Behavior Research Methods (2025) 57:299
https://doi.org/10.3758/s13428-025-02808-x

ORIGINAL MANUSCRIPT

Testing thought-probe frequency for measuring mind-wandering along with vigilance and cognitive control loss: A study with the ANTI-Vea task

María Julieta Aguirre[1] · Pablo Barttfeld[1] · Elisa Martín-Arévalo[2] · Juan Lupiáñez[2] · Fernando G. Luna[3]

Received: 14 October 2024 / Accepted: 12 August 2025

Abstract

Vigilance decrement refers to the decline in sustained attention over time during prolonged tasks, which often leads to increased errors and accidents. However, to date, there are no experimental tasks that simultaneously measure changes in vigilance, cognitive control, and mind-wandering (MW) across time-on-task. We adapted the Attentional Network Test for Interactions and Vigilance–executive and arousal components (ANTI-Vea) task to integrate mind-wandering measures along with assessments of vigilance and cognitive control. By inserting thought probes (TPs) at different frequencies per block, we aimed to identify the optimal TP rate to capture mind-wandering changes without interfering with the measurement of vigilance, thereby providing an integrative assessment of changes in mind-wandering, cognitive control, and vigilance across time. We conducted two experiments: one in the laboratory with 90 students from the National University of Córdoba, Argentina, and another online, as a replication, with 180 students from the University of Granada, Spain. Participants were divided into three groups (4, 8, 12 TPs per block) and completed the ANTI-Vea-TP task. The results revealed that the inclusion of TPs was effective in detecting changes in mind-wandering over time-on-task. Moreover, TP frequency did not have a significant effect on mind-wandering reports, vigilance, or cognitive control over time-on-task. We discuss the potential suitability of this tool for investigating the interaction between vigilance, cognitive control, and mind-wandering, in both laboratory and online environments, which is essential for evaluating different theories of vigilance decrement.

Keywords Mind-wandering · Vigilance · Cognitive control · Thought-probe frequency

Introduction

Vigilance decrement refers to the reduction in the ability to sustain attention when performing a task for an extended period (Hancock, 2017). This decrease in vigilance has been strongly associated with an increase in errors and accidents in tasks that demand sustained attention over long periods, such as working during prolonged shifts and driving environments (Edkins & Pollock, 1997; Read et al., 2012). Thomson and colleagues (2015) developed a model to account for vigilance loss that emphasized the critical role of mind-wandering (MW) and cognitive control in prolonged tasks without breaks. However, there is a growing need to develop behavioral tasks embedding several measures of vigilance and attentional components along with MW states (Luna et al., 2022; Murray et al., 2020; Thomson et al., 2015). Discrepancies and a lack of consistency have been observed among independent studies on measures used to assess MW, making it difficult to compare and generalize

✉ María Julieta Aguirre
maria.julieta.aguirre@mi.unc.edu.ar

✉ Juan Lupiáñez
jlupiane@ugr.es

✉ Fernando G. Luna
fluna@unc.edu.ar

1 Cognitive Science Group, Facultad de Psicología, Instituto de Investigaciones Psicológicas (IIPsi, CONICET-UNC), Universidad Nacional de Córdoba, Córdoba, Argentina

2 Department of Experimental Psychology, and Brain, Mind, and Behavior Research Center (CIMCYC), University of Granada, Granada, Spain

3 Facultad de Psicología, Universidad Nacional de Córdoba, Córdoba, Argentina

Published online: 03 October 2025

results (Weinstein, 2018). The present study aimed to adapt an existing and robust task, the Attentional Network Test for Interactions and Vigilance–executive and arousal components (ANTI-Vea task; Coll-Martín et al., 2023; Hemmerich et al., 2023; Luna et al., 2018; Luna, et al., 2021a; Luna et al., 2022; Luna et al., 2023a, 2023b), to integrate a measure that detects changes in MW while measuring changes in different components of vigilance and cognitive control.

In recent years, a dissociation between two components of vigilance has been proposed (Luna et al., 2018): executive and arousal vigilance. On the one hand, executive vigilance (EV) refers to the maintenance of attention to monitor the occurrence of rare but critical events that require specific responses to be detected. EV has been studied using signal-detection tasks that demand the detection of infrequent stimuli, such as the Mackworth clock test (Mackworth, 1948) or the sustained attention to response task (SART, Robertson et al., 1997). On the other hand, arousal vigilance (AV) refers to the ability to maintain an optimal state of alertness to react automatically and quickly to environmental stimuli, without the need to select specific responses (Langner & Eickhoff, 2013; Luna et al., 2018). AV is assessed through simple reaction time tasks that involve rapid responses to stimuli with minimal control over prolonged periods, such as the psychomotor vigilance test (PVT) (Dinges & Powell, 1985).

Although several theories have been developed to explain the vigilance decrement phenomenon, there is still an open debate concerning the mechanisms that lead to a progressive loss of vigilance (Esterman & Rothlein, 2019; Neigel et al., 2020). The resource depletion hypothesis posits that vigilance works through a limited pool of resources that is not automatically reloaded, and because vigilance tasks are difficult to perform, when such tasks are performed over a prolonged period, resources are progressively depleted over time and vigilance decreases (Caggiano & Parasuraman, 2004; Warm et al., 1998). Conversely, the MW hypothesis holds that vigilance tasks are instead monotonous and boring, causing attentional resources to wander from the task at hand towards task-unrelated thoughts, making it difficult to maintain attention on the external task and therefore resulting in decreased vigilance (Smallwood & Schooler, 2006).

An alternative framework has been proposed by Thomson et al. (2015)—the resource-control theory—which integrates predictions by the resource depletion and MW hypotheses, emphasizing the central role of cognitive control. According to the resource-control theory, the amount of attentional resources available is fixed and does not change over time. As MW is our default state, when performing an external task, task-irrelevant thoughts consume attentional resources that should be dedicated to the external task (Smallwood, 2010; Smallwood & Schooler, 2006). To avoid resources being devoted to task-unrelated thoughts, cognitive control is necessary to maintain attentional resources on the task at hand, thus preventing MW. Importantly, cognitive control is difficult to maintain across time and therefore tends to decrease. Cognitive control loss might cause attentional resources to be automatically diverted from the external task and progressively redirected to task-unrelated thoughts, consequently leading to decreased vigilance (Thomson et al., 2015).

To empirically test predictions by the resource-control theory, changes in vigilance, cognitive control, and MW across time should be simultaneously assessed. However, to our knowledge, no available method is suitable to simultaneously measure these three phenomena. The ANTI-Vea task seems a promising tool to advance in this direction. The ANTI-Vea is an innovative tool designed to simultaneously assess the classic attentional network components—namely phasic alertness, orienting, and cognitive control—along with changes in executive and arousal vigilance over time (Luna et al., 2021a). Indeed, the ANTI-Vea task has been successfully employed in many studies (Coll-Martín et al., 2023; Feltmate et al., 2020, 2020; Hemmerich et al., 2023; Huertas et al., 2019; Luna et al., 2018; Luna et al., 2021a; Luna et al., 2022; Luna et al., 2023a, b; Román-Caballero et al., 2021; Sanchis et al., 2020), providing a substantial corpus of data for performing different analyses, as the database of over 600 participants, in both laboratory and online settings, used to assess the reliability of the different attentional components measured by the task (Luna et al., 2021b).

In the ANTI-Vea task, the decrement in EV is observed as a progressive decrease in hits to correctly detect infrequent signals, while the decrement in AV is measured as a progressive increase in the mean and variability of reaction time (RT) (Luna et al., 2018; Luna et al., 2021b). Importantly, a decrease in cognitive control has also been observed via the ANTI-Vea, as an increase in the interference effect for selecting a target among distractors in the flanker subtask in RT and errors, and an increase over time in the inverse efficiency (IE) score of interference (Luna et al., 2022).

To test some of the predictions of resource control theory, Luna et al. (2022) analyzed data from a large sample ($N = 589$) gathered via the ANTI-Vea. The authors found that cognitive control, EV, and AV decreased over time. Most importantly, a negative correlation between changes in EV and cognitive control was observed, meaning that both components decreased with time-on-task. These results provided empirical evidence partially supporting the predictions of resource-control theory, specifically regarding the decline in cognitive control and its correlation with a decline in vigilance. However, and importantly, the task used by Luna et al. (2022) did not include a direct measure of MW. Therefore, it remained necessary to develop a task that allows for measuring changes over time in vigilance components, cognitive control, and MW, which was the main aim of the current study.

Incorporating MW measures in the ANTI-Vea may present challenges, as these measures might interrupt the vigilance decrement, thereby affecting the expected changes in EV, AV, and cognitive control. Furthermore, there is no clear consensus on the number of thought probes (TPs) or the time interval between TPs that should be used in a vigilance task to measure changes in MW (Murray et al., 2020; Weinstein, 2018).

Previous research on MW has mainly used the probe-caught method to capture changes in MW (Robison et al., 2019; Seli et al., 2013; Smallwood & Schooler, 2006), which involves interrupting the ongoing task with a TP that explicitly queries the individual about their current focus of attention (Kane et al., 2021; Weinstein, 2018). However, it is important to note that the probe-caught method is not standardized, and there is considerable variability in the rate of TPs within a task that aims to measure changes in MW (Weinstein, 2018). Such diversity in the methods for measuring MW with TPs can affect both the reporting of MW and the behavioral performance related to the ongoing task (Robison et al., 2020). Wiemers and Redick (2019) conducted a within-participant study to determine whether performance on a vigilance task (i.e., the SART) was affected by TP inclusion. The results indicated no significant differences in SART performance based on TP presence or absence. According to Wiemers and Redick, these findings suggest that TP measurement is a nonreactive method for assessing MW in attention and inhibition tasks.

Another critical factor contributing to the methodological diversity in measuring MW is the time interval between two TPs. For instance, TPs that are too close together might not allow enough time for the mind to shift from task-related to task-unrelated thoughts, whereas a long interval between two TPs may not capture differences between on-task and off-task states (Seli et al., 2013). Seli et al. (2013) examined how the rate of TPs affects the tendency to report periods of MW during a sustained attention task. Using the metronome response task, the authors pseudo-randomly distributed between 5 and 25 TPs across 600 trials, with the constraint that they had to be spaced at least 10.4 s apart. The total duration of the metronome response task was approximately 15 min. The results showed a positive relationship between the rate of probe presentation and the frequency of MW reports, suggesting that longer intervals between probes increase the likelihood of participants reporting MW. However, the authors noted that it was unclear whether this decrease was due to actual changes in MW experience or rather a reflection of reporting bias from responding to TPs in short time intervals (Seli et al., 2013).

Another aspect of methodological diversity in MW measurement is the frequency of TP presentation within the task (Murray et al., 2020). Robison et al. (2019) conducted a study to determine whether variations in TP frequency could influence behavior and MW reporting in the SART task. In their study, participants completed the semantic SART, which lasted approximately 14 min, and manipulated the frequency of TPs. The authors found no significant differences in behavioral performance or MW reporting as a function of TP frequency. Conversely, Schubert et al. (2020) showed that when TPs were presented more frequently, participants were less likely to report task-unrelated thoughts. In their study, MW was measured using TPs embedded in the SART. Participants were interrupted at either high frequency, approximately every 30 s with eight TPs per block, or low frequency, approximately every 60 s with four TPs per block, across six blocks, each containing a total of 810 trials.

