Consciousness and Cognition 131 (2025) 103870

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/yccog

Registered Report Stage II

# Unravelling the threads of thought: Probing the impact of contextual factors on mind wandering

Andreas Alexandersen [*,1], Krister Dahlberg, Gábor Csifcsák [2], Matthias Mittner [3]

*Department of Psychology, UiT The Arctic University of Norway, Norway*

A R T I C L E I N F O

*Keywords:*
Mind wandering
Contextual factors
Online experiment
Finger tapping
Approximate entropy
Behavioral variability
Performance feedback

A B S T R A C T

This study investigated the influence of contextual factors on mind wandering (MW) by leveraging an online platform for an established laboratory task. We investigated how direct performance feedback, information about task progression, and the feeling of being monitored influenced performance indices in a task used to investigate the effect of MW on executive control. Our results indicate that specific performance feedback, and not general positive feedback, consistently improved performance, while neither impacted self-reported MW. Conversely, feedback on task progression and the feeling of being monitored increased self-reported MW, possibly reflecting participant self-awareness due to contextual distractions. Intriguingly, information relaying task progression also substantially increased performance. These findings highlight the potential of performance feedback to reduce the negative effects of MW on task performance in an online setting. Additionally, the findings suggest that information about task progression, as well as the notion of being monitored during the experiment can influence task focus and should be taken into consideration when investigating fluctuations of attention during cognitive tasks.

## 1. General introduction

Inattention constitutes a prominent area of research due to its substantial impact on task performance and its association with mental disorders such as ADHD, anxiety, and depression (Seli et al., 2015; Smallwood & Schooler, 2015). Periods of inattention, where individuals' minds are engaged in thoughts unrelated to their immediate environment, are often referred to as task-unrelated thoughts (TUTs) or Mind Wandering (MW; Antrobus et al., 1970). The high frequency of MW both during experimental tasks and in day-to-day situations suggests that this phenomenon plays a central role in our daily lives (Kane et al., 2007; Klinger & Cox, 1987; Smallwood & Schooler, 2015).

Mind wandering most frequently occurs during mundane or monotonous activities. One facet of the phenomenon is that MW can be seen as a way for the brain to employ excess cognitive resources when the demands of the current task permit it (Smallwood et al., 2012; Smallwood & Schooler, 2015). For example, as you drive home from work on a familiar road, you may find that your mind has momentarily drifted to revisit memories or contemplate future events. Despite the valuable insights gained from decades of MW

* Corresponding author at: UiT The Arctic University of Norway, PO Box 6050, Langnes, N-9037 Tromsø, Norway.
*E-mail address:* andreas.alexandersen@uit.no (A. Alexandersen).
[1] https://orcid.org/0000-0002-4414-8678.
[2] https://orcid.org/0000-0001-7719-7073.
[3] https://orcid.org/0000-0003-0205-7353.

https://doi.org/10.1016/j.concog.2025.103870
Received 11 November 2024; Received in revised form 9 April 2025; Accepted 22 April 2025

research, most studies are conducted in highly controlled laboratory settings, where MW is collected via experience sampling during tasks that require sustained attention (but see; Kane et al., 2007; McVay et al., 2009).

While laboratory studies are preferred for their rigorous control over environmental factors, they implement unique conditions that are different from those encountered in the daily lives of participants, questioning the ecological validity of this approach (Kane et al., 2017). For instance, research has demonstrated that MW episodes can vary in frequency and duration depending on factors such as specific task demands, available cognitive resources, emotional state and motivation to perform a task (Kane et al., 2007, 2017; Mooneyham & Schooler, 2013). One theoretical approach to understand how variations in MW occur across different environments is the "context regulation" perspective proposed by Smallwood & Andrews-Hanna (2013). According to this view, the costs and benefits of MW are context-dependent and therefore its regulation is as well. As an example, previous research has observed that MW occurs less frequently when individuals are occupied with demanding external tasks, and conversely, boring and less demanding tasks lead to more MW episodes (Antrobus et al., 1970; Smallwood & Andrews-Hanna, 2013; Teasdale et al., 1995). This has been proposed to be a result of flexible reallocation of resources (Smallwood & Schooler, 2006; Thomson et al., 2015).

However, researchers have suggested that classical MW tasks such as the sustained attention to response task, might not require a lot of executive control (Boayue et al., 2020). Therefore, to more thoroughly investigate the relationship between executive functioning and MW, the Finger-Tapping Random Sequence Generation Task has been developed (FT-RSGT, Boayue et al., 2020). The FT-RSGT is sensitive to fluctuations in executive control with relatively high temporal resolution, and is interspersed with MW probes to elucidate how MW episodes are related to executive functioning. While the FT-RSGT is able to capture fluctuations in executive resources, it still remains unclear how different contextual factors may influence executive control and consequently, MW (Boayue et al., 2020; McVay et al., 2009; Robison & Unsworth, 2018). Therefore, to investigate how environmental factors and contextual differences affect MW, we translated the laboratory-based FT-RSGT protocol into an online setting. This approach allowed us to explore potential variations in MW and executive functioning in an environment closer to real-life scenarios (i.e., outside of the laboratory).

While there are clear benefits of conducting studies online (Reips, 2000; Sauter et al., 2020), there are some considerations that need to be addressed when replicating laboratory-based MW studies. Conducting experiments in a laboratory setting with experimenter guidance may create social pressure, which can enhance or impair performance depending on the participants' evaluation of the situation (Blascovich et al., 1999). Consequently, this pressure could lead to variations in observed task focus between the laboratory and the online versions of the task (Belletier et al., 2015). Additionally, prior research has indicated that participants tend to be less accurate in online experiments (Dandurand et al., 2008). This issue can be partially attributed to the absence of clear explanations or clarifications from readily available feedback from the experimenter (Feenstra et al., 2017; Reips, 2002). One possible solution to reduce between-subject variability caused by a lack of experimenter interaction is to use clear and elaborate instructions that have been pre-tested (Reips, 2002). Another solution is to incorporate feedback on task performance directly into the task (Burgers et al., 2015; Earley et al., 1990).

Furthermore, previous research has also emphasized the challenge of maintaining participants' attentiveness at a high level throughout the duration of an experiment, as reflected in the well-known time-on-task effect on MW frequency (Krimsky et al., 2017). While this is a concern for all experimental research, it seems to be particularly difficult in online experiments where distractions are more readily available (Saravanos et al., 2021). In MW research, it is important that participants consistently engage with and disengage from the task throughout the duration of the experiment. Tasks perceived by participants as uninteresting or unimportant tend to result in increased MW, as participants allocate mental resources to other activities such as future planning or reminiscing about past events (Mooneyham & Schooler, 2013; Smallwood & Andrews-Hanna, 2013). While previous research has highlighted the significant impact of motivation on task performance during laboratory tasks (Appel & Gilabert, 2002; Sailer et al., 2017), other studies suggest that intermittent performance feedback can keep participants engaged and adhering to instructions during online tasks (Crump et al., 2013). However, prior research has also indicated that positive feedback by itself, even if unrelated to someone's actual performance, can influence participants' perception of task meaningfulness or motivation (Burgers et al., 2015; Sailer et al., 2017). The motivational driving force behind positive encouragement and performance feedback may foster various types of motivation (Sailer et al., 2015). Positive encouragement can elicit feelings of competence and positivity, which, according to the perspectives of self-determination and emotion, may enhance motivation (Astleitner et al., 2000; Ryan and Deci, 2000). Conversely, from a behaviorist learning perspective, immediate feedback, whether positive or negative, can also increase motivation (Skinner, 1963). Additionally, another perspective on what drives motivation, the perspective of interest, suggests that providing direct feedback which enhances flow, can further boost motivation (Csikszentmihalyi et al., 2005, pp. 508-609). While motivational research encompasses numerous perspectives that do not necessarily contradict each other, they each focus on different components of what drives motivation (Park, 2017). These components can predict different motivational outcomes, and thus, different types of feedback could potentially lead to variations in task engagement (Sailer et al., 2013; Sailer et al., 2015).

Furthermore, as mentioned above, a pervasive observation in MW research is the consistent decline in performance over the course of a task (Zanesco et al., 2024). At the same time, an increasing number of studies suggest that when information about time remaining of the task is available, performance tends to increase towards the end of the experiment (Aasen et al., 2024; Drevland et al., 2024; Katzir et al., 2020). While information about remaining time is often available in laboratory studies via direct experimenter interaction, it is not always included in online studies, which potentially could influence task engagement.

Finally, research on MW relies almost exclusively on self-reports in the form of experience sampling and thought probing; thus, honesty and trustworthiness of the subjective responses is critical to MW research. Research has shown that the feeling of being monitored can influence honesty: Being observed can reduce the opportunity cost of being dishonest in anticipation of punishment or reduction in social status (e.g., how the participant is perceived by the experimenter) or by inducing guilt or shame (Beck et al., 2020). Therefore, when collecting data outside of the laboratory, a reduced feeling of being observed could lead to potential differences in self-reported MW as a consequence of more honest responding.

To summarize, the growing body of MW literature predominantly features research conducted in laboratory settings. While these environments ensure control and allow results to be comparable across laboratories, they offer limited insight into the varying contextual factors that might influence MW in everyday experiences. This study aims to explore some of these contextual factors that could potentially affect the interplay between MW and executive functioning. Specifically, this study investigates whether the FT-RSGT can be effectively translated to an online experiment and examines how different types of feedback impact task performance and MW. Additionally, the study investigates the effects of task progression information and the sensation of being observed.

## 2. General Methods

All materials, including behavioral and demographic data as well as task scripts can be found at our OSF repository (https://osf.io/wjvk2/).[4]

### 2.1. Participants

Participants were recruited through Prolific (https://www.prolific.co), an online crowd-sourcing platform known for its three-step authorization and pre-screening process, which ensures efficient and secure participant screening. The experiment was advertised with a compensation of 3£ (GBP) for completing the study, with an additional bonus payment of 2£ for adequate performance. For ethical reasons, all participants who completed the study were compensated with the full amount, i.e., received the bonus payment. Each participant was assigned an anonymized ID, preventing participation in multiple of our conditions through a blocklist. Using Prolific's prescreening facility, only participants who answered "yes" to the following questions were recruited: (1) fluent in English, (2) aged from 18 to 50 years, (3) normal or corrected eyesight. All participants self-identified as healthy adults. Additionally, they answered affirmatively: (4) no ongoing mental illness/condition and (5) no mental illness daily impact (6) no mild cognitive impairment or dementia.

Lastly, we checked the resulting data files for participants who did not comply with the instructions. Participants were informed that once the instructions and training were complete, they would have to complete the entire experiment without pausing for extended periods in order to be eligible for payment. They were informed that short breaks up to one minute were allowed between blocks. Based on these instructions, we excluded participants based on the following criteria: (1) paused tapping buttons three or more times during the experiment, (2) tapped the buttons at another frequency than once at each beep (if the number of taps exceeded the number of beeps by more than 20 %), (3) switched windows away from their browser and the experiment after the instructions but not during the breaks (more than for a total of 10 s, or more than 10 times), (4) taking long breaks (spent more than two minutes on a single break). These criteria were set to mirror a typical protocol employed in our lab (Alexandersen et al., 2022; Boayue et al., 2020).

