



Do your eyes give you away? A validation study of eye-movement measures used as indicators for mindless reading

Lena Steindorf¹ · Jan Rummel¹

Published online: 27 February 2019
© The Psychonomic Society, Inc. 2019

Abstract

Identifying eye-movement measures as objective indicators of mind wandering seems to be a work in progress. We reviewed research comparing eye movements during self-categorized episodes of normal versus mindless reading and found little consensus regarding the specific measures that are sensitive to attentional decoupling during mind wandering. To address this issue of inconsistency, we conducted a new, high-powered eye-tracking experiment and considered all previously identified mind-wandering indicators. In our experiment, only three measures (reading time, fixation count, and first-fixation duration) positively predicted self-categorized mindless reading. Aside from these single measures, the word-frequency effect was found to be generally less pronounced during mindless-reading than during normal-reading episodes. To additionally test for convergent validity between the objective and subjective mind-wandering measures, we utilized eye-movement measures as well as thought reports, to examine the effect of metacognitive awareness on mind-wandering behavior. We expected that participants anticipating a difficult comprehension test would mind wander less during reading than would those anticipating an easy test. Although we were able to induce metacognitive expectancies about task difficulty, we found no evidence that these difficulty expectancies affected either subjectively reported or objectively measured mind wandering.

Keywords Mind wandering · Task-unrelated thought · Eye movements · Reading

Imagine you are reading a book. Your eyes are moving across the page, line by line. While processing words and sentences, you immerse yourself deeper and deeper into the story. You have been reading for a while now, and your eyes keep on scanning the pages, from top to bottom, from left to right. Suddenly you realize you have just been thinking about that huge fight you had with your best friend. It seems impossible for you to reproduce the last text passages, although you remember your eyes moving across the lines. You were mind wandering. That is, your thoughts trailed off from the task at hand, which was, in this case, reading for comprehension. When people mind wander, they think about personal problems, unfulfilled tasks, or other things unrelated to their current task (Schooler et al., 2011). Mind-wandering episodes often occur without intention, or even awareness (Smallwood & Schooler, 2006). For this reason, it is possible

for you to mindlessly “read” a considerable amount of lines without noticing that your thoughts are trailing off. The fact that you maintain a reading-like eye-movement behavior while doing so makes mindless reading an attractive research topic in the mind-wandering domain. What happens to eye movements when attention is directed away from ongoing text processing? Do eye movements differ systematically between normal and mindless reading in such a way that they could serve as objective, nonintrusive indicators of mind wandering? In the present work, we tried to answer these questions by reviewing research comparing eye movements during self-categorized episodes of normal versus mindless reading. We also report a newly conducted, high-powered eye-tracking experiment, with the objective of validating previously identified mind-wandering indicators.

Mind wandering is a ubiquitous phenomenon. As much as 30%–50% of all daily thoughts are unrelated to current external events (Kane et al., 2017; Killingsworth & Gilbert, 2010; Klinger, 1999). When thoughts trail off, the perception of the external world is affected. Thoughts about a fight you had with your best friend move your mental focus away from the words and sentences on the page in front of you. During mind-wandering episodes, attention is directed away from external,

✉ Lena Steindorf
lena.steindorf@psychologie.uni-heidelberg.de

¹ Department of Psychology, Heidelberg University, Heidelberg, Germany

perceptual input toward internal processing (Schooler et al., 2011; Smallwood & Schooler, 2006; Stawarczyk, Majerus, Maquet, & D’Argembeau, 2011). When your thoughts trail off during reading, you no longer pay sufficient attention to the lines of text. This attentional decoupling process does not seem to work in an all-or-nothing fashion. Schad, Nuthmann, and Engbert (2012) proposed the *levels-of-inattention hypothesis*. Their findings support the idea of graded attentional decoupling at different levels of the cognitive hierarchy: Processing of external information can fail at early (*deep decoupling*, cascading down to later levels; see Smallwood, 2011) but also at late (*weak decoupling*) perceptual levels. We believe this extension of a dichotomous view of perceptual decoupling to be a very fruitful approach. For our purpose, however, we took a step back and critically reviewed research comparing eye movements during text-focused versus decoupled attention (mind wandering), regardless of decoupling levels.

An attention shift toward internal processing renders external information-encoding errors more likely (Smallwood, Baracua, Lowe, & Obonsawin, 2003), thus causing performance decrements in external tasks (Christoff, Gordon, Smallwood, Smith, & Schooler, 2009). During reading, mind wandering is assumed to be reflected by attentional decoupling from the semantic input, meaning the lines of text one is reading. Consequently, reading comprehension is very likely to suffer. Previous research has confirmed that thoughts often trail off during reading and that high mind-wandering frequencies are associated with poorer reading comprehension (Schooler, Reichle, & Halpern, 2004). It is assumed that readers fail to mentally connect events, or linguistic units, to create a situational model of the story during task-unrelated-thought (TUT) episodes (Smallwood, McSpadden, & Schooler, 2008).

But what does attentional decoupling have to do with eye movements, and how may eye movements aid us in detecting mindless reading? The basic idea behind eye-tracking research in general is that what people are looking at reflects what they are mentally processing, at least to some degree (Uzzaman & Joordens, 2011). During reading—a behavior that can be broken down into a sequence of fixations and saccades—a fixation on a given word is assumed to reflect the mental processing of this word. This logic becomes apparent when taking a look at the robust word-frequency effect (Inhoff & Rayner, 1986). Identifying and processing a very uncommon word such as “vat” takes longer than processing a very common word such as “cat.” Fixation times are reflective of the need for more intense processing: Low-frequency words are fixated longer than high-frequency words. Cognitive-control theories (Just & Carpenter, 1980), in contrast to oculomotor-control theories (Yang, 2006), posit a close *eye–mind link*. Models such as SWIFT (Engbert, Nuthmann, Richter, & Kliegl, 2005) and E-Z Reader (Pollatsek, Reichle, & Rayner, 2006) assume that cognitive and oculomotor processes are linked to *cognitive*

forces driving eye-movement behavior during normal reading. Therefore, cognitive-control theories predict differences in eye movements during normal versus mindless reading. The eye–mind link should be interrupted when the mind starts to wander away from external text processing toward internal processing. During mind wandering, cognitive forces are engaged in thoughts unrelated to the to-be-read text, limiting their power to drive eye movements. Mindless readers may thus move their eyes beyond the word “vat” without thoroughly processing it—and thus not taking the additional fixation time its processing usually requires. That is, when attention shifts from lines of text toward inner thoughts, eye movements should be (partially) decoupled from ongoing text processing, resulting in less sensitivity toward lexical features—for example, word frequency. Research applying the *z*-string reading paradigm (Nuthmann & Engbert, 2009; Rayner & Fischer, 1996), during which participants mindlessly read “words” only containing the letter *Z* (e.g., “Zzzzz zz zzzzzz”), suggests that readers are not only less sensitive to lexical variables, but that single eye-movement measures, such as fixation duration, are erratic (in this case, longer) when thoughts are off-task. Furthermore, Smilek, Carriere, and Cheyne (2010) noted that attentional decoupling during mindless reading could also be represented by a higher blinking frequency, thereby even physically reducing the processing of external input.

This reasoning has motivated several studies (see below) trying to identify eye-movement measures that are sensitive to attentional decoupling during mindless reading. Because people are not able to consciously control or manipulate fixations and saccades (Rayner, 1998), mind-wandering research would greatly benefit if eye-movement measures were found to be robust and stable indicators of TUT episodes. Eye movements could be used as biological markers in addition to or as a replacement for subjective self-reports. Until now, mind wandering has typically been measured via self-reports. In many mind-wandering experiments, participants are asked to briefly describe and/or classify their current thoughts’ content when a probe interrupts their present task. Although these self-reports proved to be reasonably valid (e.g., McVay & Kane, 2012), they might still be prone to memory-driven and/or classification errors. Also, the demand characteristics of a given task might influence subjectively reported mind wandering (Vinski & Watter, 2012). The apparent subjectivity of self-reports has thus led to a search for objective biological or behavioral indicators of TUTs. If found, these could be used not only to further validate subjective thought reports, but also to directly measure mind-wandering behavior without a subjective component. Besides eye movements, reaction times seem to be promising objective markers of TUTs. McVay and Kane (2012) measured attentional lapses using thought probes and found that these lapses go along with higher τ parameters, which represent the tail of reaction time distributions. In other words, more mind wandering goes along with higher

proportions of extremely long reaction times. Of course, a researcher's choice of objective markers is task-dependent. For go/no-go tasks like the SART (Robertson, Manly, Andrade, Baddeley, & Yiend, 1997), which is often used by mind-wandering researchers (McVay & Kane, 2012; Unsworth & McMillan, 2014), reaction times could be the appropriate means and seem to be a fruitful approach. For reading tasks, eye movements have started to emerge as biological mind-wandering markers. In the following sections, we review five studies that aimed to identify eye-movement measures as objective mind-wandering indicators. To this end, we will give an overview on the methods that have been applied in this line of research and discuss the results that were obtained (for an overview, see Table 1).