Noting the relevance in analyzing changes in MW and cognitive control across time while measuring the vigilance decrement (Thomson et al., 2015) and the diversity between studies regarding the frequency of TPs within a task to assess MW (Robison et al., 2020; Weinstein, 2018), we decided to conduct the present study. We adapted the ANTI-Vea task by embedding pseudo-randomized trials of TPs (ANTI-Vea-TP). To evaluate the optimal number of TPs needed to obtain an adequate measure of MW in the ANTI-Vea, we examined changes in MW along with the typical measures of the ANTI-Vea between three experimental groups that performed the same task but with varying TP frequency (i.e., 4, 8, or 12) per block. The study comprised two separate experiments: Experiment 1, conducted within a controlled laboratory environment ($N = 90$), and Experiment 2 administered online ($N = 180$), conducted as a replication of Experiment 1. Nevertheless, for the sake of conciseness and given that Experiment 2 was conducted as a direct replication of Experiment 1, we decided to report the two experiments as a single study.

The protocol for Experiment 2, including sample size estimation, procedure, data analysis plan, and hypotheses, was preregistered in the OSF after conducting preliminary analyses of Experiment 1 (please see the Wiki at https://doi.org/10.17605/OSF.IO/KNDBR). It is important to note that the cited preregistration includes additional hypotheses and analyses that will be detailed in a subsequent theoretical study, while the current paper focuses on the suitability of adding TP as a measure of MW in the ANTI-Vea and its potential effects on the measurement of vigilance and attentional functions. Methods, raw data, and data analysis scripts of the present study are publicly available at https://doi.org/10.17605/OSF.IO/6ATHX.

The hypotheses examined in this study are as follows. On the basis of our preliminary analysis, and following Robison et al. (2019), we expected no difference in MW reports according to the number of TPs administered by block. We also anticipated replicating the typical main effects and interactions for phasic alertness, orienting, and cognitive control observed with ANTI-Vea (Luna et al., 2021b), regardless of TP frequency.

Importantly, we expected the ANTI-Vea-TP task to still show a decrease in EV, AV, and cognitive control (observed as increased interference in RT, errors, and inverse efficiency score) across blocks, as found in our preliminary data and previous research with the standard ANTI-Vea (Luna et al., 2022). However, we predicted that the TP frequency would not modulate the decrease in EV or AV across blocks.

Method

Participants

Experiment 1 was conducted in the laboratory with the participation of 90 volunteers (71 women; age $M = 22.64$; $SD = 4.28$), who were undergraduate students from the National University of Córdoba, Argentina. The sample size was similar to that used in a previous study with the ANTI-Vea and three groups of participants (Luna et al., 2020). Participants were randomly assigned to one of three groups (n by group = 30), based on the frequency of TPs by block, that is, 4, 8, or 12.

Experiment 2 was performed online. In this experiment, participants were volunteer undergraduate students from the University of Granada, Spain, who were invited through an institutional email list. During the initial phase, 302 volunteers completed an online survey. Next, in a second phase, participants were asked to participate in the experimental procedure. Participants who completed the first step had the opportunity to win a financial prize through a lottery system, while those who participated in the second step received a reward of 10 euros per hour for their participation in the study.

Experiment 1 showed the effects of interest significantly with a sample size of 30 participants per group. In order to conduct a direct replication with an increasing sample size, the number used in Experiment 2 was doubled (i.e., n by group = 60). Thus, 180 participants (144 women; age $M = 23.19$; $SD = 5.22$) who had completed the online survey in the initial phase were randomly selected and invited to complete the online behavioral task based on the following criteria: age between 18 and 40 years, having completed all the questionnaires of the online survey, and having correctly answered the control questions included in the survey to ensure understanding of the items. Participants were randomly assigned to one of three groups according to the frequency of TPs, as in Experiment 1.

A power analysis was conducted in R using the SuperPower package (Lakens & Caldwell, 2021), employing 10,000 simulations, to assess the statistical power of the interaction between experiment (online vs. laboratory) as a between-participant factor and block (six levels) as a within-participant factor. The mean and standard deviation of hits per block and per experiment condition were simulated from data collected in a previous study, conducted without TPs, that included both online and laboratory samples (Luna et al., 2021b). The analysis indicated that, with a sample size of 270 participants, a decrease in hits across blocks would be observed at an α level of.05, with an effect size of $\eta^2_p = .086$ and statistical power of $1 - \beta > .99$. However, the interaction between the experiment and the change across blocks would not be significant considering an α level of.05, and would be observed with an effect size of partial eta-squared = .005 and statistical power of $1 - \beta = .73$.

Furthermore, we conducted an additional power analysis based on data from Luna et al. (2020), wherein a significant interaction on the decrement of hits was observed in a mixed design with three groups. Given the alternative hypothesis that one of the TP groups might exhibit a mitigating effect on the hit rate, we simulated 10,000 samples using a dataset that included a between-participant factor with three levels, where electrical stimulation modulated performance across blocks. With a sample size of 270 participants, this analysis indicated that the interaction between block and group for hits would have statistical power of $1 - \beta > .99$ at an α level = .05, with an effect size of $\eta^2_p = .023$.

Informed consent was obtained from all participants in both experiments, following the ethical standards established in the 1964 Declaration of Helsinki (last updated: Fortaleza, 2013). All participants had normal or corrected-to-normal vision. Experiment 1 was approved by the Ethical Committee of the Institute of Psychological Research (CEIIPsi, protocol PE41, version 2), and Experiment 2 was approved by the University of Granada's Ethical Committee (2442/CEIH/2021).

Procedure and design

Experiment 1 began with the completion of a series of self-report questionnaires. Following this, participants performed the ANTI-Vea-TP. Finally, participants answered another series of questionnaires.

In Experiment 2, participants completed several self-report questionnaires online through the Lime Survey platform. Then, participants who met the selection criteria and were invited to continue with participation performed the same procedure as in Experiment 1 but online, in a suitable location where they could access the ANTI-Vea-UGR platform (https://anti-vea.ugr.es/) using a computer.

Self-report questionnaires

In Experiment 1, participants completed the Spanish version of the MW Deliberate and Spontaneous Scales (MW-D/MW-S) (Carriere et al., 2013; Cásedas et al., 2022), which includes two subscales, each of four items. These subscales assess the inclination to engage in MW, either intentionally (e.g., "I consciously allow my thoughts to wander") or spontaneously (e.g., "My mind tends to wander even when it should have been focused on another activity"). The items are rated on a seven-point scale, ranging from 1 (e.g., "rarely") to 7 (e.g., "very much"). They also completed the short version of the National Aeronautics and Space Administration (NASA) Task Load Index (NASA-TLX) (Arger & Nogareda, 1999) and the Dundee Stress State Questionnaire (DSSQ) (Sanchez-Ruiz et al., 2015).

In Experiment 2, to achieve other objectives beyond the current study, participants completed the MW-D/MW-S, Attentional Control Scale (Derryberry & Reed, 2002), Barkley Adult ADHD Rating Scale IV–Current Symptoms (Barkley, 2011), Difficulties in Emotion Regulation Scale–Short Form (Navarro et al., 2021), Irrational Procrastination Scale (Guilera et al., 2018), NASA-TLX, and DSSQ.

In both experiments, MW-D/MW-S and the first part of the DSSQ were administered before the ANTI-Vea-TP. Additionally, in Experiment 2, the Attention Control Scale, Barkley Adult ADHD Rating Scale IV, Difficulties in Emotion Regulation Scale–Short Form, and Irrational Procrastination Scale were included before the task. After the task, in both experiments, participants completed the NASA-TLX and the second part of the DSSQ.

The purpose of collecting these questionnaires in Experiments 1 and 2 was to correlate different self-reported measures with attention, vigilance, and MW performance scores. However, this goal is part of a larger research project, and so these analyses will be reported elsewhere when data from a larger number are completed. In the present study, which aimed to validate the MW score obtained via TPs in the ANTI-Vea-TP, only data from the MW-D/MW-S scale were analyzed.

ANTI-Vea-TP

In Experiment 1, the task was designed and run using PsychoPy 2022.1.4 (Peirce et al., 2019), while in Experiment 2, the online version of the task was run through the ANTI-Vea-UGR platform (https://anti-vea.ugr.es/) (Coll-Martín et al., 2023).

The ANTI-Vea-TP comprises six experimental blocks, in which four subtasks are combined: (a) ANTI (48 trials per block), to assess the main effects and interactions of phasic alertness, orienting, and cognitive control; (b) EV (16 trials per block), a signal-detection subtask similar to the Mackworth clock (Mackworth, 1948), to assess the EV decrement; (c) AV (16 trials per block), a RT subtask similar to the PVT (Basner & Dinges, 2011), to assess the AV decrement; and (d) TPs (4, 8, or 12 trials per block), to measure changes in MW across time. Each ANTI, EV, and AV trial has a fixed duration of 4,100 ms, and each TP trial lasts twice that duration (i.e., 8,200 ms).

The stimuli and presentation sequence in each trial of the ANTI-Vea-TP task can be observed in Fig. 1. In ANTI trials (see Fig. 1a), a set of five horizontally aligned arrows appears either above or below a fixation point located at the center of the screen, pointing to either the left or right. Participants must respond to the direction pointed by the

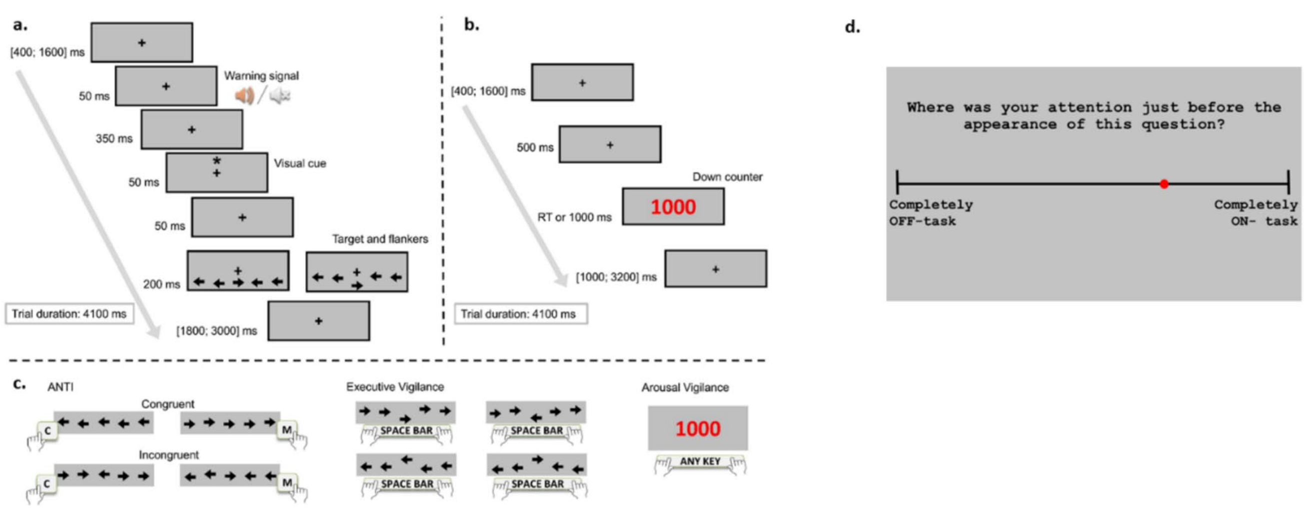

Fig. 1 Procedure for the ANTI-Vea task. **a** Stimuli sequence and timing for the ANTI and EV trials. **b** Stimuli sequence and timing for the AV trials. **c** The correct responses expected for the ANTI (see examples of congruency condition), EV, and AV trials. **d** Thought-probe trial with the continuous scale

target (i.e., the central arrow), while ignoring the direction pointed by the surrounding flankers. They are instructed to press "C" when the target points left and "M" when it points right. Additionally, randomly presented auditory warning signals and visual orientation cues can appear before the target stimulus. Phasic alertness is assessed by comparing the response in trials with (tone condition, 50% of ANTI trials) or without (no tone condition, 50% of ANTI trials) tone. Orienting is evaluated by comparing the response in trials with a valid visual cue (which predicts the correct location of the arrows regarding fixation, one third of ANTI trials), invalid visual cue (which predicts the opposite location, one third of ANTI trials), and no visual cue (one third of ANTI trials). Cognitive control is measured by comparing the response between trials where the distractor and the target point in the same (congruent trials, 50% of ANTI trials) or the opposite (incongruent trials, 50% of ANTI trials) direction.