### 2.2. Procedure

Before starting the experiment, participants adjusted their PC volume to a comfortable setting while a song was played for that purpose. To verify audio functionality, an audio test presented five animal sounds along with corresponding pictures, and participants were required to correctly map each sound to the correct image before proceeding to the main experiment (sounds were repeated in case of errors). Participants received comprehensive instructions before the task that included details on the task-requirements and thought-probes. They were also reminded about the bonus payment and provided with a warning that failure to comply with instructions would disqualify them from payment. To ensure sufficient understanding of the task instructions, participants needed to pass a quiz where they were required to correctly respond to a series of questions about the task. The experiment consisted of 20 blocks lasting approximately 60 s each, with interspersed thought probes at the end of each block, for a total duration of approximately 20–25 min depending on the time spent to answer the thought probes (Fig. 1). The experiment was programmed in JavaScript using the jsPsych library (de Leeuw et al., 2023) and was hosted on a university server using JATOS (Lange et al., 2015).

### 2.3. The experimental task

The FT-RSGT is a combination of a rhythmic finger-tapping task and the random number generation task (Baddeley et al., 1998), developed by Boayue et al. (2020). In this task, repetitive metronome sounds (beeps at 440 Hz) are presented, separated by an Inter-Stimulus Interval (ISI) of 750 ms. Participants are instructed to press one of two keys either with the left or the right index finger (keys 'f' and 'j', respectively; see Fig. 1). Participants were specifically instructed to synchronize their taps with the metronome while also striving to keep the resulting sequence of left/right taps as random as possible. Participants were provided with a written explanation of the task, and the concept of randomness was explained using the analogy of repeatedly tossing a fair coin to determine which button to press for each beep. Prior to the actual testing, participants underwent three one-minute training sessions.

*Behavioral variability*.

[4] Even though some hypotheses were pre-registered before data collection (which can be found at the OSF repository), these pre-registrations were not quality controlled because they were created in the context of a student project. We therefore disregard the pre-registrations and all analyses should be treated as exploratory.

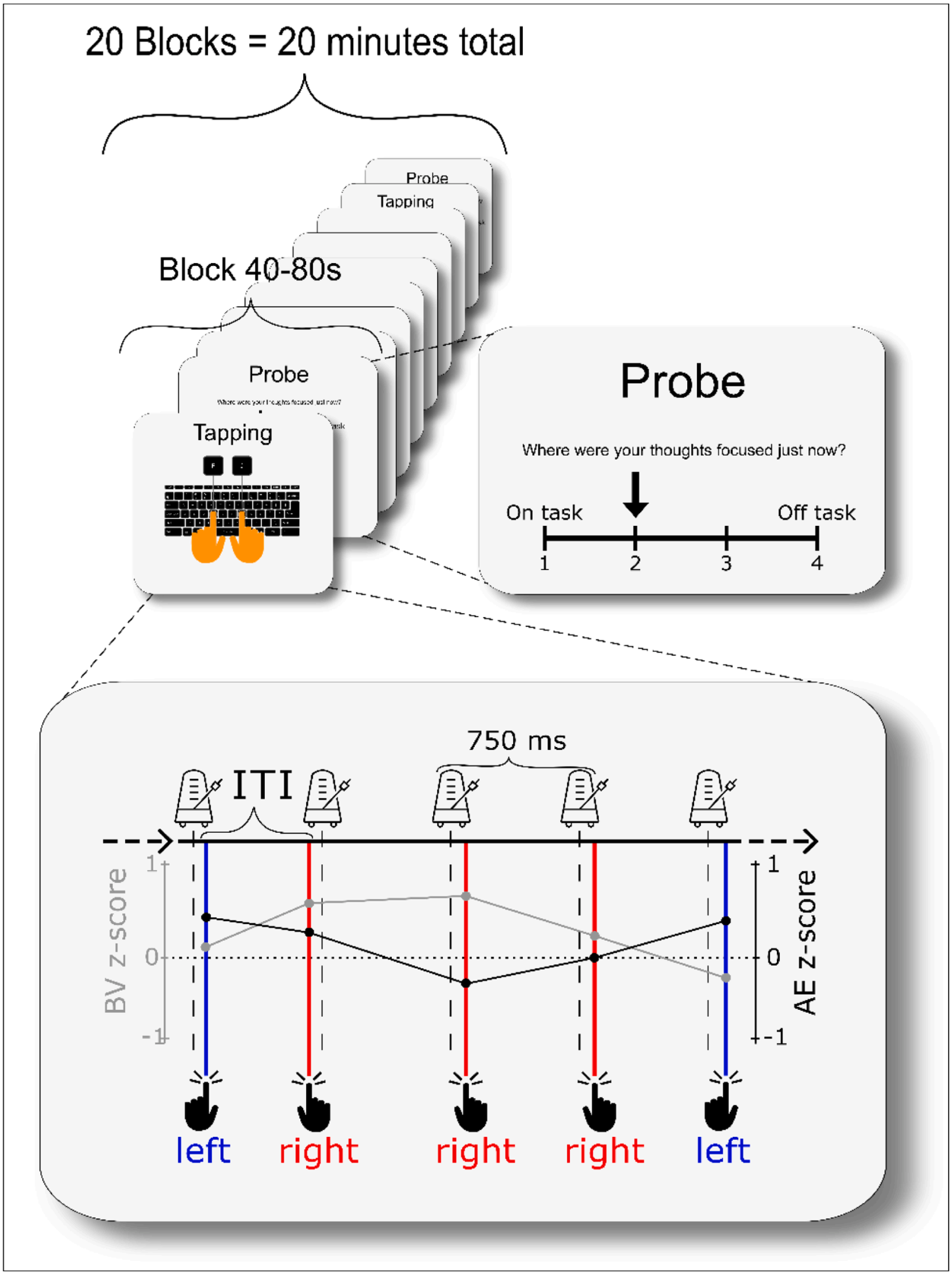


**Fig. 1.** The FT-RSGT. The figure shows the structure of the task, with probes presented at the end of each block (upper panel), as well as the relationship between the sequence and timing of taps and corresponding performance measures AE and BV (lower panel). Adapted figure by Aasen, S. R., 2024; available at https://doi.org/10.6084/m9.figshare.25471960.v1 under the CC-BY 4.0 license.

Behavioral Variability (BV) has been closely related to lapses of attention in previous studies (Seli et al., 2013) and the FT-RSGT was designed to capture this measure with high temporal resolution (Boayue et al., 2020). In this study, BV was computed as the log-transformed standard deviation (SD) of the Inter-Tap Intervals (ITI) for the 25 trials before each thought-probe. Lower BV scores indicate a more rhythmic tapping, suggesting that participants are tapping more closely in sync with the pace of the metronome.

*Approximate entropy*

The FT-RSGT was designed to capture fluctuations of executive functioning through quantifying the degree of randomness participants manage to produce in a tapping sequence, an approach that was inspired by the classical random number generation experiments by Baddeley et al. (1998). In our study, to quantify randomness in the observed sequences, we used the Approximate Entropy (AE) measure (Pincus, 1991), calculated on the binary left–right sequences coded as zero (left) and one (right). AE assesses sequence irregularity (i.e., randomness) and is sensitive to repeated patterns in the generated sequence. Higher AE values indicate greater irregularity and consequently, better executive performance. The calculation of AE involves determining the logarithmic frequency at which sequences of a specified length that are close together, remain close when the blocks are extended by one position. The AE parameter $m$, which specifies the length of the evaluated sub-sequence, was set to $m = 2$ in agreement with results by Boayue et al. (2020). The raw AE score was transformed as:

$AE_{trans} = -log(log(2) - AE_{raw})$ to yield a measure that is approximately normally distributed (Boayue et al., 2020).

*Self-reported mind wandering*

To collect self-reported mind wandering, thought probes were dispersed throughout the task to ask participants about their task focus. The task was paused every minute on average (randomly between 40–80 s) and participants were required to rate their attentional focus using a four-point Likert scale, by answering the question "Where was your attention focused just before this question?", where 1 indicated "completely focused on the task", and 4 indicated "completely unfocused on the task" (See Fig. 1).

### 2.4. Statistical Methods

*Bayesian statistics and statistical inferences*

We exclusively employed Bayesian statistics due to their numerous advantages over classical frequentist approaches (Wagenmakers et al., 2018). As opposed to traditional, frequentist null-hypothesis significance testing (NHST), Bayesian statistics does not strictly distinguish between significant and non-significant results. Rather, in Bayesian statistics, we calculate measures of evidence such as Bayes Factors, highest-density intervals (HDIs) and evidence ratios (ERs) that give a direct indication of the amount of evidence offered by the data and that are independent of thresholds set a priori (such as the significance level in NHST). To make our results more interpretable to readers who are more familiar with frequentist approaches, and for consistency in reporting, we have use fixed thresholds for labeling the rough amount of evidence carried by the effects. A description and motivation for these thresholds can be found in the statistical modeling and model selection section.

*Statistical modeling and model selection*

To investigate how contextual factors influence MW, as well investigating the correspondence between the online FT-RSGT and its laboratory counterpart, two separate analyses were conducted for each study. The first analysis was a multivariate Bayesian hierarchical linear mixed effects model, with the three indices of interest (MW, BV and AE) as outcome variable (concatenated variables). We included the predictor variables Condition (dummy coded), Variable (MW, AE, BV; dummy coded) and Trial number (1 to 20, referring to the number of probes throughout the task). The outcome variables were z-transformed using the grand mean and SD across subjects to maintain between-subject and between-condition variability.

The model also included a random slope for Variable as well as random intercepts for the different subjects. This approach resulted in a total of four different models per study (see Supplementary Methods), and the best fitting model was determined by using model weights calculated using the Leave One Out Information Criteria (LOOIC). As all models were very similar in terms of their LOOIC fit, this criterion was deemed inconclusive (the maximum difference between any two models was 8–4 units), and we chose to proceed with the full model that included all predictors and interactions in all cases. For a complete overview over the models and model selection, refer to Table 1 in the Supplementary Material.

The second analysis aimed to explore whether the online version of the FT-RSGT captured the expected relationship between self-reported MW and the behavioral indices (BV increases and AE decreases preceding MW episodes), as previously found and replicated in several laboratory studies utilizing the same task (Alexandersen et al., 2022; Boayue et al., 2020; Groot et al., 2022). This approach allowed us to compare differences in the MW-AE/BV relationship across the different contextual conditions. We chose to model the dependent variable, the ordinal self-reported MW, with a hierarchical Bayesian ordinal probit model (Brooks, 1998; Liddell & Kruschke, 2018). The predictors included in this model were AE, BV, Condition and Trial (i.e., the position of the thought probe in the task sequence), allowing both AE and BV to freely interact with Condition.

All data analyses were conducted using the "brms" package in R (Bürkner, 2017) applying HMC algorithms implemented in the Stan software (Stan Development Team, 2016) via the cmdstanr package (Gabry et al., 2023). We sampled 8000 samples using 8 parallel chains. Convergence was determined by visual inspection of the traceplots and ensuring that the $\widehat{R}$ values were below 1.05 (Vehtari et al., 2021). We summarized the posterior distribution by computing the posterior mean, 95 % Highest-Density Intervals (HDI), and Evidence Ratios (ER) for the regression coefficients. The HDI represent intervals within which the true coefficient is estimated to fall with 95 % probability, contingent on the accuracy of the model. The ER is the probability of the effect being in the expected direction. To ensure consistency and to convey the magnitude of evidence for the effects, we have established standardized labels based on the following cutoffs: ERs below 5 will be completely disregarded since the provided evidence is very weak. ERs between 5 and 10 will be described as "weak evidence" for an effect, ERs between 10 and 20 as "tentative evidence", and ERs above 20 as substantial evidence for the respective effect. It is important to emphasize that these thresholds and labels have been arbitrarily chosen and are not used to make a strict distinction into "significant" and "non-significant". Rather, these labels should guide the reader by attaching an interpretation that roughly conveys the amount of evidence carried by the different ERs. We encourage the reader to carefully consider the numerical values themselves in relation to the theoretical interpretation and plausibility of the effects.