For the present purpose, we only review studies in which participants' thoughts were randomly probed (probe-caught method; see Schooler et al., 2004) during reading. The

probe-caught method does not rely on participants' awareness of their mind-wandering behavior and is supposed to be a reliable estimate of mind-wandering frequency (Smallwood & Schooler, 2006). In the to-be-reviewed studies, self-categorized normal-reading episodes were compared to self-categorized mindless-reading episodes in order to determine changes in eye-movement behavior. This approach is especially useful in order to make eye-movement measures utilizable as objective mind-wandering indicators, because it allows the validation of possible mind-wandering indicators by means of the most widely used mind-wandering assessment methods, namely online thought probes.

Reichle et al. (2010) had four participants read the entirety of *Sense and Sensibility* by Jane Austen in up to 15-h-long sessions, while tracking their eye movements and applying the probe-caught method to assess mind-wandering behavior, with a total of 151.5 probes per participant, on average. The

Table 1 Studies comparing eye movements during self-categorized episodes of mindless reading (MR) versus normal reading (NR)

	Reichle, Reineberg, and Schooler (2010)	Smilek et al. (2010)	Uzzaman and Joordens (2011)	Foulsham, Farley, and Kingstone (2013)	Frank, Nara, Zavagnin, Touron, and Kane (2015)
Task characteristics					
<i>N</i>	4	12	22	26	29
Reading materials	<i>Sense and Sensibility</i>	Two passages of <i>A Short History of Nearly Everything</i>	16 pages from <i>War and Peace</i>	120 single sentences	Five chapters from <i>War and Peace</i>
Reading duration	13.5 h (average)	Up to 30 min		30 min	
Probe count	151.5 (average)	20	10 (average)	9 (average)	20
Area of interest	2.5-s, 5-s, 10-s, 30-s, 60-s, and 120-s intervals preceding thought probes	5-s intervals preceding thought probes	5-s intervals preceding thought probes	Target sentences	3- to 8-s intervals preceding thought probes
Eye-movement measures					
Fixation count	MR = NR (first-pass fixations)	MR < NR	MR = NR	MR > NR	MR = NR (first-pass fixations)
Fixation duration	MR = NR (first-fixation durations)	MR = NR	MR = NR	MR > NR	
Between-word regression count	MR = NR			MR = NR	MR > NR
Within-word regression count			MR < NR		
Blink count		MR > NR	MR = NR		MR > NR
Reading time	MR > NR (gaze duration & total viewing time)			MR > NR	MR = NR (gaze duration)
Word-frequency effect				MR < NR	

The < and > symbols represent significant effects with $p < .05$. For example, Uzzaman and Joordens found that there were fewer within-word regressions during mindless than during normal reading. For comparability reasons, only effects on eye-movement measures that appeared preceding probe-caught mind wandering are reported from the Reichle et al. study, although the authors also examined self-caught mind wandering. From the Frank et al. study, only effects for younger adults are reported. The authors differentiated between task-unrelated thoughts and task-related inferences, but only effects found for task-unrelated thoughts are reported here. They also advised not to use blinks as mind-wandering markers, because effects on this measure were found for only about half their participants. Uzzaman and Joordens additionally found a significant effect for a measure they called *run count*. Because it was not precisely defined and is not a commonly used eye-tracking measure, we did not include it as a measure of interest.

authors additionally applied the self-caught method, which requires participants to press a certain key whenever they catch themselves mind wandering.¹ Eye-movement measures for six different time intervals preceding the probes (2.5, 5, 10, 30, 60, and 120 s prior to a thought probe) were analyzed regarding differences between the episodes of self-classified normal versus mindless reading. In a similar study, Smilek et al. (2010) analyzed data from 12 participants who read two passages of *A Short History of Nearly Everything* by Bill Bryson while their eye movements were tracked, and their thoughts were probed 20 times. The 5-s intervals preceding thought probes were analyzed. Uzzaman and Joordens (2011) also analyzed eye-movement measures from the 5-s intervals preceding thought probes. Data from 22 participants who read 16 pages of *War and Peace* by Tolstoy were entered in their analysis. Again, the probe-caught method was applied, with participants receiving a total of ten probes apiece, on average. A slightly different method to detect differences in eye-movement behavior between normal and mindless reading was employed by Foulsham et al. (2013). In their experiment, not specific time intervals, but single sentences were analyzed. In all, 26 participants read 120 sentences, 48 of which were the to-be-examined key sentences. Thoughts were probed approximately nine times, with a probe always appearing after a key sentence. In the aforementioned three experiments, the probes appeared randomly during full-text reading, resulting in different to-be-analyzed text passages for each participant, thus increasing error variance. The single-sentence-reading method by Foulsham et al. kept the to-be-examined reading materials equal for all participants, and even allowed for manipulation of the linguistic input prior to the thought probes. Participants read sentences containing either high- or low-frequency words, so that the influence of mind wandering on the word-frequency effect could be examined, which turned out to be smaller during mindless reading in this study. A potential disadvantage of this method is, however, that presenting single sentences is less ecologically valid than full-text reading. In a study by Frank et al. (2015), participants again read chapters from Tolstoy's *War and Peace*. Twenty-nine younger adults² eye-tracking data from 3- to 8-s time intervals preceding 20 thought probes were analyzed.

By and large, the methods used in these five studies show a reasonable degree of similarity. The probe-caught method was always applied to identify passages read normally versus

mindlessly, and eye-tracking data were compared for these passages. Methodologically, these studies differed concerning the lexical input (full text vs. single sentences, different book chapters, number of lines on each screen, etc.), total reading time (from less than an hour up to 15 h), probing procedure (frequency of probing, additionally assessed self-caught mind wandering, number and selection of response options, etc.), number of participants (four to 29), and areas of interest analyzed preceding the thought probes (a single sentence vs. a specific time interval). Some of these specific features might have had an influence on the assessment of mind wandering and the eye-movement behavior. For example, differences in reported mind wandering can arise due to the framing and wording of thought probes and response options (Weinstein, 2018; Weinstein, De Lima, & van der Zee, 2018). Additionally, eye-movement behavior might have been influenced by the length of the experiment (e.g., due to fatigue) or by the lexical input (e.g., single sentences vs. full text).

At least for studies that employ similar methods, we would expect to find converging eye-movement patterns for normal versus mindless reading if eye movements were robust indicators of mind wandering. Interestingly, the five studies varied widely regarding the selected eye-movement measures and the results obtained. Table 1 shows eye-movement measures that were found to differ significantly ($p < .05$) between normal and mindless-reading episodes in at least one of the previously described studies. For comparability reasons, the table is limited to effects found for younger adults with the probe-caught method assessing TUT episodes, excluding task-related inferences, which were additionally assessed only in the Frank et al. (2015) study. Whenever a significant effect was reported for a measure in any study, we checked all other studies as to whether they reported results for this measure. As is apparent in Table 1, there seem to be large differences between the studies: Significant effects were found for several measures, but studies strongly differed as to which eye-movement measures were selected to be analyzed. Certainly, this is partially due to the theoretical focus of the respective study. However, to make eye-movement measures utilizable as objective mind-wandering indicators, a greater degree of consensus in the selection of eye-movement measures of interest would be desirable. Importantly, even if the same measure were analyzed in two or more studies, the effects could not always be replicated, and even showed opposite directions. For example, Smilek et al. (2010) found a lower fixation count during mindless reading than during normal reading, whereas Foulsham et al. (2013) found a higher fixation count during mindless reading.

Such inconsistencies in results can indicate validity problems for eye-movement measures as indicators for mindless reading. However, they can also reflect meaningful processing differences due to methodological specifics. Therefore, discrepancies in the results between studies do not have to be a

¹ Although the results differed for self- versus probe-caught mind-wandering, we focus solely on the probe-caught results in the present work. Distinguishing between self- and probe-caught mind-wandering provides important insights into awareness processes, and further research will be needed to determine the influence of different levels of mind-wandering awareness on eye-movement behavior (see also Schad et al., 2012). Still, in most of the studies we reviewed, the probe-caught method was applied exclusively. Thus, we will focus on this method alone for comparability reasons.

² The authors tested older and younger adults. For comparability reasons, we only report the effects found for younger adults.

knockout argument against the usefulness of eye movements as mind-wandering indicators. For example, single-sentence reading, as in the Foulsham et al. (2013) study, likely imposes cognitive demands different from those of full-text reading. Although different contextual demands might generate discrepancies in results between studies, objective indicators should converge when the contexts are similar. When contexts differ, boundary conditions should be identified, so that it becomes possible to predict when and why specific patterns of results will occur. To this end, studies manipulating the contextual demands of interest (e.g., single-sentence vs. full-text reading) are required.

Notably, even for studies that applied similar methods and whose contextual demands appeared to be relatively comparable, Table 1 lists discrepancies in the results between studies. To address these inconsistencies, we conducted a new, high-powered eye-tracking experiment, in which we confined the analyses to the eye-movement measures that had previously been identified as potential mind-wandering indicators (i.e., those listed in Table 1) a priori. In this study, we applied the standard approach (see above) and asked participants to read two chapters of Oliver Sacks's *Musicophilia: Tales of Music and the Brain* (Sacks, 2008) for comprehension, and we further asked them to respond to ten thought probes while reading. We tracked participants' eyes during reading so that we could compare eye-movement behavior between normal and mindless reading.