EV trials follow the same procedure as the ANTI ones, except that the target appears largely displaced either upwards or downwards from its central position (see Fig. 1a). When the target is notably displaced, participants are instructed to press the space bar upon its appearance, regardless of the arrow direction (see Fig. 1c). Successful detection of displaced targets is considered a correct response (i.e., hit), while pressing the space bar when the target is not displaced (i.e., in ANTI trials) is considered a false alarm.

In AV trials (see Fig. 1b), no warning signal or visual cues are presented. In these trials, the string of arrows is replaced by a descending millisecond counter from 1,000 to 0 ms. Participants are instructed that when the counter appears, they must press any key to stop it as quickly as possible.

Finally, in TP trials, as shown in Fig. 1d, participants have to answer the following question: "Where was your attention just before the appearance of this question?" Participants must respond by moving a red dot that appeared at the center of the line and clicking the cursor on a continuous scale ranging from "completely on-task" (extreme right, coded as 1) to "completely off-task" (extreme left, coded as −1). The TPs appeared 4, 8, or 12 times per block, immediately after the previous trial. The number of TPs was proportional to the number of vigilance trials; that is, we decided to add 25%, 50%, and 75% of the 16 vigilance trials for each component (i.e., EV or AV). After participants responded, a fixation point appeared on the screen for a variable duration (see Fig. 1a and b), and was replaced by the TP question for the 8,200-ms trial duration. TP presentation was pseudo-randomized, so that there were at least five consecutive trials of any of the other types (i.e., ANTI, EV, and/or AV) as the interval between two TP trials (minimum time interval, 20 s and 500 ms).

Before the experimental blocks, participants completed a series of practice blocks to familiarize themselves with the task. The practice blocks were similar to those of the standard ANTI-Vea task, with four progressive practice blocks with and without visual feedback. In the fourth block, TP trials were added with visual feedback to indicate whether participants responded or not.

Statistical analyses

Analyses were conducted using R 4.2.0 (R Core Team, 2024) in RStudio 2022.02.3 (Posit team, 2024). Analyses of variance (ANOVAs) were performed using the afex package (Singmann et al., 2023). Planned contrasts were performed with the emmeans package (Lenth, 2021). Effect sizes and the 95% confidence intervals around them for ANOVAs and planned contrasts were computed with the effect size package (Ben-Shachar et al., 2020). Figures were prepared with Matplotlib (Hunter, 2007) and ggplot2 (Wickham et al., 2023).

Five participants were excluded from data analysis of Experiment 2: four due to a technical issue during data acquisition that prevented us from saving responses to the TP trials, and one participant due to an incorrect task parameter configuration of stimuli presentation. Consequently, the final sample comprised 265 participants, with 87 participants in the four-TP group (30 lab; 57 online), 90 participants in the eight-TP group (30 lab; 60 online), and 88 participants in the 12-TP group (30 lab; 58 online).

Given that the online experiment was conducted as a direct replication of the lab one, we pooled data from both experiments and treated the experiment as a between-participant factor to analyze any possible modulation between online and in-lab data collection.

ANTI-Vea-TP

Following the preregistration protocol, standard analyses for the ANTI-Vea task (Luna et al., 2021b) were conducted, incorporating the group (depending on the frequency of TPs, i.e., 4, 8, or 12 TPs per block) and experiment (online, in the lab) as between-participant factors in all analyses. For the sake of conciseness, the main effects and interactions regarding the experiment factor are presented in the supplementary material (see Tables S1– S6).

In all analyses including blocks as within-participant factor, the significance of the linear component was analyzed using polynomial contrasts.

Changes in MW across time-on-task were analyzed using a mixed ANOVA, with the mean of the response on the TP trial as the dependent variable and blocks as a within-participant factor. Additionally—and although it was not anticipated in our preregistered protocol—we conducted a series of supplementary analyses. First, we calculated the percentage of times participants indicated being “on-task” (i.e., with responses on the scale > 0), by block, to assess how MW reports fluctuated over time. Next, we divided the responses into two categories: when participants reported being on-task (i.e., position reported > 0) and when they were off-task (i.e., reported position < 0). Based on these categories, we obtained two key parameters: (a) the percentage of times participants were on-task compared to the time reported as off-task (MW), and (b) the degree of concentration during the on-task state and the intensity of distraction during MW episodes. These analyses allowed us to gain a more comprehensive understanding not only of the frequency of MW but also of the intensity of focus and distraction throughout the task.

For EV trials, warning signal, visual cue, and congruency levels were not considered for the analyses, and data were collapsed across these variables. Changes in EV were analyzed through four mixed ANOVAs, considering hits (correct identification of vertically displaced targets), false alarms (incorrect identification of non-displaced targets as being vertically displaced), and nonparametric indexes of sensitivity (A') and response bias (B'') as dependent variables, with experimental blocks as the within-participant factor. False alarms were calculated following the method developed by Luna et al. (2021a).

AV trials were analyzed via three mixed ANOVAs, including the mean RT, *SD* of RT, or the percentage of lapses (i.e., RT ≥ 600 ms) as the dependent variable, and blocks as a within-participant factor.

To analyze changes in cognitive control over time-on-task, three mixed ANOVAs were conducted as in Luna et al. (2022), with blocks as a within-participant factor. Dependent variables included the interference effect (i.e., the difference between incongruent and congruent trials) for RT (only correct responses and with RT between 200 and 1,500 ms were included), the percentage of errors, and the inverse efficiency (IE) score. The IE score combines RT and accuracy to assess performance in cognitive control tasks without trade-offs between speed and accuracy (Bruyer & Brysbaert, 2011). The IE score, expressed in ms, represents the average RT in situations of perfect accuracy (i.e., when a 100% correct response rate is achieved). To calculate it, the mean correct RT is divided by the proportion of correct responses.

The main effects and interactions of phasic alertness, orienting, and cognitive control were analyzed in the ANTI trials. Trials with incorrect responses (6.91% of trials) and with RT below 200 ms or above 1,500 ms (1.80% of trials) were excluded from the analysis. Two mixed ANOVAs were conducted, one with mean correct RT and the other with percentage of errors as dependent variable. Warning signal (no tone/tone), visual cue (invalid/no cue/valid), and congruency (congruent/incongruent) factors were included as within-participant factors.

Bayesian analyses

A Bayesian approach was employed for data analysis using JASP (version 0.19.3.0) (JASP Team, 2025). Specifically, a series of Bayesian repeated-measures ANOVAs were conducted to assess the effects of the within-participant factor block (six levels), the between-participant factor group, and their interaction on the dependent variables.

To quantify the strength of evidence in favor of the null hypothesis relative to alternative models, we used the Bayes factor BF_{01} as the primary index. Model comparisons were conducted hierarchically. Additionally, the exclusion Bayes factor (BF_{excl}) was calculated to evaluate evidence against individual effects by comparing models that include a given effect with those that exclude it. This approach allows inferences to be drawn about the contribution of each factor and interaction to the overall model. This Bayesian framework offers a more informative evaluation of the data by directly quantifying the relative evidence for the null model, thereby enabling clearer conclusions about the absence of effects (Keysers et al., 2020).

Bayes factors were interpreted according to conventional thresholds: values of $BF_{01} > 3$ were taken as moderate evidence for the null hypothesis, while values greater than 10 indicated strong evidence (Jeffreys, 1961). All analyses were performed using JASP’s default priors and settings.

MW-D/MW-S

The score of each subscale was calculated as the sum of responses across items as a function of group and the experiment. Subsequently, bivariate Spearman correlations were conducted between the mean of the scores obtained in each of the subscales and the mean score in TP trials of the ANTI-Vea-TP by group. To increase the sample size when analyzing the correlations, data from the lab and online experiments were collapsed.

Results

ANTI-Vea-TP

MW

In the preregistered analysis of the TP trials (i.e., average of the response on the scale), a significant increase in MW levels, i.e., a decrease in the scale going from −1 (off-task) to +1 (on-task), across blocks was observed [F(2.47, 640.63) = 141.14, p < .001, η^2_p = .35, 95% CI (.31,.39)], with a clear linear trend [t(259) = −15.37, p < .001, η^2_p = .48, 95% CI (.39,.55)] (see Fig. 2). Importantly, as shown in Fig. 2, the main effect of group was not significant [F(2, 259) = 1.84, p = .161, η^2_p = .01, (.00,.05)], and there was no significant interaction between group and blocks [F(4.95, 640.63) = 0.63, p = .673, η^2_p = .00, (.00,.01)].

The Bayesian ANOVA (BF_{01}) further supported the absence of an effect of group, with strong evidence in favor of the null hypothesis for both the main effect of group and the interaction between group and block (see Table 1), suggesting that any potential differences were practically negligible. Additionally, the Bayes factor$_{(excl)}$ also showed far more evidence for excluding the interaction between block and group.

As can be observed in Fig. 2, the mean scale value reported decreased across blocks. However, note that the distribution also changed across blocks. As can be observed in the violin plots, the shape of the distribution of responses changed across blocks, showing less concentrated responses around the mean in the last three blocks (skewness coefficient, −0.13; kurtosis coefficient, −1.13) than in the first three blocks (skewness coefficient, −1.14; kurtosis coefficient, 0.83). This descriptive outcome motivated us to run a series of exploratory analyses regarding the proportion (right axis of Fig. 3) of on-task reports (vs. off-task) and the mean score (left axis) reported in each category, which can be observed in Fig. 3. These analyses were performed across the two between-participant factors: TP frequency group and experiment.

For the percentage of on-task responses, a significant decrease across blocks [F(2.78, 719.97) = 80.37, p < .001, η^2_p = .24, (.20,.27)] with a significant linear trend [t(259) = −12.43, p < .001, η^2_p = .37, (.29,.45)] was observed. Note in Fig. 3 (right axis) that participants started reporting being on-task in around 75% of the trials in block 1 and ended reporting being on-task on just above 50% of the times at the end of the task. Interestingly, both when considering only the trials with an on-task report (i.e., reported position in the scale > 0) and when considering only off-task reports (i.e., reported position < 0), the mean reported score decreased across blocks of trials [F(3.15, 538.33) = 52.46, p < .001, η^2_p = .23, (.19,.28)] and [F(3.12, 105.97) = 19.13, p < .001, η^2_p = .36, (.24,.45)], respectively, with a significant linear decrease in both cases [t(171) = −10.47, p < .001, η^2_p = .39, (.28,.49)] and [t(34) = −6.99, p < .001, η^2_p = .59, (.36,.73)], respectively.

