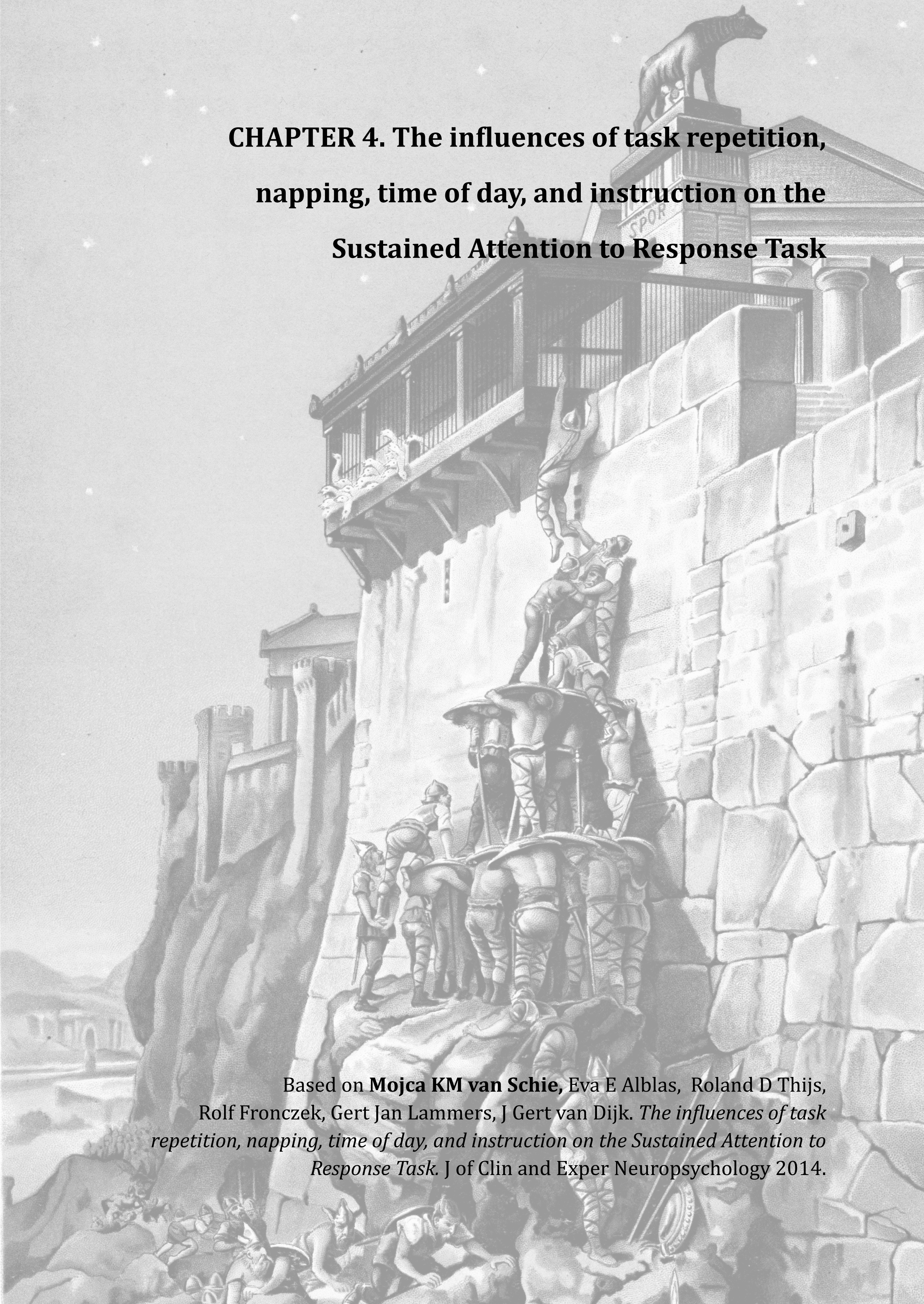
Attention please: vigilance in patients with excessive daytime sleepiness

Schie, M.K.M. van



CHAPTER 4. The influences of task repetition, napping, time of day, and instruction on the Sustained Attention to Response Task

Based on **Mojca KM van Schie,** Eva E Alblas, Roland D Thijs, Rolf Fronczek, Gert Jan Lammers, J Gert van Dijk. *The influences of task repetition, napping, time of day, and instruction on the Sustained Attention to Response Task.* J of Clin and Exper Neuropsychology 2014.

ABSTRACT

Introduction

The Sustained Attention to Response Task (SART) helps to quantify vigilance impairments. Previous studies in which five SART sessions on one day were administered, demonstrated worse performance during the first session compared to the others. The present study comprises two experiments to identify a cause of this phenomenon.

Method

Experiment 1, counting eighty healthy participants, assessed effects of repetition, napping and time of day on SART performance through a between-groups design The SART was performed twice in the morning or twice in the afternoon; half of the participants took a 20-minute nap before the second SART. A strong correlation between error count and reaction time (RT) suggested effects of test instruction. Participants gave equal weight to speed and accuracy in experiment 1; therefore, results of 20 participants were compared to those of 20 additional participants who were told to prefer accuracy (experiment 2).

Results

The average SART error count in experiment 1 was 10.1, the median RT 280 ms. Repetition nor napping influenced error count or RT. Time of day did not influence error count, but RT was significantly longer for morning than afternoon SARTs. The additional participants in experiment 2 had a 49% lower error count and a 14% higher RT than the participants in experiment 1. Error counts reduced by 50% from the first to the second session of experiment 2, irrespective of napping or time of day.

Conclusions

Preferring accuracy over speed was associated with a significantly lower error count. The data suggest that a worse performance in the first SART session only occurs when instructing participants to prefer accuracy, which is caused by repetition, not by napping or time of day.

Note

We advise to instruct participants to prefer accuracy over speed when performing the SART and to include a full practice session.

INTRODUCTION

Vigilance, which can be defined as the capability to be aware of potential changes in one's environment, is a prerequisite for adequate daytime functioning. A low vigilance may lead to cognitive mishaps, e.g. forgetting why you went up the stairs, or reading a piece of text more than once without registering the content. Low vigilance can even be dangerous, e.g. when driving a car.