Besides the validation of eye-movement measures as mind-wandering markers, the present study's secondary aim was to actually apply these markers to examine the effect of metacognitive expectancies on mind wandering. Previous research has shown that mind-wandering behavior during cognitive tasks changes in accordance with varying task demands. That is, mind wandering is usually reduced when task demands increase (e.g., Rummel & Boywitt, 2014). Considering that mind wandering is beneficial in some situations, this adjustment seems to be adaptive in everyday life. TUTs are often future-related and focus on unfulfilled tasks or personal problems. In this way, they might help people plan future actions (Klinger, 1999; Mooneyham & Schooler, 2013), come up with creative ideas (Baird et al., 2012; but see Smeekens & Kane, 2016), or keep future task goals active (Steindorf & Rummel, 2017). Given these potential positive effects, and considering the *context-regulation hypothesis* (Smallwood & Andrews-Hanna, 2013), efficient adjustment should allow people to benefit from TUTs as long as the situation allows or encourages it, while also minimizing the costs of mind wandering when working on demanding tasks.

However, directly manipulating the task demands for a reading task would have required us to change the to-be-read text between experimental conditions. Instead, we decided to focus on metacognitive awareness, the knowledge about one's own cognitive processes (Flavell, 1979), as one possible

underlying factor of mind-wandering adjustment. We manipulated participants' expectancies regarding the difficulty of an upcoming reading comprehension test. That is, all participants read the same text, but some expected a difficult reading comprehension test afterward, whereas others expected an easy test. During the reading phase, we tracked participants' eye movements and periodically probed their thoughts. All participants then worked on the same reading comprehension test. Analogous to the results found for high task demands (e.g., Rummel & Boywitt, 2014), we expected participants with high-difficulty expectations to mind wander less than low-difficulty-expectation participants. Convergent evidence for such an effect from subjective thought-reports and (some) objective eye-movement measures (see above and Table 1) would further speak for the convergent validity of the respective measures. In addition, we expected participants with high-difficulty expectations to perform better in the reading comprehension test than those with low-difficulty expectations.

Across the subsequent **Method** and **Results** sections, we report how we determined our sample size and all data exclusions, manipulations, and measures in the study (Simmons, Nelson, & Simonsohn, 2012). The study reported in this article was preregistered on June 28, 2016 (<https://osf.io/sbeqn/>), and, if not otherwise indicated, we followed the research and analysis protocol stated in the preregistration.

Method

Participants, data cleansing, and design

For multilevel-modeling approaches, there is currently no common understanding regarding power analyses. However, simulations by Maas and Hox (2005) determined what constitutes a sufficient sample size for accurate parameter estimation. The authors recommend, as stated in the preregistration, collecting data for at least 50 second-level and seven first-level units. Considering the accuracy benefits (for variance estimation) that an increasing number of second-level units might offer (beyond 50; see also Paccagnella, 2011), but also the effort of an eye-tracking study, 122 participants (second-level units) were tested at Heidelberg University, Germany. Using ten measurement points per participant, we ensured that our design also met the criteria of at least seven first-level units. To account for potential participant exclusions due to eye-tracking-related problems, we tested 22 participants more than we had preregistered.

Eye-tracking data of one participant were lost due to a software problem. The eye-tracking data quality was determined sentence by sentence, for all participants and all target sentences, by two independent raters. A sentence's eye-tracking data were categorized as unusable when technical errors were clearly evident (e.g., when calibration failed),

when a strong drift at the beginning or end of the lines was visible, or when a participant's eyes accidentally hit a thought probe trigger (see below) before reading the target sentence, causing a premature thought probe appearance. For 88.96% of all target sentences, both raters agreed on the categorizations. For the other target sentences, a liberal decision was made, and the eye-tracking data were categorized as useable. Fourteen participants were excluded from the final data set due to their producing predominantly unusable eye-tracking data (less than five useable target sentences), resulting in $N = 107$ ($M_{\text{age}} = 22.58$, $SD_{\text{age}} = 4.01$, 78.50% female, normal or corrected-to-normal vision) for all further analyses. For the final data set, 93.77% of all target sentences' eye-tracking data were categorized as usable. Unusable data were excluded from further analyses, relying on multilevel models' beneficial characteristic of being applicable to data sets with randomly missing values.

We used a one-factorial design to investigate the influence of difficulty expectations (expectation of an easy vs. a difficult comprehension test) on mind-wandering behavior and reading performance.

Apparatus

Eye movements were recorded using the SMI RED500 system (SensoMotoric Instruments, Potsdam, Germany) with a sampling rate of 500 Hz and a spatial resolution (RMS) of 0.03° . Using a remote eye-tracking system, free head movements (in a $40 \text{ cm} \times 20 \text{ cm}$ range) allowed for naturalistic reading, but participants were asked to move around as little as possible in 60- to 80-cm distance from the monitor (22-in.). The BeGaze software's (SensoMotoric Instruments, Potsdam, Germany) algorithm was used to calculate eye-movement measures, using a velocity threshold of $40^\circ/\text{s}$ to identify saccades.³

Materials

Reading task and target sentences Participants read a shortened German version of two chapters of Oliver Sacks's *Musicophilia: Tales of Music and the Brain* (Sacks, 2008). The text was presented on 58 pages in black-on-gray 32-point Arial font with double spacing and a maximum of seven lines per page, using the software Experiment Center (SensoMotoric Instruments, Potsdam, Germany).

³ From the eye-tracking and mind-wandering studies we reviewed, only one (Foulsham et al., 2013) provided details about the event detection method that was applied. Because these authors used a velocity-based algorithm and none of the other studies' authors indicated that they had deviated from this standard procedure, we also employed the standard algorithm and parameter values that our software recommends, for comparability reasons. We acknowledge, however, that more sophisticated methods for event detection are currently under development (e.g., Nyström & Holmqvist, 2010) that will eventually replace these methods in the near future.

Participants moved through the text autonomously, using the space bar to turn pages. On each page, a participant's fixations on a trigger cross (dwell time = 500 ms) in the top left corner made the text appear. This procedure ensured that, for each page, eye-movement recordings always started at the top left corner position. Ten target sentences (always presented in the middle of the page, $M_{\text{word-count}} = 31.30$, $SD_{\text{word-count}} = 5.27$) were determined before data collection, five containing a high-frequency ($M_{\text{freq}} = 1,892.80$, $SD_{\text{freq}} = 1,309.78$), and the other five containing a low-frequency ($M_{\text{freq}} = 8.40$, $SD_{\text{freq}} = 10.90$), target word. Word frequencies were extracted from the dlex database (Heister et al., 2011), and target words were matched for length (see the Appendix Table 6 for a full list of the target words, including translations and information on frequency and length). After a participant had read a target sentence, fixations on the first words of the following sentence (dwell time = 300 ms) triggered the appearance of a thought probe. After responding to the thought probe, participants were redirected to the page they had been reading before the probe appeared. In doing this, we were able to define target sentences with specific features (high- vs. low-frequency words) and to keep the to-be-examined passages equal for all participants (Foulsham et al., 2013). Still, participants could engage in natural reading, in contrast to the Foulsham et al. study, in which single sentences were presented sequentially.

Thought probes To subjectively assess mind wandering, we probed participants' thoughts after each target sentence, resulting in a total of ten probes. Participants were asked to categorize their current thoughts as being task-related thoughts ("I am thinking about the text I am reading"; later referred to as *TRTs*), task-triggered intrusions ("I am thinking about things related to the text I am reading"; later referred to as *TTIs*), or TUTs ("I am thinking about things unrelated to the text").

Procedure

In individual sessions, participants signed a consent form before receiving detailed written information about the eye-tracking procedure and instructions for the reading phase and the comprehension test. Furthermore, the concept of mind wandering and the response options of the thought probes were explained. A five-point calibration of the eye-tracker was applied prior to the reading phase and was repeated once in the middle of the reading phase, as well as whenever this was considered necessary by the experimenter who monitored eye movements on a separate screen, nonvisible to the participants. Once the eye-tracker was successfully calibrated, participants practiced turning pages in a self-paced fashion on a passage of instruction text, by pressing the space bar and fixating the trigger area on the following page. They also

practiced the thought-probing procedure once before receiving critical instructions for the experimental manipulation. The participants in the high-difficulty expectation condition were told that the upcoming comprehension test was going to be very difficult for a student population, whereas the participants in the low-difficulty expectation condition were informed that the test was going to be fairly easy. This information was embedded in a cover story referring to previous research, to render the manipulation more trustworthy. Following these instructions, the reading phase began, in which participants autonomously read the two chapters within up to 30 min and responded to a total of ten thought probes, with each probe occurring after a target sentence. After the reading task, participants completed a retrospective mind-wandering questionnaire, categorizing the entirety of their thoughts during the reading phase into the three thought-probe response categories (TRTs, TTIs, TUTs) using percentage scores. We then asked participants to rate how difficult they expected the upcoming reading comprehension test to be on a 10-point scale, from *very easy* to *very difficult*. The following test consisted of 15 multiple-choice items (four response options each, only one of which was correct) that had been experienced as moderately difficult in a pretest. After completion of the comprehension test, participants rated how difficult they had perceived the test to be, on the previously used 10-point scale, and answered demographic questions. Finally, participants were debriefed and dismissed.