EV

As usually observed with the ANTI-Vea (Luna et al., 2018; Luna et al., 2021b), the EV decrement (Fig. 4) was observed as a significant decrease in hits across blocks [F(4.45, 1152.88) = 13.13, p < .001, η^2_p = .05, (.03,.07)], with a significant linear component [t(259) = −5.76, p < .001, η^2_p = .11, (.05,.19)] (see Fig. 4). A significant decrease in FA across blocks [F(4.54, 1175.53) = 4.34, p = .001, η^2_p = .02,

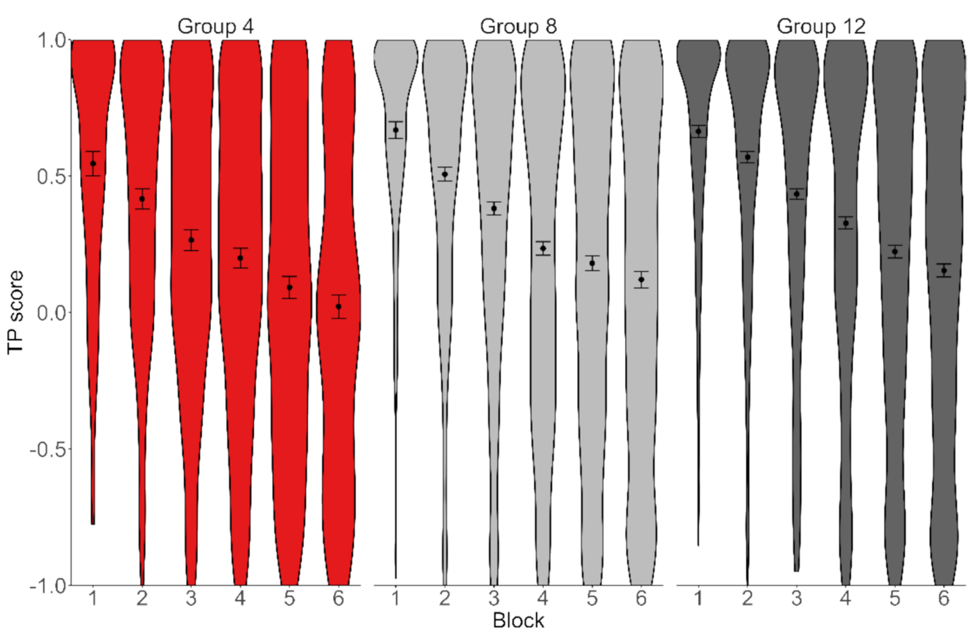

Fig. 2 Distribution of TP scores across blocks for groups as a function of 4, 8, and 12 TPs by block. The dot within each violin plot represents the mean, and the bars indicate the 95% confidence intervals of the mean

Table 1 Bayes factor for the models including block, group, and block*group terms

Dependent variable		Model	BF_{01}	Effects	BF_{excl}
EV	Hits	Block + Group + Block * Group	15,432.578	Block	1.835×10^{-10}
				Group	10
				Block * Group	4,436
	FA	Block + Group + Block * Group	2,250.189	Block	1.265
				Group	49.757
				Block * Group	1,068.609
	A′	Block + Group + Block * Group	115,113.497	Block	13.274
				Group	15.565
				Block * Group	35,127.821
	B″	Block + Group + Block * Group	22,570.713	Block	0.313
				Group	41.756
				Block * Group	7,063.955
AV	RT of AV	Block + Group + Block * Group	12,279.659	Block	5.329×10^{-15}
				Group	3.041
				Block * Group	4,583.984
	SD RT of AV	Block + Group + Block * Group	18,764.918	Block	0.000
				Group	18.680
				Block * Group	5,067.692
	Lapses	Block + Group + Block * Group	7,396.099	Block	1.615×10^{-14}
				Group	1.406
				Block * Group	3,582.805
Cognitive control	IE Cognitive control	Block + Group + Block * Group	234.593	Block	9.027×10^{-5}
				Group	29.953
				Block * Group	61.339
	Interference effect RT	Block + Group + Block * Group	16,836.720	Block	0.462
				Group	49.078
				Block * Group	5,674.524
	Interference effect errors	Block + Group + Block * Group	4,976.362	Block	0.890
				Group	47.679
				Block * Group	2,045.244
MW	TPs (mean)	Block + Group + Block * Group	4,992.380	Block	0.000
				Group	2.440
				Block * Group	2,015.257

(.00,.03)], linear component [$t(259) = -3.62$, $p < .001$, $\eta^2_p = .05$, (.01,.11)] was also observed. In addition, there was a significant decrement across blocks of *A′* [$F(4.66, 1206.42) = 2.41$, $p = .039$, $\eta^2_p = .00$, (.00,.02)], linear trend [$t(259) = -2.56$, $p = .010$, $\eta^2_p = .02$, (.00,.07)], and a significant increment across blocks of *B″* [$F(4.92, 1275.04) = 4.47$, $p < .001$, $\eta^2_p = .02$, (.00,.03)], linear trend [$t(259) = 3.74$, $p < .001$, $\eta^2_p = .05$, (.01,.11)].

The main effect of group was not significant for hits [$F(2, 259) = 0.95$, $p = .389$, $\eta^2_p = .00$, (.00,.04)], FA [$F(2, 259) = 0.26$, $p = .774$, $\eta^2_p = .00$, (.00,.02)], *A′* [$F(2, 259) = 0.55$, $p = .575$, $\eta^2_p = .00$, (.00,.03)], or *B″* [$F(2, 259) = 0.21$, $p = .808$, $\eta^2_p = .00$, (.00,.02)]. Most importantly, group did not modulate the effect of blocks for hits [$F(8.90, 1152.88) = 0.66$, $p = .747$, $\eta^2_p = .00$, (.00,.01)], FA [$F(9.08, 1175.53) = 1.70$, $p = .084$, $\eta^2_p = .01$, (.00,.02)], *A′* [$F(9.32, 1206.42) = 0.83$, $p = .591$, $\eta^2_p = .00$, (.00,.01)], or *B″* [$F(9.85, 1275.04) = 1.35$, $p = .200$, $\eta^2_p = .00$, (.00,.02)]. It is worth noting that the decline in EV was consistently observed across all conditions, regardless of the number of TPs presented. As detailed in the supplementary material, no statistically significant differences were found between groups with different TP frequencies and a large sample with no TPs from data of a previous study (Luna et al., 2021b), suggesting that the presence of TPs did not significantly affect task performance. Also, the Bayesian ANOVA provided strong evidence in favor of the null hypothesis for both effects, the main group effect and the interaction, as shown in Table 1, indicating that any observed difference is insignificant. Moreover, the $BF_{(excl)}$ indicated that there was significantly

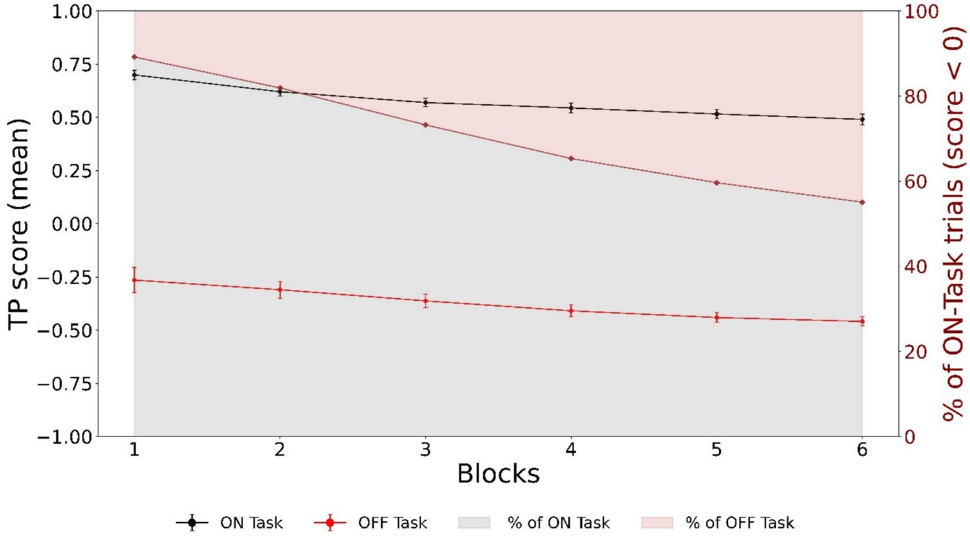

Fig. 3 TP reports over time on-task. The percentage of on-task trials is represented in the right axis, indicating the percentage of times participants remained focused on the task, i.e., with scores > 0. Thus, the red area represents the percentage of MW across blocks, whereas the gray area represents the percentage of on-task reports across blocks. The mean value reported within each category is represented by the red and black lines, respectively. Thus, the values of off-task (in red) and on-task (in black) represent the mean (left axis) raw score of the MW report, ranging from −1 to 0 and from 0 to 1, respectively. Error bars represent 95% CI of the mean

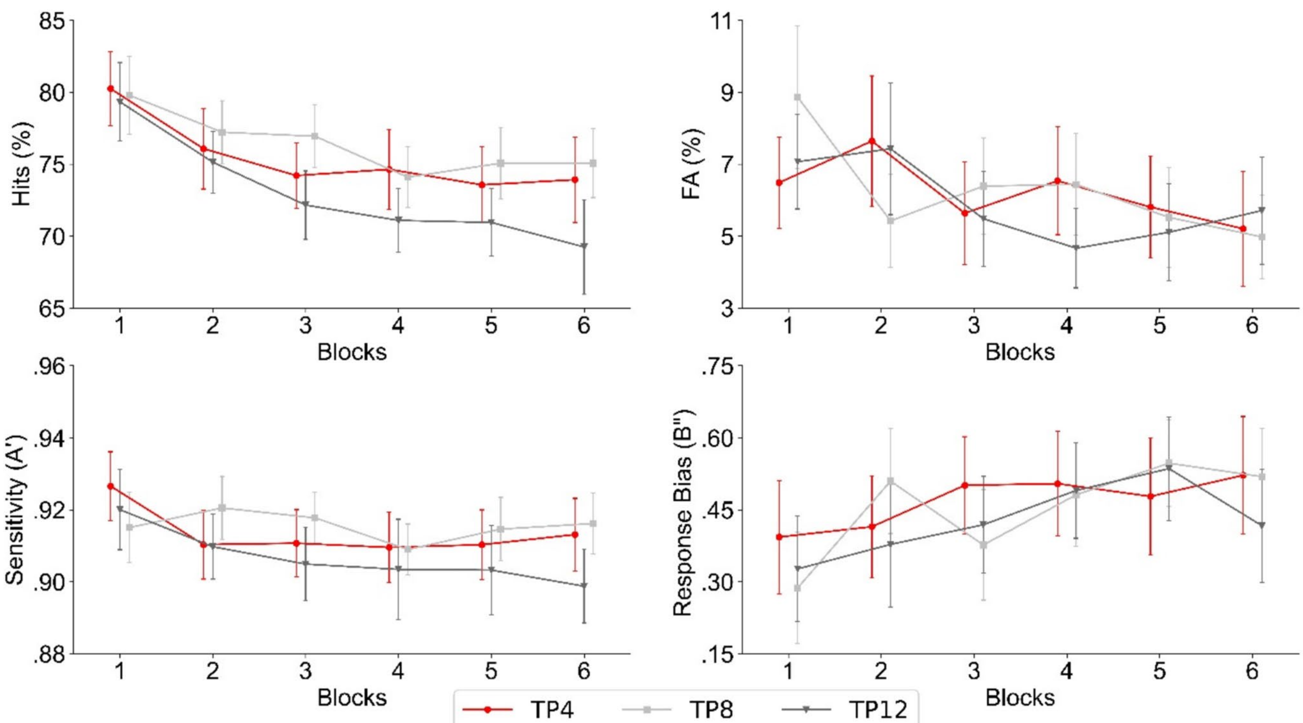

Fig. 4 Executive vigilance performance as a function of time-on-task. Error bars represent 95% CI of the mean

more evidence in favor of excluding the interaction between block and group.

AV

As shown in Fig. 5, the AV decrement across blocks was observed as a significant increase in mean RT [F(3.61, 934.24) = 20.30, p <.001, η^2_p =.07, (.05,.10)], with a significant linear component [t(259) = 6.88, p <.001, η^2_p =.15, (.08,.24)], SD of RT [F(3.88, 1003.95) = 26.95, p <.001, η^2_p =.09, (.06,.12)], with a significant linear component [t(259) = 9.39, p <.001, η^2_p =.25, (.17,.34)], and the percentage of lapses [F(3.81, 987.54) = 18.93, p <.001, η^2_p =.07, (.04,.09)], also with a significant linear component [t(259) = 6.64, p <.001, η^2_p =.15, (.08,.23)].