The sleep disorder narcolepsy is an excellent model of disturbed vigilance (Fronczek, Middelkoop, van Dijk, & Lammers, 2006; Valley & Broughton, 1981; Van Schie et al., 2012). Patients with narcolepsy experience severe vigilance problems in daily life, for instance not being able to recall the content of a conversation, not being able to finish a book, or being unable to concentrate on studies or work. While several wake-promoting drugs aim to improve vigilance and reduce sleepiness, their efficacy depends largely on patient's subjective reports and is difficult to determine objectively.

There are several methods to estimate the level of vigilance, ranging from subjective visual-analog scales through pupillography (Morad, Lemberg, Yofe, & Dagan, 2000) and quantified electro-encephalography (EEG) (Coenen, 1995), to a variety of response tasks assessing sustained attention. Such tests can be regarded as an operationalization of vigilance. One such test is the Sustained Attention to Response Task (SART) (Robertson, Manly, Andrade, Baddeley, & Yiend, 1997), a 4-minute 19-second computer task in which participants should withhold presses to one out of nine stimuli. The SART has shown an ability to quantify vigilance impairment in patients with excessive daytime sleepiness, for instance narcolepsy (Fronczek et al., 2006; Van Schie et al., 2012). Its primary outcome measure is the error count, consisting of key presses when no key should be pressed (commission errors), and absent presses when a key should have been pressed (omission errors).

In previous SART studies in healthy controls and patients with narcolepsy or other sleep disorders such as idiopathic hypersomnia or obstructive sleep apnea, we performed the SART five times per day prior to each session of a Multiple Sleep Latency Test (MSLT). The MSLT measures the tendency to fall asleep and is a routine part of the diagnostic work-up of excessive daytime sleepiness. This test consists of five 20-minute sessions at 1.5-hour intervals, in which participants are requested to try to fall asleep. The average sleep latency of five sessions is the main outcome parameter. Average sleep latency < 8 minutes indicates excessive daytime sleepiness. In this combined SART-MSLT design the SART error score of the first session was higher than of subsequent sessions in healthy controls (Fronczek et al., 2006) and in patients with excessive daytime sleepiness (Van Schie et al., 2012).

This difference may reflect a true change in vigilance level over the day (Manly, Lewis, Robertson, Watson, & Datta, 2002; Schmidt, Collette, Cajochen, & Peigneux, 2007), or the result of napping between MSLT sessions, or a learning effect. Understanding such sources of variability may improve the reliability of the SART as a tool to measure vigilance, and therefore lower the number of SART sessions needed.

Possible reasons for an increased error score on the first SART session in previous studies were the early hour and the lack of a nap before that session. The lower error scores in later sessions could represent a learning effect, even though this does not seem likely for the following reasons: participants had always received a separate practice session before the first formal SART test; a learning effect is not expected in vigilance tasks; and finally none has been found in the SART before (Manly, Robertson, Galloway, & Hawkins, 1999; McAvinue, O'Keeffe, McMackin, & Robertson, 2005; Robertson et al., 1997).

The design of our first experiment was constructed to unravel the effects of three factors in healthy controls: the influence of repetition, a nap occasion before the SART and time of day.

EXPERIMENT 1

METHODS

Participants

Eighty healthy participants were recruited by advertisement. Reasons for exclusion were a diagnosis of a sleep disorder, any disorder significantly affecting attention, use of psychotropic medication, complaints of sleepiness, an Epworth Sleepiness Scale (ESS) > 10 (Johns, 1991), an irregular sleep pattern including night shift work and traveling through time zones in the two weeks before inquiry, or poor sleep hygiene. The latter was defined as an average night sleep < 6 hours or highly variable bed times (variability in bed time of > 2 hours for ≥ 3 days per week). Twenty men and 60 women with a mean age of 26.2 ± 9.7 years (range 18 to 55 years) were included. The sex ratio was due to recruitment in a faculty with more female students and employees. Participants were assigned to one of four groups matched for age, sex, and level of education. Table 1 shows the baseline characteristics of each group. The groups did not differ significantly in ESS score, sleep duration, or the number of days per week with a deviation from habitual bed time with more than one hour.

The study was approved by the medical ethical committee of Leiden University Medical Centre, and written informed consent was obtained from all participants prior to the study. Participants were given a small incentive to compensate for their time.

Design

The conditions "time of day" and "napping opportunity" led to a two-by-two design with four groups to which participants were assigned quasi-randomly, i.e. they were matched for age, sex and level of education. The two morning groups were tested between 9:15 and 12:15 hours, the two afternoon groups between 13:00 and 16:00 hours. All groups performed two SART sessions with a 1.5-hour break in between the two sessions. A 20-minute nap period similar to an MSLT session was offered to two groups (N+), but not to the two other groups (N-). The N+ groups were after the fact divided into those who actually slept (S+) and who did not (S-).

Upon arrival to the sleep laboratory, participants provided information about their sleep pattern of the last seven days. This was followed by the placement of electrodes for electro-encephalography (EEG, see below) before starting the SART.

Table 1 - Baseline characteristics of the study groups

	AM N+ ***		AM N-	PM N+		PM N-
	S+ (n = 13)	S- (n = 6)	(n = 20)	S+ (n = 16)	S- (n = 4)	(n = 20)
N of males	2	3	5	2	3	5
Age in years	24.6 ± 8.8	29.0 ± 12.8	25.9 ± 10.1	25.2 ± 8.7	30.5 ± 12.9	26.4 ± 10.0
ESS	5.1 ± 3.0	5.0 ± 3.3	5.0 ± 2.0	5.7 ± 2.8	4.8 ± 3.3	4.1 ± 2.9
Average nighttime sleep (hrs.)*	8.7 ± 0.8	8.2 ± 0.5	8.6 ± 0.9	8.5 ± 0.7	8.6 ± 0.7	8.5 ± 0.8
for week days*	8.4 ± 0.9	8.0 ± 0.8	8.3 ± 0.9	8.2 ± 0.7	8.7 ± 0.6	8.2 ± 0.9
for weekend days*	9.3 ± 0.7	8.7 ± 1.2	9.3 ± 1.2	9.2 ± 0.8	8.4 ± 1.7	9.2 ± 1.1
Nr of days per week with deviation from habitual bed time**	1.2 ± 0.9	1.3 ± 0.8	0.9 ± 0.9	0.8 ± 0.9	1.5 ± 0.6	0.9 ± 0.9
last 7 days	1.0 ± 1.3	1.2 ± 1.0	1.3 ± 1.3	1.3 ± 1.0	1.0 ± 0.8	1.5 ± 1.2
Sleep latency nap	10.3 ± 5.5	NA	NA	12.1 ± 4.1	NA	NA

