



RESEARCH ARTICLE

Wandering eyes: Eye movements during mind wandering in video lectures

Han Zhang¹  | Kevin F. Miller^{1,2,3} | Xin Sun¹  | Kai S. Cortina^{1,2,3}

¹Combined Program in Education and Psychology, University of Michigan, Ann Arbor, Michigan

²School of Education, University of Michigan, Ann Arbor, Michigan

³Department of Psychology, University of Michigan, Ann Arbor, Michigan

Correspondence

Han Zhang, Combined Program in Education and Psychology, University of Michigan, Ann Arbor, Michigan, 610 E. University Avenue, Room 1413, Ann Arbor, MI 48109.
Email: hanzh@umich.edu

Summary

Video lectures are increasingly prevalent, but they present challenges to learners. Students' minds often wander, yet we know little about how mind wandering affects attention during video lectures. This paper presents two studies that examined eye movement patterns of mind wandering during video lectures. In the studies, mind wandering reports were collected by either self-caught reports or thought probes. Results were similar across the studies: mind wandering was associated with an increased allocation of fixations to the instructor's image. For fixations on the slides, the average duration increased but the dispersion decreased. Moreover, preliminary evidence suggested that fixation duration and dispersion can diminish soon after self-caught reports of mind wandering. Overall, these findings help advance our understanding of how learners' attention is affected during mind wandering and may facilitate efforts in objectively identifying mind wandering. Future research is needed to determine if these findings can extend to other instructional formats.

KEYWORDS

attention, eye tracking, growth curve analysis, mind wandering, video lecture

1 | ATTENTION DURING VIDEO LECTURES

Educational researchers have long expressed concern that students have difficulty maintaining attention during lectures (e.g., Bloom, 1953; Schoen, 1970). With the rapid expansion of online courses, the problem of keeping attention extends beyond traditional classrooms. Although video lectures make learning more accessible, they often rely on learners to regulate attention and thus leave room for attention failures (Schacter & Szpunar, 2015).

Video lectures may be particularly permissive of mind wandering (MW), an internal shift in the contents of thought away from the ongoing task to self-generated, task-irrelevant thoughts (Smallwood & Schooler, 2015, 2006). MW had been an underrecognized factor in education until recently (Smallwood, Fishman, & Schooler, 2007). Recent development of experience-sampling methods has rendered MW a measurable phenomenon (Smallwood & Schooler, 2015). The self-caught method asks participants to report MW whenever they catch themselves doing so, and the probe-caught method occasionally

interrupts participants during a task to report their current mental state. Both methods have been proven feasible and reasonably valid to obtain MW data (McVay & Kane, 2012; Risko, Anderson, Sarwal, Engelhardt, & Kingstone, 2012; Varao-Sousa & Kingstone, 2019).

Recent studies showed that MW was highly prevalent during video lectures. For example, Risko et al. (2012) found that participants on average reported MW to about 43% of the thought probes in three video lectures with different topics (Psychology, Economics, and Classics); similarly, Kane et al. (2017) reported a 45% MW rate in a 52-min statistics video lecture. Using a much shorter lecture (a 21-min statistics lecture), Szpunar, Khan, and Schacter (2013) found an MW rate of about 40%. Notably, all these studies showed that MW rate was negatively associated with lecture comprehension. Some evidence also suggested that MW increased over the course of a lecture, which may lead to worse retention for latter parts of the lecture material (Beserra, Nussbaum, & Oteo, 2019; Farley, Risko, & Kingstone, 2013; Risko et al., 2012; Wammes & Smilek, 2017; Young, Robinson, & Alberts, 2009).

2 | EXAMINING EYE MOVEMENTS OF MW DURING VIDEO LECTURES

Despite MW's prevalence, its covert and internal nature presents challenges to studying how information processing is affected during this mental state. Eye tracking may provide important insights into this question. Due to constraints on our visual system, successful visual processing often requires the viewer to gather information from different places. Therefore, visual attention can be reflected in two fundamental but related aspects of eye movements: when to move the eyes and where to move the eyes (Rayner, 1998). Both the when and the where aspects may determine when attention is no longer on the task. A few recent studies (Hutt et al., 2017; Hutt et al., 2017; Hutt, Mills, White, Donnelly, & D'Mello, 2016) used eye tracking to study MW from a data-mining perspective. For example, Hutt, Hardey, et al. (2017) tracked eye movements during a video lecture and probed participants to report MW/on-task. They built algorithms on the basis of a large number of features to predict self-report and were able to achieve above-chance performance. This very encouraging result showed a clear link between MW and eye movements in a video lecture context. However, in some cases, it was unclear which features were useful or generalizable across videos. And it may be difficult to interpret some of the features (e.g., minimum fixation duration; Hutt et al., 2016) in the context of existing theories.