The main effect of group was significant only in the percentage of lapses [F(2, 259) = 3.52, p =.031, η^2_p =.03, (.00,.07)], without showing a significant main effect in mean RT [F(2, 259) = 2.56, p =.079, η^2_p =.02, (.00,.06)] or SD of RT [F(2, 259) = 1.44, p =.239, η^2_p =.01, (.00,.04)]. Importantly, no significant interactions between the group and blocks were observed for mean RT [F(7.21, 934.24) = 0.39, p =.913, η^2_p =.00, (.00,.00)], SD of RT [F(7.75, 1003.95) = 0.74, p =.655, η^2_p =.00, (.00,.01)], or the percentage of lapses [F(7.63, 987.54) = 0.67, p =.708, η^2_p =.00, (.00,.01)]. The presence of TPs did not result in a significant difference in the AV performance. As indicated by additional analyses in the supplementary material, the observed effects were not statistically different when TPs were used compared to the no-TP condition. Importantly, strong evidence in support of the null hypothesis was observed through the Bayesian ANOVA for both the main group effect and the interaction, as detailed in Table 1, suggesting that the differences are not significant. Also, the exclusion Bayes factor revealed substantially more evidence for excluding the interaction between block and group.

Cognitive control

Cognitive control decreased over time-on-task, as demonstrated by a significant increase across blocks in the interference effect for mean RT [F(4.75, 1225.81) = 4.51, p <.001, η^2_p =.02, (.00,.03)], with a significant linear component [t(258) = 3.56, p <.001, η^2_p =.05, (.01,.11)], percentage of errors [F(4.94, 1278.17) = 4.34, p <.001, η^2_p =.02, (.00,.03)], with a significant linear component [t(259) = 3.96, p <.001, η^2_p =.06, (.01,.12)], and the IE score [F(4.75, 1200.99) = 9.78, p <.001, η^2_p =.04, (.02,.06)], also with a marginal linear component [t(253) = 7.081, p <.001, η^2_p =.01, (.00,.25)] (Fig. 6).

The main effect of group was not significant for the interference effect in mean RT [F(2, 258) = 0.03, p =.967, η^2_p =.00, (.00,.00)], percentage of errors [F(2, 259) = 0.78, p =.461, η^2_p =.00, (.00,.03)], or the IE score [F(2, 253) = 1.66, p =.193, η^2_p =.01, (.00,.05)]. Moreover, no significant interactions were found between the group and blocks for mean RT [F(9.50, 1225.81) = 0.76, p =.663, η^2_p =.00, (.00,.01)] and the percentage of errors [F(9.87, 1278.17) = 1.70, p =.078, η^2_p =.01, (.00,.02)].

However, a significant small interaction was observed between group and blocks for the IE [F(9.49, 1200.99) = 2.16, p =.020, η^2_p =.02, (.00,.03)]. Nevertheless, despite the significant interaction between groups and blocks, pairwise comparisons of the linear component between groups showed no significant differences among them, as follows. The increase in IE was not significantly

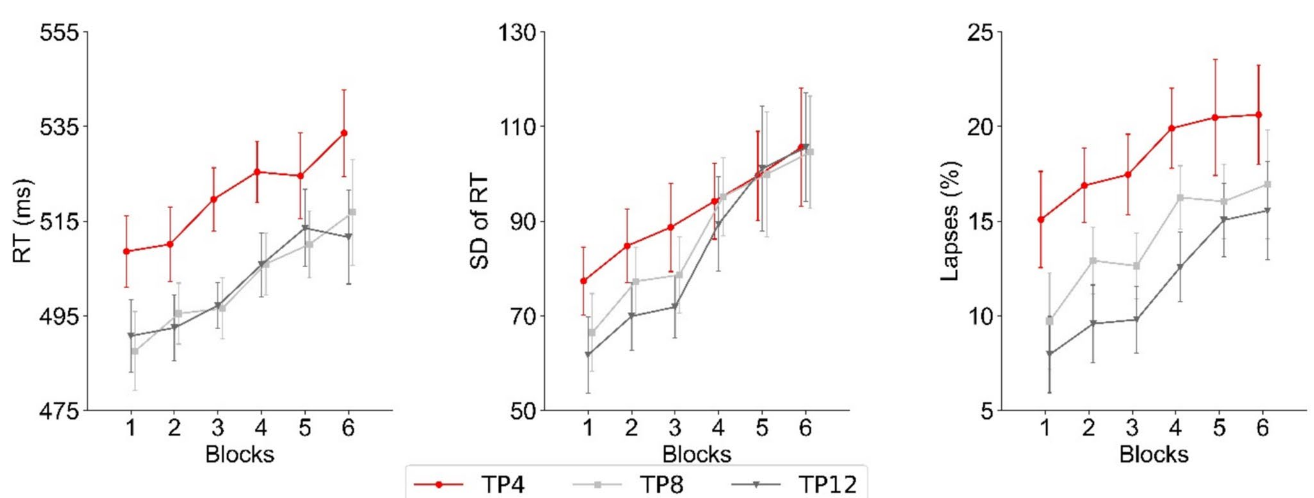

Fig. 5 Arousal vigilance performance as a function of time-on-task. Error bars represent 95% CI of the mean

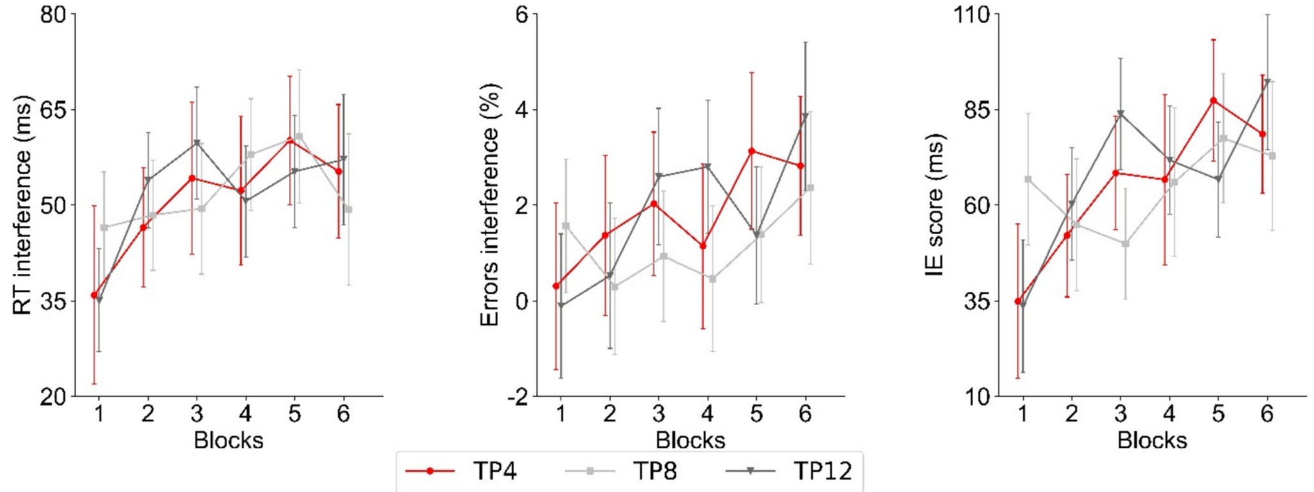

Fig. 6 Cognitive control performance as a function of time-on-task. IE = inverse efficiency. Error bars represent 95% CI of the mean

different between the group with four TPs and the group with eight TPs [$t(253) = 1.36$, $p = .363$, $d = 0.09$, (−0.04, 0.21)]. Similarly, comparisons between the group with four TPs and the group with 12 TPs [$t(253) = 0.00$, $p = .999$, $d = 0.00$, (−0.12, 0.12)] and between the group with eight TPs and the group with 12 TPs [$t(253) = -1.36$, $p = .410$, $d = -0.09$, (−0.21, 0.04)] did not reveal significant differences in the linear component of IE across blocks. The results indicate that performance in cognitive control declines, with no statistically significant differences between the conditions with varying TP frequencies and the no-TP condition (see supplementary material). For all cognitive control analyses, Bayesian ANOVA showed strong evidence in favor of the null hypothesis for both the main effect of group and the interaction, as shown in Table 1, implying that any potential differences are too small to have practical significance. Furthermore, the exclusion Bayes factor provided much stronger evidence in favor of excluding the interaction between block and group.

Phasic alertness, orienting, and cognitive control

All the typical main effects of the classic attentional functions measured in the ANTI-Vea (Luna et al., 2021a; Luna et al., 2018, 2022) were observed as significant in the ANTI-Vea-TP (Fig. 7). Regarding warning signal, the significant main effect for mean RT [$F(1, 259) = 283.19$, $p < .001$, $\eta^2_p = .52$, (.44,.59)] and errors [$F(1, 259) = 125.13$, $p < .001$, $\eta^2_p = .33$, (.24,.41)] showed that responses were faster and more accurate in the tone than in the no tone condition (Fig. 7, left graphs).

For the visual cueing effect, the main effect was significant for mean RT [$F(1.91, 495.39) = 303.13$, $p < .001$, $\eta^2_p = .54$, (.49,.59)] and errors [$F(1.95, 505.38) = 7.08$, $p = .001$, $\eta^2_p = .03$, (.00,.06)]. As depicted in Fig. 7 (see center graphs), the typical validity {invalid > valid: only for RT [$t(259) = 23.51$, $p < .001$, $d = 1.46$, 95% CI (1.28, 1.63)]; not for errors: [$t(259) = 1.67$, $p = .216$, $d = 0.10$, (−0.02, 0.23)]}, benefits {no cue > valid: RT [$t(259) = 14.08$, $p < .001$, $d = 0.87$, (0.73, 1.02)]; but not for errors [$t(259) = -2.18$, $p = .076$, $d = -0.14$, (−0.26., −0.01)]}, and costs {invalid > no cue: RT [$t(259) = 11.47$, $p < .001$, $d = 0.71$, (0.57, 0.85)]; and errors: [$t(259) = 3.66$, $p < .001$, $d = 0.23$, (0.10, 0.35)]} effects were observed.

Lastly, the congruency effect showed that responses were significantly faster and more accurate in the congruent than in the incongruent condition [RT: $F(1, 259) = 434.01$, $p < .001$, $\eta^2_p = .63$, (.56,.68); errors: $F(1, 259) = 26.74$, $p < .001$, $\eta^2_p = .09$, (.04,.17)].