## 3. Study 1 – Conceptual replication and performance feedback

### 3.1. Introduction

MW is a phenomenon that is heavily dependent on motivation and its measurement may therefore be compromised by the availability of external distractions (e.g., smart phones, other people, music, etc.) when direct supervision is absent. Therefore, it can be challenging to investigate MW in an online experiment. Nevertheless, prior research has effectively employed online adaptations of MW protocols utilizing the standard sustained attention to response task (SART; Cheyne et al., 2009) and a metronome-based finger-tapping task (Seli et al., 2013). However, unlike the simpler version of the metronome task that mainly relies on sustained attention and audio-motor coordination, the FT-RSGT also depends heavily on executive control necessary to generate random sequences of left–right taps and thus, is suitable for assessing the interplay between MW and the executive system under various experimental conditions. Therefore, the primary objective of Study 1 was to validate the online adaptation of the FT-RSGT in capturing behavioral indicators of MW.

Previous research suggests that online experiments often exhibit reduced accuracy, which may be related to inter-individual differences in interpreting task-instructions (Dandurand et al., 2008). Since the FT-RSGT requires participants to produce random sequences, it is crucial that they have a solid understanding of the concept of randomness as understood in the context of our study. Generally, a formal understanding of randomness and probability can vary significantly among the general public, as these concepts are inherently difficult to grasp (Batanero, 2015).

To address this issue, one experimental condition, the training feedback condition (TFB) introduced performance feedback during the training session to compensate for the absence of direct experimenter supervision, and to investigate whether this modification would enhance performance or alter self-reported MW. Introducing performance feedback during the training allowed participants to get immediate feedback on their performance relative to the requirements of the task and hence facilitated their understanding of the behavior that was expected of them.

Furthermore, previous studies have found that implementing graphical performance feedback positively affects the perceived meaningfulness and motivation of the task, which subsequently could improve performance (Burgers et al., 2015; Sailer et al., 2017). Specifically, feedback from receiving a high score, or performing better than previous scores have shown to increase motivation in games by making participants feel more competent and autonomous (Burgers et al., 2015). Additionally, previous studies have shown that increased task motivation reduces self-reported MW (Seli et al., 2019).

Therefore, we introduced a new experimental condition, the performance feedback condition (PFB), where performance feedback was presented throughout the experiment to investigate whether continuously getting performance feedback throughout the task would further increase performance. Importantly, Study 1 contrasted training feedback and performance feedback against a baseline condition, the no feedback (NoFB) condition. This condition mirrors the laboratory version of the FT-RSGT, without any form of feedback. Across all conditions, we anticipated that the previously reported relationship between BV, AE and MW would still be intact in the online version of the task: AE would be reduced during MW, BV would be increased during MW, and finally, MW would increase with time spent on task. We further expected that, training feedback and performance feedback would improve performance (increase AE, reduce BV and MW) relative to the NoFB condition.

### 3.2. Methods

*Participants*

A total of 120 participants were recruited to participate in the study (40 per condition). Nine participants had to be excluded upon investigating the data (five participants took a long break from the experiment mid-task, three were excluded for tapping multiple times per stimulus, and one participant was excluded due to minimizing the experiment multiple times mid-task). The final sample was

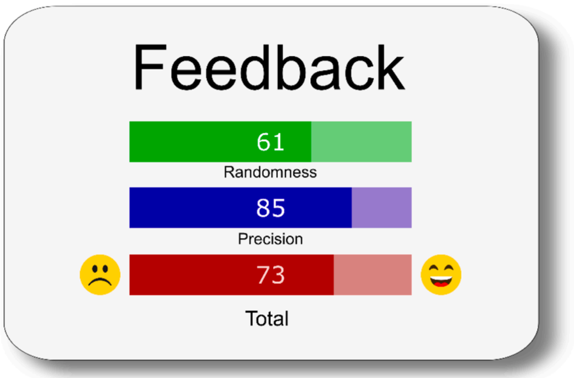


**Fig. 2.** *Feedback scales shown to the participants after training. Note*: Example of a total feedback score presented during the last five blocks of training. For the first ten training blocks, only the precision or randomness scales were shown, respectively.

N = 111, aged between 18–50 years (43 females, mean age = 28.7, SD = 7.99). Out of the 111 participants, 37 participants (16 females, aged 20–43, mean age = 27.7, SD = 7.11) completed the FT-RSGT without any feedback (NoFB condition), another 35 (15 females, aged 18–50, mean age = 29, SD = 7.83) received performance feedback after the training session only (TFB condition), while the final 39 (12 females, aged 19 – 50, mean age = 29.5, SD = 8.96) received performance feedback during training, as well as after every thought probe (PFB condition).

*Performance training and feedback*

The training session consisted of multiple short sessions, each divided into five blocks focusing on AE (feedback only on the randomness of the sequence) and another five on BV (feedback only on the precision of the taps). The final portion consisted of five blocks that assessed performance across both scales, providing participants with feedback based on their measurements for AE, BV and a joint score. The measurements are based on the participants' performance and represented as a visual analogue scale ranging from zero to one hundred (Fig. 2). A full score on the BV scale was achieved when the SD of the ITI was no more than 30 ms. Conversely, a score of zero was assigned when the SD exceeded 200 ms, while values between SD = 200 ms and SD = 30 ms were linearly interpolated, yielding values from zero to one hundred percent. To quantify performance on the randomness scale, the raw AE values were normalized by dividing by log(2), which is the maximum value of AE, and then multiplied by 100 to represent participants' estimated performance on a percentage scale. The total feedback score integrated both the randomness and precision scores (it was calculated as the mean of BV and AE) and was presented on a separate scale together with the individual scores (see Fig. 2). The feedback was presented in the middle of the screen, at the end of each training block (TFB), or after each thought probe (PFB).

### 3.3. Results

*Effects of feedback*

Our Bayesian linear mixed-model analysis including all three behavioral variables revealed no evidence that self-reported MW differed substantially between the three task conditions (see Fig. 3). We found no evidence of decreased self-reported MW in the TFB compared to the NoFB condition (NoFB – TFB: $\beta = 0.06$ [-0.27, 0.39], $ER^{+} = 1.80$), nor in the PFB condition (NoFB – PFB: $\beta = 0.11$ [-0.21, 0.43], $ER^{+} = 3.04$).

We observed no evidence for a difference in BV between NoFB and TFB (NoFB – TFB: $\beta = 0.03$, [-0.24, 0.31], $ER^{+} = 1.48$). However, we observed substantial evidence that there was a noticeable reduction in BV in PFB compared to NoFB (PFB – NoFB: $\beta = -0.32$, [-0.59, −0.06], $ER_{-} = 119.30$), and TFB (PFB – TFB: $\beta = -0.29$, [-0.56, −0.02], $ER_{-} = 50.86$). Additionally, we observed weak evidence of increased AE between the TFB condition and the NoFB condition (TFB – NoFB: $\beta = 0.14$, [-0.13, 0.41,], $ER^{+} = 5.78$), and this trend continued when comparing the PFB to TFB (PFB – TFB: $\beta = 0.14$, [-0.12, 0.40], $ER^{+} = 5.85$). As a result, we found substantial evidence for an increase in AE in the PFD condition compared to the NoFB condition (PFB – NoFB: $\beta = 0.28$, [0.02, 0.54], $ER^{+} = 59.95$).

We also observed an effect for task duration (Trial) on all three task markers in line with previous work on sustained attention, as well as previous work with the FT-RSGT (Alexandersen et al., 2022; Boayue et al., 2020). The results indicate substantial evidence that MW increased as time on task increased in the NoFB condition ($\beta = 0.12$, [0.06, 0.17], $ER^{+} = 15999$), and as well as in the TFB condition ($\beta = 0.15$, [0.09, 0.21] $ER^{+} = \infty$), and in the PFB condition ($\beta = 0.18$, [0.12, 0.23], $ER^{+} = \infty$), with tentative evidence of the effect being stronger during PFB (PFB – NFB: $\beta = 0.06$, [-0.02, 0.14], $ER^{+} = 14.04$; Fig. 3). As expected, we also observed substantial evidence that BV increased as task duration increased in the NoFB condition ($\beta = 0.07$, [0.01, 0.13], $ER^{+} = 95.39$). In the TFB condition, we observed no evidence that BV increased over time ($\beta = 0.03$, [-0.03, 0.09], $ER^{+} = 4.78$), while we found weak evidence of BV increasing at a slower rate than in the NoFB condition (NoFB − TFB: $\beta = 0.07$, [-0.4, 0.19], $ER^{+} = 8.56$). This was, however, not true for the PFB condition. Instead, we observed weak evidence for BV to decrease as time increased ($\beta = -0.03$, [-0.09, 0.03], $ER_{-} = 6.26$), with tentative evidence that it differed from the TFB condition (PFB − TFB: $\beta = -0.09$, [-0.20, 0.03], $ER_{-} = 14.27$), and substantial evidence that BV developed differently as time increased in the PFB condition compared to the NoFB condition (PFB − NoFB: $\beta = -0.16$, [-0.28, −0.05], $ER_{-} = 379.95$). Finally, we observed substantial evidence for the expected relationship between Trial and AE

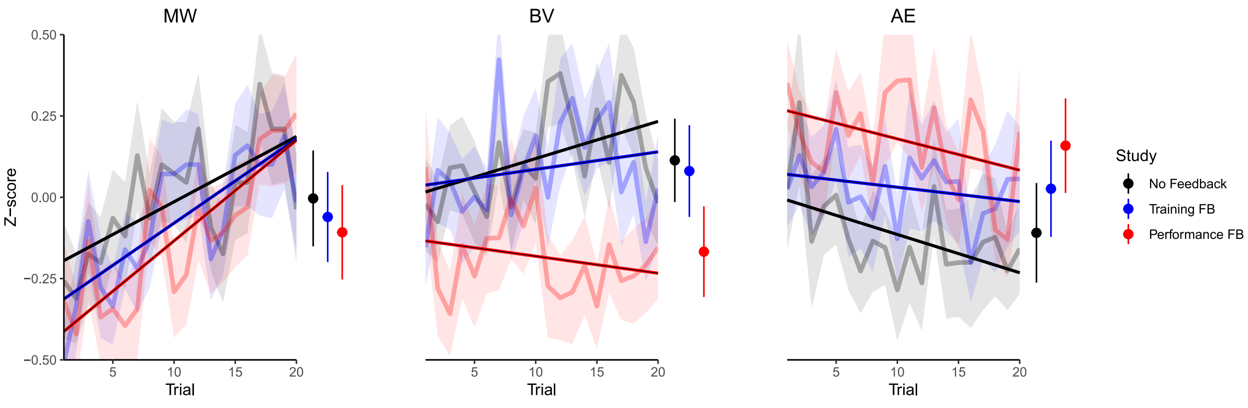


**Fig. 3.** *Differences in AE, BV and MW across conditions as a function of trial number. Marginal effects are depicted on the right side of each plot; error bars and ribbons represent the standard error of the mean. Note:* MW = Mind wandering, BV = Behavioral variability, AE = Approximate entropy.

where AE decreased with increasing Trial in the NoFB condition ($\beta$ = -0.07, [-0.12, −0.01], ER$_{-}$ = 73.07), as well as in the PFB condition ($\beta$ = -0.05, [-0.11, 0.00], ER$_{-}$ = 30.56). We observed no evidence that this effect was present in the TFB condition ($\beta$ = -0.03, [-0.09, 0.03], ER$_{-}$ = 4.5), however, we did not observe any evidence that it was different from the NoFB condition (TFB − NoFB: $\beta$ = 0.04, [-0.04, 0.12], ER$^{+}$ = 4.36) or the PFB condition (TFB − PFB: $\beta$ = 0.03, [-0.06, 0.11], ER$^{+}$ = 2.76).