Legend: AM: morning group; PM: afternoon group; N+/-: group with/without napping opportunity; S+/-: slept/remained wake during nap opportunity; ESS: Epworth Sleepiness Scale.
* Calculated from a subjective report of bed times and wake-up times.
** Deviation is defined as > 1.0 hour earlier or later than habitual bed time.
*** EEG recording of one subject in this group is missing due to a technical problem, so that appropriate classification of this subject in either S+ or S- was not possible.

Sustained Attention to Response Task

This test, lasting 4 minutes and 19 seconds, displays the numbers 1 to 9 25 times (225 in total) in random order on a black computer screen. Participants had to respond to the appearance of each number by pressing a button, except for the number 3, which occurred 25 times in all. Participants had to press the button before the next number appeared and were instructed to give equal importance to accuracy and speed in performing the task

(Manly *et al.*, 1999; Robertson *et al.*, 1997). We used the total error count, the average reaction time (RT) and the coefficient of variation (CV) of the RT as SART parameters.

Two SART sessions with a 1.5-hour break were administered to all participants using rooms, body positions, lights and equipment as described previously (Fronczek et al., 2006; Van Schie et al., 2012). Participants performed a short SART training before the first session. The nap occasion was provided to the N+ groups directly after the first SART session. Participants were allowed to go for short walks in the hospital and eat or drink between this nap and the second SART session, or between the two SART sessions for the N- groups. They were instructed to abstain from coffee, coca cola and energy drinks.

Questionnaires

The Stanford Sleepiness Scale (SSS) was administered prior to each SART session to assess the momentary level of sleepiness. Two 100 mm visual-analogue scales (VAS) were presented to the participants following each SART session. Participants were asked to evaluate their own performance concerning (1) accuracy ($VAS_{accuracy}$, from very poor to very good) and (2) response speed (VAS_{RT}, from very slow to very fast). This judgment was compared with their objective performance to obtain an estimation of their approach towards the task, i.e. whether the participants indeed felt they had complied with the instruction to pay equal attention to accuracy and speed.

Electro-encephalography

EEG electrodes were applied to all participants; for the N+ groups to record whether sleep occurred, and for the N- groups to confirm that the participants had remained awake. A total of 9 recording sites (Fpz, Z, Cz, Pz, Oz, Fp1, Fp2, Pg1 and Pg2) were measured with gold-plated 10 mm electrodes placed according to the 10-20 system (Jasper, 1958). Data were acquired using a portable polysomnography recorder (Titanium; Embla Systems, Broomfield, CO) and scored in Somnologica Studio 5.1.1.1684 (Embla Systems). Sleep was scored in 30-second epochs according to the AASM criteria (Littner et al., 2005) by one sleep technician.

Statistical analysis

Data were analyzed using IBM® SPSS® Statistics version 20. The analyses consisted of two parts. Firstly, the accuracy and RT measures of the SART sessions were compared between groups and conditions, i.e. time of day and opportunity to nap, using the paired t-test or the Mann-Whitney U-test when data were not normally distributed. Multiple testing was accounted for by Holm-Bonferroni correction. Delta scores for all SART accuracy and RT measures, as well as for SSS, were calculated by subtracting the score of the first session from the score of the second one. Correlations between delta scores were assessed using Pearson's r or Spearman's ρ, depending of the distribution of the data.

SART performance of the morning groups was compared to that of the afternoon groups, and SART performance of the N+ groups to that of N- groups. The latter comparison comprised the second session and delta score corrected for the first session. As the nap was provided after the first session, comparing this session was not considered useful to assess the influence of a nap occasion. The comparison between N+ and N- groups was secondarily broken down into a comparison of S+ and S- participants.

Additionally, linear mixed effect models (LMMs) were used to assess the combined effects of the conditions on SART performance, i.e. main effects of repetition, time of day, and napping, as well as their interactions, and to correct the analysis of accuracy measures for RT, CV of RT and SSS. Significance of model parameters was determined after Holm-Bonferroni correction for multiple comparisons.

RESULTS

Participants

EEG recordings confirmed that all participants from the N- groups stayed awake during and between SART sessions. Thirteen participants in the morning N+ group fell asleep during the napping opportunity, compared to 16 in the afternoon N+ group.

No differences in baseline characteristics were observed between those who did and did not fall asleep.

SART performance

SART error count and CV of RT are presented as mean ± SD, and SART RT as median with 25th-75th percentiles. The mean error count of all participants was 10.1 ± 4.5, the median RT was 280 ms (261 – 303 ms) and the mean CV of RT was 0.24 ± 0.06. A significant correlation was found between error count and RT (r_s = -0.53, p < 0.01 for session 1, and r_s = -0.58, p < 0.01 for session 2), as well as between error count and CV of RT (r_s = 0.31, p < 0.01 for session 1, and r_s = 0.34, p < 0.01 for session 2). $VAS_{accuracy}$ was significantly and negatively correlated with SART error count (r_s = -0.43, p < 0.01), and positively with average RT (r_s = 0.28, p < 0.01) and VAS_{RT} (r_s = 0.20, p = 0.01), but not with CV of RT; this indicated that a perceived higher accuracy was associated with a lower error count, higher RT, and, paradoxically, with a higher perceived response speed. Figure 1 presents SART data per group for both error count and RT.