Furthermore, the typical interactions between the classic attentional functions were also observed as significant, as previously reported with the ANTI (Callejas et al., 2004) and ANTI-Vea (Luna et al., 2018, 2021a, 2021b) tasks. The interaction between warning signal and congruency was significant for RT [$F(1, 259) = 28.39$, $p < .001$, $\eta^2_p = .10$, (.04,.17)] and errors [$F(1, 259) = 4.43$, $p = .036$, $\eta^2_p = .02$, (.00,.06)]. The interaction between visual cue and congruency was significant for RT [$F(1.98, 513.37) = 19.41$, $p < .001$, $\eta^2_p = .07$, (.03,.11)] and errors [$F(2, 517.57) = 5.12$, $p = .006$, $\eta^2_p = .02$, (.04,.05)]. The interaction between warning signal and visual cue was only significant for RT [$F(2, 516.72) = 84.53$, $p < .001$, $\eta^2_p = .25$, (.18,.30)], but not for errors [$F(2, 516.82) = 2.01$, $p = .135$, $\eta^2_p = .00$, (.00,.03)]. Lastly, a significant three-way interaction was observed in the analysis of RT [F(2,

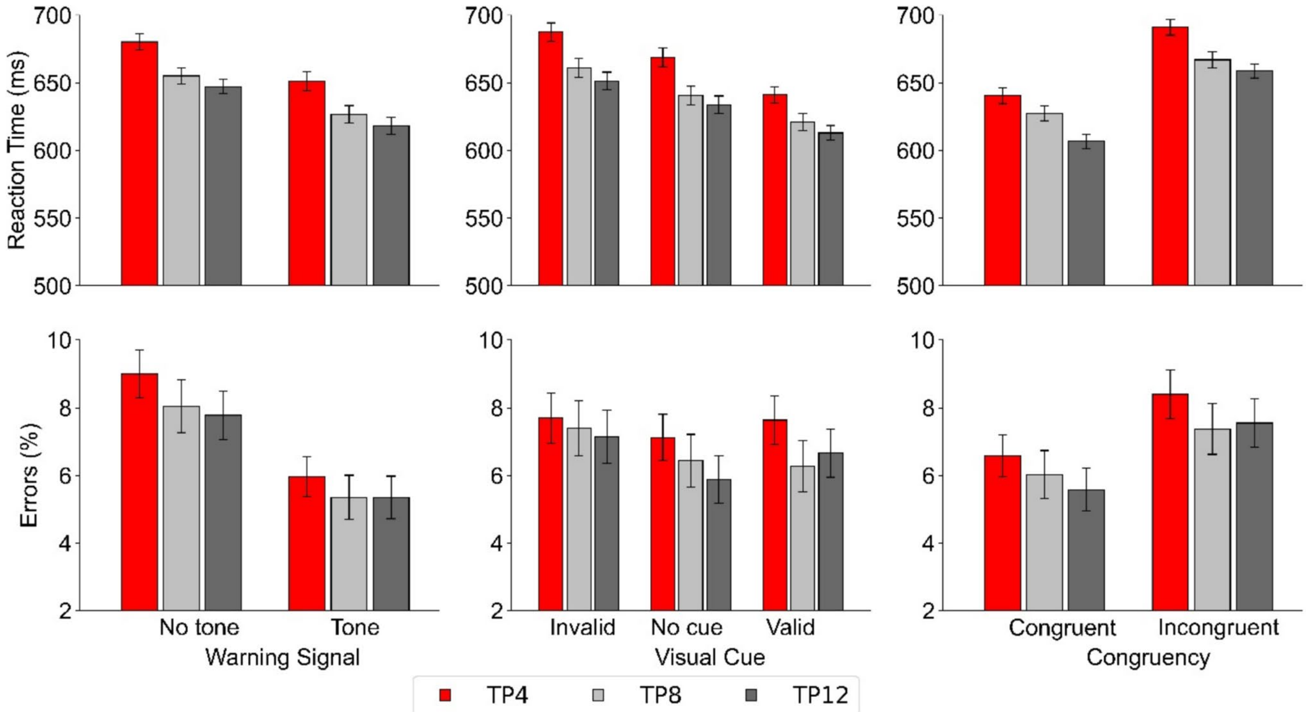

Fig. 7 Mean of RT (top graphs) and percentage of errors (bottom graphs) for the warning signal (left), visual cue (center), and congruency (right) conditions, as a function of the group (4, 8, or 12 TPs by block). Error bars represent 95% CI of the mean

517.17) = 5.73, p = .003, η^2_p = .02, (.00,.05)], but not for errors [F(1.98, 512.59) = 0.20, p = .817, η^2_p = .00, (.00,.01)].

A significant main effect of group (see Fig. 7) was observed for RT [F(2, 259) = 3.67, p = .027, η^2_p = .03, (.00,.07)], indicating that the group with four TPs showed significantly larger RT than the group with 12 TPs [t(259) = 2.60, p = .026, d = 0.16, (−0.04, −0.28)]. However, there were no significant differences between the groups with four TPs and eight TPs [t(259) = 1.96, p = 0.124, d = 0.12, (0.00, −0.24)] or between the groups with eight TPs and 12 TPs [t(259) = −0.65, p = 0.791, d = 0.04, (−0.08, 0.16)]. No significant main effect of group was observed for errors [F(2, 259) = 0.43, p = .648, η^2_p = .00, (.00,.02)]. Importantly, no significant interactions were found between the group and the warning signal [F(2, 259) = 0.03, p = .968, η^2_p = .00, (.00,.00)], visual cue [F(3.83, 495.39) = 2.30, p = .061, η^2_p = .02, (.00,.04)], or congruency [F(2, 259) = 0.04, p = .965, η^2_p = .00, (.00,.00)] factors for RT. Similarly, for errors, there were no significant interactions between group and warning signal [F(2, 259) = 0.29, p = .749, η^2_p = .00, (.00,.02)], visual cue [F(3.90, 505.38) = 2.50, p = .051, η^2_p = .02, (.00,.04)], or congruency [F(2, 259) = 0.46, p = .634, η^2_p = .00, (.00,.03)], showing therefore that TP frequency did not modulate the main effects assessed in ANTI trials.

MW-D/MW-S

Table 2 shows the mean and *SD* of MD scores as a function of group and experiment. Spearman correlations between overall scores in the MW-S subscale and mean TP score in ANTI-Vea-TP showed a positive and significant correlation (rho = .19, p = .001). However, no significant correlation was observed between the MW-D score and mean TP score in the ANTI-Vea (rho = .11, p = .063). Table 3 presents Spearman correlations as a function of group and MW-S/MW-D scales

Table 2 Descriptive statistics of MW-deliberate and MW-spontaneous subscales as a function of group and experiment

Group	Experiment	MW-D		MW-S	
		M	*SD*	*M*	*SD*
4 TPs	Lab	14.60	6.39	32.47	9.66
	Online	17.61	6.07	34.67	9.97
8 TPs	Lab	15.80	5.65	33.20	9.267
	Online	18.90	5.70	36.55	8.13
12 TPs	Lab	16.33	6.02	34.10	10.53
	Online	17.91	5.56	34.16	8.82

M: mean; *SD*: standard deviation; TP: thought probe; MW-D: mind-wandering deliberate; MW-S: mind-wandering spontaneous

Table 3 Spearman's correlations between mean TP score in the ANTI-Vea-TP task and overall score in the MW-D and MW-S subscales, as a function of group and collapsed data

Group	TP mean and MW-D		TP mean and MW-S	
	Spearman's rho	*p*	Spearman's rho	*p*
4 TP	−0.05	.618	0.02	.833
8 TP	0.26	**.011**	0.29	**<.01**
12 TP	0.15	.142	0.27	**<.01**
Overall	.11	.063	0.19	**<.01**

TP: thought probe; MW-D: mind-wandering deliberate; MW-S: mind-wandering spontaneous. Significant outcomes are highlighted in bold

Discussion

The present study aimed to investigate whether it was possible to assess MW changes in different TP frequencies within the ANTI-Vea task, while still measuring changes in cognitive control and vigilance components. The ANTI-Vea has proven to be an effective task for simultaneously measuring decrements in EV, AV, and cognitive control in a single session, thus providing simultaneous assessment of several attention and vigilance components (Luna et al., 2021b). The present version of the task with TPs, while further contributing to the understanding of methodological considerations in measuring MW, critically provides a unique tool to investigate the interaction between vigilance, cognitive control, and MW, in both the lab and online environments, which is critical for testing different theories of vigilance decrement (Esterman & Rothlein, 2019).

Importantly, all the typical effects usually measured via the ANTI-Vea were significantly observed despite the interruptions introduced with TP trials. Moreover, no significant differences were found between the groups with TP and a no-TP group, as shown in the supplementary material. Furthermore, consistently with previous research, we observed the usual increase in the frequency of MW across time on-task (Zanesco et al., 2025), with TP frequency not modulating the overall report of MW or the change in MW across blocks (Robison et al., 2020). These outcomes suggest that the ANTI-Vea-TP task is a useful tool for measuring increased MW across time on-task, and that frequency of TP may not influence reported levels of MW. Several studies have found no association between the rate of TP and an increase in MW (Robison et al., 2019; Wiemers & Redick, 2019). In contrast, other studies have reported lower levels of MW when the TP rate was higher (Schubert et al., 2020; Seli et al., 2013). Our findings align with those of Robison et al. (2019) and Wiemers and Redick (2019), wherein the number of TP seems to have no effect on the reporting of MW. Varying the rate of TP per block did not produce statistically significant differences in reports of MW. This result contrasts with that of Welhaf et al. (2022), who found that correlations between the rate of TP and other constructs, such as working memory capacity, attentional control ability, and disorganized schizotypy, stabilized when using eight probes.

Furthermore, Seli et al. (2013) found that longer intervals between probes were associated with increased reported MW. In their study, using the metronome response task, they presented TP in a pseudorandomized way, ensuring an interval of at least 10 s and 400 ms between two TPs. Their results suggest that MW might decrease with very short intervals between probes. In our case, contrary to their findings, we did not observe differences in MW reports between groups based on the number of TPs, with a minimum interval of 20 s and 500 ms. Furthermore, in spite of the reported changes in MW, Seli et al. (2013) and Schubert et al. (2020) found no change in typical vigilance task performance measures as a function of frequency of probes.

Our results showed a linear increase in MW over time-on-task. Measuring MW via a continuous scale allows for the evaluation of moment-to-moment fluctuations in attentional states, which may go unnoticed in dichotomous response modes (e.g., "on-task" vs. "off-task") (Arnicane et al., 2021). Additionally, theories attempting to explain associations between MW and vigilance, as the resource-control model (Thomson et al., 2015), must consider not only the increased frequency of off-task reports provided by dichotomous TP but also the small fluctuations in MW, as individuals likely resort to a variety of experiences to classify their focus of attention on a continuum from fully engaged to completely off-task (Zanesco et al., 2020). Previous research suggests that categorical assessment, especially dichotomous assessment, could bias estimates of MW rates, inflating this measure (Arnicane et al., 2021; Seli et al., 2018), thus highlighting the importance of employing a continuous scale.

The present study might open the possibility that both considerations are possible—that there can be a dichotomous state between being focused on-task and MW off-task, and there can also be a degree within each state (more or less focused on-task and more or less off-task engaged in MW). Such interpretation is driven by two pieces of information from the present study. On the one hand, as shown in Fig. 2, in the initial blocks of trials, MW reports seem to cluster around a central value, suggesting participants were predominantly focused on the task. However, in later blocks, the distribution of MW reports appears more spread out, with a bimodal trend emerging. On the other hand, interestingly, as shown in Fig. 3, when dividing trials according to whether participant reports are on-task or off-task, in both cases the main reported value decreased across time-on-task. In other words, the data seem to suggest that participants do not always score on the ends of the scale, reporting being either on- or off-task, but more gradual changes in their on/off-task engagement. Although future research should replicate and further analyze dichotomic and gradual changes within a task, there seems to be an alternation between opposite states (task-focus vs. MW) and different degrees within each state.

Regarding vigilance measures, our study revealed a decrease in EV over time, indicated by changes in signal detection theory indices (i.e., hits, FA, A', B''). The decrease in EV reflects the challenge of maintaining attention on rare but critical events during prolonged tasks (Luna et al., 2018; Luna, Roca, et al., 2021a, 2021b). The presence and number of TPs did not significantly influence EV measures, suggesting that the frequency of MW measurement does not directly affect the decline in EV. Our results replicated similar patterns of data observed in other studies of vigilance decrement in signal detection tasks (Hancock, 2017; Hemmerich et al., 2023; Lara et al., 2014; Luna, Roca, et al., 2021a, 2021b; Martínez-Pérez et al., 2023) Similarly, regarding AV, as shown by mean RT, *SD* of RT, and lapses, we also observed a typical decrement over time (Luna, Barttfeld, et al., 2021a, 2021b; Luna et al., 2018). Although the number of TPs per block only modulated the overall percentage of lapses, it should be noted that, most importantly, TP frequency did not modulate changes in AV in any of the dependent variables. As in EV outcomes, the absence of significant interactions between TP groups and blocks for AV measures suggests that different TP frequency may not affect measuring AV changes in the ANTI-Vea-TP task.

