Computers in Human Behavior 112 (2020) 106453

Contents lists available at ScienceDirect

Computers in Human Behavior

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

Full length article

Wandering minds, wandering mice: Computer mouse tracking as a method to detect mind wandering

M.R. Dias da Silva *, M. Postma

Tilburg University Cognitive Science and Artificial Intelligence Department, Warandelaan 2, 5037AB Tilburg, The Netherlands

A R T I C L E I N F O

Keywords:
Mouse tracking
Mind wandering
Attention
Embodied cognition
Working memory
Arousal

A B S T R A C T

Mind wandering is a state in which an individual's attention is decoupled from the task at hand. Mind wandering affects performance in many tasks requiring focused attention, including online learning. Previous studies have focused primarily on investigating mind wandering in contexts which are conducive to mind wandering, that is, for highly repetitive and monotonous tasks or during tasks with very low attentional demands. However, mind wandering also occurs during highly engaging and demanding tasks. In this study, we examine whether mouse tracking can be used to predict mind wandering in an engaging task involving classical computer interfaces. Assuming that mouse trajectories towards a particular response on the screen are continuously updated by time-dependent and temporally-dynamic cognitive processes, as a behavioral methodology, mouse tracking can provide unique insights into attentional processes. In our experiment, a total of 272 students completed a mouse-based operation span task, during which their thoughts were probed and their mouse movements recorded. Naive Bayes, Linear Discriminant Analyses, K-Nearest Neighbors, Tree Bag, and Random Forest classifiers were able to predict mind wandering with F1-scores of up to 15% above a random-chance baseline. The results show that hand reach movements can be tracked to detect mind wandering in a user-independent manner in online tasks, thus providing a viable alternative to self-report methods and (neuro)physiological measures. Our finding has relevant implications for a variety of user interfaces which require hand and finger movements for the purposes of human–computer interactions.

1. Introduction

There is an ever-going conflict between pieces of information that compete for our critical attentional resources. While navigating the world around us, we must allocate attentional resources to both our external environment and to our internal thoughts and feelings (Smallwood, 2013; Smallwood & Schooler, 2015) as we seek out meaning from the stimuli around and within us. Indeed, an essential part of our human interaction with the world involves the constant interplay between externally-oriented attention and internally-oriented attention, commonly referred to as mind wandering (Mills, Herrera-Bennett, Faber, & Christoff, 2018; Smallwood & Andrews-Hanna, 2013). Often conceptualized as a decoupling of attention from the here and now towards internal thoughts and feelings (Smallwood & Schooler, 2015), mind wandering has proven to be important for planning and goal-setting (Klinger, 2013), creating an integrated sense of identity (Smallwood & Andrews-Hanna, 2013) and fostering creativity (Baird et al., 2012). However, it can also be detrimental to performance in a wide variety of contexts.

From a population point of view, the inability to regulate adaptively both the frequency and content of mind wandering thoughts under different contexts has been associated with a variety of disorders (Andrews-Hanna, Smallwood, & Spreng, 2014; Watkins, 2008). Such disorders of content tend to have negative consequences for cognitive functioning and well-being in both clinical and non-clinical populations. Excessive negative thoughts are characteristic of depression (Marchetti, Van de Putte, & Koster, 2014), while excessively grandiose and positive thoughts may be characteristic of mania (Watkins, 2008). Forms of thinking that are too focal could be indicative of autism, while forms of thought that are too scattered could be reflective of Attention Deficit Disorder (Franklin, Mooneyham, Baird, & Schooler, 2014).

With respect to the activities during which mind wandering episodes may occur, they seem to be particularly frequent during tasks that require sustained attention and engage working memory (Mrazek et al., 2012; Randall, Oswald, & Beier, 2014; Unsworth & Robison, 2016). Individual differences in working memory have been found to predict performance on a wide range of measures, from low-level attention

* Corresponding author.
E-mail address: m.r.diasdasilva@tilburguniversity.edu (M.D.d. Silva).

https://doi.org/10.1016/j.chb.2020.106453
Received 14 March 2020; Received in revised form 3 June 2020; Accepted 8 June 2020
Available online 12 June 2020

tasks to higher level reasoning (Cowan et al., 2005). A distinction is made between resource limited tasks – in which using focused attentional resources is necessary for performance, versus data – limited tasks, in which investment of attentional resources is irrelevant for performance (Randall et al., 2014). Very easy or very difficult[1], as well as highly practiced, well learned tasks would be data-limited and leave more room for mind wandering than resource-limited tasks do. According to this view, individuals with higher working memory capacity would be better able to regulate their attentional resources, such that in more demanding, complex tasks, they would mind wander less; while in easier tasks, they would mind wander more (Randall et al., 2014; Smallwood, 2013). As such, there is a U-shaped relationship between mind wandering and task demands, as mind wandering occurs more frequently in low-demand and in high demand tasks than in moderately demanding tasks, with higher levels of working memory being associated with a reduced likelihood to mind wander as task demands increase. Intuitively, mind wandering is more harmful to performance of high- relative to low-demand tasks, and more so for individuals with lower working memory capacity.

The negative association between mind wandering and performance, particularly on demanding tasks, warrants the need to develop technologies that are able help individuals to down-regulate the content and frequency of their thoughts whenever necessary (Faber, Bixler, & D'Mello, 2017). In order to reduce the negative effects of mind wandering, it is important to be able to identify when individuals are mind wandering. The challenge in studying mind wandering, however, stems from the fact that it cannot easily be induced in laboratory settings (Dias da Silva, Rusz, & Postma-Nilsenová, 2018; McVay & Kane, 2013). Moreover, its detection relies on self-reports, by means of experience sampling methods such as thought probes and retrospective measures, which are inherently subjective (Smallwood & Schooler, 2015). Compared to thought probes, retrospective measures are useful in that they do not interrupt the natural flow of the task. However, retrospective measures are only able to make general estimations about the total frequency of mind wandering on a task, while thought probes are better at pinpointing specific instances of mind wandering within a task. Past studies show that to a certain degree, both are subject to incorrect estimations from participants (Seli, Carriere, Levene, & Smilek, 2013; Smallwood & Schooler, 2006). Because of this, additional behavioral measures such as reaction times (McVay & Kane, 2009), reading speed (Mills, D 'mello, Bosch, & Olney, 2015), fidgeting (Carriere, Seli, & Smilek, 2013; Seli et al., 2014), and (neuro)physiological responses such as brain activity and eye movements (Faber et al., 2017; Franklin, Broadway, Mrazek, Smallwood, & Schooler, 2013; Mittner et al., 2014; Smallwood, Brown, Tipper, Giesbrecht, Franklin, Mrazek, et al., 2011), have been explored to distinguish periods of focused attention (FA) from periods of mind wandering (MW). These neuro(physiological) approaches, however, interfere with the natural performance on a primary task (Grimes & Valacich, 2015) in that they require additional measuring instruments, which introduce a certain level of discomfort for the participant. In this paper, we examined mouse movement behavior as a method that could be used to detect the occurrence of mind wandering unobtrusively.

1.1. Mouse movements as a behavioral measure of mind wandering

It has been shown that in relation to various domains, including decision making, attention, and learning (Freeman, Dale, & Farmer, 2011; Grimes & Valacich, 2015; Lins & Schöner, 2019; Papesh & Goldinger, 2012), computer mouse tracking effectively traces the evolution of internal cognitive processes through action execution. As a natural and practiced visuo-motor response, various populations, from young children (Hermens, 2018) to older adults (Seelye et al., 2015) can easily perform mouse-based tasks. In a typical mouse tracking paradigm, alternative choices can be represented in front of a participant, and the evolution of reach trajectories towards a target can be visualized as a representation of how competing cognitive states are resolved over time (Song & Nakayama, 2009). A mouse is a primary means of interacting with computers in everyday tasks, ranging from writing email to navigating a page while reading an article. As almost all computers are equipped with mice, measuring mouse movements is affordable and widely accessible, and can be effectively run as a background process during mouse-based tasks.

Can computer mouse movements predict the occurrence of mind wandering? Previous research was able to predict engagement – a concept often contrasted with mind wandering – from mouse features in an unsupervised manner (Arapakis, Lalmas, & Valkanas, 2014). There is also evidence of a direct link between changes in hand movement behavior during mind wandering. Kam et al. (2012) found increased tracking errors during periods of mind wandering relative to periods of focused attention in a simple visuomotor ball tracking task. They describe such errors to be a consequence of attenuated sensory processing. Alternatively, changes in hand movements during mind wandering may be explained by embodied cognition theory, which suggests a variety of cognitive activities are reflected in bodily states, such as posture, arm and hand movements (Barsalou, 2008). In particular, the situated action view of embodied cognition assumes a close coupling of perception and action during goal achievement (Barsalou, 2008). It posits that the way we move our body, how we are standing, or what we are touching or holding can both provide information about and also influence the way that we feel, think about or evaluate a situation. Thus, from an embodied cognition perspective, bodily movements can be viewed as extensions of cognitive and attentional processes. As such, mind wandering can be better described as an embodied experience which is shaped by movements.

As extensions of attentional processes, it is plausible that hand reach movements with a computer mouse may be able to reflect episodes of mind wandering. Research on mind wandering and attention indicates that periods of off-task thought are associated with changes in arousal (Robison & Unsworth, 2019; Unsworth & Robison, 2017), that is, changes in physiological activation which indicates responsiveness to sensory stimulation (Eysenck, 1982). More specifically, both high or low levels of arousal are related to lower attentional control, more lapses in attention, and consequently, a greater susceptibility to mind wandering. Meanwhile moderate levels of arousal are associated with optimal task engagement and task performance (Cohen, Aston-Jones, & Gilzenrat, 2004; Kahneman, 1973; Lenartowicz, Simpson, & Cohen, 2013; Mittner, Hawkins, Boekel, & Forstmann, 2016; Yerkes & Dodson, 1908). Differences in the arousal associated with mind wandering and the accompanying physiological signature likely reflect distinct stages of the experience. For example, mind wandering may begin during an underarousing, monotonous circumstance. As we try extricate ourselves from this, arousal levels rise, reaching a peak when our efforts to engage in stimulating activities fail, accompanied by feelings of restlessness (Danckert, Hammerschmidt, Marty-Dugas, & Smilek, 2018). In line with this, pupillary measures suggest mind wandering to be associated with changes in arousal states (Unsworth & Robison, 2017, 2018).