Repetition

SART error count, RT and CV of RT of the second SART session did not differ significantly from those of the first over all participants (mean difference SART error count = 0.86 ± 5.53, *95% C.I.* -0.27 – 1.99, *r* = 0.17, median difference average RT = 5 ms, -15 – 22 ms, *ns*, *r* = -0.14, mean difference CV = 0.01 ± 0.06, *95% C.I* -0.01 – 0.02, *r* = 0.11). Delta scores for accuracy and RT did not differ significantly between the four groups.

Fig. 1a SART error count **Fig. 1b SART RT**

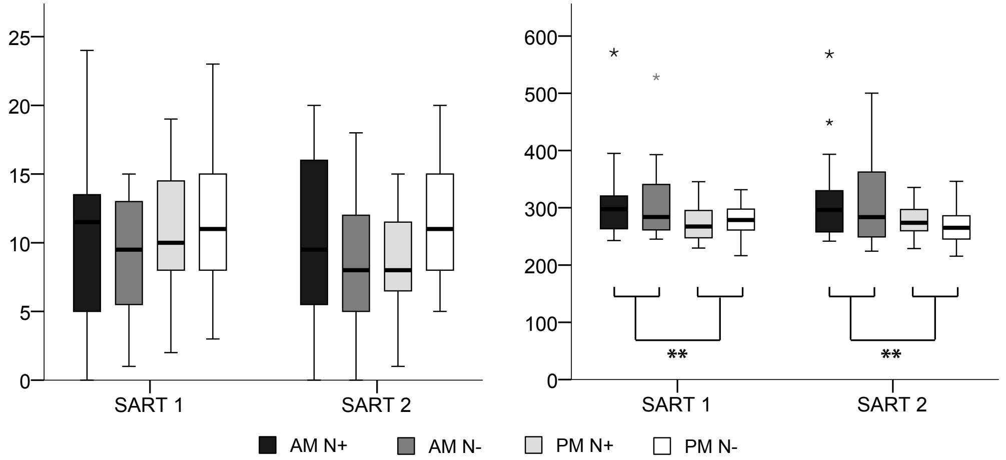


Figure 1 – A. SART error count per group, mean ± S.E. B. SART RT in ms per group.
Legend: AM: morning group; PM: afternoon group; N+/-: group with/without napping opportunity.
** indicate significant differences.

Time of day

The morning and afternoon groups did not differ significantly in SART error count between both sessions (mean difference first session = -1.43 ± 6.99, *95% C.I.* -3.66 – 0.81, *r* = -0.20, mean difference second session = -0.35 ± 6.65, *95% C.I.* -2.48 – 1.78, *r* = -0.05), nor in CV of RT (mean difference first session = 0.01 ± 0.07, *95% C.I.* -0.02 – 0.03, *r* = 0.08, mean difference second session = 0.00 ± 0.10, *95% C.I.* -0.03 – 0.04, *r* = 0.04). However, the morning groups had a longer RT compared to the afternoon groups on both session 1 (median difference average RT = 20 ms, -9 – 67 ms, *p* = 0.01, *r* = -0.40) and 2 (median difference average RT = 28 ms, -13 – 65 ms, *p* < 0.01, *r* = -0.44).

Napping

Neither SART error count (mean difference = 0.65 ± 7.15, *95% C.I.* -1.64 – 2.94, *r* = 0.09) nor average RT (median difference = -15 ms, -46 – 31 ms, *r* = 0.16) or CV of RT (mean difference = 0.02 ± 0.11, *95% C.I.* -0.02 – 0.05, *r* = 0.17) differed significantly between the second sessions of the N+ and N- groups. After correction for the first session by taking the difference with the second session (delta score), SART outcome measures still did not differ between the groups.

Subsequent analysis revealed that these measures did not differ either between the second SART sessions of the participants from the N+ groups who had actually slept (S+), compared to their matched counterparts in the N- group, nor to the participants from the nap groups who had remained awake (S-).

Table 2 – Linear Mixed Models

Model parameters	Non-adjusted model		Adjusted for RT		Adjusted for RT*T	
Basis	***Beta / S.E. / p***					
Intercept	8.88/1.00/0.00	*	18.46/2.17/0.00	*	15.55/2.42/0.00	*
Target factors	***Beta / S.E. / p***					
Time of day (T)	2.78/1.41/0.05		0.34/1.04/0.74		18.08/4.32/0.00	*
Nap occasion (N)	1.58/1.41/0.27		2.02/1.01/0.05		1.98/0.99/0.05	
Test session (S)	N.A.		N.A.		0.04/0.55/0.94	
Interactions	***Beta / S.E. / p***					
T*N	-3.78/1.99/0.06		-2.83/1.44/0.05		-2.66/1.41/0.06	
T*S	N.A		N.A		-1.54/0.78/0.05	
Covariates	***Beta / S.E. / p***					
SSS	N.A.		N.A.		N.A.	
RT	N.T.		-0.06/0.01/0.00	*	-0.05/0.01/0.00	*
CV	N.T.		33.02/4.79/0.00	*	33.74/4.58/0.00	*
RT*T	N.T.		N.T.		-0.06/0.01/0.00	*

Legend: RT: reaction time; SSS: Stanford Sleepiness Scale; CV: coefficient of variation of RT; Beta: regression coefficient derived from the linear mixed model; S.E: standard error of the regression coefficient; N.A: not available, i.e. no significant contribution to the final model; N.T: not tested in the model.

Model building strategy: *Compound symmetry* was chosen as model for the covariance matrix. The model of the mean was created from a saturated model including all target factors and possible interactions between them, followed by removing non-significant parameters as long as the model fit was not significantly impaired. The interactions N*S and T*N*S did not contribute significantly to any of the tested models and were therefore omitted from this table.
Models including the interaction of RT with all three factors were tested, but only the interaction of RT*T resulted in a different model and was therefore displayed. Three-way or higher-order interactions with RT were not modeled. Asterisks flag significant LMM coefficients.