Taking all the above into consideration, the presence of the TPs allowed the measurement of changes in MW across time on-task, without impacting the measurement of the decrease in vigilance components, as can be seen in the supplementary material. It seems that participants consider TP reports as another aspect of a complex task like the ANTI-Vea, rather than as small rests which could have eliminated the vigilance decrement across blocks. Thus, the decision of whether a larger or smaller number of TPs are used in future research with the ANTI-Vea-TP could be based on total task duration (the more TPs used, the longer the task) or whether MW needs to be tested more or less frequently, knowing that the number of TPs would only minimally affect the measures when a minimum of ~20 s is maintained between TPs. A larger number of TPs by block could be more useful, for instance, in psychophysiological research, wherein a large set of trials is usually necessary to compute some physiological indices (Luna et al., 2023a, b).

Importantly, a decline in cognitive control over time was also observed, evidenced by an increase in the interference effect in mean RT, errors, and IE score, as in Luna et al. (2022). In contrast, Zholdassova et al. (2021) found no changes in cognitive control over time using the attention network test (ANT) task, which does not measure vigilance components or MW. Similarly, Satterfield et al. (2019) did not observe a modulation of cognitive control state in the decline of vigilance. Our results replicate the findings of Luna et al. (2022) and provide additional evidence of a decline in cognitive control over time. Furthermore, again, TP frequency did not affect cognitive control measures in the ANTI-Vea-TP task.

Note that, as reported in the supplementary material, the experiment (online vs. in-lab) had a significant main effect on overall scores of MW [$F(1, 259) = 7.33$, $p = .007$, $\eta^2_p = .03$, (.00,.08)], showing that participants who completed the experiment online ($M = 0.22$, $SD = 0.54$) reported more MW than those who completed the experiment in the laboratory ($M = 0.37$, $SD = 0.51$). This effect was also observed in other variables such as FA and B'' for EV trials, mean RT and lapses for AV trials, and mean RT and errors for ANTI trials (see Tables S1–S5 in the supplementary material). Moreover, we found an interaction between blocks of IE and the experiment. However, it should be noted from Tables S1 to S5 that these effects refer to the overall data and are small. A previous study conducted by our lab (Luna et al., 2021b) showed no differences between the attention and vigilance measures taken online and in-lab, and both methods demonstrated high reliability for assessing vigilance and attentional components.

Moreover, following the study by Wiemers and Redick (2019), which demonstrated that there are no significant differences in SART performance with or without TPs, we observed the typical main effects and interactions commonly reported with the ANTI-Vea task (Luna et al., 2021a). Again, this suggests that individuals may not perceive it as an interruption to the main ANTI subtask, but rather as an additional subtask.

Finally, in this study, the concurrent validity of our new measure of MW in the ANTI-Vea-TP task was evaluated by correlations of MW scores with trait self-reported scores via the MW-D/MW-S questionnaire. The results showed a significant correlation between the mean of the TP score and MW-S for the groups of eight and 12 TPs, providing evidence of strong concurrent validity in these groups. However, for the group of four TPs, no significant correlation was found. The lack of significant correlations in the group of four TPs could be due to the smaller number of trials to compute the MW score, compared to the groups of eight and 12 TPs, and therefore a lower reliability of the MW measure.

Although our study provides valuable insights into measuring MW along with vigilance components and cognitive control, it is important to acknowledge certain limitations of the present research. Variables such as the experiment and TPs per block, were manipulated between participants. Nevertheless, considering the length of the tasks and the potential for participant familiarization, this design was chosen over a fully within-participant design. Another potential limitation of our study is the use of both a sliding scale and subsequent dichotomization for TP responses. While this approach allowed us to capture a more nuanced representation of MW reports, it also introduced inconsistency in measurement that may affect its interpretation. Kane et al. (2021) argue that Likert-type or sliding scale measures of MW may be less valid than categorical options, as they could introduce additional variance unrelated to the underlying cognitive state.

Our study contributes to the ongoing discussion regarding the methodological variability in measuring MW, especially with the use of TPs. The results indicate that, at least within the ANTI-Vea-TP task, TP frequency when considering a minimum interval of ~ 20 s between two TPs may not affect the reported levels of MW. Additionally, TP frequency does not seem to alter the nature of the ANTI-Vea task, as the results found here replicated those reported in several previous studies (Luna et al., 2018, 2022; Luna et al., 2021b).

In conclusion, our findings emphasize the importance of task-specific considerations in research on MW and provide valuable insights for future studies investigating the interaction between vigilance, cognitive control, and MW. This study represents an initial step focused on a methodological analysis of embedding TPs in the ANTI-Vea task. The ANTI-Vea-TP task allows for the measurement of the decrement in EV, AV, and cognitive control, as well as changes in MW, within a single session, in both the lab and online environments. This method enables future studies to analyze theoretical models of the vigilance decrement phenomenon (Thomson et al., 2015). Additionally, the ANTI-Vea-TP provides a versatile tool to investigate how individual and contextual factors influence fluctuations in vigilance, thus contributing to research on strategies to mitigate vigilance decrement and optimize performance in prolonged tasks.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13428-025-02808-x.

Author contributions MJA: conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, writing—original draft, review, and editing. PB: conceptualization, data curation, formal analysis, funding acquisition, investigation, project administration, methodology, resources, software, supervision, validation, visualization, and writing—review and editing. EM-A: conceptualization, formal analysis, methodology, software, supervision, validation, and writing review and editing. JL: conceptualization, formal analysis, methodology, project administration, funding acquisition, software, supervision, validation, and writing—review and editing. FL: conceptualization, data curation, formal analysis, investigation, methodology, resources, software, validation, visualization, project administration, and writing—review and editing. All authors contributed to the article and approved the submitted version.

Funding This study was supported by the Spanish Ministerio de Ciencia, Innovación y Universidades, with research grants PID2023-148421NB-I00 and PID2020-114790GB-I00 funded by MICIU/AEI/https://doi.org/10.13039/501100011033, and by ESF+, CEX2023-001312-M by MCIN/AEI/https://doi.org/10.13039/501100011033 and UCE-PP2023-11 by the University of Granada. In addition, MJA received PhD scholarship support from the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina, and a mobility scholarship from the Asociación Universitaria Iberoamericana de Posgrado (AUIP).

Data availability All data (https://doi.org/10.17605/OSF.IO/63B59) and materials (https://doi.org/10.17605/OSF.IO/A27RV) are publicly available in the Open Science Framework repository.

Code availability The code used for data processing and statistical analyses, along with the datasets generated and analyzed in this study, is available in the Open Science Framework repository (https://doi.org/10.17605/OSF.IO/JAFYS).

Declarations

Consent to participate Informed consent was obtained from all participants of the study.

Consent for publication Only participants who did not indicate that their anonymized data should not be used for analysis and publication were included.

Ethics approval Signed informed consent was obtained from all participants in both experiments, following the ethical standards established in the 1964 Declaration of Helsinki (last updated: Fortaleza, 2013). All participants had normal or corrected-to-normal vision. Experiment 1 was approved by the Ethical Committee of the Institute of Psychological Research (CEIIPsi, protocol PE41, version 2), and Experiment 2 was approved by the University of Granada's Ethical Committee (2442/CEIH/2021).

Conflicts of interest The authors declare no conflict of interest. The funders had no role in the design of the tasks and the tool, in the writing of the manuscript, or in the decision to publish the paper. Fernando G. Luna is employed by Beway Consulting S.L., which was not involved in the study design, the data collection or interpretation, the writing of this article, or the decision to submit it for publication.

References

Arger, I., & Nogareda, C. (1999). *Estimación de la carga mental de trabajo: El método NASA TLX. NTP 544.* INSHT. https://www.insst.es/documents/94886/327064/ntp_544.pdf/0da348cc-7006-4a8a-9cee-25ed6f59efdd

Arnicane, A., Oberauer, K., & Souza, A. S. (2021). Validity of attention self-reports in younger and older adults. *Cognition, 206*, Article 104482. https://doi.org/10.1016/j.cognition.2020.104482

Barkley, R. A. (2011). *Barkley Adult ADHD Rating Scale-IV (BAARS-IV)*. Guilford Press.

Basner, M., & Dinges, D. F. (2011). Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. *Sleep, 34*(5), 581–591. https://doi.org/10.1093/sleep/34.5.581

Ben-Shachar, M., Lüdecke, D., & Makowski, D. (2020). Effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software, 5*(56), Article 2815. https://doi.org/10.21105/joss.02815

Bruyer, R., & Brysbaert, M. (2011). Combining speed and accuracy in cognitive psychology: Is the inverse efficiency score (IES) a better dependent variable than the mean reaction time (RT) and the percentage of errors (PE)? *Psychologica Belgica, 51*(1), Article 5. https://doi.org/10.5334/pb-51-1-5

Caggiano, D. M., & Parasuraman, R. (2004). The role of memory representation in the vigilance decrement. *Psychonomic Bulletin & Review, 11*(5), 932–937. https://doi.org/10.3758/BF03196724

Callejas, A., Lupiáñez, J., & Tudela, P. (2004). The three attentional networks: On their independence and interactions. *Brain and Cognition, 54*(3), 225–227. https://doi.org/10.1016/j.bandc.2004.02.012

Carriere, J. S. A., Seli, P., & Smilek, D. (2013). Wandering in both mind and body: Individual differences in mind wandering and inattention predict fidgeting. *Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale, 67*(1), 19–31. https://doi.org/10.1037/a0031438

Cásedas, L., Cebolla, A., & Lupiáñez, J. (2022). Individual differences in dispositional mindfulness predict attentional networks and vigilance performance. *Mindfulness, 13*(4), 967–981. https://doi.org/10.1007/s12671-022-01850-6

Coll-Martín, T., Román-Caballero, R., Martínez-Caballero, M. D. R., Martín-Sánchez, P. D. C., Trujillo, L., Cásedas, L., Castellanos, M. C., Hemmerich, K., Manini, G., Aguirre, M. J., Botta, F., Marotta, A., Martín-Arévalo, E., Luna, F. G., & Lupiáñez, J. (2023). The ANTI-Vea-UGR platform: A free online resource to measure attentional networks (alertness, orienting, and executive control) functioning and executive/arousal vigilance. *Journal of Intelligence, 11*(9), Article Article 181. https://doi.org/10.3390/jintelligence11090181

Derryberry, D., & Reed, M. A. (2002). Anxiety-related attentional biases and their regulation by attentional control. *Journal of Abnormal Psychology, 111*(2), 225–236. https://doi.org/10.1037/0021-843X.111.2.225

Dinges, D. F., & Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments, & Computers, 17*(6), 652–655. https://doi.org/10.3758/BF03200977

Edkins, G. D., & Pollock, C. M. (1997). The influence of sustained attention on railway accidents. *Accident Analysis and Prevention, 29*(4), 533–539. https://doi.org/10.1016/S0001-4575(97)00033-X

Esterman, M., & Rothlein, D. (2019). Models of sustained attention. *Current Opinion in Psychology, 29*, 174–180. https://doi.org/10.1016/j.copsyc.2019.03.005

Feltmate, B. B. T., Hurst, A. J., & Klein, R. M. (2020). Effects of fatigue on attention and vigilance as measured with a modified attention network test. *Experimental Brain Research, 238*(11), 2507–2519. https://doi.org/10.1007/s00221-020-05902-y

Fortaleza. (2013). *World Medical Association. Declaration of Helsinki – Ethical principles for medical research involving human subjects.* WMA.