*Relationship between MW and task performance.*

In our ordinal regression analysis treating MW as dependent variable, we observed strong evidence for all hypotheses regarding the behavioral indices. More specifically, we observed substantial evidence that MW probe responses increased with BV in all three conditions, with the strongest relationship in the NoFB condition ($\beta$ = 0.22, [0.11, 0.34], ER$^{+}$ = $\infty$). This relationship was also present both in the PFB condition ($\beta$ = 0.17, [0.06, 0.28], ER$^{+}$ = 799) and in the TFB condition ($\beta$ = 0.09, [-0.02, 0.20], ER$^{+}$ = 21.97), although compared to the NoFB condition, we found no evidence that this relationship was different in the PFB condition (BV x PFB: $\beta$ = -0.05, [-0.21, 0.10], ER$_{-}$ = 3.01), while we found tentative evidence that this relationship was weaker in the TFB condition (BV x TFB: $\beta$ = -0.13, [-0.29, 0.03], ER$_{-}$ = 18.90).

We also observed substantial evidence of MW probe response to decrease with AE, i.e., lower randomness predicted off-task focus in the NoFB condition ($\beta$ = -0.12, [-0.23, 0.02], ER$_{-}$ = 110.11), in the PFB Condition ($\beta$ = -0.10, [-0.22, 0.01], ER$_{-}$ = 26.40) and tentative evidence of this relationship in the TFB condition ($\beta$ = -0.10, [-0.22, 0.02], ER$_{-}$ = 15.81). We found no evidence that the negative relationship between AE and MW was different across the conditions (AE x TFB: $\beta$ = 0.03, [-0.13, 0.18], ER$^{+}$ = 1.69; AE x PFB: $\beta$ = 0.02, [-0.13, 0.17], ER$^{+}$ = 1.48).

### 3.4. Discussion

Our findings demonstrate the successful replication of the lab-based FT-RSGT in an online environment as indicated by the expected relationship between MW and the behavioral performance measures AE and BV, even in the absence of the stringent experimental control typically afforded by laboratory settings. In addition, the incorporation of performance feedback during training increased overall performance, which further increased when performance feedback was extended throughout the entire task. However, this enhancement in performance (enhanced tapping precision and randomness), did not impact the frequency of self-reported MW. While we also observed a somewhat stronger time-on-task effect on MW in the PFB condition, this was likely a result of increased task focus at the start of the task, as self-reported MW scores were lower at task onset, yet they ended up at roughly at the same levels nearing the end of the task (Fig. 3).

The simultaneous increase in performance while maintaining similar levels of task focus point to the performance feedback having a beneficial impact on task understanding and/or motivation. Recent research utilizing the psychomotor vigilance task has demonstrated improvements in response times without corresponding changes in self-reported focus when using content-free cues and feedback (Unsworth et al., 2024). Similarly, Robison et al., (2021) observed that performance feedback not only enhanced response times but also reduced intentional mind-wandering. These findings, alongside our current results showing improved performance without alterations in self-reported task focus, underscore the potential limitations of self-reported measures in capturing subtle changes in attentional states. Behavioral performance may serve as a more sensitive and direct indicator of attentional lapses, as self-reports often display increased variability, making them less sensitive to subtle shifts in task focus. While our study employed thought probes contrasting on-task with off-task thoughts, previous research with more detailed probes has reported increased task-related interference in feedback conditions, indicating that feedback mechanisms might enhance *meta*-awareness and prompt participants to reflect more on their performance (Robison et al., 2021; Unsworth et al., 2024). This heightened *meta*-awareness could contribute to the observed improvements in behavioral metrics but may also interfere with how participants report their task focus, depending on the dimensions provided by the thought probes. Consequently, it is plausible that the improved behavioral performance observed in our study was due to fewer lapses in attention caused by performance feedback, similar to the findings reported by Robison et al., (2021) and Unsworth et al., (2024). Yet, this reduction in attentional lapses was not detectable through self-report due to the lack of sensitivity in our thought probes, highlighting the importance of integrating objective performance measures alongside self-reports to achieve a comprehensive understanding of attentional dynamics.

As to why performance is increased following performance feedback, previous research has found that as participants accumulate experience in task-solving strategies over time, and as executive control diminishes because of reduced vigilance, participants can adopt less effortful strategies of performing the task to offset the increased opportunity cost of remaining on the task (Kurzban et al., 2013; Pavlova, 2024). In the absence of performance indicators, participants may assume that they are continuing to perform well throughout the task and erroneously divert resources away from the task, leading to reduced performance. In our study, participants can continuously update their beliefs about their performance in real-time, facilitated by the performance feedback that fosters an improved comprehension of what constitutes a “random” and “precise” tapping pattern. This process effectively narrows the disparity between participants’ beliefs regarding randomness and precision and the actual performance of these measures.

Furthermore, while the expected relationship between BV and MW was present in both the TFB and PFB conditions, this relationship was weaker compared to the condition without any performance feedback (NoFB). This suggests that performance feedback may have simplified the task. According to the executive resource use hypothesis (Smallwood & Schooler, 2006), an easier task could leave surplus executive resources available to participants, which could then be utilized for either engaging in MW or solving the task. The observed increase in performance, without a corresponding increase in MW propensity, suggests that participants might have redirected these extra resources towards task-solving strategies as the task became easier with performance feedback.

Another explanation for the increased task performance when performance feedback was available is heightened motivation. As an example, Burgers et al. (2015) found that both positive and negative feedback enhanced engagement and willingness to continue solving a brain-training game. Additionally, the impact of regular and engaging feedback on motivation and task performance has been explored within the context of "gamification"—the incorporation of game design elements into non-game tasks (Deterding et al., 2011). In our study, performance feedback can be viewed as a gamification element. More specifically, it falls under the category of performance graphs, which typically inform participants about their performance relative to previous attempts, thereby allowing them to evaluate their progress over time (Sailer et al., 2013). Research suggests that this type of feedback fosters a mastery orientation, which can enhance intrinsic motivation, prolonging task engagement (Sailer et al., 2013, 2017). Although there is some ambiguity regarding the effectiveness of gamification elements in boosting task engagement, the general consensus suggests that gamification tends to produce positive effects (Sailer et al., 2017). Therefore, it is possible that the observed improvement in performance in the "gamified" version of the task, could be attributable to motivational increase through mastery orientation, where the motivation stems from trying to beat one's previous score (Sailer et al., 2013). A potential increase in motivation could increase the total amount of executive resources available to the participants, which then, according to the executive function use hypothesis (Smallwood & Schooler, 2006) could have been allocated to task solving strategies.

However, this interpretation of our results (increased performance due to an increase in resources) disagrees with the executive failure hypothesis suggested by McVay and Kane (2010), which suggests that executive resources are directly linked to MW and are spent to keep intrusive thoughts away to remain focused on the task. According to the executive failure hypothesis, if executive resources were increased by the performance feedback, this should have resulted in a reduction in overall MW, regardless of whether it was motivation or task comprehension that increased executive resources.

Instead, our results are in accord with the executive function use hypothesis (Smallwood & Schooler, 2006) and the executive resource control account (Thomson et al., 2015), which suggests that participants can independently allocate cognitive resources either to task-unrelated thoughts or to task-solving efforts. Specifically, if the cognitive resource demands of the task are reduced through enhanced understanding from specific performance feedback, or if motivation to beat one's previous score boosts the total available resources through mastery orientation, more executive resources might become available for implementing task-solving strategies (Smallwood & Schooler, 2006; Thomson et al., 2015).

In summary, we conclude that the observed improvement in randomness and tapping precision with performance feedback likely results from two primary factors: 1) The motivational enhancement provided by gamification elements, which increases overall effort to maintain high performance, and/or 2) improved task comprehension, potentially reducing the difficulty of the task, which narrows the disparity between participants' beliefs regarding randomness and precision and the actual performance of these measures. Hence, our findings indicate that direct performance feedback can be a useful addition and/or substitution for direct experimenter feedback, potentially increasing task comprehension and increasing performance without unduly affecting switches in attentional focus.

To investigate whether the results obtained in Study 1 was due to motivation from the gamified version of the task, or if was caused by increased task comprehension, we conducted a study that compared general positive feedback with performance feedback and feedback only during training.

## 4. Study 2 – Indiscriminate positive feedback

### 4.1. Introduction

Prior research has underscored that merely providing general positive feedback or encouragement can be sufficient to enhance task performance (Astleitner, 2000). Positive feedback can foster an increase in self-perceived competency and autonomy, which in turn can heighten intrinsic motivation (Burgers et al., 2015; Sailer et al., 2017). Additionally, studies investigating the adaptability of executive resources revealed individuals' ability to strategically select moments for MW (Seli et al., 2018), and allocate resources based on their motivation to do so (Robison & Unsworth, 2018), suggesting that an increase in motivation can reduce self-reported MW (Seli et al., 2019). Therefore, we extended the current study to discern whether the observed increase in performance in Study 1 was attributable primarily to the specificity of performance feedback, or predominantly driven by a motivational boost from


**Fig. 4.** The feedback screen presented after every probe in the PosFB condition.

encouragement through positive feedback. To disentangle the effects of performance feedback from those of general positive feedback, we introduced a novel experimental condition featuring indiscriminate general positive feedback following each thought probe. We then compared this condition to both the training feedback and the performance feedback conditions from the previous study. We hypothesized that indiscriminate positive feedback (PosFB condition) would lead to reduced performance in terms of AE and BV compared to the PFB group but would not differ from the TFB group. Additionally, we hypothesized that self-reported MW would be unchanged between all the groups.

### 4.2. Methods

In this study the feedback presented after every probe was a picture of a “Thumbs up”, and the text “Great work!” (Fig. 4). Additionally, performance feedback was presented during the training before the experiment started, just like in the TFB and PFB conditions.

#### 4.2.1. Participants

Out of the 40 new participants recruited for the PosFB condition, 7 participants had to be excluded after inspecting the datafiles (four took long breaks during the experiment, two tapped exceedingly, and one participant switched windows away from the task multiple times during the task). The final sample (N = 33) was aged from 20 to 47 years (Mean = 28.5, SD = 7.41), with 23 males and 10 females. Detailed demographics data displaying acceptance/rejection, sex, nationality, the inclusion criteria are available at the study repository (osf.io/tf5zw).

