jamil, T. *et al.* Default "Gunel and Dickey" Bayes factors for contingency tables. *Behav. Res. Methods* **49**, 638–652 (2017).

51.   Mathôt, S. A simple way to reconstruct pupil size during eye blinks. *Figshare* **17**, 491–496 (2013).

52.   Mathôt, S., Fabius, J., Van Heusden, E. & Van der Stigchel, S. Safe and sensible preprocessing and baseline correction of pupil-size data. *Behav. Res. Methods* **50**, 94–106 (2018).

53.   Franco, A., Neves, C. M., Quintão, C., Vigário, R. & Vieira, P. Singular Spectrum Analysis of Pupillometry Data. Identification of the Sympathetic and Parasympathetic Activity. *Procedia Technol.* **17**, 273–280 (2014).

54.   Leal, G., Neves, C. & Vieira, P. M. Pupillometry: Development of equipment for studies of autonomic nervous system. *IFIP Adv. Inf. Commun. Technol.* **372 AICT**, 553–562 (2012).

55.   Mathôt, S. Pupillometry: Psychology, Physiology, and Function. **1**, 1–23 (2018).

56.   Watson, C., Kirkcaldie, M. & Paxinos, G. *The Brain. The Brain* (Elsevier Inc., 2010). doi:10.1016/C2009-0-01579-7