Guilera, G., Barrios, M., Penelo, E., Morin, C., Steel, P., & Gómez-Benito, J. (2018). Validation of the Spanish version of the Irrational Procrastination Scale (IPS). *PLoS One, 13*(1), Article e0190806. https://doi.org/10.1371/journal.pone.0190806

Hancock, P. A. (2017). On the nature of vigilance. *Human Factors, 59*(1), 35–43. https://doi.org/10.1177/0018720816655240

Hemmerich, K., Lupiáñez, J., Luna, F. G., & Martín-Arévalo, E. (2023). The mitigation of the executive vigilance decrement via HD-tDCS over the right posterior parietal cortex and its association with neural oscillations. *Cerebral Cortex, 33*(11), 6761–6771. https://doi.org/10.1093/cercor/bhac540

Huertas, F., Ballester, R., Gines, H. J., Hamidi, A. K., Moratal, C., & Lupiáñez, J. (2019). Relative age effect in the sport environment. Role of physical fitness and cognitive function in youth soccer players. *International Journal of Environmental Research and Public Health, 16*(16), Article 2837. https://doi.org/10.3390/ijerph16162837

Hunter, J. D. (2007). Matplotlib: A 2D graphics enviroment. *Computing in Science & Engineering, 9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

JASP Team (2025). JASP (version 0.19.3.0). Retrieved from https://jaspstats.org/

Jeffreys, H.(1961). Theory of Probability, 3rd edn. Oxford University Press, Oxford

Kane, M. J., Smeekens, B. A., Meier, M. E., Welhaf, M. S., & Phillips, N. E. (2021). Testing the construct validity of competing measurement approaches to probed mind-wandering reports. *Behavior Research Methods, 53*(6), 2372–2411. https://doi.org/10.3758/s13428-021-01557-x

Keysers, C., Gazzola, V., & Wagenmakers, E.-J. (2020). Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nature Neuroscience, 23*(7), 788–799. https://doi.org/10.1038/s41593-020-0660-4

Lakens, D., & Caldwell, A. R. (2021). Simulation-based power analysis for factorial analysis of variance designs. *Advances In Methods And Practices In Psychological Science, 4*(1), 251524592095150. https://doi.org/10.1177/2515245920951503

Langner, R., & Eickhoff, S. B. (2013). Sustaining attention to simple tasks: A meta-analytic review of the neural mechanisms of vigilant attention. *Psychological Bulletin, 139*(4), 870–900. https://doi.org/10.1037/a0030694

Lara, T., Madrid, J. A., & Correa, Á. (2014). The vigilance decrement in executive function is attenuated when individual chronotypes perform at their optimal time of day. *PLoS One, 9*(2), Article e88820. https://doi.org/10.1371/journal.pone.0088820

Lenth, R. (2021). *emmeans: Estimated marginal means, aka least-squares means*. Available at: https://cran.r-project.org/package=emmeans

Luna, F. G., Marino, J., Roca, J., & Lupiáñez, J. (2018). Executive and arousal vigilance decrement in the context of the attentional networks: The ANTI-Vea task. *Journal of Neuroscience Methods, 306*(May), 77–87. https://doi.org/10.1016/j.jneumeth.2018.05.011

Luna, F. G., Román-Caballero, R., Barttfeld, P., Lupiáñez, J., & Martín-Arévalo, E. (2020). A high-definition tDCS and EEG study on attention and vigilance: Brain stimulation mitigates the executive but not the arousal vigilance decrement. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2020.107447

Luna, F. G., Barttfeld, P., Martín-Arévalo, E., & Lupiáñez, J. (2021a). The ANTI-Vea task: Analyzing the executive and arousal vigilance decrements while measuring the three attentional networks. *Psicológica Journal, 42*(1), 1–26. https://doi.org/10.2478/psicolj-2021-0001

Luna, F. G., Roca, J., Martín-Arévalo, E., & Lupiáñez, J. (2021b). Measuring attention and vigilance in the laboratory vs. online: The split-half reliability of the ANTI-Vea. *Behavior Research Methods, 53*(3), 1124–1147. https://doi.org/10.3758/s13428-020-01483-4

Luna, F. G., Tortajada, M., Martín-Arévalo, E., Botta, F., & Lupiáñez, J. (2022). A vigilance decrement comes along with an executive control decrement: Testing the resource-control theory. *Psychonomic Bulletin & Review, 29*(5), 1831–1843. https://doi.org/10.3758/s13423-022-02089-x

Luna, F. G., Aguirre, M. J., Martín-Arévalo, E., Ibáñez, A., Lupiáñez, J., & Barttfeld, P. (2023a). Different oscillatory rhythms anticipate failures in executive and arousal vigilance. *Frontiers in Cognition, 2*, Article 1128442. https://doi.org/10.3389/fcogn.2023.1128442

Luna, F. G., Aguirre, M. J., Martín-Arévalo, E., Ibáñez, A., Lupiáñez, J., & Barttfeld, P. (2023b). Event-related potentials associated with attentional networks evidence changes in executive and arousal vigilance. *Psychophysiology, 60*(8), Article e14272. https://doi.org/10.1111/psyp.14272

Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search. *The Quarterly Journal of Experimental Psychology, 1*(1), 6–21. https://doi.org/10.1080/17470214808416738

Martínez-Pérez, V., Andreu, A., Sandoval-Lentisco, A., Tortajada, M., Palmero, L. B., Castillo, A., Campoy, G., & Fuentes, L. J. (2023). Vigilance decrement and mind-wandering in sustained attention tasks: Two sides of the same coin? *Frontiers in Neuroscience, 17*, Article 1122406. https://doi.org/10.3389/fnins.2023.1122406

Murray, S., Krasich, K., Schooler, J. W., & Seli, P. (2020). What's in a task? Complications in the study of the task-unrelated-thought variety of mind wandering. *Perspectives on Psychological Science, 15*(3), 572–588. https://doi.org/10.1177/1745691619897966

Navarro, O., Restrepo-Ochoa, D., Rommel, D., Ghalaret, J.-M., & Fleury-Bahi, G. (2021). *Validation of a brief version of the Difficulties in Emotion Regulation Scale with a Spanish speaking population (DERS-S SF)*.

Neigel, A. R., Claypoole, V. L., Smith, S. L., Waldfogle, G. E., Fraulini, N. W., Hancock, G. M., Helton, W. S., & Szalma, J. L. (2020). Engaging the human operator: A review of the theoretical support for the vigilance decrement and a discussion of practical applications. *Theoretical Issues In Ergonomics Science, 21*(2), 239–258. https://doi.org/10.1080/1463922X.2019.1682712

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Posit team. (2024). *RStudio: Integrated Development Environment for R [Computer software]*. Posit Software, PBC. http://www.posit.co/

R Core Team. (2024). *R: A language and environment for statistical computing (R version 4.2.0)*. R Foundation for Statistical Computing https://www.r-project.org

Read, G. J. M., Lenné, M. G., & Moss, S. A. (2012). Associations between task, training and social environmental factors and error types involved in rail incidents and accidents. *Accident Analysis and Prevention, 48*, 416–422. https://doi.org/10.1016/j.aap.2012.02.014

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. https://doi.org/10.1016/S0028-3932(97)00015-8

Robison, M. K., Miller, A. L., & Unsworth, N. (2019). Examining the effects of probe frequency, response options, and framing within the thought-probe method. *Behavior Research Methods, 51*(1), 398–408. https://doi.org/10.3758/s13428-019-01212-6

Robison, M. K., Miller, A. L., & Unsworth, N. (2020). A multi-faceted approach to understanding individual differences in mind-wandering. *Cognition, 198*, Article 104078. https://doi.org/10.1016/j.cognition.2019.104078

Román-Caballero, R., Martín-Arévalo, E., & Lupiáñez, J. (2021). Attentional networks functioning and vigilance in expert musicians and non-musicians. *Psychological Research, 85*(3), 1121–1135. https://doi.org/10.1007/s00426-020-01323-2

Sanchez-Ruiz, M.-J., Pérez-González, J. C., Romo, M., & Matthews, G. (2015). Divergent thinking and stress dimensions. *Thinking Skills and Creativity, 17*, 102–116. https://doi.org/10.1016/j.tsc.2015.06.005

Sanchis, C., Blasco, E., Luna, F. G., & Lupiáñez, J. (2020). Effects of caffeine intake and exercise intensity on executive and arousal vigilance. *Scientific Reports, 10*(1), 8393. https://doi.org/10.1038/s41598-020-65197-5

Satterfield, K., Harwood, A. E., Helton, W. S., & Shaw, T. H. (2019). Does depleting self-control result in poorer vigilance performance? *Human Factors: The Journal of the Human Factors and Ergonomics Society, 61*(3), 415–425. https://doi.org/10.1177/0018720818806151

Schubert, A.-L., Frischkorn, G. T., & Rummel, J. (2020). The validity of the online thought-probing procedure of mind wandering is not threatened by variations of probe rate and probe framing. *Psychological Research, 84*(7), 1846–1856. https://doi.org/10.1007/s00426-019-01194-2

Seli, P., Beaty, R. E., Cheyne, J. A., Smilek, D., Oakman, J., & Schacter, D. L. (2018). How pervasive is mind wandering, really? *Consciousness and Cognition, 66*, 74–78. https://doi.org/10.1016/j.concog.2018.10.002

Seli, P., Carriere, J. S. A., Levene, M., & Smilek, D. (2013). How few and far between? Examining the effects of probe rate on self-reported mind wandering. *Frontiers in Psychology*, *4*(JUL). https://doi.org/10.3389/fpsyg.2013.00430

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben- Shachar, M. S. (2023). *afex: Analysis of Factorial Experiments (R package version 1.3-0) [Computer software]*. https://cran.r-project.org/package=afex

Smallwood, J. (2010). Why the global availability of mind wandering necessitates resource competition: Reply to McVay and Kane (2010). *Psychological Bulletin, 136*(2), 202–207. https://doi.org/10.1037/a0018673

Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin, 132*(6), 946–958. https://doi.org/10.1037/0033-2909.132.6.946

Thomson, D. R., Besner, D., & Smilek, D. (2015). A resource-control account of sustained attention: Evidence from mind-wandering and vigilance paradigms. *Perspectives on Psychological Science, 10*(1), 82–96. https://doi.org/10.1177/1745691614556681

Warm, J. S., Dember, W. N., & Hancock, P. A. (1998). Workload and vigilance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 42*(10), 769–771. https://doi.org/10.1177/154193129804201025

Weinstein, Y. (2018). Mind-wandering, how do i measure thee with probes? Let me count the ways. *Behavior Research Methods, 50*(2), 642–661. https://doi.org/10.3758/s13428-017-0891-9

Welhaf, M. S., Meier, M. E., Smeekens, B. A., Silvia, P. J., Kwapil, T. R., & Kane, M. J. (2022). A "Goldilocks zone" for mind-wandering reports? A secondary data analysis of how few thought probes are enough for reliable and valid measurement. *Behavior Research Methods, 55*(1), 327–347. https://doi.org/10.3758/s13428-021-01766-4

Wiemers, E. A., & Redick, T. S. (2019). The influence of thought probes on performance: Does the mind wander more if you ask it? *Psychonomic Bulletin & Review, 26*(1), 367–373. https://doi.org/10.3758/s13423-018-1529-3

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D.(2023). *dplyr: A Grammar of Data Manipulation. (R packageversion 1.1.4.) [Computer software]*. https://CRAN.R-project.org/package=dplyr

Zanesco, A. P., Denkova, E., & Jha, A. P. (2025). Mind-wandering increases in frequency over time during task performance: An individual-participant meta-analytic review. *Psychological Bulletin, 151*(2), 217–239. https://doi.org/10.1037/bul0000424

Zanesco, A. P., Denkova, E., Witkin, J. E., & Jha, A. P. (2020). Experience sampling of the degree of mind wandering distinguishes hidden attentional states. *Cognition, 205*, Article 104380. https://doi.org/10.1016/j.cognition.2020.104380

Zholdassova, M., Kustubayeva, A., & Matthews, G. (2021). The ANT executive control index: No evidence for temporal decrement. *Human Factors, 63*(2), 254–273. https://doi.org/10.1177/0018720819880058

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.