*Statistical modeling and analysis*.

For Study 2, we used the same analysis strategy as in Study 1, except that we replaced the NoFB condition with the new PosFB condition, using the TFB condition as reference.

### 4.3. Results

*Effects of feedback*

The results indicate no evidence that self-reported MW was different in the PosFB condition and TFB where no feedback was given during the task (PosFB − TFB: $\beta$ = -0.07, [-0.41, 0.26], $ER_-$ = 1.88), nor when compared to the PFB condition (PosFB − PFB: $\beta$ = -0.01, [-0.34, 0.32], $ER_-$ = 1.09). Additionally, we observed no evidence for a difference in BV between the PosFB and TFB conditions (PosFB − TFB: $\beta$ = -0.06, [-0.35, 0.23], $ER_-$ = 2.03), while there was tentative evidence that BV was slightly increased compared to the PFB condition (PosFB − PFB: $\beta$ = 0.20, [-0.08, 0.49], $ER^+$ = 11.53). Similarly, we observed no evidence of a difference in AE between the PosFB and the TFB conditions (PosFB − TFB: $\beta$ = -0.12, [-0.40, 0.16], $ER_-$ = 4.05), contrasting it from the PFB condition (PosFB − PFB: $\beta$ = -0.27, [-0.54, 0.01], $ER_-$ = 31.72), where we found substantial evidence that AE was increased. In line with our hypotheses and results from Study 1, these results indicate that executive performance and BV in the PosFB condition were similar to the those in the TFB condition, resulting in overall reduced performance in the PosFB condition when compared the PFB condition.

Finally, we observed no evidence that self-reported MW developed differently over time in the PosFB condition compared to the TFB condition (PosFB − TFB: $\beta$ = 0.00, [-0.09, 0.08], $ER_-$ = 1.1) or the PFB condition (PosFB − PFB: $\beta$ = -0.03, [-0.11, 0.05], $ER_-$ = 3.53). We also observed no evidence that BV increased over the course of the experiment at a different rate compared to the TFB condition (PosFB − TFB: $\beta$ = 0.01, [-0.11, 0.13], $ER^+$ = 0.74), which, as shown in Study 1, is in the opposite direction compared to the PFB condition. That is, we found tentative evidence that BV was increased in the PosFB condition (PosFB − PFB: $\beta$ = 0.06, [-0.02, 0.15], $ER_-$ = 16.13). Finally, we found no evidence that AE developed differently over time in the PosFB condition compared to the TFB condition (PosFB − TFB: $\beta$ = 0.00, [-0.12, 0.12], $ER_-$ = 1.07), and the PFB condition (PosFB − PFB: $\beta$ = 0.03, [-0.05, 0.11], $ER^+$ = 2.97).

*Relationship between MW and task performance*

Similarly to the results of Study 1, we found substantial evidence that BV was predicting MW in the PosFB condition with the same relationship, i.e. an increase in self-reported MW is predicted by an increase in BV ($\beta$ = 0.19, [0.08, 0.30], $ER^+$ = 1332), and we found no evidence that the strength of the relationship was different from the PFB condition (PosFB − PFB: $\beta$ = 0.01, [-0.15, 0.17], $ER^+$ = 1.17). However, we found weak evidence that this relationship was somewhat stronger than in the TFB condition (PosFB − TFB: $\beta$ = 0.08, [-0.08, 0.25], $ER^+$ = 5.94).

Additionally, the relationship between AE and MW was also similar in the PosFB condition, namely, we found tentative evidence that an increase in MW was preceded by a decrease in AE ($\beta$ = -0.10, [-0.22, 0.02], $ER_-$ = 15.53), and we found no evidence that the strength of the relationship was different compared to the TFB condition (PosFB − TFB: $\beta$ = 0.00, [-0.17, 0.17], $ER^+$ = 1.07), or the PFB condition (PosFB – PFB: $\beta$ = 0.01, [-0.1, 0.16], $ER^+$ = 1.17).

### 4.4. Discussion

Our results support the hypothesis that indiscriminate positive feedback did not affect performance or MW compared to the control condition, where performance feedback was given during the training session. Hence, indiscriminate positive feedback resulted in reduced performance relative to the PFB condition. This finding supports our notion that the improved performance observed in Study 1 for PFB was primarily driven by enhanced task comprehension and/or a motivation derived from mastery orientation through gamification elements. While indiscriminate positive feedback might foster intrinsic motivation through encouragement, previous research suggests that although intrinsic motivation and mastery orientation are interrelated, the relationship between them is not unidirectional (Bieg et al., 2017; Harackiewicz et al., 2008). However, it is important to note that the indiscriminate positive feedback does not reflect real performance and can therefore be misleading as participants are led to believe that even bad performance is adequate. In contrast, our PFB condition implemented actual feedback reflecting performance and was therefore not exclusively positive.

Building on the insight gained from how performance feedback influences performance over time, the next study explored another crucial aspect of task dynamics, the impact of the participants' awareness of time spent in the task. Specifically, we investigated how participants' performance is influenced by providing them with information about the remaining duration of a task.

## 5. Study 3 – Progress bar

### 5.1. Introduction

Research on online surveys indicates that participants commonly express a preference for having access to progression information, where participants often perceive it as a motivating factor that encourages sustained involvement (Villar et al., 2013). In addition, a study manipulating progress bars during online surveys, found that using a fake progress bar did not impact dropout rates. Instead, the dropout rates were contingent on the information conveyed by the progress bar (the speed of the progress bar), highlighting participants' attentiveness to the available progression information (Villar et al., 2013). Therefore, inclusion of information regarding the progress in the experiment might be beneficial for studying MW as a tool to increase task engagement and to prevent detachment from the task over time.

Additionally, in experiments run in a laboratory where experimenter interaction is present, information about how much is left of the experiment is often implicitly or explicitly available. Interestingly, recent studies using the FT-RSGT have noted a slight upturn in performance nearing the end of the experiment, coinciding with participants being reminded that the experiment is about to end (Aasen et al., 2024; Drevland et al., 2024). Therefore, we designed a study to explore whether providing participants with information regarding the remaining time on task, conveyed through a continuously updating progress bar, would influence the allocation of executive resources throughout the experiment. Given that the FT-RSGT is normally void of any visible elements on screen during the task, adding the progress bar involved including a potentially distracting visual element to the task, potentially leading to an increase in self-reported MW as participants spend more time reflecting on the time left in the experiment.

Furthermore, the online version of the FT-RSGT does not provide an opportunity to ask for time remaining of the experiment, unlike the lab version of the task where experimenters are present to answer questions. Therefore, providing participants with task progress information might serve to provide information similar to what can be expected in a laboratory study where experimenters are present. We compared the new sample with that from the TFB condition, since both underwent training under the same conditions (with feedback), while participants in the new condition (ProgFB), were also presented with an updating progress bar throughout the entire experiment. We hypothesized that a progress bar would cause participants to report an overall increase in MW relative to the TFB condition, which would be strongest at the start of the task, before any potential motivation derived from the progress bar would counteract the distraction. Additionally, we hypothesized that BV and AE would be unchanged between conditions.

### 5.2. Method

In the ProgFB condition, a real time updating progress bar var visible at the top of the screen reflecting how much time was left of the experiment. Additionally, similar to the previous studies, participants received performance feedback during the training part of the experiment also in the ProgFB condition.

*Participants*

Out of the 40 participants recruited, upon datafile inspection, three were excluded due to excessive amount of tapping resulting in a

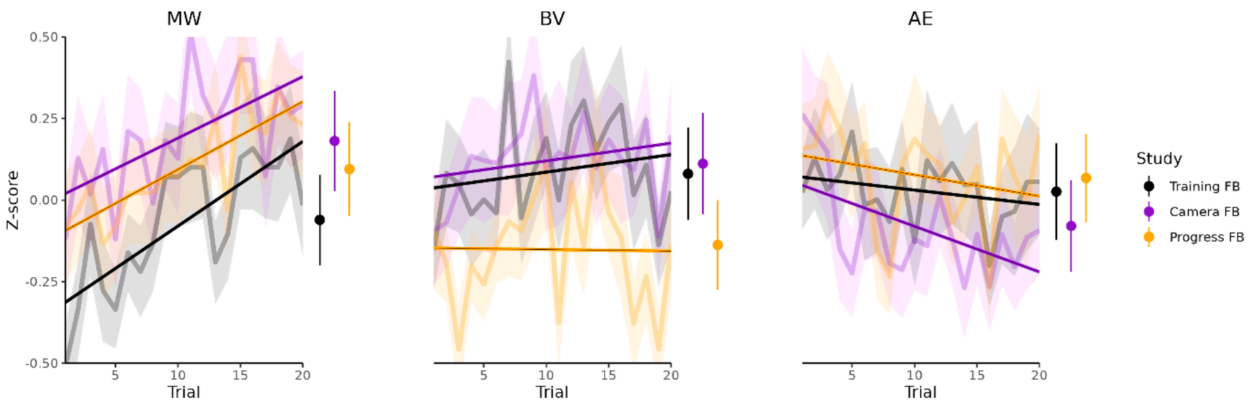


**Fig. 5.** *Differences in AE, BV and MW across conditions as a function of trial number. Marginal effects are depicted on the right side of each plot; error bars and ribbons represent the standard error of the mean. Note:* MW = Mind wandering, BV = Behavioral variability, AE = Approximate Entropy.

final sample of 37. The final sample was aged between 18 and 47 years (Mean = 27.8, SD = 7.46), with 19 males and 18 females.

*Statistical modeling and analysis*

To investigate differences in how progress feedback affects MW and task performance, the new sample was combined and compared with the TFB condition from Study 1. We employed the same analysis plan as in the previous studies.

### 5.3. Results

*Effects of feedback*

The results (shown in Fig. 5) indicate weak evidence that self-reported MW was somewhat higher in the ProgFB condition than the TFB Condition ($\beta = 0.18$, [-0.15, 0.51], $ER^+ = 6.1$), while we found substantial evidence that BV was lower ($\beta = -0.24$, [-0.53, 0.04], $ER_- = 21.04$), and no evidence that AE was different ($\beta = 0.04$, [-0.25, 0.33], $ER^+ = 1.61$). With respect to the temporal evolution of the three indices throughout the task (the effect of Trial), we observed substantial evidence that self-reported MW increased in the ProgFB condition ($\beta = 0.12$, [0.7, 0.18], $ER^+ = 3199$), and no evidence that it developed differently than in the TFB condition (ProgFB − TFB: $\beta = 0.03$, [-0.05, 0.11], $ER_- = 3.47$). We observed no evidence that BV increased with time in the ProgFB condition ($\beta = 0.00$, [-0.6, 0.05], $ER^+ = 0.95$), which means we also observed no evidence that BV developed differently over time compared to the TFB condition (ProgFB − TFB: $\beta = 0.00$, [-0.11, 0.11], $ER_- = 0.94$). However, we did observe tentative evidence suggesting that AE decreased over time in the ProgFB condition ($\beta = -0.04$, [-0.09, 0.02], $ER_- = 11.03$), with no evidence that this decrease was different from the decrease in the TFB condition (ProgFB − TFB: $\beta = 0.02$, [-0.09, 0.14], $ER^+= 0.56$).