Interestingly, high arousal has been associated with decreased fine motor control and increases in neuromotor noise during hand reach movements (Grimes, Jenkins, & Valacich, 2013), while low arousal has been associated with automatism (Kahneman, 1973; Morsella, Larson, & Bargh, 2010). Considering the relationship between mind wandering and arousal, and the influence of arousal on motor control, we could expect that mind wandering episodes would be associated with changes in motor behavior necessary to move the computer mouse. We consider two models in order to describe changes in motor control with computer mouse movements during mind wandering: the *stochastic optimized submovement (SOS) model* and the *response activation model*. According to the *SOS model* (Meyer, Abrams, Kornblum,

[1] Such that the level of difficulty is beyond one's ability to complete a task.

Wright, & Smith, 1988, 1990) mouse movements towards a target are described as having two parts — an initial high-velocity phase, which although fast, tends to be imprecise, and a subsequent deceleration phase, which is corrective in nature, where speed decreases, but accuracy increases (Graham & MacKenzie, 1996; Grimes & Valacich, 2015). When a target is approached, a tradeoff in speed is necessary to increase precision of movement, as there is limited information capacity for motor control (Fitts, 1954). The mind attempts to minimize the total movement by optimizing the velocity and number of submovements towards the target; however, as neuromotor noise (i.e., from high arousal) is introduced into the model, there are fewer resources available for the intended corrective movements, leading to slower and less precise movements (Meyer et al., 1988, 1990; van Beers, 2004).

Complementary to the *SOS* model, the *response activation model* describes what happens to motor movements as additional alternative targets are introduced. It proposes that motor movements represent an aggregation of all potential movements that could arise from all potentially actionable cognitions (Welsh & Elliott, 2004). When competing cognitions are introduced, motor movements become less precise and response times slower, reflecting disruptions in fine motor control as necessary cognitive resources are consumed (Hick, 1952).

While the *SOS model* describes what happens during movements towards one particular target, the *response activation model* also takes into account the influence of competing cognitions (for instance, debating whether to choose between two or more answers to a question). Both the predictions concerning the speed–accuracy trade off and arousal introduced by the *SOS model* and predictions concerning competing cognitions introduced by the *response activation model* are relevant for predicting changes in motor control during mind wandering. As competing cognitions are introduced by mind wandering, mouse movements would likely change depending on the state of mind wandering. They could reflect an underaroused state, or alternatively, a restless state. As such, consequent changes in arousal likely increase the amount of noise and uncertainty during a task, potentially leading to slower, less precise, more complex and more variable computer mouse movements.

To sum up, the challenge in studying mind wandering arises both from the over-reliance on self-reports as well as in intrusive neuro(physiological) measures which interfere with the natural performance of a task. As such, it is important to find objective measures for the detecting mind wandering that corroborate self-reports and do not interrupt the flow of a task. Used to continuously track a variety of cognitive processes, mouse-tracking seems to be a promising candidate for objectively detecting mind wandering.

1.2. Current study

The focus of the current study is to explore if mind wandering during a complex cognitive task (a working memory test, i.e., an operation span task) can be detected from mouse movements. Due to their ability to capture cognitive processes in real time, computer mouse movements may actually provide valuable insight into the temporal cognitive dynamics underlying mind wandering. Research has shown that the adverse consequences of mind wandering are greater in tasks with higher cognitive demand (e.g.working memory tasks; McVay & Kane, 2012; Mrazek et al., 2012; Rummel & Boywitt, 2014; Smallwood & Andrews-Hanna, 2013). During the task (Fig. 1), participants need to shift between an unrelated processing task while updating contents of working memory (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Engle, Laughlin, Tuholski, & Conway, 1999; Unsworth & Engle, 2005; Unsworth, Redick, Lakey, & Young, 2010). Specifically, the evolution of mouse trajectories can be traced during the processing portion of the operation span task. Consolidating previous research on mind wandering and arousal, as well as mouse movements and arousal, we will explore whether various mouse movement measures can be indicators of mind wandering. In this paper, we will examine how computer mouse movement features are related to distracted thought. We will train various classifiers to predict differences in participants' locus of attention. Previous mind wandering studies have distinguished between focused attention (FA), task-unrelated thought (TUT), and task-related interference (TRI). During task-unrelated thought, thought content is irrelevant to the task at hand. During task-related interference, thought content involves a preoccupation with performance on the task at hand (Matthews et al., 1999). The latter two categories reflect different types of self-generated thought and involve a decoupling of attention from the external environment towards internal thoughts and feelings. Assuming that mouse trajectories towards a particular response on the screen are continuously updated by cognitive processes, we expect that mind wandering will be evident in computer mouse movements during a working memory task. Building on the *SOS model*, we propose mind wandering to be a source of neuromotor noise, leading to slower and less precise corrective mouse movements. According to the *response activation model*, computer mouse movements should become slower and more erratic as more choices are introduced into the model. Building on this, we would expect that during mind wandering, additional internally oriented cognitions would also compete with external choices. Changes in arousal and consequent disruptions in motor control would lead to slower, less precise, more complex and more variable hand reach movements.

2. Methods

2.1. Participants

In total, 274 participants, recruited from the university student pool, 180 female, 17 to 41 years of age ($M = 22.09$, $SD = 3.25$),[2] took part in this experiment in exchange for course credit. The sample size was selected on the basis of sample sizes reported in previous studies (Mrazek et al., 2012; Yamauchi & Xiao, 2017). Two participants were excluded because the experimental environment crashed, leaving 272 remaining participants in our analyses. All subjects were native Dutch speakers who could use a computer mouse. The study was approved by the Tilburg University Ethics Committee (identification code: REC#2017/06), and informed consent was obtained from each participant at the beginning of the experimental session. After signing the consent form, participants filled out a questionnaire collecting their demographics and completed the Operation Span Task, which took on average 20 min to complete. Standard procedure was followed for the Operation Span (OSPAN) (see Fig. 1) task (Conway et al., 2005).

2.2. Material

The task required participants to maintain access to memory items (letters) while completing an unrelated processing task (math equations) with an individualized response deadline ($M+2.5SD$), calculated during 15 processing-task-only items (Unsworth, Heitz, Schrock, & Engle, 2005). This allowed each participant to set their own pace and ensured that they did not rehearse the to-be-recalled letters by limiting the amount of time they had to solve the math operations. Participants viewed a compound math equation on the computer screen, and once they had solved it, they were instructed to click on the start button. If participants took longer than their average time + $2.5SD$ to click on the start button, the trial was marked as an error. On the center of the next screen, they saw a number, as well as a TRUE and FALSE box on the top left corner and top right corner of the screen, respectively. If the number they saw corresponded to the correct answer to the math equation, participants were instructed to click on the TRUE button, and if not, on the FALSE button. A capital letter appeared for 1000 ms

[2] Mean age for females = 21.60, $SD = 3.26$, and mean age for males = 23, $SD = 3.03$.

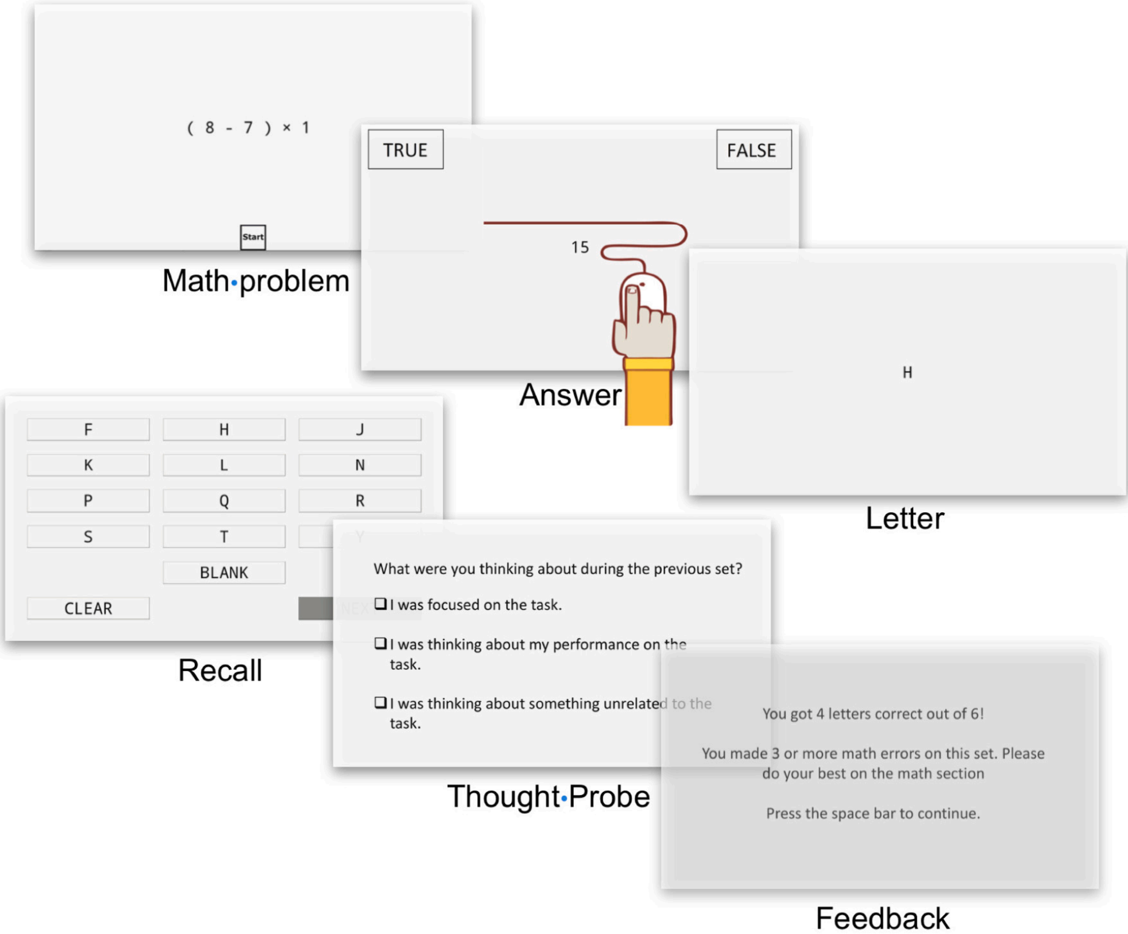

Fig. 1. Operation span task.

after the math operation. After 3–7 compound equations–letter pairs, all 12 letters appeared on the screen and participants were required to identify (by clicking) the letters that were presented in the trial in serial order. Each set length (3–7) was presented 3 times, randomly ordered for each participant, for a total of 75 trials (15 sets). Mind wandering was assessed by thought probes embedded throughout the task after each set (Mrazek et al., 2012). In order to prevent participants from devoting all their processing time to remembering the letters, they were instructed to aim for an accuracy of at least 85% on the math operations,[3] so that letter recall would not come at the expense of performance on the math operations. The program calculated the sum of all correctly recalled set sizes (OSPAN score), the total number of letters recalled in the correct position, and the total number of math errors (Unsworth et al., 2005). At the end of the task, participants received feedback concerning their performance.