Results

The analyses reported in this section were executed as preregistered, if not indicated otherwise. We employed confirmatory strategies of analyses and conducted one-tailed tests (Cho & Abe, 2013) whenever we had preregistered a directional hypothesis—that is, for all analyses concerning the difficulty-expectation manipulation. For these analyses, we considered any effect in the direction opposite from our preregistered expected outcomes to be nonsignificant. For those analyses concerning validation of the eye-movement measures (e.g., because previous research had found opposing results), we employed a standard nondirectional hypothesis-testing strategy.

Behavioral measures

Performance on the reading comprehension test was calculated as the percentage of correctly answered comprehension questions. The amount of online-reported TUTs and TTIs during the reading phase was defined as the sum of probes in which participants categorized their thoughts as being either task-unrelated or task-triggered, respectively.

Table 2 shows the means, standard deviations, and correlations for the behavioral measures. Mean comprehension

accuracy was satisfactory, implying that participants generally paid a decent amount of attention to the text, although, on average, TUTs or TTIs were reported in almost half of all thought-probe occurrences. Retrospective thought reports converged with the online thought reports, as indicated by the moderate to strong positive correlations of the two. Comprehension accuracy weakly correlated with TUTs (online and retrospective); the better performance in the comprehension test was, the fewer TUTs were reported. The comprehension test's expected difficulty correlated weakly with retrospectively reported TUTs: The fewer TUTs were reported, the more difficult the test was expected to be. A closer look at this pattern revealed that it was only present for participants in the high-difficulty-expectation group. For this group, difficulty expectations negatively correlated with online TUTs, $r(52) = -.31, p = .011$ (one-tailed), and, numerically identically, with retrospectively reported TUTs, $r(52) = -.31, p = .011$ (one-tailed). For the low-difficulty-expectation group, we found close-to-zero correlations.

To test for further group differences in the behavioral measures, we first ran a manipulation check. A 2×2 analysis of variance (ANOVA) with experimental condition (high- vs. low-difficulty expectation) as a between-participants factor and point in time (before vs. after test) as a within-participants factor for comprehension test difficulty estimates revealed a significant difference between expected (before test, $M = 6.42, SD = 2.01$) and perceived (after test, $M = 5.46, SD = 1.71$) difficulty, $F(1, 105) = 16.68, p < .001, \eta_p^2 = .14$. The main effect of experimental condition was also significant, $F(1, 105) = 33.22, p < .001, \eta_p^2 = .24$, with higher difficulty estimates for high-difficulty-expectation participants ($M = 6.57, SD = 1.12$) than for low-difficulty-expectation participants ($M = 5.30, SD = 1.17$). A significant interaction, $F(1, 105) = 72.49, p < .001, \eta_p^2 = .19$, and additional simple effects revealed that a group difference was only present for expected difficulty estimates, $F(1, 105) = 62.27, p < .001, \eta_p^2 = .37$, and not for perceived difficulty estimates, $F(1, 105) < 1, p = .734, \eta_p^2 < .01$. High-difficulty-expectation participants ($M = 7.63, SD = 1.41$) expected the comprehension test to be more difficult than did low-difficulty-expectation participants ($M = 5.19, SD = 1.77$), indicating a successful expectation manipulation. However, contradictory to our predictions, we did not find any significant group differences in online TUTs, online TTIs, retrospective TUTs, retrospective TTIs, or comprehension accuracy, all $t_s \leq 1.1$, all $p_s > .270$. Taken together, although the participants in the high-difficulty-expectation condition expected the comprehension test to be more difficult than did those in the low-difficulty-expectation condition, we did not find corresponding differences regarding either their mind-wandering behavior during the reading phase or their test performance. Therefore, this manipulation was not successful in influencing mind-wandering behavior in

Table 2 Means, standard deviations, and correlations for the behavioral measures

Measure	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7
1. Comprehension accuracy (%)	80.87	13.18	–	[–.21, .17]	[–.48, –.13]	[–.37, –.03]	[–.07, .28]	[–.34, .03]	[–.22, .17]
2. Expected difficulty (1 to 10)	6.42	2.01	–.03	–	[–.22, .18]	[–.33, .06]	[–.06, .32]	[–.34, .01]	[–.08, .29]
3. Perceived difficulty (1 to 10)	5.46	1.71	–.31*	–.02	–	[–.12, .26]	[–.24, .15]	[–.18, .23]	[–.33, .04]
4. Online task-unrelated thoughts (TUTs, %)	17.57	14.46	–.19*	–.15	.08	–	[–.30, .11]	[.58, .79]	[–.14, .25]
5. Online task-triggered intrusions (TTIs, %)	31.96	13.35	.12	.15	–.05	–.11	–	[–.22, .16]	[.42, .65]
6. Retrospective TUTs (%)	16.42	13.74	–.16*	–.17*	.04	.70*	–.04	–	[–.16, .23]
7. Retrospective TTIs (%)	27.75	13.68	–.03	.11	–.15	.06	.54*	.02	–

Values in parentheses indicate the scale of the measure. Values for the comprehension test's expected and perceived difficulty relate to a 10-point scale from *very easy* (1) to *very difficult* (10). Only in this table, for comparability reasons, we report online TUTs and TTIs in the form of percentages. Pearson correlations are displayed below, and bootstrapped 95% confidence intervals for the correlation coefficients are displayed above the diagonal. The numbers in the top row of the table refer to the corresponding measure in the first column. * $p \leq .050$ (one-tailed, in the predicted direction)

Italic values indicate descriptive statistics for the single measures. Bold values highlight significant correlations between measures

any way.⁴ On that account, we refrained from testing the preregistered mediating effect of mind wandering on reading comprehension.

Eye-tracking analyses

Comparison of eye-movement measures during normal reading versus TUT episodes We used the BeGaze software's algorithm to calculate all eye-movement measures of interest (Table 3) from the raw eye-tracking data. For each eye-tracking measure, up to ten data points (depending on data exclusions) per participant were considered for the following analyses. For the within-word regression count, between-word regression count, blink count, and fixation count, the number of observed occurrences of the respective event (i.e., regressions, blinks, or fixations) during a given target sentence and for a given participant represents one data point. For reading time, the time it took a participant to read a given target sentence represents one data point. Last, for first-fixation duration, the mean of the entirety of a participant's first fixations in one target sentence represents a data point. That is, all values are to be interpreted sentence-wise (e.g., two blinks per sentence), except for first-fixation duration, which is to be interpreted word-wise (e.g., an average first-fixation duration of 150 ms for all words in one target sentence).

Table 3 shows the means (aggregated across target sentences and participants) and standard deviations for all relevant unstandardized and uncorrected eye-movement measures for TRT and TUT episodes, respectively. Before the sentence-wise calculated eye-movement measures were entered in the analysis, they were divided by the respective target sentence's character count (to account for different sentence lengths) and z -standardized (to make the estimated coefficients comparable). The first-fixation duration measure was also z -standardized.

⁴ An additional multilevel regression analysis for TUT occurrences, with the predictors condition (effect-coded), difficulty expectation (as continuous predictor), and their interaction, also showed no effect of difficulty expectancies on TUTs. We thank Jonathan Smallwood for suggesting this additional analysis.

To test for differences in eye-movement measures between TUT and TRT episodes, we used a multilevel modeling approach. We chose this approach for several reasons: First, it can account for dependencies in the data due to repeated measures. In our case, eye-movement measures and thought-probe responses were nested in target sentences and participants. Second, because mind wandering is known to become more likely, the more time is spent on a task (e.g., Foulsham et al., 2013; McVay & Kane, 2009; Rummel & Nied, 2017; Steindorf & Rummel, 2017), we also considered time on task as an additional predictor. Because, analogously to previous studies, a strong time-on-task effect was apparent in the present data, we decided, although this step was not preregistered, to *detrend* the present data (i.e., to remove the trend from a time series; Wang & Maxwell, 2015; Wu, Huang, Long, & Peng, 2007). A third advantage of multilevel models is that parameters can be estimated despite missing data—for example, due to unusable eye-tracking data. Finally, using a logit-link function, we can account for the dichotomous nature of our dependent variable (TRT vs. TUT responses). Also in contrast to our preregistration, we did not specify a model containing all eye-movement measures of interest as mind-wandering predictors, due to multicollinearity problems. Instead, we specified six separate multilevel models, one for each eye-movement measure.

We included crossed random intercepts varying with target sentences and participants⁵ and regressed TUTs⁶ (with 1 indicating TUTs and 0 indicating TRTs) on the respective eye-

⁵ Adding random slopes did not improve the model fits, and thus we chose the more parsimonious approach.