*Relationship between MW and task performance*

We also observed substantial evidence that the positive relationship between self-reported MW and BV was present in ProgFB condition ($\beta = 0.13$, [0.03, 0.23], $ER^+ = 284.71$), with no evidence that it was different compared to the TFB condition (ProgFB − TFB: $\beta = 0.03$, [-0.12, 0.18], $ER^+ = 1.98$). Interestingly, we observed no evidence that self-reported MW episodes were preceded by a reduction in AE ($\beta = -0.02$, [-0.13, 0.08], $ER_- = 0.49$), with weak evidence suggesting this relationship was differed from the TFB condition (ProgFB − TFB: $\beta = 0.08$, [-0.08, 0.24], $ER^+ = 5.45$), contrasting it from the findings of Studies 1 and 2.

### 5.4. Discussion

The results indicate that participants who had access to progress information conveyed through an updating progress bar during the experiment generally reported slightly elevated levels of MW, while performance was also increased or unaffected (reduced BV, while AE remained unchanged). One possible explanation for the increase in self-reported MW is that the progress bar itself served as a visual cue, prompting participants to engage in thoughts about remaining time on task. As per our instructions, such thoughts that are normally classified as "task-related interference" (Stawarczyk et al., 2011), would be classified as MW. Furthermore, MW episodes might have been qualitatively different in the progress bar condition compared to the training feedback condition based on our finding that MW episodes were not preceded by reduced AE during ProgFB. In other words, in the presence of the progress bar, participants managed to allocate executive resources to create random patterns while still engaging in MW episodes equally often. This suggests that MW episodes in this study might have been less resource-intensive, leaving more cognitive capacity for random sequence generation. The current study used thought probes that did not differentiate between qualitatively different types off-task states, such as task-related interference or external distraction (e.g., Stawarczyk et al., 2011) and these explanations remain therefore speculative. The main focus of this study was to contrast the effects of contextual factors on on-task vs. off-task states and we therefore kept the thought probes identical across conditions as any changes made to these probes might themselves have led to differences in MW or behavioral performance (for example through changes in *meta*-awareness; (Smallwood et al., 2007; 2008). Future studies should explore this idea further by explicitly manipulating the wording and number of thought probes, to investigate whether these would also manifest in changes to MW and/or behavioral performance.

In line with our results, a recent study investigating cognitive performance explored the impact of manipulating access to remaining time information on task performance (Katzir et al., 2020). Their findings revealed that participants exhibited enhanced performance when provided with task progression information. The authors suggest that this improvement in performance can be attributed to a decrease in the opportunity cost associated with alternative tasks as the current task nears completion. Essentially, as tasks approach their conclusion, the interest in attending to other tasks diminishes, potentially leading to a reduction in the motivation to switch tasks rather than an increase in motivation to complete the current task. Moreover, the study highlights that participants become increasingly sensitive to each step toward task completion as they approach the goal, as each step eliminates more of the total remaining time on the task (Katzir et al., 2020).

In terms of our performance markers in the FT-RSGT, two recent laboratory studies utilizing the FT-RSGT in combination with brain stimulation found the well-established phenomenon of declining performance over time (Aasen et al., 2024; Drevland et al., 2024). However, the researchers observed a noteworthy upturn in performance as well as reduced MW during the final round of the task, coinciding with the participants being notified of the impending end of the experiment. The researchers hypothesized that the observed performance boost could be attributed to a 'surge' in motivation stemming from the imminent completion of the experiment, relieving the need to conserve resources. One key distinction between the previously mentioned laboratory studies and our study is that the laboratory studies utilized four 10-minute blocks, whereas our study employed a single 20-minute block. Despite this difference, as illustrated in Fig. 5, BV exhibits a reverse-U shaped pattern, which is more pronounced in the progress bar condition where information about progress is more readily accessible. This could suggest that the information conveyed by the progress bar resembles the laboratory studies where participants were reminded that the experiment was about to end, just before the last block.

Additionally, in terms of MW research, as time on task increases, vigilance typically diminishes, a phenomenon that results in executive resources being disproportionately allocated to MW rather than the primary task. However, the exact mechanism behind this phenomenon has been widely debated (McVay & Kane, 2010; Smallwood & Schooler, 2006; Thomson et al., 2015). Notably, another recent study examining resource allocation and MW found that participants adeptly allocated resources to MW when the task was predictable, underscoring the role of attention to task progression in resource allocation (Seli et al., 2018). Our results suggest that access to task progression information might protect against the deleterious effects of diminishing vigilance, at least when it comes to BV during the FT-RSGT. We suggest that providing participants with access to remaining time on task information might eliminate uncertainty regarding the necessity for resource reservation, potentially explaining why we did not observe a clear time-on-task effect regarding the BV measure in the progress bar condition. These results align with executive function use hypothesis, as well as the more recent resource control account hypothesis, which proposes that participants can flexibly allocate resources based on task demands and motivation (Smallwood & Schooler, 2006; Thomson et al., 2015).

Our results demonstrate how information about task progress can influence the allocation of executive resources and self-reported MW. This, in turn, affects the ability to adapt performance to meet task demands and influences engagement in MW. A final contextual factor that might vary both between laboratory studies, but also between the laboratory and real life, is the feeling of being monitored. It is often common practice for experimenters to be able to monitor participants during laboratory tasks to ensure participants are complying with task instructions, but this is not routinely done in online experiments. Therefore, our final study aimed to investigate how this feeling of being monitored would affect performance indicators in the FT-RSGT, as well as self-reported MW.

## 6. Study 4 – Camera monitoring

### 6.1. Introduction

Perhaps the most striking difference between online and lab-based studies is the participants' perception of being observed. Early research indicates that the presence of an authoritative figure, such as a scientist or medical professional, can alter participants' behavior and even physiological responses, a phenomenon often referred to as the "white coat effect" (Pickering et al., 2002). Consequently, online studies conducted outside of the traditional laboratory setting lack an experimenter's direct oversight and therefore can introduce differences in participants' behavior. For instance, if participants believe that the experimenter is closely monitoring their performance, they might alter their self-reporting to ensure that their reported task focus more accurately reflects their actual task focus, driven by a fear of being caught in a discrepancy.

However, recent advancements in web-based eye-tracking software have facilitated the utilization of participants' integrated web cameras to exert control during online experimentation (e.g., Papoutsaki et al., 2016). This software not only allow to monitor participants throughout the experiment, but also enables the monitoring of participants' gaze. We hypothesized that the introduction of web-cam based eyetracking would introduce the feeling of being monitored, simulating the experience of being in a laboratory. Furthermore, we hypothesized that this feeling of being watched would heighten participants' self-reported MW, as they would likely engage more frequently in reflections about their performance or the purpose behind the monitoring. To explore this hypothesis, we incorporated "sham" camera monitoring to mimic the sensation of being observed, similar to a laboratory setting. Each participant proceeded with the experiment after aligning their face within a displayed camera monitor box on their screen, ostensibly for eye-tracking equipment calibration. Notably, no data from the camera were recorded either during or after these adjustments even though the camera light was on during the complete duration of the experiment.

### 6.2. Methods

To simulate the feeling of being monitored, we employed the WebGazer software (Papoutsaki et al., 2016) to simulate a fake eye tracking calibration. Before the study commenced, participants were shown their own face on the screen and instructed to calibrate the camera for eye tracking. This involved looking at various parts of the screen, mimicking the standard procedure for eye-tracking setup (Papoutsaki et al., 2016). Additionally, the task script also turned on the integrated camera-LED if it was available on the participants computer. Similarly to the previous studies, participants received performance feedback during the training part of the experiment.

### 6.3. Participants

Out of the 40 new participants recruited, one was excluded for pausing in the middle of the experiment. The final sample (N = 39) was aged from 18 to 49 years (M = 26.1, SD = 6.63), with 31 males and 8 females.

*Statistical modeling and analysis*

To investigate differences how the sense of being monitored through a camera affects indices of MW, a sample of 37 new participants were combined and compared with the TFB condition. We used the same rationale for analysis plan and model selection as in Studies 1, 2 and 3, with the only difference being the new CamFB condition.

### 6.4. Results

*Effects of feedback*

The results suggest tentative evidence that self-reported MW was higher in the CamFB condition compared to the TFB condition

(CamFB – TFB: $\beta = 0.26$, [-0.08, 0.60], $ER^{+} = 14.61$), however, unlike in Study 3, we did not observe any evidence that this increase in self-reported MW was accompanied by a decrease in BV (TFB − CamFB: $\beta = 0.04$, [-0.28, 0.35], $ER^{+} = 1.45$). Our analysis also suggested no evidence that AE was different in the CamFB condition compared to the TFB condition ($\beta = -0.12$, [-0.42, 0.19], $ER_{-} = 3.46$). In the CamFB condition, as in the previous studies, we observed substantial evidence that MW increased as time on task increased ($\beta = 0.11$, [0.05, 0.16], $ER^{+} = 3199$), yet we found weak evidence that this increase over time was not as strong as in the TFB condition (CamFB − TFB: $\beta = -0.04$, [-0.12, 0.4], $ER_{-} = 5.14$). Additionally, we also found weak evidence that BV increased over the course of the experiment in the CamFB condition ($\beta = 0.03$, [-0.03, 0.9], $ER^{+} = 6.1$), with no evidence that this increase over time was different in the CamFB condition compared to the TFB condition (CamFB – TFB: $\beta = -0.04$, [-0.16, 0.07], $ER_{-} = 3.67$). Finally, we also observed substantial evidence that AE was reduced over time in the CamFB condition ($\beta = -0.08$, [-0.14, −0.03], $ER_{-} = 560.40$), resulting in no evidence that AE developed differently over time in CamFB compared to the TFB condition (CamFB – TFB: $\beta = -0.01$, [-0.12, 0.10], $ER_{-} = 1.39$). For a graphical overview of these results see Fig. 5.

*Relationship between MW and task performance*

As in Study 3, we found substantial evidence that BV predicted MW in the CamFB condition ($\beta = 0.12$ [0.01, 0.23], $ER^{+} = 63.52$) with no evidence that the strength of the relationship was different compared to the TFB condition ($\beta = -0.01$ [-0.17, 0.14], $ER_{-} = 1.33$). Additionally, similar to Study 3, we did not observe any evidence that self-reported MW was preceded by reduced AE in the CamFB condition ($\beta = 0.00$, [-0.11, 0.11], $ER^{+} = 1$), with weak evidence suggesting that this relationship was different from the TFB condition (CamFB − TFB: $\beta = 0.10$ [0.06, −0.26], $ER^{+} = 8.05$).

### 6.5. Discussion

Our results indicate that introducing a fake camera to induce the feeling of being monitored increased self-reported MW. Inducing a feeling of being monitored can be viewed as an “internal” distraction, which can potentially lead to an increase in *meta*-cognitive, evaluative processes (e.g., “Am I performing as expected?”) potentially introducing more MW. Interestingly however, this increase in self-reported MW was not accompanied by a reduction in performance. Therefore, it is possible that the observed increase in MW is merely a product of more candid reporting, where participants were afraid of being caught in a lie, as they believe they are being closely monitored. This could explain why performance is similar to the TFB condition while self-reported MW is increased. However, the expected relationship where MW reports are preceded by a reduction in AE was not present in the CamFB condition. One possible explanation is that the sensation of being watched may have triggered more “intrusive” MW episodes. This could have prevented participants from choosing when to engage in MW, resulting in episodes that occurred independently of performance. Thus, in contrast to Study 1 and 2, the sensation of being monitored might not have allowed participants to strategically schedule their MW episodes during periods of low AE, specifically when they were less motivated to maintain performance.