2.3. Instrumentation

The Operation Span Task was programmed on Opensesame (Mathôt, Schreij, & Theeuwes, 2012), version 3.1.6, using a modified version of the script provided by Eoin Travers.[4] The experiment was run in full screen mode on a P2210 Dell monitor, 22 inch (55.88 cm), with a resolution of 1366 by 768 pixels on a Windows 7 operating system. The desktop computer was placed on a table so that enough space was available to move the mouse around without hitting the keyboard or the edge of the table. Mouse settings were left at their default values (acceleration on and medium speed). A Dell USB 3 Button Scrollwheel wired Optical Mouse was used to record cursor coordinates for the math verification portion of the experiment. There was enough space available for participants to move the mouse without hitting the keyboard or the edge of the table.

Mouse movements were recorded during the math verification part of the task towards one of two alternatives (TRUE or FALSE, on the uppermost right and left sides of the screen). Upon clicking on the start button, mouse movements started to be recorded. The dimensions of the TRUE and FALSE buttons were of 279 by 157 pixels, and dimensions of the start button were of 80 x 80 pixels. Cursor coordinates were recorded every 30 ms. To ensure that any effects were not due to the direction of movement, we reversed the positions of the TRUE and FALSE buttons for 88[5] (out of 272) participants.[6]

[3] In the standard operation span task, participants who do not answer at least 85% of all math questions correctly are excluded from analysis. Although we still instructed participants to aim to achieve at least 85% accuracy, the exclusion criterion was irrelevant for the purposes of this study, and we thus kept all participants for analysis.

[4] https://github.com/EoinTravers/QuickstartMousetracking.

[5] Originally 90 (out of 274), but two participants were excluded due to a procedural error.

[6] We designed our experiment according to original implementation of the OSPAN, in which TRUE is displayed on the left of the screen and FALSE on the right side. However, after collecting data from 184 participants, we deemed it necessary to collect more data with the location of the FALSE button now on the left and the location of the TRUE button on the right to ensure that any mouse movement effects found were not due to the location of the response buttons.

In the instructions, participants were informed that after each set, they would be asked a question about their thoughts during the previous set. They were also informed that it is normal for people's minds to wander off task or to thoughts about their performance on the task. After each set, participants were asked, *What were you thinking about during the previous set?*, and had to choose from 3 alternatives, namely, *(1) I was focused on the task, (2) I was focused on my performance on the task*, and *(3) I was thinking about something unrelated to the task.* Alternative 1 denoted all instances in which participants were focused on the task; alternative 2 denoted all instances in which participants experienced task-related interference; and alternative 3 denoted all instances in which participants experienced task-unrelated thought (Robison, Miller, & Unsworth, 2019; Stawarczyk, Majerus, Maj, Van der Linden, & D'Argembeau, 2011).

2.4. Data processing

Individual raw data files were merged and read into R version 3.4.1 (R Core Team, 2013). Because of the individualized response times during the processing part of the OSPAN task, trials in which participants took longer than their average time to click on the start button were discarded. If participants took longer than their average time to complete the math response, the experiment would issue an error message and move on to the following trial (429 trials out of 20400 trials). As the end coordinates of such trials were not comparable to the rest of the trials, we did not include them in our analyses. Mouse tracking data were then imported and processed using the library "mousetrap" (Kieslich, Henninger, Wulff, Haslbeck, & Schulte-Mecklenbeck, 2019) on R. Trajectories were measured from the moment the start button was pressed to the moment either the TRUE or FALSE response was clicked on. If participants took longer than 10 s to select an alternative, mouse coordinates were not recorded. This happened 3 times. All trajectories aligned to a common starting position and were remapped onto one side, and various measures were computed for each trajectory.

2.5. Class imbalance

In order to observe the distribution of answers to the mind wandering probes, we calculated the percentage of probes in which participants responded to be focused, to be having task-related interference, and to be having task-unrelated thought. Proportions of focused attention (FA), task-related interference (TRI) and task-unrelated thought (TUT) were 0.69, 0.23, and 0.09, respectively (Fig. 2).

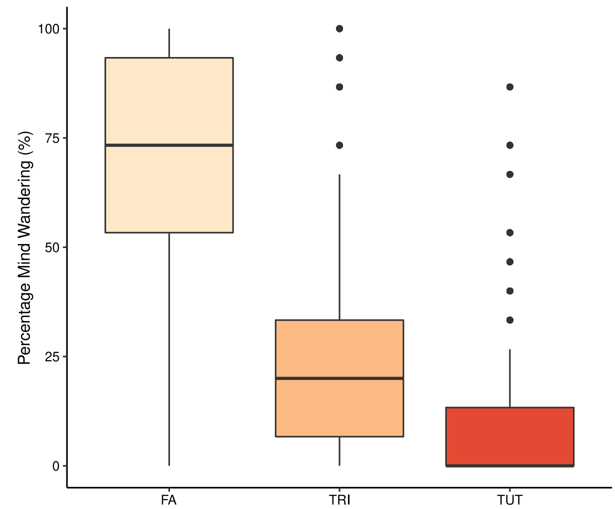

Fig. 2. Distribution of mind wandering scores.

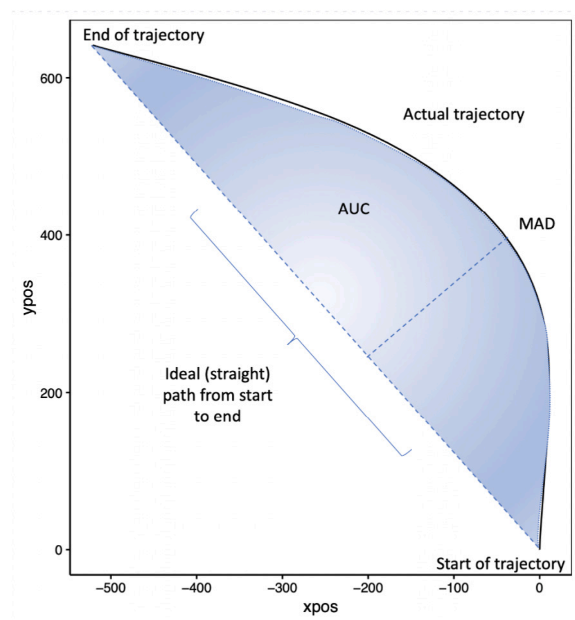

Fig. 3. Visual representation of trajectory measures.

2.6. Extracting features

We extracted 27 mouse features (Kieslich & Henninger, 2017) from the x and y-coordinates recorded for each participant. As sets after which thought probes were placed varied from 3 (easiest) to 7 trials (hardest) in size, we also included set size as a contextual feature in our analyses (Mrazek et al., 2012). Descriptive statistics of features are presented in Table 1. Features were first aggregated per participant before calculation of the mean and standard deviation. *MAD* refers to the signed maximum absolute deviation connecting the direct path between the start and end point of the trajectory (straight line) (Fig. 3). If the *MAD* occurs above the direct path, it has a positive value; if it occurs below, then a negative value. *MAD time* refers to the time at which the maximum absolute deviation was first reached. *MD above* and *MD below* refer to the maximum deviation above and below the direct path, respectively, while *MD above time* and *MD below time* refer to the time at which the maximum deviation above and below was reached first, respectively. *AD* refers to the average deviation from the direct path. *AUC* (Fig. 3) refers to the geometric area between the actual trajectory and the direct path (where areas below the direct path have been subtracted). *x-* and *y-pos flips* refer to the number of directional changes along x- and *y*-axis, respectively. *x* and *y-pos reversals* refer to the number of crossings of the x- and *y*-axis, respectively.

All features were aggregated per set (15 sets per participant, varying in size from 3 to 7 trials), yielding a total of 4080 instances.

2.7. Dimensionality reduction

Pearson's correlations (Fig. 4) between the mouse-tracking features indicate that some features may be measuring nearly identical underlying constructs (e.g. *MAD* and *MD above*, $r = 0.99$, *vel-max time* and *acc-min time*, $r = 0.96$, *total dist* and *x-pos max* $r = 0.90$).

Therefore, PCA (Fig. 5) was used to reduce the dimensionality of the data, removing any multicollinearity. Four components were used that cumulatively accounted for 26%, 49%, 61%, and 68% of the variance in the mouse-tracking data, respectively.

Analysis of the components on the basis of the pattern matrix (Table 2) indicates that the first component (spatial) concerns the space and path covered, indicating deviations from the optimal trajectory, the

Table 1
Unstandardized means and standard deviations of mouse tracking features and set size for each class (FA — focused attention, TRI — task-related interference, TUT — task-unrelated thought).