⁶ We conducted our analyses excluding all target sentences categorized as TTI episodes. In the literature, there is no common agreement as to how to categorize and treat TTIs (they might reflect a thought mode between on-task and off-task thoughts). Additionally, the studies that have examined mind wandering using eye-tracking have not differentiated between TUTs and TTIs. Using a compound TUTs + TTIs score as criterion, one more eye-movement measure proved to be a significant predictor (between-word regression count, $b = 0.15$, Wald $Z = 2.10$, $p = .036$) than in the analyses with only TUTs as criterion. The results for the other eye-movement measures, however, did not change.

Table 3 Descriptive statistics for the eye-movement measures of interest for target sentences categorized as either task-related thought (TRT) or task-unrelated thought (TUT) episodes

Eye-tracking measure	M_{TRT}	SD_{TRT}	M_{TUT}	SD_{TUT}
Within-word regression count	1.52	1.60	1.49	1.68
Between-word regression count	3.17	2.60	3.16	2.41
Blink count	2.86	2.85	2.69	2.30
Reading time (ms)	9,526.67	3,624.36	9,947.95	3,249.09
Fixation count	33.26	12.16	34.88	11.26
First-fixation duration (ms)	198.46	32.15	205.68	31.25

Values are to be interpreted sentence-wise (e.g., on average, participants exhibited 34.88 fixations per mindlessly read target sentence), except those for first-fixation duration, which is to be interpreted word-wise (e.g., on average, participants' first-fixation durations on mindfully read single words lasted 198.46 ms). To calculate the means, we aggregated across participants and target sentences that were categorized as TRT or TUT episodes, respectively.

movement measure (Level 1), time on task (Level 1; a variable ranging from 0 to 9, depicting the order of thought collection), and the effect-coded experimental condition (Level 2). We included time on task as a Level 1 predictor in order to detrend (see above and Wang & Maxwell, 2015) the outcome variable, to investigate the relationship between the eye-movement measures and TUT occurrences above and beyond the systematic increase of TUT occurrences over time. The coefficients, estimated using the `glmer` function of the R package `lme4` (Bates, Maechler, Bolker, & Walker, 2015) and the `bobyqa` optimizer for some of the models (if this was necessary to achieve convergence; Powell, 2009), are depicted in Table 4. Reading time, fixation count, and first-fixation duration significantly predicted TUT occurrences. Positive b coefficients for these measures indicated longer reading times, longer first-fixation durations, and more fixations during TUT than during TRT episodes, which is also descriptively apparent in Table 3. Within-word regression count, between-word regression count, and blink count did not predict TUT occurrences. In all multilevel logistic regression models, time on task significantly predicted TUTs (all p values < .001) with the likelihood for a TUT occurrence generally increasing over time. The effect-coded experimental condition remained a nonsignificant predictor in all specified models (all p values > .05). Thus, comparable to our results on the behavioral measures, we did not find evidence that the difficulty expectation manipulation affected the eye-movement measures.

Word-frequency effect for normal reading versus TUT episodes To test whether the word-frequency effect was less pronounced during TUT episodes, we compared eye movements on high- versus low-frequency target words. We expected an interaction between word frequency (low vs. high) and thought mode (TUT vs. TRT), in terms of smaller differences in the eye-movement measures between high- and low-frequency target words during TUT than during TRT episodes. We employed multilevel regression models using the `lme` function in the R package `nlme` (Pinheiro, Bates, DebRoy,

Sarkar, & R Development Core Team, 2010). Word frequency and thought mode (all effect-coded) entered the analyses as fixed predictors for the eye-movement measures, while allowing intercepts to vary with participant and target word identity (cf. Baayen, Davidson, & Bates, 2008). In doing so, we simultaneously controlled for participant-specific and target-word-specific variability in the eye-movement measures. Again, we applied separate regression models for the prediction of gaze duration (sum of all durations of all fixations on a target word during the first pass), total viewing time (sum of all durations of all fixations on a target word, including fixations following a regression), fixation count (total number of fixations on a target word), and regressions-into-target-word count (total number of regressions back to a target word). Table 5 shows estimates for the main and interaction effects. For thought mode, we did not find significant main effects. A significant main effect of word frequency, with negative b coefficients, represents the word-frequency effect. High-frequency target words, in comparison to low-frequency target words, were fixated for a shorter time (see gaze duration and total viewing time in the table) and less often (see fixation count). Therefore, we found the typical word-frequency effect with three eye-movement measures. Regressions into target words were rare (see Fig. 1) and did not show a reliable word-frequency effect.

To test whether the word-frequency effect became smaller during TUT as compared to TRT episodes, the interaction between word frequency and thought mode is crucial (see Table 5). Figure 1 additionally illustrates the interaction patterns. Descriptively, the word-frequency effect appeared to be somewhat smaller during TUT than during TRT episodes for all analyzed eye-movement measures.⁷ However, only the fixation count measure yielded a significant interaction.

⁷ For the regressions-into-target-word count, the word-frequency effect reversed during TUT episodes. Because these regressions were very rare and the measure did not produce a reliable word-frequency effect, we refrain from strongly interpreting this finding.

Table 4 Eye-tracking measures (Level 1) predicting TUT occurrences in separate multilevel logistic regressions, also including time on task (Level 1) and effect-coded experimental condition (Level 2) as further predictors

Predictor	<i>b</i>	<i>SE_b</i>	95% CI	Wald Z	<i>p</i>
Within-word regression count	0.03	0.11	[- 0.19, 0.24]	0.29	.769
Between-word regression count	0.11	0.12	[- 0.12, 0.35]	0.98	.328
Blink count	- 0.11	0.13	[- 0.37, 0.14]	- 0.82	.413
Reading time (ms)	0.27	0.12	[0.03, 0.52]	2.24	.025
Fixation count	0.35	0.12	[0.12, 0.60]	2.95	.003
First-fixation duration (ms)	0.24	0.11	[0.01, 0.46]	2.11	.035

Intercepts were allowed to vary with target sentences and participants. In three of the six regression models, the eye-tracking measure of interest (reading time, fixation count, or first-fixation duration) positively predicted TUT occurrences. In all of the models, time on task significantly predicted TUTs (all *p* values < .001), with the likelihood for a TUT occurrence generally increasing over time. The effect-coded experimental condition remained nonsignificant in all of the models (all *p* values > .05). *SE* = standard error, *CI* = confidence interval

Discussion

Identifying objective measures of mind wandering in addition to widely used subjective self-reports is an important endeavor but is still a work in progress: Different authors have determined a large variety of eye-movement measures as mind-wandering indicators. However, as we reviewed in the introduction, there is little consensus as to the specific measures that are sensitive to attentional decoupling during mindless reading. To address this inconsistency problem, we conducted a new, high-powered eye-tracking experiment in which we reanalyzed all previously identified mind-wandering indicators. In the present work, only three measures (reading time, fixation count, and first-fixation duration) positively predicted self-categorized mindless reading. Interestingly, aside from these single measures, the word-frequency effect was found

to be less pronounced during mindless-reading than during normal-reading episodes.

In addition to our validation efforts, we tested the idea that mind wandering might be adjusted to expectancies about task demands. More precisely, we expected participants in expectation of a difficult reading comprehension test to mind wander less while reading than would participants with low-difficulty expectations. Furthermore, we expected less mind wandering to result in better performance on the test. Indeed, we found that having fewer off-task thoughts went along with higher comprehension accuracy in the reading test, but we did not find significant group differences in test performance. Additionally, although we did successfully induce opposing expectancies about the task demands, the groups did not differ regarding their mind-wandering behavior. However, for participants anticipating a difficult test, higher difficulty expectations (assessed prior to the

Table 5 Main and interaction effects of four multilevel regressions, each predicting an eye-movement measure by word frequency (low vs. high), thought mode (TUT vs. TRT), and the interaction of both

Dependent variable	Predictor	<i>b</i>	<i>SE_b</i>	95% CI	<i>t</i>	<i>p</i>
Gaze duration	Word frequency	- 45.50	8.38	[- 61.96, - 29.04]	- 5.43	< .001
	Thought mode	- 9.49	8.59	[- 26.37, 7.40]	- 1.10	.270
	Word frequency × Thought mode	- 7.94	8.43	[- 24.49, 8.62]	- 0.94	.347
Total viewing time	Word frequency	- 71.26	10.45	[- 91.79, - 50.73]	- 6.82	< .001
	Thought mode	6.22	10.94	[- 15.29, 27.73]	0.57	.570
	Word frequency × Thought mode	- 16.93	10.56	[- 37.68, 3.81]	- 1.60	.109
Fixation count	Word frequency	- 0.24	0.05	[- 0.33, - 0.15]	- 5.23	< .001
	Thought mode	0.00	0.05	[- 0.09, 0.09]	- 0.00	.996
	Word frequency × Thought mode	- 0.12	0.05	[- 0.21, - 0.03]	- 2.63	.009
Regressions into target word	Word frequency	- 0.01	0.01	[- 0.04, 0.02]	- 0.53	.597
	Thought mode	- 0.02	0.01	[- 0.05, 0.01]	- 1.54	.124
	Word frequency × Thought mode	- 0.02	0.01	[- 0.05, 0.01]	- 1.41	.159