Nevertheless, our results suggest that introducing the feeling of being monitored could serve as a tool to increase the reliability of self-reports in MW research. However, at this time, this remains a speculation based on the observed pattern of the results. Most importantly, our results suggest that self-reported MW tends to increase in contexts where participants feel that they are being watched. Therefore, the perception of being monitored could act as a confounding factor in replicating MW research and should be carefully considered during study design.

## 7. General discussion

Overall, our findings from four studies investigating the impact of various contextual factors on MW and task performance indicate that both self-reported MW and objective measures of executive performance are sensitive to contextual influences. More specifically, providing participants with direct performance feedback enhanced task performance without altering self-reported task focus. We suggest that performance feedback might increase motivation and fosters an increased comprehension of task requirements in the FT-RSGT.

However, the fact that the increase in performance in the gamified version of the task (PFB condition) was not accompanied by reduced self-reported MW suggests that the improvement was most likely not due to increased motivation. It seems improbable that participants were more motivated to solve the task in the PFB condition while still reporting the same level of task focus. This interpretation is further supported by findings from Study 2, which suggest that encouragement alone does not enhance performance, as evidenced by the unchanged performance in the PosFB condition. Therefore, while our results do not provide definitive conclusions, it is more plausible that performance feedback enhanced available executive resources by reducing task difficulty, rather than by boosting motivation. Nevertheless, our results support the resource-control account proposed by Thomson et al. (2015), which suggests that executive resources can be flexibly allocated to either MW or task demands. In our study, it appears that the extra resources made available by reduced task demands were likely allocated towards further enhancing performance.

Furthermore, the availability of task progression information also improved performance in terms of tapping precision. We suggest that access to progress information likely alleviated the need to reserve resources and reduced the opportunity cost of switching to other tasks (Katzir et al., 2020), which primarily affected the easier aspect of performance (i.e., tapping precision, but not sequence randomness). This effect was noticeable as the FT-RSGT approached completion, where the typical decline in performance appears to have been counteracted by the presence of the progress bar.

Interestingly, enhancement in performance during ProgFB occurred alongside slightly increased levels of self-reported MW. We suggest that a potential increase in self-reported MW may be attributed to the information conveyed by the progress bar, which could serve as a distracting element that steers thoughts toward more task-unrelated content, thereby increasing MW. Two recent studies found that participants who received performance feedback during a vigilance task reported more task-related inference (Robison et al., 2021; Unsworth et al., 2024). Consistent with these findings, our findings indicate that such guided thoughts (e.g., wondering how many minutes are left, or how you are currently performing), which can be classified as task-related yet stimulus-independent, might be less detrimental to performance than self-generated spontaneous MW (Stawarczyk et al., 2011).

However, our version of the FT-RSGT does not differentiate between such task-related interference thoughts and task-unrelated thoughts, requiring participants to classify both as MW. Future studies should aim to disentangle these different types of off-task cognition by incorporating more detailed thought probes to determine whether the increase in MW is due to task-related interference or task-unrelated MW.

Introducing the sense of being monitored through sham camera monitoring also increased self-reported MW. A possible explanation for the increase in self-reported MW is that the induced notion of being monitored might have introduced task-unrelated thoughts that are more centered around the self, resembling the content of typical self-generated MW episodes (Smallwood & Schooler, 2015). These self-centered MW episodes may have seemed more unrelated to the experiment than to other task-related interference thoughts, which could explain the elevated levels of self-reported MW in the camera monitoring condition.

On the other hand, while the notion of being monitored increased reports of MW, it did not lead to a change in performance. Participants might have assumed that their performance was closely monitored, thereby changing the way they reported their task focus. It is possible that the sense of being monitored did not actually increase MW, but instead, it removed false reports of task focus. This could explain why performance remained unchanged despite elevated levels of self-reported MW. However, as in the progress bar condition, we did not observe the expected relationship of reduced executive functioning (reflected by AE scores) preceding reports of MW. Therefore, it is more likely that MW episodes were qualitatively different, and not just a result of a different response strategy to the thought probes. We propose that internal distractions, such as contemplating the remaining time or questioning the purpose of being monitored, might encourage more intrusive and spontaneous MW episodes. Consistent with prior research on selective MW (Seli et al., 2018), if participants experienced more intrusive episodes of MW, they might not have been able to strategically schedule MW episodes during periods of low task engagement characterized by low AE. Consequently, this could lead to MW episodes occurring more randomly throughout the task, disrupting the typically stable relationship of MW episodes being preceded by reduced AE. It is possible that the relationship between AE and MW requires scheduled periods of low task engagement (deliberate MW and low effort behavior), which we commonly find in the FT-RSGT. In the presence of more spontaneous episodes of MW such as brief distractions causing by a progress bar or camera monitoring, which might resemble task-related interference, this otherwise stable relationship between AE and MW is weakened or completely abolished.

Our results indicate that in the presence of distractors, which may induce task-related interference or elevated levels of task-unrelated thought, executive performance on the FT-RSGT does not reliably predict MW episodes. In Studies 1 and 2, where executive performance was predictive of self-reported MW, we did not observe elevated levels of MW. Consequently, it remains uncertain whether the relationship between AE and MW is influenced by heightened levels of self-reported MW, or if it is disrupted by qualitatively different MW episodes.

The intricate relationship between performance indicators and self-reports underscores the complexity of MW and emphasize the importance of employing multiple indicators in conjunction to capture behavioral markers of MW. Additionally, it is crucial to consider the context in which participants are placed. Overall, these studies illuminate the nuanced interplay between MW, task performance and contextual factors, offering valuable insights that will inform future research designs and interpretations in the field of MW research, particularly when conducted online.

## 8. Conclusion

When attempting to capture the elusive phenomenon of the internal train of thoughts, caution must be exercised regarding underlying assumptions. Minor contextual shifts can yield significant impacts on both probe-caught MW as well as its behavioral indicators. Online studies offer effective means of data collection beyond the confines of the traditional laboratory setting, with performance feedback serving as a viable alternative to direct experimenter interaction as well as shielding against the negative effects of MW on overall cognitive performance. Moreover, mind wandering episodes are notably associated with heightened BV and diminished AE during the FT-RSGT, with the latter relationship dissipating in the presence of visual distractors and/or when feeling of being monitored.

## CRediT authorship contribution statement

**Andreas Alexandersen:** Writing – review & editing, Writing – original draft, Validation, Investigation, Formal analysis. **Krister Dahlberg:** Investigation. **Gábor Csifcsák:** Writing – review & editing, Validation, Supervision, Methodology. **Matthias Mittner:** Writing – review & editing, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Conceptualization.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.concog.2025.103870.

## Data availability

The data that support the findings of this study are available on OSF (https://osf.io/wjvk2/).

## References


Aasen, S. R., Drevland, R. N., Csifcsák, G., & Mittner, M. (2024). Increasing mind wandering with accelerated intermittent theta burst stimulation over the left dorsolateral prefrontal cortex. *Neuropsychologia, 204*, Article 109008. https://doi.org/10.1016/j.neuropsychologia.2024.109008

Alexandersen, A., Csifcsák, G., Groot, J., & Mittner, M. (2022). The effect of transcranial direct current stimulation on the interplay between executive control, behavioral variability and mind wandering: A registered report. *Neuroimage: Reports, 2*(3), Article 100109. https://doi.org/10.1016/j.ynirp.2022.100109

Antrobus, J. S., Singer, J. L., Goldstein, S., & Fortgang, M. (1970). Section of Psychology: mindwandering and cognitive structure. *Transactions of the New York Academy of Sciences, 32*(2 Series II), 242–252. https://doi.org/10.1111/j.2164-0947.1970.tb02056.x

Appel, C., & Gilabert, R. (2002). Motivation and task performance in a task-based web-based tandem project. *ReCALL, 14*(1), 16–31. https://doi.org/10.1017/S0958344002000319

Astleitner, H. (2000). Designing emotionally sound instruction: the FEASP-approach. *Instructional Science, 28*(3), 169–198. https://doi.org/10.1023/A:1003893915778

Baddeley, A., Emslie, H., Kolodny, J., & Duncan, J. (1998). Random generation and the executive control of working memory. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology, 51*(4), 819–852. https://doi.org/10.1080/713755788

Batanero, C. (2015, February). Understanding randomness: Challenges for research and teaching. *CERME 9 - Ninth Congress of the European Society for Research in Mathematics Education*. Charles University in Prague, Faculty of Education. Prague, Czech Republic. https://hal.science/hal-01280506.

Beck, T., Bühren, C., Frank, B., & Wiederhold, E. (2020). Can honesty oaths, peer interaction, or monitoring mitigate lying? *Journal of Business Ethics, 163*. https://doi.org/10.1007/s10551-018-4030-z

Belletier, C., Davranche, K., Tellier, I. S., Dumas, F., Vidal, F., Hasbroucq, T., & Huguet, P. (2015). Choking under monitoring pressure: Being watched by the experimenter reduces executive attention. *Psychonomic Bulletin & Review, 22*(5), 1410–1416. https://doi.org/10.3758/s13423-015-0804-9

Bieg, S., Reindl, M., & Dresel, M. (2017). The relation between mastery goals and intrinsic motivation among university students: A longitudinal study. *Educational Psychology, 37*(6), 666–679. https://doi.org/10.1080/01443410.2016.1202403

Blascovich, J., Mendes, W. B., Hunter, S. B., & Salomon, K. (1999). Social "facilitation" as challenge and threat. *Journal of Personality and Social Psychology, 77*(1), 68–77. https://doi.org/10.1037/0022-3514.77.1.68

Boayue, N. M., Csifcsák, G., Kreis, I. V., Schmidt, C., Finn, I., Hovde Vollsund, A. E., & Mittner, M. (2020). The interplay between executive control, behavioural variability and mind wandering: Insights from a high-definition transcranial direct-current stimulation study. *European Journal of Neuroscience, 53*(5), 1489–1516. https://doi.org/10.1111/ejn.15049

Brooks, S. (1998). Markov chain Monte Carlo method and its application. *Journal of the Royal Statistical Society: Series D (The Statistician), 47*(1), 69–100. https://doi.org/10.1111/1467-9884.00117

Burgers, C., Eden, A., van Engelenburg, M. D., & Buningh, S. (2015). How feedback boosts motivation and play in a brain-training game. *Computers in Human Behavior, 48*, 94–103. https://doi.org/10.1016/j.chb.2015.01.038

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software, 80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Cheyne, J. A., Solman, G. J. F., Carriere, J. S. A., & Smilek, D. (2009). Anatomy of an error: A bidirectional state model of task engagement/disengagement and attention-related errors. *Cognition, 111*(1), 98–113. https://doi.org/10.1016/j.cognition.2008.12.009

Csikszentmihalyi, M., Abuhamdeh, S., & Nakamura, J. (2005). Flow. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (598–608). Guilford Publications.

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating amazon's mechanical turk as a tool for experimental behavioral research. *PLOS ONE, 8*(3), Article e57410. https://doi.org/10.1371/journal.pone.0057410

Dandurand, F., Shultz, T. R., & Onishi, K. H. (2008). Comparing online and lab methods in a problem-solving experiment. *Behavior Research Methods, 40*(2), 428–434. https://doi.org/10.3758/BRM.40.2.428

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification". Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. Association for Computing Machinery, New York, NY. https://doi.org/10.1145/2181037.2181040.