Features[a]	FA		TRI		TUT	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
x-pos max	135.30	70.29	138.36	73.08	151.79	92.36
x-pos min	−531.58	20.64	−533.09	24.28	−539.14	27.67
y-pos max	655.10	14.45	655.30	16.06	654.00	16.66
y-pos min	−0.44	1.23	−0.36	1.09	−0.42	1.23
MAD	280.24	112.25	282.32	112.41	300.91	141.95
MAD time	559.81	146.36	597.51	183.51	628.99	202.97
MD above	290.52	108.20	292.21	108.62	312.15	133.81
MD above time	537.05	140.07	577.84	179.27	604.97	210.59
MD below	−19.90	11.48	−19.25	13.48	−22.21	16.95
MD below time	445.83	150.33	481.89	239.95	521.80	252.19
AD	82.71	40.25	83.68	41.22	88.37	46.32
AUC[b] x 10^3	109.00	39.00	111.00	42.00	117.00	51.00
x-pos flips	1.48	0.48	1.50	0.55	1.54	0.66
y-pos flips	0.90	0.45	0.92	0.52	0.97	0.84
x-pos reversals	0.84	0.27	0.84	0.31	0.89	0.41
y-pos reversals	0.06	0.11	0.06	0.14	0.07	0.17
RT	1104.96	273.81	1174.78	322.65	1224.99	412.23
initiation time	195.39	91.76	206.08	107.33	221.85	107.90
idle time	380.30	212.13	432.66	251.67	459.71	320.02
total dist	1218.71	203.69	1223.86	195.08	1267.82	260.48
vel max	6.43	1.01	6.45	1.14	6.47	1.18
vel max time	587.86	167.52	630.69	205.12	656.09	221.68
acc max	0.10	0.02	0.10	0.02	0.10	0.03
acc max time	545.63	163.04	590.49	207.60	621.96	234.80
acc min	−0.10	0.02	−0.10	0.02	−0.10	0.03
acc min time	635.16	163.01	677.91	203.20	710.68	221.35
sample entropy	0.10	0.03	0.10	0.03	0.10	0.03
set size	4.82	0.35	5.33	0.89	5.46	1.08

[a] All time related values are presented in milliseconds (ms), all position related values are presented in pixels (px), area (AUC) is displayed in px^2, and all speed related variables are presented in pixels/ms.

[b] Exact values for AUC *Mean* and *SD* for each class are displayed respectively. (FA)*Mean*: 109,243.80 px^2, *SD*: 38,620.08 px^2, (TRI) *Mean*: 110,536.80 px^2, *SD*: 41,572.50 px^2, and (TUT) *Mean*: 117,172.10 px^2, and *SD*: 50,747.45 px^2.

Table 2
PCA pattern matrix with values for the highest loading component, with ultimate cutoff point of 0.35.

Variables	PC1	PC2	PC3	PC4
MAD	0.94			
MD above	0.92			
AD	0.92			
AUC	0.88			
x-pos max	0.87			
total dist	0.75			
sample entropy	0.70			
x-pos reversals	0.65			
MD below	0.48			
y-pos flips	0.48			
MAD time		0.92		
vel max time		0.90		
acc max time		0.89		
acc min time		0.88		
MD above time		0.88		
RT		0.86		
idle time		0.80		
initiation time		0.61		
MD below time		0.58		
vel max			0.84	
acc min			−0.81	
acc max			0.77	
x-pos min			−0.67	
y-pos max			0.35	
y-pos reversals				0.79
y-pos min				−0.76
x-pos flips				0.38

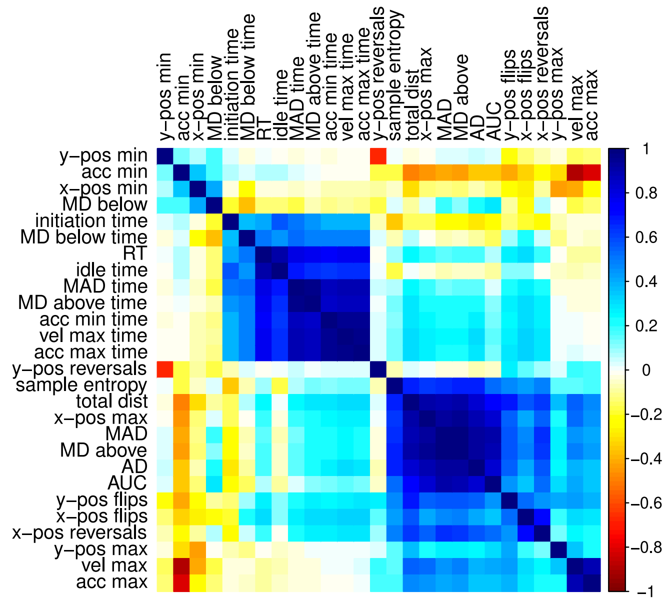

Fig. 4. Correlations between mouse tracking variables.

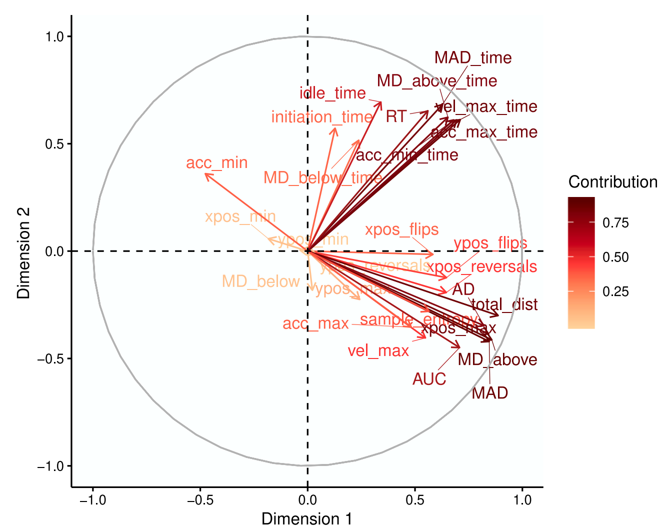

Fig. 5. Plot of variables contributing to PCA components. Positively correlated variables point to the same direction in the plot. Negatively correlated variables point to opposite directions in the plot.

second (time-related) consists primarily of variables that have to do with the temporal occurrence of cursor movements, the third (speed-related) consists of velocity and acceleration of movements, and finally, the fourth component (y-positions) consists of primarily y-position related features.

2.8. Training and testing

We used a *leave-one-participant-out* cross-validation procedure, where eighty percent of participants were in the training/validation set, and 20% of participants in the test set. A *leave-one participant-out* cross-validation procedure is used to train classifiers in a user-independent manner (Yatani & Truong, 2012) and has been previously applied for predicting mind wandering (Pham & Wang, 2015). Each model was trained on $N - 1$ participants, and one participant was held out for validation, for N folds, where N is the number of participants in the training/validation set. This was then compared to a corresponding random chance baseline[7] for all three classes (FA — focused attention, TRI — task-related interference and TUT — task-unrelated thought) separately.

2.9. Model building

Several studies have been able to predict mind wandering with machine learning classification methods in a user-independent fashion from behavioral and (neuro)physiological data. Using eye movement features (Bixler & D'Mello, 2016), trained 20 different machine learning algorithms and were able to predict mind wandering best using a Naive Bayes classifier. Pham and Wang (2015) also used various classifiers to detect mind wandering in online lectures from heart rate features and found the k-nearest neighbor classifier to perform best. Zhang and Kumada (2017) trained 6 machine learning algorithms and found Linear Discriminant Analysis and AdaBoost to predict mind wandering best using driving behavior information as features. Although various classifiers have been used to predict mind wandering from different behavioral and physiological features and in different contexts, there is not yet a consensus concerning which one is best. Therefore, we trained 5 classifiers which have been previously used in predicting mind wandering (NB — Naive Bayes, KNN — K-Nearest Neighbor, LDA — Linear Discriminant Analysis, TB — Tree Bag, and RF — Random Forest) on the recorded features in order to predict whether participants were focused on the task at hand (FA), were having task-related interference (TRI), or were having task-unrelated thought (TUT). We consider these different classifiers because we have no a priori prediction about the type of model that is best suited for this classification task. We use both the standardized mouse tracking features (z-score standardized after aggregating per participant) as well as PCA-reduced features as input for different models (Zhang & Kumada, 2017).

2.10. Data preparation

Participants reported being focused on 69% of trials, to be having task-related interference on 23% of trials and to be focused on 9% of trials. Considering the level of demand and engagement required to complete the operation span task, it is not surprising that participants spent the majority of the time focused on the task. Because of this, we observed a bias towards predictions of the majority class with all classifiers (Table 3). In addition to the class imbalance, there was also a considerable difference in the amount of participants who reported having task-unrelated thought (TUT) or task-related interference (TRI). Out of the 272 participants, only 107 responded at least once to all three mind wandering probes. In order to remove the imbalance in responses across participants present in the full data set, we performed further analyses on the 107 participants who had variation in responses for all 3 classes. Furthermore, we randomly undersampled the majority class(es) to match the instances in the minority class, leaving one observation for each class (3 per participant), for a total of overall 321 observations ($3x107$). We used a nested cross-validation procedure for the balanced data, where an inner CV loop was used for training/validation and an outer loop to compute a generalized estimate of performance. Eighty percent of participants were in the training/validation set (inner loop), and 20% of participants in the test set. We performed a *leave-one-participant-out* cross validation on the training/validation set (inner loop). Each model was trained on $N-1$ participants, and one participant was held out for validation, for N folds. Model performance was then assessed on the test data. In order to estimate how well each model generalizes, we repeated the cross-validation process 15 times with random[8] training/validation and test split. We then obtained an average accuracy over the 15 iterations (outer loop). This was then compared to a corresponding random chance baseline.[9]

3. Results

We report the classification performance in terms of overall accuracy of all classes (*Acc.*), followed by the *F1 score* (balanced average of precision and recall), *Sensitivity* (true positive rate, also known as recall), *Specificity* (true negative rate), and the *Balanced Accuracy* (representing an average between *Sensitivity* and *Specificity*) for each class.

Results from the *leave-one-participant-out* cross validation procedure on the imbalanced data ($N_{total} = 4080$, $FA = 2805$, $TRI = 923$, $TUT = 352$) are displayed in Table 3. When testing the models on unseen data, we observe above chance overall accuracies for all models. As expected, all classifiers were able to predict FA above a random chance baseline, indicating a clear bias towards prediction of the majority class, as high *Sensitivities* are accompanied by low *Specificities*. We also found that the *F1* and *Sensitivity* scores for most classifiers predicting task-unrelated thought (TUT) and task-related interference (TRI) fall below a random chance baseline, with *Specificities* above chance. Only the Naive Bayes (NB) classifier was able to predict task-unrelated thought (TUT) above a random chance baseline ($F1 = 0.14$, $Sensitivity = 0.11$), while no classifiers were able to predict task-related interference (TRI) above chance levels. We found nearly identical predictions for models using PCA components as features. As such, we do not display them here.