Linear Mixed Models of SART error count corrected for RT

Table 2 presents the model parameters of three LMMs of the combined effects of repetition, time of day and napping on SART error count, firstly unadjusted, secondly adjusted for RT measures, and thirdly including the interaction of RT and time of day. As the latter model fitted our data best, conclusions about the investigated conditions were based upon this model. In line with the correlations mentioned above, the adjusted LMMs indicated a significant effect of CV of RT on SART error count and a significant inverse effect of RT on SART error count. In other words, the higher the CV or the shorter the RT, the more errors were made ($p < 0.01$). The latter was even more pronounced for the afternoon groups, as demonstrated by the interaction effect of RT and time of day ($p = 0.01$): for every additional 25 ms of RT the error count decreased with 1.25 errors for the morning groups, compared to 2.75 errors for the afternoon groups. As the model including this interaction also showed a significant positive effect of time of day on SART error count ($ß = 18.04$, $p = 0.01$), combining these findings indicates that the afternoon groups made more errors than the morning groups for RTs below 301 ms, but fewer errors when RT exceeded this value: for an RT of 280 ms, the error count was 2.3 points higher for the afternoon groups compared to the morning groups.

DISCUSSION OF EXPERIMENT 1

SART performance in the first session was similar to performance in the second session for all four groups; no differences were found between N+ and N- groups or between morning and afternoon groups concerning accuracy. A significant speed-accuracy trade-off was observed.

Time of day

The groups performing the SART in the morning had longer RT but preserved accuracy compared to the afternoon groups. Since the expected speed-accuracy trade-off (Wickelgren, 1977) was indeed present, we corrected LMM analyses of SART error count for RT. The final LMM containing the interaction of RT with time of day indicated that the speed-accuracy trade-off was stronger for the afternoon groups: for every additional 25 ms 1.5 error less was made compared to morning groups. The intercept of the model was also higher for the afternoon groups after correction for RT. The size of this difference was rather small though: at a RT of 280 ms, afternoon groups made 2.3 more errors than morning groups. At a RT of 301 ms this difference disappeared.

Napping in between test sessions

The N+ and N- groups did not differ in either error rate or RT. The same was found for S+ participants compared to their matched counterparts in the N- group. As such, a napping opportunity as provided in an MSLT does not interfere with performance of healthy participants on a SART session 70 minutes later. Since our study did not comprise sleep-deprived participants or sleep-disordered patients, no inferences of the influence of a nap for these populations can be made based on this study.

Subjective sleepiness and SART performance

No correlations were found between SART performance and scores of the SSS administered immediately preceding the SART session. This appears to contrast with the study by Manly and colleagues, who found an inverse correlation between errors of commission and SSS rating (Manly et al., 2002). However, their correlation is derived from large differences in both SSS and SART measurements at 13:00 and 19:00 on the one hand versus 01:00 and 07:00 on the other hand. As we did not observe significant differences in SART performance or SSS score between different times of day within our shorter time interval, possible correlations between those measurements would have to be rather strong to reach significance. Moreover, strong correlations were not expected in the present study, because all conditions that produce marked sleepiness were excluded. Despite all this, the lack of a correlation could also indicate that the SSS and SART measure different phenomena: the SSS reflects subjectively experienced sleepiness while the SART reflects sustained attention.

Repetition

No effect of repetition was observed in any group, nor in any of the sleep or time-of-day conditions in which two groups were combined. Although this is consistent with previous research by McAvinue in healthy controls (McAvinue et al., 2005), it contrasts with earlier findings from our study group in data of controls (Fronczek et al., 2006) and patients with excessive daytime sleepiness (Van Schie et al., 2012).

As such, the current experiment failed to disentangle the mechanism responsible for this prior observation, even though the possible contributions of time of day and napping were separated, and two consecutive sessions were administered to SART-naïve participants to be able to catch possible improvements due to a learning effect.

SART error count

In addition to the absence of a difference in SART error count between the first and the second session, the average error count was, at 10.1 errors (median error count at 10.0) rather higher than the median error count of 2.0 from the healthy control participants in our previous study (Fronczek et al., 2006). The currently observed error count even approached the range previously observed for patients with narcolepsy (median error count of 10.6-11.2 (Fronczek et al., 2006; Van Schie et al., 2012)). A possible explanation might reside in task instruction: in the current experiment, participants were instructed to give equal importance to accuracy and speed in performing the task, which is how the instruction was originally proposed by Robertson and colleagues in 1997 (Robertson et al., 1997). In contrast, participants were instructed to perform the SART as accurately as possible in the study by Fronczek (Fronczek et al., 2006).

Until now, we had not considered test instruction critically important for the

SART error count, as healthy participants in earlier studies (Manly et al., 1999; Robertson et al., 1997; Zordan, Sarlo, & Stablum, 2008) who received the original instruction, had similarly low error counts as found in Fronczek's study. When there is a significant speed-accuracy trade-off, error counts may, however, only be compared between studies when RT is taken into account; the RTs of these studies were either similarly high as in Fronczek's study (373 ms, (Robertson et al., 1997) compared to 393 ms, (Fronczek et al., 2006)), or not presented (Manly et al., 1999; Zordan et al., 2008). Based on the current results, a high error count with a low RT of 280 ms, we hypothesized that the instruction difference might have affected the results. Since participants with a long RT made fewer errors, it appeared plausible that changing the instruction towards not paying attention to RTs would lead to longer RTs but fewer errors, as a result of the speed-accuracy trade-off.

A second experiment was therefore conducted to investigate whether the instruction that was given to healthy participants on how to perform the SART influenced SART performance, in particular the size of the error count.

EXPERIMENT 2

METHODS

Participants

Five healthy participants were randomly chosen from each of the four groups of experiment 1, and new participants were recruited to match these 20 controls on age, sex and level of education. The new participants provided written informed consent prior to the study and received the same incentive as the controls had received. Characteristics of the I_M and I_0 groups are presented in Table 3. The groups did not differ significantly in ESS score, sleep duration, or the number of days per week with a deviation from habitual bed time with more than one hour. All participants had either completed or were following higher education at the time of testing (not indicated in the table).