Drevland, R. N., Aasen, S. R., Csifcsák, G., & Mittner, M. (2024). Reducing mind wandering using continuous theta burst stimulation. Preprint. Open Science Framework. https://doi.org/10.31234/osf.io/u5j7s.

Earley, P., Northcraft, G., Lee, C., & Lituchy, T. (1990). Impact of process and outcome feedback on relation of goal setting to task performance. *Academy of Management Journal, 33*. https://doi.org/10.2307/256353

Feenstra, H. E. M., Vermeulen, I. E., Murre, J. M. J., & Schagen, S. B. (2017). Online cognition: Factors facilitating reliable online neuropsychological test results. *The Clinical Neuropsychologist, 31*(1), 59–84. https://doi.org/10.1080/13854046.2016.1190405

Gabry, J., Češnovar, R., & Johnson, A. (2023). *cmdstanr: R Interface to "CmdStan"* [Computer software]. https://mc-stan.org/cmdstanr/.

Groot, J. M., Csifcsák, G., Wientjes, S., Forstmann, B. U., & Mittner, M. (2022). Catching wandering minds with tapping fingers: Neural and behavioral insights into task-unrelated cognition. *Cerebral Cortex, 32*(20), 4447–4463. https://doi.org/10.1093/cercor/bhab494

Harackiewicz, J. M., Durik, A. M., Barron, K. E., Linnenbrink-Garcia, L., & Tauer, J. M. (2008). The role of achievement goals in the development of interest: Reciprocal relations between achievement goals, interest, and performance. *Journal of Educational Psychology, 100*(1), 105–122. https://doi.org/10.1037/0022-0663.100.1.105

Kane, M. J., Brown, L. H., McVay, J. C., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007). For whom the mind wanders, and when: An experience-sampling study of working memory and executive control in daily life. *Psychological Science, 18*(7), 614–621. https://doi.org/10.1111/j.1467-9280.2007.01948.x

Kane, M. J., Gross, G. M., Chun, C. A., Smeekens, B. A., Meier, M. E., Silvia, P. J., & Kwapil, T. R. (2017). For whom the mind wanders, and when, varies across laboratory and daily-life settings. *Psychological Science, 28*(9), 1271–1289. https://doi.org/10.1177/0956797617706086

Katzir, M., Emanuel, A., & Liberman, N. (2020). Cognitive performance is enhanced if one knows when the task will end. *Cognition, 197*, Article 104189. https://doi.org/10.1016/j.cognition.2020.104189

Klinger, E., & Cox, W. M. (1987). Dimensions of thought flow in everyday life. *Imagination, Cognition and Personality, 7*(2), 105–128. https://doi.org/10.2190/7K24-G343-MTQW-115V

Krimsky, M., Forster, D. E., Llabre, M. M., & Jha, A. P. (2017). The influence of time on task on mind wandering and visual working memory. *Cognition, 169*, 84–90. https://doi.org/10.1016/j.cognition.2017.08.006

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

Lange, K., Kühn, S., & Filevich, E. (2015). "Just Another Tool for Online Studies" (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLOS ONE, 10*(6), Article e0130834. https://doi.org/10.1371/journal.pone.0130834

de Leeuw, J. R., Gilbert, R. A., & Luchterhandt, B. (2023). jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software, 8*(85), 5351. https://doi.org/10.21105/joss.05351

Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology, 79*, 328–348. https://doi.org/10.1016/j.jesp.2018.08.009

McVay, J. C., & Kane, M. J. (2010). Does mind wandering reflect executive function or executive failure? Comment on Smallwood and Schooler (2006) and Watkins (2008). *Psychological Bulletin, 136*(2), 188–207. https://doi.org/10.1037/a0018298

McVay, J. C., Kane, M. J., & Kwapil, T. R. (2009). Tracking the train of thought from the laboratory into everyday life: An experience-sampling study of mind wandering across controlled and ecological contexts. *Psychonomic Bulletin & Review, 16*(5), 857–863. https://doi.org/10.3758/PBR.16.5.857

Mooneyham, B. W., & Schooler, J. W. (2013). The costs and benefits of mind-wandering: A review. *Canadian Journal of Experimental Psychology, 67*(1), 11–18. https://doi.org/10.1037/a0031569

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). WebGazer: Scalable webcam eye tracking using user interactions. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence - IJCAI 2016*. https://par.nsf.gov/biblio/10024076-webgazer-scalable-webcam-eye-tracking-using-user-interactions.

Park, S. W. (2017). *Motivation theories and instructional design*. https://pressbooks.pub/lidtfoundations/chapter/motivation-in-lidt-by-seungwon-park/.

Pavlova, M. K. (2024). A dual process model of spontaneous conscious thought. *Consciousness and Cognition, 118*, Article 103631. https://doi.org/10.1016/j.concog.2023.103631

Pickering, T. G., Gerin, W., & Schwartz, A. R. (2002). What is the white-coat effect and how should it be measured? *Blood Pressure Monitoring, 7*(6), 293.

Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences, 88*(6), 2297–2301. https://doi.org/10.1073/pnas.88.6.2297

Reips, U. D. (2000). The Web Experiment Method: Advantages, disadvantages, and solutions. In: Birnbaum, M.H., *Psychological experiments on the Internet*. San Diego, CA: Acadmic Press, 89-118. https://doi.org/10.5167/uzh-19760.

Reips, U. D. (2002). Standards for internet-based experimenting. *Experimental Psychology, 49*, 243–256. https://doi.org/10.1026//1618-3169.49.4.243

Robison, M. K., & Unsworth, N. (2018). Cognitive and contextual correlates of spontaneous and deliberate mind-wandering. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(1), 85–98. https://doi.org/10.1037/xlm0000444

Robison, M. K., Unsworth, N., & Brewer, G. A. (2021). Examining the effects of goal-setting, feedback, and incentives on sustained attention. *Journal of Experimental Psychology: Human Perception and Performance, 47*(6), 869–891. https://doi.org/10.1037/xhp0000926

Ryan, R., & Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American Psychologist, 55*, 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Sailer, M., Hense, J. U., Mandl, H., & Klevers, M. (2013). Psychological perspectives on motivation through gamification. *Interaction Design and Architecture(s) Journal, 19*, 18–37. https://doi.org/10.55612/s-5002-019-002

Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior, 69*, 371–380. https://doi.org/10.1016/j.chb.2016.12.033

Saravanos, A., Zervoudakis, S., Zheng, D., Stott, N., Hawryluk, B., & Delfino, D. (2021). The hidden cost of using amazon mechanical turk for research. *HCI International 2021—Late Breaking Papers: Design and User Experience* (147–164). Springer International Publishing. https://doi.org/10.1007/978-3-030-90238-4_12.

Sauter, M., Draschkow, D., & Mack, W. (2020). Building, hosting and recruiting: A brief introduction to running behavioral experiments online. *Brain Sciences, 10*(4). https://doi.org/10.3390/brainsci10040251

Seli, P., Carriere, J., 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*. https://www.frontiersin.org/article/10.3389/fpsyg.2013.00430.

Seli, P., Carriere, J. S. A., Wammes, J. D., Risko, E. F., Schacter, D. L., & Smilek, D. (2018). On the clock: Evidence for the rapid and strategic modulation of mind wandering. *Psychological Science, 29*(8), 1247–1256. https://doi.org/10.1177/0956797618761039

Seli, P., Cheyne, J. A., & Smilek, D. (2013). Wandering minds and wavering rhythms: Linking mind wandering and behavioral variability. *Journal of Experimental Psychology: Human Perception and Performance, 39*(1), 1–5. https://doi.org/10.1037/a0030954

Seli, P., Schacter, D. L., Risko, E. F., & Smilek, D. (2019). Increasing participant motivation reduces rates of intentional and unintentional mind wandering. *Psychological Research, 83*(5), 1057–1069. https://doi.org/10.1007/s00426-017-0914-2

Seli, P., Smallwood, J., Cheyne, J. A., & Smilek, D. (2015). On the relation of mind wandering and ADHD symptomatology. *Psychonomic Bulletin & Review, 22*(3), 629–636. https://doi.org/10.3758/s13423-014-0793-0

Skinner, B. F. (1963). Operant behavior. *American Psychologist, 18*(8), 503–515. https://doi.org/10.1037/h0045185

Smallwood, J., & Andrews-Hanna, J. (2013). Not all minds that wander are lost: The importance of a balanced perspective on the mind-wandering state. *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00441

Smallwood, J., Brown, K., Baird, B., & Schooler, J. W. (2012). Cooperation between the default mode network and the frontal–parietal network in the production of an internal train of thought. *Brain Research, 1428*, 60–70. https://doi.org/10.1016/j.brainres.2011.03.072

Smallwood, J., McSpadden, M., & Schooler, J. W. (2007). The lights are on but no one's home: Meta-awareness and the decoupling of attention when the mind wanders. *Psychonomic Bulletin & Review, 14*(3), 527–533. https://doi.org/10.3758/BF03194102

Smallwood, J., McSpadden, M., Luus, B., & Schooler, J. (2008). Segmenting the stream of consciousness: The psychological correlates of temporal structures in the time series data of a continuous performance task. *Brain and Cognition, 66*(1), 50–56. https://doi.org/10.1016/j.bandc.2007.05.004

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

Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering: Empirically navigating the stream of consciousness. *Annual Review of Psychology, 66*(1), 487–518. https://doi.org/10.1146/annurev-psych-010814-015331

Stan Development Team, 2016. RStan: The R interface to Stan. http://mc-stan.org/.

Stawarczyk, D., Majerus, S., Maquet, P., & D'Argembeau, A. (2011). Neural correlates of ongoing conscious experience: Both task-unrelatedness and stimulus-independence are related to default network activity. *PLOS ONE, 6*(2), Article e16997. https://doi.org/10.1371/journal.pone.0016997

Teasdale, J. D., Segal, Z., & Williams, J. M. G. (1995). How does cognitive therapy prevent depressive relapse and why should attentional control (mindfulness) training help? *Behaviour Research and Therapy, 33*(1), 25–39. https://doi.org/10.1016/0005-7967(94)E0011-7

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: A Journal of the Association for Psychological Science, 10*(1), 82–96. https://doi.org/10.1177/1745691614556681

Unsworth, N., Robison, M. K., & Miller, A. L. (2024). Mobilizing effort to reduce lapses of sustained attention: Examining the effects of content-free cues, feedback, and points. *Cognitive, Affective, & Behavioral Neuroscience.*. https://doi.org/10.3758/s13415-024-01254-0

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved R-hat for assessing convergence of MCMC (with Discussion). *Bayesian Analysis, 16*, 667–718. https://doi.org/10.1214/20-BA1221

Villar, A., Callegaro, M., & Yang, Y. (2013). Where am I? A meta-analysis of experiments on the effects of progress indicators for web surveys. *Social Science Computer Review, 31*(6), 744–762. https://doi.org/10.1177/0894439313497468

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review, 25*(1), 35–57. https://doi.org/10.3758/s13423-017-1343-3

Zanesco, A. P., Denkova, E., & Jha, A. P. (2024). Mind-wandering increases in frequency over time during task performance: An individual-participant meta-analytic review. *Psychological Bulletin*. https://doi.org/10.1037/bul0000424