3.1. Removing class imbalance

Once removing the influence of class imbalance, we see a considerable improvement in performance measures. More specifically, results from the nested cross-validation on the matched data ($N = 321$) indicate that K-Nearest Neighbors (KNN) and Random Forest (RF) were able to predict task-unrelated thought (TUT) best ($F1 = 0.41$, $Sensitivity = 0.44$ for KNN; $F1 = 0.37$, $Sensitivity = 0.39$ for RF). All classifiers predicted focused attention (FA) above chance level ($F1$ and $Sensitivity > 0.30$). However, no classifiers were able to predict task-related interference (TRI) well (see *F1* and *Sensitivity* scores for TRI in Table 4).

When using PCA components as features, all models performed above a random chance accuracy baseline (Table 5). We also see a particular improvement in the *Balanced Accuracies*, indicating an improvement in both *Sensitivity* and *Specificity* for all classes, such that all values were above chance (except for LDA — Linear Discriminant Analysis when predicting TRI — task-related interference). NB — Naive Bayes, TB — Tree Bag, and RF — Random Forest were the best classifiers, with overall accuracies (*Acc.*) of 0.47, 0.42, 0.40, respectively. *F1* scores (for each model per class) are displayed in Fig. 6.

The *F1* scores for these top 3 classifiers (Naive Bayes, Tree Bag, Random Forest) for each class (TUT — task-unrelated thought, TRI — task-related interference, and focused attention) were as follows: TUT (0.47, 0.46, 0.39), TRI (0.44, 0.39, 0.42), and FA (0.48, 0.41, 0.37), respectively. Baseline *F1* scores for task-unrelated thought, task-related interference and focused attention were 0.32, 0.33, and 0.30, respectively. Both the Linear Discriminant Analysis and K-Nearest Neighbors classifiers predicted task-related thought and focused attention well,

[7] A random chance baseline was computed by randomly sampling classes from the test set.

[8] We used random splits of the data because of the relatively few observations after removing class imbalance. Note that there was some overlap between the test sets in each of the 15 iterations (Mean overlap of 19% ($SD = 7\%$).

[9] A random chance baseline was obtained by randomly sampling classes from the test set. This was also averaged over 15 iterations.

Table 3
Performance of multiple models on matched data using 27 mouse tracking features and set size for each class (task-unrelated thought (TUT), task-related interference (TRI), and focused attention (FA); N = 4080). Above chance performance is presented in bold font.

	Acc.	*F1*			*Sensitivity*			*Specificity*			*Balanced Acc.*		
		TUT	TRI	FA	TUT	TRI	FA	TUT	TRI	FA	TUT	TRI	FA
Random	.54	.07	.22	.69	.07	.22	.69	.92	.78	.31	.50	.50	.50
NB	**.63**	**.14**	.14	**.78**	**.11**	.10	**.89**	**.95**	**.93**	.19	**.53**	**.51**	**.54**
LDA	**.66**	.02	.01	**.79**	.01	.01	**.99**	**1.00**	**.99**	.01	**.51**	.50	.50
KNN	**.64**	.00	.14	**.78**	.00	.10	**.94**	**1.00**	**.93**	.11	.50	**.51**	**.52**
TB	**.64**	.00	.13	**.78**	.00	.08	**.93**	**1.00**	**.93**	.08	.50	**.51**	**.51**
RF	**.66**	.00	.01	**.79**	.00	.01	**.99**	**1.00**	**.99**	.00	.50	.50	.49

Table 4
Performance of multiple models on matched data using 27 mouse tracking measures and set size as features (N = 321). Above chance performance is presented in bold font.

	Acc	*F1*			*Sensitivity*			*Specificity*			*Balanced Acc.*		
		TUT	TRI	FA	TUT	TRI	FA	TUT	TRI	FA	TUT	TRI	FA
Random	.32	.32	.33	.30	.32	.33	.30	.66	.66	.65	.49	.50	.47
NB	**.38**	**.34**	**.35**	**.42**	.30	**.34**	**.50**	**.78**	**.69**	.60	**.54**	**.52**	**.55**
LDA	**.34**	**.33**	.29	**.40**	.32	.28	**.43**	**.69**	**.70**	.63	**.51**	.49	**.53**
KNN	**.37**	**.41**	.31	**.39**	**.44**	.30	**.39**	.66	**.71**	**.69**	**.55**	.50	**.54**
TB	**.34**	**.33**	.32	**.36**	.32	.32	**.37**	**.68**	.64	**.68**	.50	.48	**.53**
RF	**.35**	**.37**	.31	**.37**	**.39**	.29	**.38**	.65	**.71**	**.68**	**.52**	.50	**.53**

Table 5
Performance of multiple models on matched data using 4 PCA components and set size as features (N = 321). Above chance performance is presented in bold font.

	Acc	*F1*			*Sensitivity*			*Specificity*			*Balanced Acc.*		
		TUT	TRI	FA	TUT	TRI	FA	TUT	TRI	FA	TUT	TRI	FA
Random	0.32	.32	.33	.30	.32	.33	.30	.66	.66	.65	.49	.50	.47
NB	**0.47**	**.47**	**.44**	**.48**	**.50**	**.41**	**.48**	**.69**	**.77**	**.73**	**.60**	**.59**	**.61**
LDA	**0.37**	**.40**	.18	**.45**	**.42**	.14	**.54**	.65	**.82**	.58	**.54**	.48	**.56**
KNN	**0.38**	**.39**	**.35**	**.39**	**.40**	.33	**.40**	**.68**	**.74**	.65	**.54**	**.54**	**.53**
TB	**0.42**	**.46**	**.39**	**.41**	**.47**	**.39**	**.40**	**.71**	**.71**	**.71**	**.59**	**.55**	**.56**
RF	**0.40**	**.39**	**.42**	**.37**	**.38**	**.44**	**.38**	**.74**	**.67**	**.69**	**.56**	**.56**	**.53**

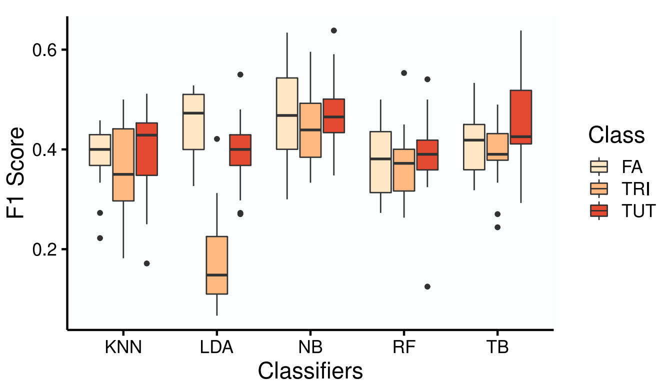

Fig. 6. Performance of different models with PCA components as predictors.

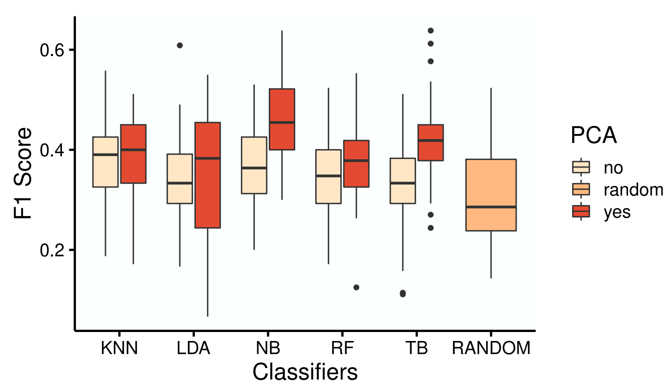

Fig. 7. Comparison of performance of models with standardized mouse tracking features versus with PCA components.

but not task-related interference. Specifically, the Linear Discriminant Analysis classifier predicted task-related interference below chance level, while K-Nearest Neighbors predicted task-related interference just at chance level.

When comparing the performance of models with standardized mouse tracking features to models with PCA components (Fig. 7), we observe that reducing the dimensionality of the data leads to improvements in *F1* performance, especially for Naive Bayes (NB), Random Forest (RF), and Tree Bag (TB).

3.2. Importance of features

We explore the importance of features in predicting mind wandering on the balanced data using the Random Forest algorithm, which is suitable for selecting a large number of features with few observations (Yamauchi, 2013). Random Forest performs ensemble learning, as 500 or more decision trees are formed by randomly selecting observations and variables. It then estimates likelihoods of the dependent variable with the importance scores of features. We repeat this procedure across the 15 training/validation sets and average the importance

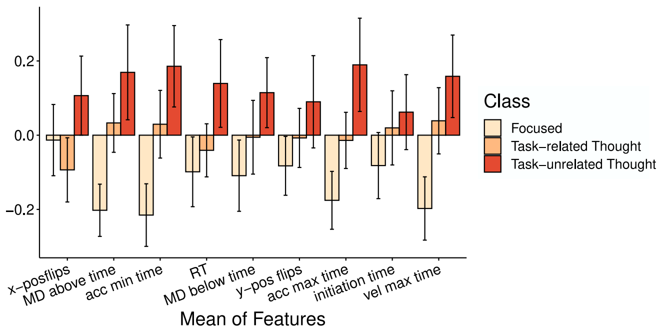

Fig. 8. Z-score standardized means of top ten features for predicting task-unrelated thoughts.

of features across the different folds. Results from this procedure (Table 6) indicate that the top five features pertain primarily to task difficulty (set size), followed by *vel max time, MD below time, RT*, and *MD above time*. Out of these top ten features, we observe that 7 pertain to the time to reach maximum velocity, maximum (and minimum) acceleration, and maximum deviation (both above and below). Time to reach maximum velocity provides information about the onset of cognitive processes involved in committing towards a response (Hehman, Stolier, & Freeman, 2015). Time to reach maximum deviation reflects the time to reach the greatest distance from the ideal trajectory and is positively correlated with the time to reach maximum velocity and acceleration ($r = .89$). Therefore, these features provide valuable information concerning the influence of mind wandering on selection of an alternative during the math processing part of the OSPAN task. The remaining last 2 features (*total distance* and *x-pos reversals*) of the top ten pertain to the space and distance covered by trajectories. These features provide important information concerning how mind wandering may influence the complexity of mouse movements, as demonstrated by differences in trajectory paths.