Intercepts were allowed to vary with target words and participants. Three of the four regression models showed a word-frequency effect. That is, word frequency predicted the eye-movement measure of interest (gaze duration, total viewing time, or fixation count). Critically, the interaction between word frequency and thought mode was a significant predictor of fixation count, indicating a reduced word-frequency effect during TUT as compared to TRT episodes (see Fig. 1). Regressions into the target words were rare and did not show a reliable word-frequency effect. *SE* = standard error, *CI* = confidence interval

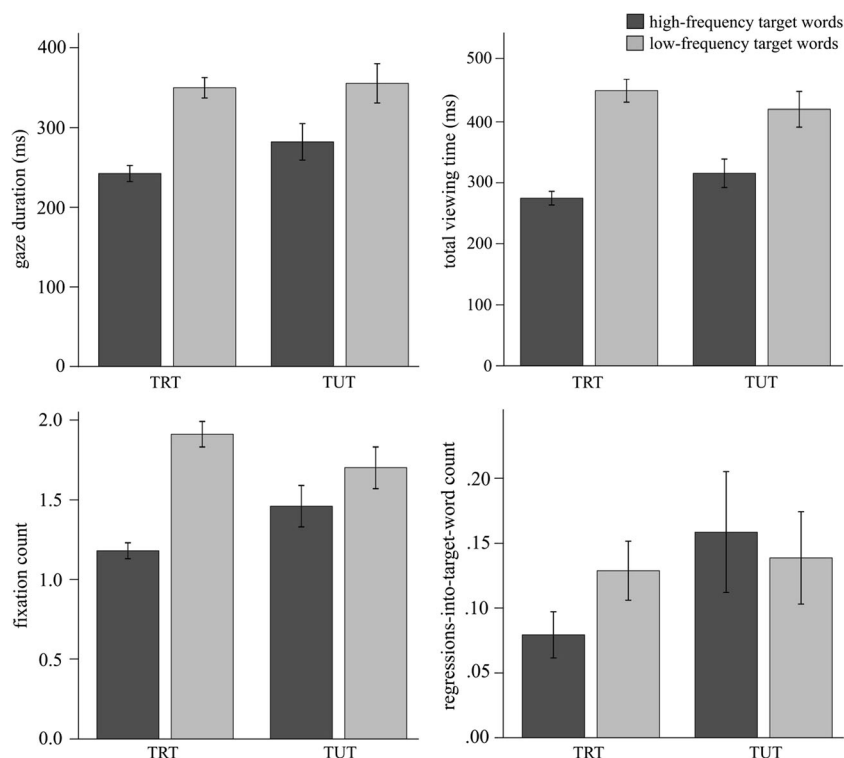


Fig. 1 Mean values for gaze duration (duration of all fixations on a target word during the first pass), total viewing time (sum of all fixations on a target word, including fixations following a regression), fixation count (total number of fixations on a target word), and regressions-into-target-word count (total number of regressions back to a target word) for high- and low-frequency target words during TRT and TUT episodes. Higher

bars for low-frequency than for high-frequency target words represent the word-frequency effect, which is descriptively smaller during TUT episodes. Because regressions into the target word were very rare and because the measure did not produce a reliable word-frequency effect, this measure was not interpreted further. Error bars represent standard errors of the means

test) were associated with less mind wandering during the reading task. In this group, their thoughts might have been influenced by difficulty expectations as an extrinsic motivational factor. Lacking extrinsic motivation to focus on the text, low-difficulty-expectation participants' thoughts might have been primarily influenced by intrinsic motivational factors such as interest in the topic, so that we could not find a relationship between difficulty expectancies and mind wandering. However, future research will be necessary to test this assumption. Indeed, the relationship between meta-awareness and mind-wandering behavior might be more complicated than has been assumed. In a study by Sanders, Wang, Schooler, and Smallwood (2017), directly enhancing meta-awareness of mind wandering improved task focus, but not reading comprehension, and only after a period of self-focus.

Regarding the candidate mind-wandering indicators, all three eye-movement measures that we found to be sensitive to attentional decoupling are related to reading speed. While reading mindlessly, our participants seemed to “slow down”: Their fixation durations were longer, and they exhibited more fixations. Consequently, they needed more time to “read” a sentence. Changes in reading/processing time are often found to be related to mind wandering. For example, Reichle et al. (2010) also reported increased gaze durations and increased total viewing times

during off-task thought episodes. Additionally, research applying the z-string reading paradigm (Nuthmann & Engbert, 2009; Rayner & Fischer, 1996) has suggested that fixation durations are longer during mindless reading. Recently, other authors have made use of these findings and developed a machine-learned model that is able to discriminate between (self-caught) mindful and mindless reading on the basis of eye movements (Faber, Bixler, & D’Mello, 2018). Such deceleration of cognitive processes during mind-wandering episodes is observed not only for reading tasks. During word-learning tasks, off-task processing, or decoupling from the task, was found to be indexed by slower reaction times (Smallwood, O’Connor, Sudbery, & Obonsawin, 2007). Also, during simple go/no-go tasks, mind wandering has a prolonging effect on processing time and goes along with higher proportions of extremely long reaction times (McVay & Kane, 2012). Franklin, Smallwood, and Schooler (2011) used the relationship between mind wandering and temporal changes to develop an algorithm that successfully predicted participants’ mind-wandering reports during word-by-word reading. However, the authors found that participants sped up during mind-wandering episodes. They proposed that this substantial difference from the results reported by Reichle et al. (2010) might be due to paradigmatic differences (word-by-word reading vs. naturalistic reading). A similar speed-up is often found in the

SART when participants are mind wandering (Smallwood et al., 2004). Thus, the evidence has been mixed concerning the direction of the mind-wandering–processing-time relationship, and further research will be needed.

Franklin et al.'s (2011) algorithm is based on the idea of different sensitivities to the lexical qualities of words during mind wandering than during on-task attention. They found people to be less affected by word features such as length and frequency during mindless reading. This idea is also supported by Reichle et al. (2010), Foulsham et al. (2013), and now, additionally, by our results. We also found a weaker word-frequency effect during mindless reading, although this reduction was only significant for the fixation-count measure. Also, from a theoretical standpoint, a reduced word-frequency effect should be a good indicator of attentional decoupling during mindless reading: While people's minds wander, they naturally disengage from the lexical input in front of them and are thus likely to ignore word features that usually affect processing time. Please note, however, that we selected target words for the present study that differed considerably in their frequency but were matched in other respects (length). Whereas this approach is optimal for demonstrating a mind-wandering-related reduction of the word-frequency effect, future research is needed to determine to which degree the effects observed in this study are stimulus dependent and whether a meaningful reduction would still occur with weaker word-frequency manipulations.

In most studies examining mindless reading, probes appeared randomly during full-text reading, resulting in different to-be-analyzed text passages for each participant. Only in the Foulsham et al. (2013) study, single-sentence reading allowed for the manipulation of linguistic input. These authors compared high- and low-frequency target words during mindless versus normal reading. We applied a similar approach and also defined target sentences with specific features (high- vs. low-frequency words). However, in contrast to the Foulsham et al. study, our participants were able to engage in naturalistic reading. This was achieved by letting participants' fixations on the first words of the sentence following a target sentence trigger thought probe appearances. Consequently, we could analyze predefined target sentences and words even though our participants engaged in full text reading. By keeping the to-be-examined passages equal for all participants while still enabling naturalistic reading, our method combines the advantages of both full-text and single-sentence reading. We therefore recommend this method for future research.

We also recommend using multilevel models in future research. In addition to their general advantages (see the [Results](#) section), they allow for the specific modeling of time-on-task effects that are often found for mind-wandering behavior (e.g., Foulsham et al., 2013; McVay & Kane, 2009; Rummel & Nied, 2017; Steindorf & Rummel, 2017). As a result, the effects of the variables of interest can be examined while controlling for systematic changes due solely to time on task. Additionally, mind-

wandering research has to deal with data that are not normally distributed. This issue can be resolved by specifying responses to thought probes as a dichotomous (or categorical) dependent variable in the multilevel framework.

As we discussed in our literature review in the introduction, the research using eye-movement measures as mind-wandering indicators is far from being coherent: Previous studies have differed regarding the lexical input that was used, the reading time, the number of participants, the analyzed areas of interest, the thought-probing procedures, and the types of statistical analyses. However, even when the methods were relatively comparable, different authors analyzed different measures and found divergent effects, and even effects in different directions, for the same measure. We hope that the present work will aid in the development of robust and stable indicators of mindless reading, and we believe that it also emphasizes the importance of validation research. In general, we believe that when a field is growing as rapidly as the mind-wandering area (Schooler et al., 2014; Seli, Risko, Smilek, & Schacter, 2016), researchers should not only focus on conducting more research, but also on better connecting new research with the existing research. An eye-movement measure that only proves to be a valid mind-wandering predictor in one study under unique conditions will not help the field to move forward.