Design

The study design for the new participants was exactly the same as for the participants from experiment 1 (i.e. 5 participants per combination of time of day/napping) with only one difference: participants were instructed to pay attention to accuracy only (modified instruction, I_M), instead of giving equal importance to accuracy and speed in performing the task (original instruction, I_0).

Statistical analysis

Data were analyzed using IBM® SPSS® Statistics version 20. Firstly, the analyses consisted of paired *t*-tests or a non-parametric counterpart to investigate SART performance between the two instruction groups. Secondly, the analysis concerned LMMs to assess the combined effects of instruction and the other conditions on SART performance, i.e. main effects of instruction, repetition, time of day, and napping, as well as their interactions. SART error count was again corrected for RT measures in the analysis and the Holm-Bonferroni adjustment was used to account for multiple testing. LMM analysis was also used to compare participants' judgments about their performance with their objective performance across instruction groups.

Table 3 - Baseline characteristics of the study groups

	I_0 (N = 20)	I_M (N = 20)
N of males	6	6
Age in years	26.7 ± 11.4	28.0 ± 11.5
ESS	4.9 ± 2.4	5.1 ± 2.7
Average nighttime sleep (hrs.)*	8.5 ± 0.8	8.4 ± 0.6
for weekdays*	8.2 ± 0.9	8.2 ± 0.7
for weekend days*	9.3 ± 1.1	9.0 ± 0.8
Nr of days per week with deviation from habitual bed time**	1.0 ± 0.9	1.1 ± 0.9
last 7 days	1.3 ± 1.2	1.1 ± 0.8

Legend: I_0: original instruction; I_M: modified instruction; ESS: Epworth Sleepiness Scale; Nr: number.
* Calculated from a subjective report of bed times and wake-up times.
** Deviation is defined as > 1.0 hour earlier or later than habitual bed time.

RESULTS

SART performance

The mean error count of the I_0 group was 10.5 ± 4.3, the median RT was 280 ms (261 - 303) and the mean CV of RT was 0.23 ± 0.06 ms. These values were 5.2 ± 2.7, 319 ms (282 - 409 ms) and 0.23 ± 0.07 ms for the I_M group. Figure 2 presents SART data per group for both error count and RT. The error count was significantly lower in the I_M group than in the I_0 group for both session 1 (mean difference = 3.55 ± 7.12, *95% C.I.* 0.22 – 6.88, $p = 0.04$, $r = 0.46$) and session 2 (mean difference = 6.90 ± 6.14, *95% C.I.* 4.03 – 9.77, $p < 0.01$, $r = 0.76$). SART RT was significantly higher in the I_M group than in the I_0 group for both session 1 (median difference average RT = 35 ms, -2 – 149 ms, $p = 0.02$, $r = 0.54$) and session 2 (median difference average RT = 56 ms, 1 – 209 ms, $p = 0.02$, $r = 0.54$). CV of RT did not significantly differ between I_M and I_0 groups (mean difference for the first session = -0.03 ± 0.11, *95% C.I.* -0.08 – 0.03, $r = 0.16$, for the second session = 0.02 ± 0.11, *95% C.I.* -0.03 – 0.08, $r = 0.14$). $VAS_{accuracy}$ was significantly and negatively correlated with SART error count (-2.93, *C.I.* -3.87 – -1.99, $p < 0.01$), and positively with VAS_{RT} (0.30, *C.I.* 0.10 – 0.50, $p < 0.01$), irrespective of instruction; this indicated that a perceived higher accuracy was associated with a lower error count and a higher perceived response speed in both instruction groups.

Repetition

In contrast to the I_0 group, SART error count decreased from the first to the second session (mean difference = 3.55 ± 3.25, *95% C.I.* 2.03 – 5.07, $p < 0.01$, $r = 0.75$), as did CV of RT (mean difference = 0.04 ± 0.05, *95% C.I.* 0.02 – 0.06, $p < 0.01$, $r = 0.63$). While error count and CV of RT differed between the first and second session of the I_M group, RT did not (median difference average RT = 10 ms, -27 – 43 ms, $r = 0.16$).

Table 4 – Linear Mixed Models

Model parameters	**Non-adjusted model**		**Adjusted for RT**		**Adjusted for RT*I**	
Basis	***Beta / S.E. / p***					
Intercept	10.55/0.95/0.00	*	10.88/1.87/0.00	*	15.78/2.35/0.00	*
Target factors	***Beta / S.E. / p***					
Instruction (I)	-3.55/1.34/0.01	*	-2.98/1.12/0.01		-9.26/2.62/0.00	*
Time of day (T)	N.A.		1.53/1.10/0.17		0.56/0.93/0.55	
Nap occasion (N)	N.A.		3.20/1.33/0.02		3.15/0.99/0.00	*
Test session (S)	-0.20/1.01/0.84		1.14/1.13/0.32		-1.08/0.56/0.06	
Interactions	***Beta / S.E. / p***					
I*N	N.A.		1.22/1.43/0.40		N.A.	
I*S	-3.35/1.42/0.02	*	-1.32/1.16/0.26		N.A.	
T*N	N.A.		-3.96/1.35/0.01	*	-4.09/1.33/0.01	*
T*S	N.A.		-1.73/1.13/0.14		N.A.	
N*S	N.A.		-1.32/1.14/0.26		N.A.	
Covariates	***Beta / S.E. / p***					
SSS	N.A.		N.A.		N.A.	
RT	N.T.		-0.03/0.00/0.00	*	-0.05/0.01/0.00	*
CV	N.T.		32.10/4.87/0.00	*	31.96/4.50/0.00	*
RT*I	N.T.		N.T.		0.02/0.01/0.02	*

Legend: RT: reaction time; SSS: Stanford Sleepiness Scale; CV: coefficient of variation of RT; Beta: regression coefficient derived from the linear mixed model; S.E: standard error of the regression coefficient; N.A: not available, i.e. no significant contribution to the final model; N.T: not tested in the model.