When exploring the direction of features in predicting task-unrelated thought (Fig. 8), we observe more *x-position flips*, longer *time to reach maximum deviation* and *maximum acceleration*, longer *RTs*, more *y-position flips*, longer *time to reach maximum acceleration* and *velocity*, and longer *initiation times* during task-unrelated thoughts (see Fig. 8). If we compare the difference in means in Fig. 8 to the means of features in Table 1, we notice consistency between mean values for the entire dataset and values in the downsampled data. The top mouse tracking feature for predicting task unrelated thought, *x-position flips*, were the number of times participants moved the mouse cursor along the x-axis (Dale, Roche, Snyder, & McCall, 2008). More *x-position flips* represent increased complexity of movement along the decision axis, indicating more changes of mind. Next in importance were *time to reach maximum deviation* and *time to reach minimum acceleration*, indicating the time it takes to commit towards a response (Duran, Dale, & McNamara, 2010). Greater values indicate a delay in this commitment during task-unrelated thoughts.

4. Discussion

This study explores an objective method for inferring mind wandering. This is of importance for future research using mouse tracking methods for monitoring participant's attention during an online task beyond the laboratory. Such research is attractive in terms of its low costs, scalability, and ecological validity. Importantly, our findings provide support for a dynamic and embodied view of cognition, in which mind wandering and hand reach movements are seen as functionally interdependent (Freeman et al., 2011). In a typical mouse-tracking task, additional competing choice options are presented as additional buttons on the screen — here they were internal cognitions. As such, mouse-tracking measures that typically reflect response conflict between choice buttons (i.e. *MAD* and *AUC*) were not strongly affected by mind wandering. Instead, we build on the *response activation model* to explain how mind wandering can be a source of competing cognitions which lead to slower, less precise and more complex mouse movements.

The importance of set size in predicting mind wandering indicates that task difficulty plays a role in predicting task-unrelated thought. This is supported by the *context-regulation hypothesis* (Smallwood & Andrews-Hanna, 2013), for which there is ample evidence demonstrating that more mind wandering occurs under low task demands, and less mind wandering occurs under high task demands. Yet, once task demands become too high, conditions facilitate mind wandering once again (Adam & Vogel, 2017). Closely following set size, mouse tracking features containing information about the length of the initial phase of a trajectory as well as information regarding changes of mind were most important in predicting mind wandering. Temporal features such as *time to reach maximum deviation* and time to reach maximum acceleration give us insight into the influence of mind wandering on the amount of time required for making a decision for (or against) a particular response. Knowing this, we build on the *SOS model* to explain how mind wandering may actually lead to changes in arousal and consequent disruptions in motor control, leading to a slower initial (high-velocity) phase of a trajectory. In addition to temporal features, features relating to the complexity of trajectories, *x- and y-position flips*, are most important for predicting mind wandering. More flips along the horizontal and vertical axes indicate that more corrective movements took place during trajectories where participants were having task-unrelated thoughts. Taken together, these findings support our expectations in that mind wandering led to slower, more erratic, and more complex trajectories.

Our goal in this study was to explore whether we could find objective differences between all three attentional states: focused attention, task-related interferences, and task-unrelated thought. Although we describe task-related interferences to be a particular type of mind wandering, it is important to note that task-related interferences have been often found to be an ambiguous state to interpret (both by researchers and participants), falling somewhere in between task focus and task-unrelated thought, but being neither here nor there (McVay & Kane, 2009). Conceptually, it makes sense to distinguish task-related interferences from task-unrelated thought. However, in practice, this is still a challenge. In a study comparing how often participants attributed task-related interference to on-task thoughts and mind-wandering, respectively, they estimated about 1/3 of task-related interference to be attributed to mind-wandering, and about 2/3 to be attributed to on-task thoughts (Robison et al., 2019). Because of this ambiguity, task-related interferences have been excluded from analysis in previous research (McVay & Kane, 2009, 2012; Robison et al., 2019; Unsworth & Mcmillan, 2012). It may also be that because of their ambiguity, task-related interferences were most poorly predicted by our algorithms (often below or just at chance level).

Some research demonstrates that participants have a tendency to respond more slowly after making an error either because of a depletion in cognitive resources or because of a strategic increase in response caution (Ceccarini & Castiello, 2018). However, participants were encouraged to respond as accurately as possible to the math operations in this task according to their average response time, making the amount of math errors across conditions negligible (4.03% for focused attention, 5.69% for task-related interference, and 7.51% for task-unrelated thought). Moreover, such post-error slowing has been investigated primarily in speeded-RT tasks which consist of button or key presses. In contrast, the operation span task used in this study was mouse-based, which is by nature slower than button or key-based tasks. As we were interested in observing the evolution of responses over time during mind wandering, it was important that we keep all trials in our analyses, irrespective of potential post-error slowing. As such, we found that the most important mouse features in predicting mind wandering involve time features related to the initial high-velocity phase of a trajectory.

Table 6
Importance of features.

Overall ranking of features		Ranking of features per class					
		TUT		TRI		FA	
set size	60.39	set size	99.16	RT	63.08	y-pos flips	45.45
vel max time	54.19	x-pos flips	68.32	vel max time	63.01	set size	44.90
MD below time	53.69	MD above time	64.82	MAD time	55.29	total dist	44.47
RT	51.52	acc min time	64.27	acc max time	52.92	MD above	43.16
MD above time	50.76	RT	55.30	acc min time	52.73	MD below time	42.49
acc max time	50.42	MD below time	54.13	MD below time	48.28	x-pos reversals	42.44
acc min time	47.81	y-pos flips	53.94	set size	48.13	MAD	41.28
MAD time	46.01	acc max time	53.78	MD above time	40.37	x-pos flips	39.22
x-pos reversals	44.09	initiation time	49.83	idle time	36.01	x-pos max	37.23
total dist	41.12	vel max time	49.31	AD	34.99	acc max time	37.19

4.1. Relation to previous research

Although this study is the first to use mouse movements to predict mind wandering, it is not the first to attempt to predict mind wandering during a working memory task. Recent research has used eye tracking features and task performance in order to predict mind wandering during a similar task — namely, during a spatial complex working memory span task (Huijser, Taatgen, & Van Vugt, 2017). They found task performance (working memory span score) to be the strongest predictor of distracted thought and found a considerable drop in classification accuracy when excluding task performance, with *Accuracy* falling just above chance level, and *Sensitivity* dropping considerably below chance level. In our study, task performance was also the strongest predictor of mind wandering; however our primary interest was investigating the predictive power of mouse movements independently of task performance. Moreover, while Huijser et al. (2017) only included eye movement features extracted from a 2 s interval before each probe (*after* participants had already completed a set), we recorded mouse movements *while* participants completed the processing portion of the operation span task. As such, we extracted a much larger amount of data during the sets preceding each thought probe, which is seemingly able to provide insight into the continuous temporal dynamics of attention and mind wandering during a task.

4.2. Limitations

A possible caveat in our study is the fact that our probes were placed at the end of each set, which varied in level of difficulty (3 to 7 items per set). Participants' answers to the probe were extrapolated to the entire set of varying difficulty, while mouse movements were only measured during the processing portion of the task but not during letter recall (immediately before the probe). A possible solution would be to embed probes during the processing portion of the task (before letter recall). However, this would radically change the nature of the task, because doing so would disrupt the flow of this task and consequently, one's ability to recall the letters in serial order. Alternatively, mouse movements could also rather be recorded during letter recall, or during breaks between sets, but this would lead mouse movements not to be comparable across trials, as start and end positions would differ considerably.

Secondly, the class imbalance both between and within participants made this classification task a particularly difficult one. More instances of the minority class(es) would have likely facilitated our predictions. However, as we already collected data from 274 participants, it is unlikely that an increase in sample size would lead to more balanced classes. As our task was both demanding and engaging, in line with the *context-regulation hypothesis*, it makes sense that under high task demands, mind wandering would be minimal. Prior research using the operation span task (Mrazek et al., 2012) found mind wandering rates close to 25%. Note that in their study, thought probes were not categorical but rather required participants to indicate to what extent their attention was either on-task or on task-unrelated concerns using a 5-point scale. Similarly, in visual working memory tasks, Adam and Vogel (2017) found mind wandering rates of approximately 27%. In the current study mind wandering rates were only 9%. This discrepancy between this number and previous studies' mind wandering rates may be due to several reasons: First, the intermittent probes during our task were retrospective, supposedly reflecting thought content during the entire previous set. The categorical probes forced participants to make a choice about their preceding states of mind, although they may have neither been fully mind-wandering nor fully focused. Instead, a continuous probe may have allowed participants to quantify the degree to which they had been mind wandering, as demonstrated by Mrazek et al. (2012). Second, as attentional resources were likely consumed by the engagement demands of the task for the majority of people in most of the sets, there were likely few resources left either for meta-awareness of mind wandering or for mind wandering itself. Third, it may be that providing feedback with regards to performance after every set may have biased participants to either respond that they were focused (Awh & Vogel, 2015) or that they were concerned with their performance on the task.[10] Finally, probing participants after every set may have induced participants to be more alert and to focus on their performance more intently (Seli, Cheyne, & Smilek, 2013). In Mrazek et al. (2012)'s study, mind wandering was probed only after 25% of sets while in Adam and Vogel (2017)'s study, mind wandering was probed only after 20% of sets. Given the low mind wandering rates in this study, future research should replicate and extend the ideas presented here with different tasks that likely result in higher rates of mind wandering. In order to reduce the class imbalance while maintaining the ecological validity of the study, recommendations for future studies would be to increase the number of trials each individual must complete, reduce the number of probes per participants, and investigate the difference between continuous and categorical probes.

5. Conclusion

Taken together, our findings demonstrate that mouse movements have information that can be used to detect mind wandering. The fact that we had imbalanced classes made this a difficult classification task. Not only that, we attempted to predict mind wandering in a participant-independent manner, which makes classification even more challenging due to individual differences in hand movements. When accounting for class imbalance by downsampling instances of the majority class(es) to match instances of the minority class, we were able to predict mind wandering above chance level in a participant-independent manner. We observed a considerable improvement in predictions of task-unrelated thought while focused attention was predicted consistently above chance level by all classifiers. After reducing the dimensionality of features, performance improved for all classes, such that accuracy was above chance for all classifiers.