Special considerations apply to the use of within-word regressions as a mind-wandering indicator. Uzzaman and Joordens (2011) observed fewer within-word regressions during mindless reading than during normal reading, whereas we did not detect any such differences. Of course, this inconsistency might be due to differences in the specific methods applied (language, sample, font size, etc.). However, because a valid mind-wandering indicator should prove stable across such conditions (as long as the methods are fairly comparable), the present study at least suggests that within-word regressions might be a less reliable indicator of mind wandering. Additionally, with our monitor characteristics and font size, medium-sized words had a length of 2°–3°. Since we applied a velocity threshold of 40°/s to identify saccades, the detection of within-word regressions could be problematic, due to technical issues. Some words in our target sentences might have simply been too small for us to detect backward movements, rendering the measure unreliable.⁸ We conclude that future research will be needed to determine the usefulness of within-word regressions to predict mind wandering, but that measures that strongly depend on technical characteristics will probably not become first-choice indicators.

In sum, our study is only one among those that have assessed eye movements in the hope of detecting mind-wandering indicators, wherefore our results must be considered as preliminary. Moreover, our manipulation of TUT levels via task-demand expectancies, which could have provided a further test of the convergent validity of the eye-movement mind-wandering

⁸ We thank Stephanie Huetten for raising this point.

indicators, was not successful. We therefore encourage researchers who plan on conducting mind-wandering studies and assessing eye movements to incorporate better convergent-validity tests. Furthermore, we encourage them not only to report the data for indicators that turned out to be significant in their particular study, but to report results for all candidate indicators, as we did here. At the current stage of this field of research, nonsignificant results are just as important as significant results in determining whether eye-tracking methods are a valid tool for mind-wandering assessment.

Whereas the early studies were important for identifying eye movements as a potential mind-wandering proxy, future research will also have to focus on the role that methodological details (e.g., those listed in Table 1) play in the divergent results observed so far. Valid indicators of mind wandering will, in the end, have to converge across studies imposing comparable contextual demands. Additionally, plausible boundary conditions need to be identified and tested for those cases in which indicators do *not* converge. To achieve this goal, full transparency concerning methods used and the (expected) results—optimally, preregistered before a study is conducted—will be key. In line with these recommendations, we hope for future research to fill in the gaps that still exist concerning the validation of eye-movement measures as indicators for mindless reading.

We believe that mind-wandering research would greatly benefit if eye-movement measures were found to be robust and stable indicators of mindless reading. So far, subjective self-reports are the state-of-the-art assessment tool for mind wandering. Not only might answers to thought probes be prone to memory-driven and classification errors, but the probing procedure itself might also interrupt a participant's train of thought and impair task performance. Eye-movement measures as

nonintrusive, objective mind-wandering indicators could circumvent these problems for reading tasks. Similarly, pupillometry represents a promising method for predicting TUTs during reading. Pupil dilation has been found to be higher prior to mindless reading than in normal-reading episodes (Franklin, Broadway, Mrazek, Smallwood, & Schooler, 2013; Smallwood et al., 2011; but see Grandchamp, Braboszcz, & Delorme, 2014). Eye-movement measures and pupil dilation could bring us one step closer to real-time detection of disruptive mind-wandering behavior during reading.

Additionally, such measures may even have the potential to be used to prevent mind wandering in situations in which it is not only disruptive, but even dangerous. Driver inattention due to mind wandering, for example, poses a serious threat to road safety: Driving simulation studies have shown that driver behavior is indeed affected by participants' thought mode (Baldwin et al., 2017). Aside from mind wandering, fatigue is a major cause of traffic accidents, and eye-state biofeedback methods have already been tested in cars to prevent drivers from falling asleep (e.g., Devi & Bajaj, 2008). Detecting TUT episodes by means of eye movements or pupillometry, in addition to fatigue detection, could be in the interest of road safety. Still, as we pointed out above, the validation of these biological markers is a work in progress. We hope that our work will drive the validation process forward, so that eye movements can be used as indicators for mindless reading in the not-too-distant future.

Author note The present research was supported by a grant from the German Research Foundation (DFG) to the second author (Grant No. RU1996/1-1). The authors thank Nadine Gronewold, Amelie Haindl, Fynn Ole Wöstenfeld, Fabian Dittmar, and Holly Hammerton for their help in data collection, with special thanks to Holly for also proofreading the manuscript.

Appendix

Table 6 Target words used to examine the word-frequency effect during episodes of mindless versus normal reading, including their English translation and information on frequency and character count

Frequency	Target word (in German)	English translation	Type frequency	Character count	Mean frequency	Mean character count
Low	Chanukka	Hanukkah	4	8	8.40	10.60
	Musikologe	musicologist	1	10		
	Imaginieren	imagination	9	11		
	Pathologien	pathologies	27	11		
	Frontallappen	frontal lobe	1	13		
High	Gehirn	brain	2,146	6	1,892.80	10.40
	Eifersucht	jealousy	882	10		
	Strukturen	structures	2,229	10		
	Gesundheit	health	3,773	10		
	Eigentümlichkeit	peculiarity	434	16		

Word frequencies (type frequencies) were extracted from the dlex database (obtained under <http://dlexdb.de/>), which constitutes a corpus of the German language consisting of different text sources such as newspapers, fiction books, and scientific works (Heister et al., 2011)

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

References

- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Baird, B., Smallwood, J., Mrazek, M. D., Kam, J. W. Y., Franklin, M. S., & Schooler, J. W. (2012). Inspired by distraction: Mind wandering facilitates creative incubation. *Psychological Science*, *23*, 1117–1122. <https://doi.org/10.1177/0956797612446024>
- Baldwin, C. L., Roberts, D. M., Barragan, D., Lee, J. D., Lemer, N., & Higgins, J. S. (2017). Detecting and quantifying mind wandering during simulated driving. *Frontiers in Human Neuroscience*, *11*, 406. <https://doi.org/10.3389/fnhum.2017.00406>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Cho, H.-C., & Abe, S. (2013). Is two-tailed testing for directional research hypotheses tests legitimate? *Journal of Business Research*, *66*, 1261–1266. <https://doi.org/10.1016/j.jbusres.2012.02.023>
- Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., & Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proceedings of the National Academy of Sciences*, *106*, 8719–8724. <https://doi.org/10.1073/pnas.0900234106>
- Devi, M. S., & Bajaj, P. R. (2008, July). *Driver fatigue detection based on eye tracking*. Paper presented at the 2008 First International Conference on Emerging Trends in Engineering and Technology, Nagpur, India.
- Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, *112*, 777–813. <https://doi.org/10.1037/0033-295X.112.4.777>
- Faber, M., Bixler, R., & D'Mello, S. K. (2018). An automated behavioral measure of mind wandering during computerized reading. *Behavior Research Methods*, *50*, 134–150. <https://doi.org/10.3758/s13428-017-0857-y>
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*, 906–911. <https://doi.org/10.1037/0003-066X.34.10.906>
- Foulsham, T., Farley, J., & Kingstone, A. (2013). Mind wandering in sentence reading: Decoupling the link between mind and eye. *Canadian Journal of Experimental Psychology*, *67*, 51–59. <https://doi.org/10.1037/a0030217>
- Frank, D. J., Nara, B., Zavagnin, M., Touron, D. R., & Kane, M. J. (2015). Validating older adults' reports of less mind-wandering: An examination of eye movements and dispositional influences. *Psychology and Aging*, *30*, 266–278. <https://doi.org/10.1037/pag0000031>
- 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*, 2289–2294. <https://doi.org/10.1080/17470218.2013.858170>
- Franklin, M. S., Smallwood, J., & Schooler, J. W. (2011). Catching the mind in flight: Using behavioral indices to detect mindless reading in real time. *Psychonomic Bulletin & Review*, *18*, 992–997. <https://doi.org/10.3758/s13423-011-0109-6>
- Grandchamp, R., Braboszcz, C., & Delorme, A. (2014). Oculometric variations during mind wandering. *Frontiers in Psychology*, *5*, 31. <https://doi.org/10.3389/fpsyg.2014.00031>
- Heister, J., Würzner, K. M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., & Kliegl, R. (2011). dlexDB—A lexical database for the psychological and linguistic research. *Psychologische Rundschau*, *62*, 10–20. <https://doi.org/10.1026/0033-3042/a000029>
- Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, *40*, 431–439. <https://doi.org/10.3758/bf03208203>
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*, 329–354. <https://doi.org/10.1037/0033-295X.87.4.329>
- 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*, 1271–1289. <https://doi.org/10.1177/0956797617706086>
- Killingsworth, M. A., & Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science*, *330*, 932–932. <https://doi.org/10.1126/science.1192439>
- Klinger, E. (1999). Thought flow: Properties and mechanisms underlying shifts in content *At play in the fields of consciousness: Essays in honor of Jerome L. Singer* (pp. 29–50). Mahwah, NJ, US: Erlbaum.
- Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*, 86–92. <https://doi.org/10.1027/1614-2241.1.3.86>
- 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*, 196–204. <https://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*, 525–549. <https://doi.org/10.1037/a0025896>
- Mooneyham, B. W., & Schooler, J. W. (2013). The costs and benefits of mind-wandering: A review. *Canadian Journal of Experimental Psychology*, *67*, 11–18. <https://doi.org/10.1037/a0031569>
- Nuthmann, A., & Engbert, R. (2009). Mindless reading revisited: An analysis based on the SWIFT model of eye-movement control. *Vision Research*, *49*, 322–336. <https://doi.org/10.1016/j.visres.2008.10.022>
- Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior Research Methods*, *42*, 188–204. <https://doi.org/10.3758/brm.42.1.188>
- Paccagnella, O. (2011). Sample size and accuracy of estimates in multi-level models. *Methodology*, *7*, 111–120. <https://doi.org/10.1027/1614-2241/a000029>
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Development Core Team. (2010). nlme: Linear and nonlinear mixed effects models. Retrieved from <https://CRAN.R-project.org/package=nlme>
- Pollatsek, A., Reichle, E. D., & Rayner, K. (2006). Tests of the E-Z Reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology*, *52*, 1–56. <https://doi.org/10.1016/j.cogpsych.2005.06.001>
- Powell, M. J. D. (2009). The BOBYQA algorithm for bound constrained optimization without derivatives (Cambridge NA Report 2009/NA06, pp. 26–46). Cambridge, UK: Cambridge Centre for Mathematical Sciences.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422. <https://doi.org/10.1037/0033-2909.124.3.372>
- Rayner, K., & Fischer, M. H. (1996). Mindless reading revisited: Eye movements during reading and scanning are different. *Perception & Psychophysics*, *58*, 734–747. <https://doi.org/10.3758/bf03213106>