Model building strategy: *Compound Symmetry* was chosen as model for the covariance matrix. The model of the mean was created from a saturated model including all target factors and possible interactions between them, followed by removing non-significant parameters as long as the model fit was not significantly impaired. The interactions I*T, I*T*S, I*N*S, T*N*S, I*T*N and I*T*N*S did not contribute significantly to any of the tested models and were therefore omitted from this table.
Models including the interaction of RT with all three factors were tested, but only the interaction of RT*instruction resulted in a different model and was therefore displayed. Three-way or higher-order interactions with RT were not modeled. Asterisks flag significant LMM coefficients.

Linear Mixed Models of SART error score corrected for RT

Table 4 presents the model parameters of three LMMs of the combined effects of instruction, repetition, time of day and napping on SART error score, firstly unadjusted, secondly adjusted for RT measures, and thirdly including the interaction of RT and instruction. The latter model had the best fit and was used. A significant effect of CV of RT on SART error count and a significant inverse effect of RT on SART error count were observed for both instruction groups, but the latter was less pronounced for the I_M group: for every additional 25 ms, 1.25 errors less were made in the I_0 group compared to 0.75 errors in the I_M group. After correction for RT, LMM indicated that SART error count was still lower in the I_M group than in the I_0 group (*95% C.I.* -14.56 – -3.96): at an RT of 300 ms, for instance, the size of the difference was modeled to be 3.26 errors. The non-significant contributions of the interaction effects of instruction with either or both time of day and napping indicated that the differences between I_0/I_M groups were similar for participants tested in the morning versus afternoon, and with or without napping opportunity. The combined effect of napping with the interaction effect of napping and time of day, irrespective of instruction, indicated that the error count was higher in the morning nap group.

4

Fig. 2a SART error count **Fig. 2b SART RT**

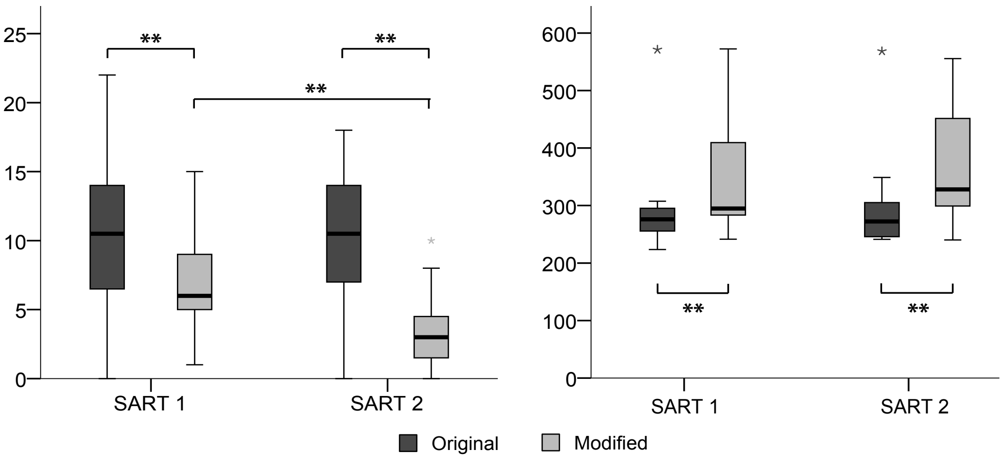


Figure 2– SART performance separated for the original and modified instruction groups. A. Mean SART error count ± S.E. B. SART RT in ms per group. ** indicate significant differences.

DISCUSSION OF EXPERIMENT 2

Participants who had been instructed to only pay attention to accuracy made significantly fewer errors but performed more slowly and with less variation in reaction times than those instructed to pay equal attention to accuracy and speed. The lower error count remained significant after adjusting for RT.

Instruction

The modified instruction was strongly associated with a more than two-fold decrease in SART error score on the second session (effect size = 0.76). As effect sizes > 0.70 correspond to clinically relevant differences, the instruction thus significantly and relevantly influenced SART performance (Cohen, 1988).

The speed-accuracy trade-off was less pronounced but still present for the I_M group, indicating that response strategy may still influence the error count. However, the remaining net effect was small: to reduce the error count by 1 error the I_M group had to prolong responses by 33 ms. As the decrease in error count was accompanied by a small non-significant increase in RT, it is possible that the improved SART performance of the second session in the first experiment is due to participants developing a 'slower but more accurate' strategy. Interestingly, participants were not aware of such a strategy: the higher participants estimated their accuracy, the faster they estimated their response speed. This yielded for both instruction groups. Apart from reaction times per se, participants in the I_M groups showed less variation in their reaction times. This could possibly be interpreted as a learning effect.

Effect size of instruction

The error counts of the I_M group resembled those of controls in the study of Fronczek et al: 7.0 in the current experiment compared to the previous 6.0 for the first session, and 3.5 compared to 2.0 for the second session. The modified instruction thus likely accounted for the difference in height of error score between the participants from the first experiment and our previous study, as well as for the effect of repetition. Comparing the error count of our I_0 group to that in studies with the same instruction, the mean error count of the I_0 group (10.5) resembled one of them, a study by 't Hart et al (9.7 errors, (Hart et al., 2012)), but is somewhat higher compared to a third study (5.9, (Zordan et al., 2008)).

GENERAL DISCUSSION

This study investigated the influences of time of day, napping, repetition, and instruction on the performance on two consecutive sessions of the SART in healthy participants.

The aim was to unravel the mechanism responsible for a decrease in SART error count from the previously found marked drop from the first to the second session. Our results demonstrated that such an improvement is only found when participants are instructed to pay attention to accuracy and to ignore response speed. The improvement is likely attributed to an effect of repetition, i.e. a learning effect, although participants were not aware of such an effect, given their own performance judgments. The link to instruction also explained why one of our previous studies did find a learning effect (Fronczek et al., 2006), in contrast to other studies (McAvinue et al., 2005).