[10] Note that in Mrazek et al. (2012)'s findings, feedback was provided, while in Adam and Vogel (2017)'s study, no feedback was provided after each set.

Mind wandering is an important aspect of the human condition; that is, the ability to decouple from the present environment and represent situations and thoughts that are unrelated to the here and now enable planning, goal orientation, and creativity. There are, however, situations in which mind wandering is detrimental to performance and successful navigation through some of our day-to-day activities. Therefore, identifying cues to mind wandering may enable us to catch it before it does any harm. The fact that we were able to trace mind wandering by means of computer mouse movements is a further step in understanding how cognition leaks into action. This study demonstrates that mouse tracking features are a promising objective measure for predicting mind wandering from hand movements in online user-interfaces.

CRediT authorship contribution statement

M.R. Dias da Silva: Conceptualization, Methodology, Software, Data curation, Writing - original draft, Formal analysis, Visualization, Investigation, Project administration. **M. Postma:** Supervision, Conceptualization, Methodology, Project administration.

References

Adam, K. C., & Vogel, E. K. (2017). Confident failures: Lapses of working memory reveal a metacognitive blind spot. *Attention, Perception, and Psychophysics*, *79*(5), 1506–1523. http://dx.doi.org/10.3758/s13414-017-1331-8.

Andrews-Hanna, J. R., Smallwood, J., & Spreng, R. N. (2014). The default network and self-generated thought: Component processes, dynamic control, and clinical relevance. *Annals of the New York Academy of Sciences*, *1316*(1), 29–52. http://dx.doi.org/10.1111/nyas.12360.

Arapakis, I., Lalmas, M., & Valkanas, G. (2014). Understanding within-content engagement through pattern analysis of mouse gestures. In *CIKM* (pp. 1439–1448). Shanghai, China: http://dx.doi.org/10.1145/2661829.2661909.

Awh, E., & Vogel, E. K. (2015). Attention: Feedback focuses a wandering mind. *Nature Neuroscience*, *18*(3), 327–328. http://dx.doi.org/10.1038/nn.3962.

Baird, B., Smallwood, J., Mrazek, M. D., Kam, J. W. Y., Franklin, M. S., & Schooler, J. W. (2012). Inspired by distraction. *Psychological Science*, *23*(10), 1117–1122. http://dx.doi.org/10.1177/0956797612446024.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*(1), 617–645. http://dx.doi.org/10.1146/annurev.psych.59.103006.093639.

Bixler, R., & D'Mello, S. (2016). Automatic gaze-based user-independent detection of mind wandering during computerized reading. *User Modeling and User-Adapted Interaction*, *26*(1), 33–68. http://dx.doi.org/10.1007/s11257-015-9167-1.

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

Ceccarini, F., & Castiello, U. (2018). The grasping side of post-error slowing. *Cognition*, *179*, 1–13. http://dx.doi.org/10.1016/J.COGNITION.2018.05.026.

Cohen, J. D., Aston-Jones, G., & Gilzenrat, M. S. (2004). A systems-level perspective on attention and cognitive control: Guided activation, adaptive gating, conflict monitoring, and exploitation versus exploration.. In *Cognitive neuroscience of attention* (pp. 71–90). New York: Guilford Press.

Conway, A. R., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*(2), 163–183. http://dx.doi.org/10.1016/S0160-2896(01)00096-4.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786.

Cowan, N., Elliott, E. M., Saults, S. J., Morey, C. C., Mattox, S., Hismjatullina, A., et al. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*(1), 42–100. http://dx.doi.org/10.1016/j.cogpsych.2004.12.001.

Dale, R., Roche, J., Snyder, K., & McCall, R. (2008). Exploring action dynamics as an index of paired-associate learning. In O. Sporns (Ed.), *PLoS One*, *3*(3), e1728. http://dx.doi.org/10.1371/journal.pone.0001728.

Danckert, J., Hammerschmidt, T., Marty-Dugas, J., & Smilek, D. (2018). Boredom: Under-aroused and restless. *Consciousness and Cognition*, *61*, 24–37.

Dias da Silva, M. R., Rusz, D., & Postma-Nilsenová, M. (2018). Ruminative minds, wandering minds: Effects of rumination and mind wandering on lexical associations, pitch imitation and eye behaviour. In M. Faber (Ed.), *PLoS One*, *13*(11), http://dx.doi.org/10.1371/journal.pone.0207578.

Duran, N. D., Dale, R., & McNamara, D. S. (2010). The action dynamics of overcoming the truth. *Psychonomic Bulletin and Review*, *17*(4), 486–491. http://dx.doi.org/10.3758/PBR.17.4.486.

Engle, R. W., Laughlin, J. E., Tuholski, S. W., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309–331. http://dx.doi.org/10.1037/0096-3445.128.3.309.

Eysenck, M. (1982). *Attention and arousal: Cognition and performance* (1st ed.). Berlin: Springer Science & Business Media.

Faber, M., Bixler, R., & D'Mello, S. K. (2017). An automated behavioral measure of mind wandering during computerized reading. *Behavior Research Methods*, 1–17. http://dx.doi.org/10.3758/s13428-017-0857-y.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*(6), 381–391. http://dx.doi.org/10.1037/h0055392.

Franklin, M. S., Broadway, J. M., Mrazek, M. D., Smallwood, J., & Schooler, J. W. (2013). Window to the wandering mind: Pupillometry of spontaneous thought while reading. *Quarterly Journal of Experimental Psychology*, *66*(12), 2289–2294. http://dx.doi.org/10.1080/17470218.2013.858170.

Franklin, M. S., Mooneyham, B. W., Baird, B., & Schooler, J. W. (2014). Thinking one thing, saying another: The behavioral correlates of mind-wandering while reading aloud. *Psychonomic Bulletin & Review*, *21*(1), 205–210. http://dx.doi.org/10.3758/s13423-013-0468-2.

Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in motion reveals mind in motion. *Frontiers in Psychology*, *2*(APR), http://dx.doi.org/10.3389/fpsyg.2011.00059.

Graham, E. D., & MacKenzie, C. L. (1996). Physical versus virtual pointing. In *Proceedings of the SIGCHI conference on human factors in computing systems common ground - CHI '96* (pp. 292–299). http://dx.doi.org/10.1145/238386.238532.

Grimes, G. M., Jenkins, J. L., & Valacich, J. S. (2013). Exploring the effect of arousal and valence on mouse interaction. In *International conference on information systems* (pp. 1–15). Milan.

Grimes, G. M., & Valacich, J. S. (2015). Mind over mouse: The effect of cognitive load on mouse behavior mind over mouse. In *International conference on information systems* (pp. 1–13). Fort Worth.

Hehman, E., Stolier, R. M., & Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes & Intergroup Relations*, http://dx.doi.org/10.1177/1368430214538325.

Hermens, F. (2018). When do arrows start to compete? A developmental mouse-tracking study. *Acta Psychologica*, *182*, 177–188. http://dx.doi.org/10.1016/j.actpsy.2017.11.015, URL: http://linkinghub.elsevier.com/retrieve/pii/S0001691817301828.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, *4*(1), 11–26. http://dx.doi.org/10.1080/17470215208416600.

Huijser, S. ., Taatgen, N. ., & Van Vugt, M. (2017). Distracted in a demanding task: A classification study with artificial neural networks. In B. Verheij, & M. Wiering (Eds.), *Proceedings of BNAIC 2017* (pp. 199–212). APA, URL: http://www.rug.nl/research/portal.

Kahneman, D. (1973). Attention and effort. *The American Journal of Psychology*, http://dx.doi.org/10.2307/1421603.

Kam, J. W., Dao, E., Blinn, P., Krigolson, O. E., Boyd, L. A., & Handy, T. C. (2012). Mind wandering and motor control: Off-task thinking disrupts the online adjustment of behavior. *Frontiers in Human Neuroscience*, *6*(DEC), 329. http://dx.doi.org/10.3389/fnhum.2012.00329.

Kieslich, P. J., & Henninger, F. (2017). Mousetrap: An integrated, open-source mouse-tracking package. *Behavior Research Methods*, *49*(5), 1652–1667. http://dx.doi.org/10.3758/s13428-017-0900-z.

Kieslich, P. J., Henninger, F., Wulff, D. U., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (2019). Mouse-tracking: A practical guide to implementation and analysis. In M. Schulte-Mecklenbeck, A. Kühberger, & J. G. Johnson (Eds.), *A handbook of process tracing methods* (pp. 111–130). New York, NY.: Routledge, URL: https://github.com/pascalkieslich/mousetrap-os.

Klinger, E. (2013). Goal commitments and the content of thoughts and dreams: Basic principles. *Frontiers in Psychology*, *4*(JUL), 1–17. http://dx.doi.org/10.3389/fpsyg.2013.00415.

Lenartowicz, A., Simpson, G. V., & Cohen, M. S. (2013). Perspective: causes and functional significance of temporal variations in attention control. *Frontiers in Human Neuroscience*, *7*, 381. http://dx.doi.org/10.3389/fnhum.2013.00381.

Lins, J., & Schöner, G. (2019). Computer mouse tracking reveals motor signatures in a cognitive task of spatial language grounding. *Attention, Perception, and Psychophysics*, *81*(7), 2424–2460. http://dx.doi.org/10.3758/s13414-019-01847-9.

Marchetti, I., Van de Putte, E., & Koster, E. H. W. (2014). Self-generated thoughts and depression: from daydreaming to depressive symptoms. *Frontiers in Human Neuroscience*, *8*(March), 1–10. http://dx.doi.org/10.3389/fnhum.2014.00131.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). Opensesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. http://dx.doi.org/10.3758/s13428-011-0168-7.

Matthews, G., Joyner, L., Gilliland, K., Campbell, S., Falconer, S., & Huggins, J. (1999). Validation of a comprehensive stress state questionnaire: Towards a state "big three?" In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (7th ed.). (pp. 335–350). Tilburg: Tilburg University Press.

McVay, J. C., & Kane, M. J. (2009). Conducting the train of thought: Working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(1), 196–204. http://dx.doi.org/10.1037/a0014104.