- Reichle, E. D., Reineberg, A. E., & Schooler, J. W. (2010). Eye movements during mindless reading. *Psychological Science*, *21*, 1300–1310. <https://doi.org/10.1177/0956797610378686>
- Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). 'Oops!': Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, *35*, 747–758. [https://doi.org/10.1016/S0028-3932\(97\)00015-8](https://doi.org/10.1016/S0028-3932(97)00015-8)
- Rummel, J., & Boywitt, C. D. (2014). Controlling the stream of thought: Working memory capacity predicts adjustment of mind-wandering to situational demands. *Psychonomic Bulletin & Review*, *21*, 1309–1315. <https://doi.org/10.3758/s13423-013-0580-3>
- Rummel, J., & Nied, L. (2017). Do drives drive the train of thought?—Effects of hunger and sexual arousal on mind-wandering behavior. *Consciousness and Cognition*, *55*, 179–187. <https://doi.org/10.1016/j.concog.2017.08.013>
- Sacks, O. (2008). Fantasie nach Noten: Über musikalische Vorstellungskraft und die Macht der Töne. *Gehirn und Geist*, *6*, 48–55.
- Sanders, J. G., Wang, H.-T., Schooler, J., & Smallwood, J. (2017). Can I get me out of my head? Exploring strategies for controlling the self-referential aspects of the mind-wandering state during reading. *Quarterly Journal of Experimental Psychology*, *70*, 1053–1062. <https://doi.org/10.1080/17470218.2016.1216573>
- Schad, D. J., Nuthmann, A., & Engbert, R. (2012). Your mind wanders weakly, your mind wanders deeply: Objective measures reveal mindless reading at different levels. *Cognition*, *125*, 179–194. <https://doi.org/10.1016/j.cognition.2012.07.004>
- Schooler, J. W., Mrazek, M. D., Franklin, M. S., Baird, B., Mooneyham, B. W., Zedelius, C., & Broadway, J. M. (2014). The middle way: Finding the balance between mindfulness and mind-wandering. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 60, pp. 1–33). Orlando, FL: Academic Press.
- Schooler, J. W., Reichle, E. D., & Halpern, D. V. (2004). Zoning out during reading: Evidence for dissociations between experience and metacognition. In D. T. Levin (Ed.), *Thinking and seeing: Visual metacognition in adults and children* (pp. 204–226). Cambridge, MA: MIT Press.
- Schooler, J. W., Smallwood, J., Christoff, K., Handy, T. C., Reichle, E. D., & Sayette, M. A. (2011). Meta-awareness, perceptual decoupling and the wandering mind. *Trends in Cognitive Sciences*, *15*, 319–326. <https://doi.org/10.1016/j.tics.2011.05.006>
- Seli, P., Risko, E. F., Smilek, D., & Schacter, D. L. (2016). Mind-wandering with and without intention. *Trends in Cognitive Sciences*, *20*, 605–617. <https://doi.org/10.1016/j.tics.2016.05.010>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 word solution. *Dialogue: The Official Newsletter of the Society for Personality and Social Psychology*, *26*(2), 4–7. <https://doi.org/10.2139/ssrn.2160588>
- Smallwood, J. (2011). Mind-wandering while reading: Attentional decoupling, mindless reading and the cascade model of inattention. *Language and Linguistics Compass*, *5*, 63–77. <https://doi.org/10.1111/j.1749-818X.2010.00263.x>
- 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*, 441. <https://doi.org/10.3389/fpsyg.2013.00441>
- Smallwood, J., Baracaia, S. F., Lowe, M., & Obonsawin, M. (2003). Task unrelated thought whilst encoding information. *Consciousness and Cognition*, *12*, 452–484. [https://doi.org/10.1016/S1053-8100\(03\)00018-7](https://doi.org/10.1016/S1053-8100(03)00018-7)
- Smallwood, J., Brown, K. S., Tipper, C., Giesbrecht, B., Franklin, M. S., Mrazek, M. D., . . . Schooler, J. W. (2011). Pupillometric evidence for the decoupling of attention from perceptual input during offline thought. *PLoS ONE*, *6*, e18298. <https://doi.org/10.1371/journal.pone.0018298>
- Smallwood, J., Davies, J. B., Heim, D., Finnigan, F., Sudberry, M., O'Connor, R., & Obonsawin, M. (2004). Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness and Cognition*, *13*, 657–690. <https://doi.org/10.1016/j.concog.2004.06.003>
- Smallwood, J., McSpadden, M., & Schooler, J. W. (2008). When attention matters: The curious incident of the wandering mind. *Memory & Cognition*, *36*, 1144–1150. <https://doi.org/10.3758/mc.36.6.1144>
- Smallwood, J., O'Connor, R. C., Sudbery, M. V., & Obonsawin, M. (2007). Mind-wandering and dysphoria. *Cognition and Emotion*, *21*, 816–842. <https://doi.org/10.1080/02699930600911531>
- Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin*, *132*, 946–958. <https://doi.org/10.1037/0033-2909.132.6.946>
- Smeekens, B. A., & Kane, M. J. (2016). Working memory capacity, mind wandering, and creative cognition: An individual-differences investigation into the benefits of controlled versus spontaneous thought. *Psychology of Aesthetics, Creativity, and the Arts*, *10*, 389–415. <https://doi.org/10.1037/aca0000046>
- Smilek, D., Carriere, J. S. A., & Cheyne, J. A. (2010). Out of mind, out of sight: Eye blinking as indicator and embodiment of mind wandering. *Psychological Science*, *21*, 786–789. <https://doi.org/10.1177/0956797610368063>
- 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*, e16997. <https://doi.org/10.1371/journal.pone.0016997>
- Steindorf, L., & Rummel, J. (2017). "I should not forget the apples!"—Mind-wandering episodes used as opportunities for rehearsal in an interrupted recall paradigm. *Applied Cognitive Psychology*, *31*, 424–430. <https://doi.org/10.1002/acp.3328>
- Unsworth, N., & McMillan, B. D. (2014). Similarities and differences between mind-wandering and external distraction: A latent variable analysis of lapses of attention and their relation to cognitive abilities. *Acta Psychologica*, *150*, 14–25. <https://doi.org/10.1016/j.actpsy.2014.04.001>
- Uzzaman, S., & Joordens, S. (2011). The eyes know what you are thinking: Eye movements as an objective measure of mind wandering. *Consciousness and Cognition*, *20*, 1882–1886. <https://doi.org/10.1016/j.concog.2011.09.010>
- Vinski, M. T., & Watter, S. (2012). Priming honesty reduces subjective bias in self-report measures of mind wandering. *Consciousness and Cognition*, *21*, 451–455. doi.org/10.1016/j.concog.2011.11.001
- Wang, L., & Maxwell, S. E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological Methods*, *20*, 63–83. <https://doi.org/10.1037/met0000030>
- Weinstein, Y. (2018). Mind-wandering, how do I measure thee with probes? Let me count the ways. *Behavior Research Methods*, *50*, 642–661. <https://doi.org/10.3758/s13428-017-0891-9>
- Weinstein, Y., De Lima, H. J., & van der Zee, T. (2018). Are you mind-wandering, or is your mind on task? The effect of probe framing on mind-wandering reports. *Psychonomic Bulletin & Review*, *25*, 754–760. <https://doi.org/10.3758/s13423-017-1322-8>
- Wu, Z., Huang, N. E., Long, S. R., & Peng, C.-K. (2007). On the trend, detrending, and variability of nonlinear and nonstationary time series. *Proceedings of the National Academy of Sciences*, *104*, 14889–14894. <https://doi.org/10.1073/pnas.0701020104>
- Yang, S.-N. (2006). An oculomotor-based model of eye movements in reading: The competition/interaction model. *Cognitive Systems Research*, *7*, 56–69. <https://doi.org/10.1016/j.cogsys.2005.07.005>