SART in sleep medicine

The associations between instruction and error count and between instruction and learning effect need not necessarily hold to the same degree for patients with sleep disorders. The error rate of patients with narcolepsy who received the instruction to pay attention to accuracy only was similar to that of patients with narcolepsy who received the instruction to pay equal attention to both accuracy and speed (Fronczek et al., 2006; Van Schie et al., 2012). It seems likely that patients with narcolepsy already function at maximum task capacity when instructed to pay equal attention to both accuracy and speed: their long RT (mean of 337 ms) suggests that speed was already sacrificed at the expense of accuracy (Van Schie et al., 2012), so that dropping the speed condition would not result in a better accuracy. Their low level of accuracy compared to controls (Fronczek et al., 2006), i.e. their inability to sustain attention to a 4-minute lasting task, may very well reflect the problems patients with narcolepsy face in daily life when trying to follow a conversation or read a book.

Our previous study that used the original instruction (Van Schie et al., 2012), also investigated SART performance in patients with idiopathic hypersomnia and obstructive sleep apnea. Their performance did not significantly differ from that of patients with narcolepsy. It would be interesting to investigate the modified instruction in these conditions as well. Again, patients with these disorders might function at their maximum capacity when instructed to pay equal attention to both accuracy and speed.

Implications for the use of the SART

The results indicate that the SART discriminates better between healthy controls and patients with narcolepsy when the instruction is given to prefer accuracy to speed, than when accuracy and speed are considered equally important. We therefore recommend the “accuracy first” instruction. To minimize the consequences of the learning effect that has been observed when using this instruction, we strongly recommend the use a full practice session, i.e. a 225-trial session instead of the 30-trial session that was used as practice session in both our experiments and the manuscript by Fronczek et al (Fronczek et al., 2006). The rationale for this recommendation is the observation that the higher error count in the first session of Fronczek's study was followed by stable, lower error counts in the second to the fifth session. In other words, regarding this first session as practice session minimizes the consequences of the learning effect.

The present study also showed that a nap opportunity of 20 minutes more than 1 hour prior to a SART session did not strongly influence the error count of that session in healthy participants. As such, SART sessions obtained from an MSLT design are likely to be suitable for comparison with separate SART sessions from a follow-up occasion. Before doing so, the question whether patients with sleep disorders profit or not from a nap should be answered.

The time of day had no clear effect on the SART error count, and if such an effect exists at all, it is rather small and occurred only following the original instruction. The modified instruction allows a comparison between SART sessions administered at different times of the day during normal working hours.

To conclude, instructing healthy participants to perform the SART as accurately as possible leads to a lower error count with lower between-subject variability, and is thus the preferred instruction to assess the best performance in terms of error count that a subject can achieve.

ACKNOWLEDGEMENTS

We thank P.P. Buijserd-Amesz, P.J. van Someren, W.W.M. Huyser, A. Janson and S. van Gent for their help in acquiring the data, J.J.M. Witteman for the scoring of sleep and R.H.A.M. Reijntjes for his help in the processing of the data.

REFERENCES

1. Coenen, A. M. (1995). Neuronal activities underlying the electroencephalogram and evoked potentials of sleeping and waking: implications for information processing. *Neurosci Biobehav Rev, 19*(3), 447-463.
2. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
3. Fronczek, R., Middelkoop, H. A., van Dijk, J. G., & Lammers, G. J. (2006). Focusing on vigilance instead of sleepiness in the assessment of narcolepsy: high sensitivity of the Sustained Attention to Response Task (SART). *Sleep, 29*(2), 187-191.
4. Hart, E. P., Dumas, E. M., Reijntjes, R. H., van der Hiele, K., van den Bogaard, S. J., Middelkoop, H. A., . . . van Dijk, J. G. (2012). Deficient sustained attention to response task and P300 characteristics in early Huntington's disease. *J Neurol, 259*(6), 1191-1198. doi: 10.1007/s00415-011-6334-0
5. Johns, M. W. (1991). A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep, 14*(6), 540-545.
6. Littner, M. R., Kushida, C., Wise, M., Davila, D. G., Morgenthaler, T., Lee-Chiong, T., . . . Kramer, M. (2005). Practice parameters for clinical use of the multiple sleep latency test and the maintenance of wakefulness test. *Sleep, 28*(1), 113-121.
7. Manly, T., Lewis, G. H., Robertson, I. H., Watson, P. C., & Datta, A. (2002). Coffee in the cornflakes: time-of-day as a modulator of executive response control. *Neuropsychologia, 40*(1), 1-6.
8. Manly, T., Robertson, I. H., Galloway, M., & Hawkins, K. (1999). The absent mind: further investigations of sustained attention to response. *Neuropsychologia, 37*(6), 661-670.
9. McAvinue, L., O'Keeffe, F., McMackin, D., & Robertson, I. H. (2005). Impaired sustained attention and error awareness in traumatic brain injury: implications for insight. *Neuropsychol Rehabil, 15*(5), 569-587. doi: 10.1080/09602010443000119
10. Morad, Y., Lemberg, H., Yofe, N., & Dagan, Y. (2000). Pupillography as an objective indicator of fatigue. *Curr Eye Res, 21*(1), 535-542.
11. Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). 'Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia, 35*(6), 747-758.
12. Schmidt, C., Collette, F., Cajochen, C., & Peigneux, P. (2007). A time to think: circadian rhythms in human cognition. *Cogn Neuropsychol, 24*(7), 755-789. doi: 10.1080/02643290701754158
13. Valley, V., & Broughton, R. (1981). Daytime performance deficits and physiological vigilance in untreated patients with narcolepsy-cataplexy compared to controls. *Rev Electroencephalogr Neurophysiol Clin, 11*(1), 133-139.

14. Van Schie, M. K., Thijs, R. D., Fronczek, R., Middelkoop, H. A., Lammers, G. J., & Van Dijk, J. G. (2012). Sustained attention to response task (SART) shows impaired vigilance in a spectrum of disorders of excessive daytime sleepiness. *J Sleep Res, 21*(4), 390-395. doi: 10.1111/j.1365-2869.2011.00979.x
15. Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychol (Amst), 41*, 67-85.
16. Zordan, L., Sarlo, M., & Stablum, F. (2008). ERP components activated by the “GO!” and “WITHHOLD!” conflict in the random Sustained Attention to Response Task. *Brain Cogn, 66*(1), 57-64. doi: 10.1016/j.bandc.2007.05.005