McVay, J. C., & Kane, M. J. (2012). Drifting from slow to "d'oh!": Working memory capacity and mind wandering predict extreme reaction times and executive control errors. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *38*(3), 525–549. http://dx.doi.org/10.1037/a0025896.

McVay, J. C., & Kane, M. J. (2013). Dispatching the wandering mind? Toward a laboratory method for cuing "spontaneous" off-task thought. *Frontiers in Psychology*, *4*(SEP), http://dx.doi.org/10.3389/fpsyg.2013.00570.

Meyer, D. E., Abrams, R. A., Kornblum, S., Wright, C. E., & Smith, J. E. (1988). Optimality in human motor performance: Ideal control of rapid aimed movements. *Psychological Review*, *95*(3), 340–370. http://dx.doi.org/10.1037/0033-295X.95.3.340.

Meyer, D. E., Abrams, R. A., Kornblum, S., Wright, C. E., & Smith, J. E. (1990). Speed-accuracy tradeoffs in aimed movements: Toward a theory of rapid voluntary action. In M. Jeannerod (Ed.), *Attention and performance 13: Motor representation and control* (pp. 173–226). Hillsdale, NJ.: Lawrence Erlbaum Associates, Inc..

Mills, C., D 'mello, S., Bosch, N., & Olney, A. M. (2015). Mind wandering during learning with an intelligent tutoring system. In *International conference on artificial intelligence in education* (pp. 267–276). http://dx.doi.org/10.1007/978-3-319-19773-9{_}27.

Mills, C., Herrera-Bennett, A., Faber, M., & Christoff, K. (2018). Why the mind wanders: How spontaneous thought's default variability may support episodic efficiency and semantic optimization. In K. C. Fox, & K. Christoff (Eds.), *The Oxford Handbook of spontaneous thought: Mind-wandering, creativity, and dreaming* (pp. 11–22). Oxford University Press, http://dx.doi.org/10.1093/oxfordhb/9780190464745.013.42.

Mittner, M., Boekel, W., Tucker, A. M., Turner, B. M., Heathcote, A., & Forstmann, B. U. (2014). When the brain takes a break: A model-based analysis of mind wandering. *The Journal of Neuroscience*, *34*(49), 16286–16295. http://dx.doi.org/10.1523/JNEUROSCI.2062-14.2014.

Mittner, M., Hawkins, G. E., Boekel, W., & Forstmann, B. U. (2016). A neural model of mind wandering. *Trends in Cognitive Sciences*, *20*(8), 570–578. http://dx.doi.org/10.1016/j.tics.2016.06.004.

Morsella, E., Larson, L. R. L., & Bargh, J. A. (2010). Indirect cognitive control, working-memory-related movements, and sources of automatisms. In E. Morsella (Ed.), *Expressing oneself/ expressing one's self: Communication, Cognition, language, and identity* (pp. 61–85). New York: Psychology Press.

Mrazek, M. D., Smallwood, J., Franklin, M. S., Chin, J. M., Baird, B., & Schooler, J. W. (2012). The role of mind-wandering in measurements of general aptitude. *Journal of Experimental Psychology: General*, *141*(4), 788–798. http://dx.doi.org/10.1037/a0027968.

Papesh, M. H., & Goldinger, S. D. (2012). Memory in motion: Movement dynamics reveal memory strength. *Psychonomic Bulletin & Review*, *19*(5), 906–913. http://dx.doi.org/10.3758/s13423-012-0281-3.

Pham, P., & Wang, J. (2015). Attentivelearner: Improving mobile MOOC learning via implicit heart rate tracking. In C. Conati, & et al. (Eds.), *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*: *Vol. 9112*, (pp. 367–376). Switzerland: Springer International Publishing Switzerland, http://dx.doi.org/10.1007/978-3-319-19773-9{_}37.

R Core Team *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Randall, J. G., Oswald, F. L., & Beier, M. E. (2014). Mind-wandering, cognition, and performance: A theory-driven meta-analysis of attention regulation. *Psychological Bulletin*, *140*(6), 1411–1431. http://dx.doi.org/10.1037/a0037428.

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

Robison, M. K., & Unsworth, N. (2019). Pupillometry tracks fluctuations in working memory performance. *Attention, Perception, and Psychophysics*, *81*(2), 407–419. http://dx.doi.org/10.3758/s13414-018-1618-4.

Rummel, J., & Boywitt, C. D. (2014). Controlling the stream of thought: Working memory capacity predicts adjustment of mind-wandering to situational demands. *Psychonomic Bulletin and Review*, *21*(5), 1309–1315. http://dx.doi.org/10.3758/s13423-013-0580-3.

Seelye, A., Hagler, S., Mattek, N., Howieson, D. B., Wild, K., Dodge, H. H., et al. (2015). Computer mouse movement patterns: A potential marker of mild cognitive impairment. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *1*(4), 472–480. http://dx.doi.org/10.1016/j.dadm.2015.09.006.

Seli, P., Carriere, J. S. A., Levene, M., & Smilek, D. (2013). How few and far between? Examining the effects of probe rate on self-reported mind wandering. *Frontiers in psychology*, *4*, 430. http://dx.doi.org/10.3389/fpsyg.2013.00430.

Seli, P., Carriere, J. S., Thomson, D. R., Cheyne, J. A., Ehgoetz Martens, K. A., & Smilek, D. (2014). Restless mind, restless body. *Journal of Experimental Psychology: Learning Memory and Cognition*, *40*(3), 660–668. http://dx.doi.org/10.1037/a0035260.

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. http://dx.doi.org/10.1037/a0030954.

Smallwood, J. (2013). Distinguishing how from why the mind wanders: A process-occurrence framework for self-generated mental activity. *Psychological Bulletin*, *139*(3), 519–535. http://dx.doi.org/10.1037/a0030010.

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*(AUG), 1–6. http://dx.doi.org/10.3389/fpsyg.2013.00441.

Smallwood, J., Brown, K. S., Tipper, C., Giesbrecht, B., Franklin, M. S., Mrazek, M. D., et al. (2011). Pupillometric evidence for the decoupling of attention from perceptual input during offline thought. In S. Gilbert (Ed.), *PLoS One*, *6*(3), e18298. http://dx.doi.org/10.1371/journal.pone.0018298.

Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin and Review*, *132*(6), 946–958. http://dx.doi.org/10.1037/2326-5523.1.S.130.

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. http://dx.doi.org/10.1146/annurev-psych-010814-015331, arXiv:0402594v3.

Song, J. H., & Nakayama, K. (2009). Hidden cognitive states revealed in choice reaching tasks. *Trends in Cognitive Sciences*, *13*(8), 360–366. http://dx.doi.org/10.1016/j.tics.2009.04.009.

Stawarczyk, D., Majerus, S., Maj, M., Van der Linden, M., & D'Argembeau, A. (2011). Mind-wandering: Phenomenology and function as assessed with a novel experience sampling method. *Acta Psychologica*, *136*(3), 370–381. http://dx.doi.org/10.1016/j.actpsy.2011.01.002.

Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between operation span and raven. *Intelligence*, *33*(1), 67–81. http://dx.doi.org/10.1016/J.INTELL.2004.08.003.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. http://dx.doi.org/10.3758/BF03192720.

Unsworth, N., & Mcmillan, B. D. (2012). Mind wandering and reading comprehension: Examining the roles of working memory capacity, interest, motivation, and topic experience. *Journal of Experimental Psychology*, *39*(3), 832–842. http://dx.doi.org/10.1037/a0029669.

Unsworth, N., Redick, T. S., Lakey, C. E., & Young, D. L. (2010). Lapses in sustained attention and their relation to executive control and fluid abilities: An individual differences investigation. *Intelligence*, *38*(1), 111–122. http://dx.doi.org/10.1016/j.intell.2009.08.002.

Unsworth, N., & Robison, M. K. (2016). The influence of lapses of attention on working memory capacity. *Memory and Cognition*, *44*(2), 188–196. http://dx.doi.org/10.3758/s13421-015-0560-0.

Unsworth, N., & Robison, M. K. (2017). The importance of arousal for variation in working memory capacity and attention control: A latent variable pupillometry study. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *43*(12), 1962–1987. http://dx.doi.org/10.1037/xlm0000421.

Unsworth, N., & Robison, M. K. (2018). Tracking arousal state and mind wandering with pupillometry. *Cognitive, Affective and Behavioral Neuroscience*, *18*(4), 638–664. http://dx.doi.org/10.3758/s13415-018-0594-4.

van Beers, R. J. (2004). The role of execution noise in movement variability. *Journal of Neurophysiology*, *91*(2), 1050–1063. http://dx.doi.org/10.1152/jn.00652.2003.

Watkins, E. R. (2008). Constructive and unconstructive repetitive thought. *Psychological Bulletin*, *134*(2), 163–206. http://dx.doi.org/10.1037/0033-2909.134.2.163.

Welsh, T. N., & Elliott, D. (2004). Movement trajectories in the presence of a distracting stimulus: Evidence for a response activation model of selective reaching. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *57*(6), 1031–1057. http://dx.doi.org/10.1080/02724980343000666.

Yamauchi, T. (2013). Mouse trajectories and state anxiety: Feature selection with random forest. In *Humaine association conference on affective computing and intelligent interaction mouse* (pp. 399–404). http://dx.doi.org/10.1109/ACII.2013.72.

Yamauchi, T., & Xiao, K. (2017). Reading emotion from mouse cursor motions: Affective computing approach. *Cognitive Science*, 1–49. http://dx.doi.org/10.1111/cogs.12557.

Yatani, K., & Truong, K. N. (2012). Bodyscope: A wearable acoustic sensor for activity recognition. In *UbiComp'12 - Proceedings of the 2012 ACM conference on ubiquitous computing* (pp. 341–350). Pittsburgh, Pennsylvania: http://dx.doi.org/10.1145/2370216.2370269.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*(5), 459–482. http://dx.doi.org/10.1002/cne.920180503.

Zhang, Y., & Kumada, T. (2017). Relationship between workload and mindwandering in simulated driving. In J. Xu (Ed.), *PLoS One*, *12*(5), 1–12. http://dx.doi.org/10.1371/journal.pone.0176962.