The when-and-where question during MW has been primarily examined in the context of reading comprehension. For example, the usual association between looking times and word properties (e.g., word frequency) was reduced during MW (Foulsham, Farley, & Kingstone, 2013; Reichle, Reineberg, & Schooler, 2010; Steindorf & Rummel, 2019); readers also tended to skip more words and perform fewer horizontal eye movements during MW (Bixler & D'Mello, 2016; Faber, Bixler, & D'Mello, 2018). These findings not only inspired theoretical accounts of how reading is disrupted during MW (e.g., Smallwood, 2011) but also facilitated the development of

algorithms to detect MW during natural reading (e.g., Bixler & D'Mello, 2016; Faber et al., 2018). These successes point to the importance to expand this line of research to various lecture settings to better understand how the learner's attention is disrupted and how we can help learners recover from MW.

2.1 | LOOKING AT THE INSTRUCTOR

Most video lectures are multimedia experiences that integrate multiple sources of information. In a popular video lecture setup that presents an instructor and slides (the setup used in this study, see Figure 1), these two sources of information can vary in their informativeness and appeal. Visual attention is easily attracted by other people's faces (Gullberg & Holmqvist, 2006), but most of the learning-relevant information is presented on the slides. This setup presents a question for the learner of where to look at (cf. Mayer, 2005). The competition between the instructor and the slides provides a good opportunity to explore the where aspect of eye movement control. Given the impaired top-down control of comprehension during MW, visual attention may be less associated with the importance or the relatedness of information presented during video lectures.

2.2 | LOOKING AT THE SLIDES

Because the slides usually occupy a large portion of the visual field, both the when and the where decisions are highly relevant to this area. If visual processing of the slides indeed decreases during MW, we must look further at how the temporal and the spatial aspects of the fixations are affected. One common measure for examining the when question is fixation duration. Previous studies found that MW was associated with longer mean fixation duration in reading (Bixler & D'Mello, 2016; Foulsham et al., 2013; Reichle et al., 2010) and scene

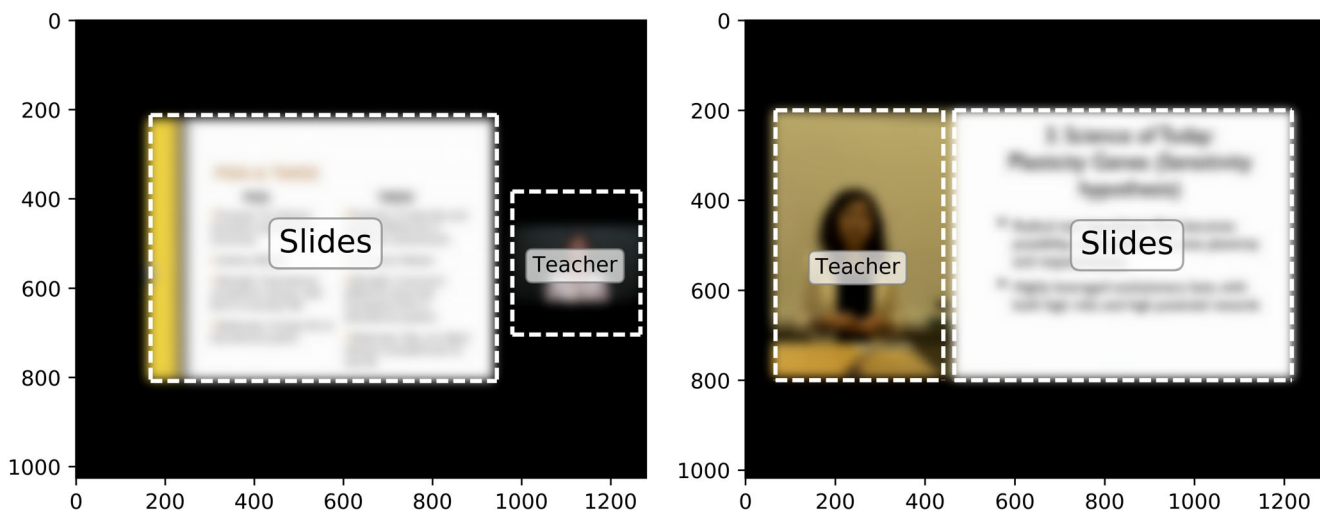


FIGURE 1 An illustration of the lecture videos used (Study 1 at left, education; and Study 2 at right, genetics). Areas of interest were noted by dotted lines. The x and y axes represents the screen pixel coordination [Colour figure can be viewed at wileyonlinelibrary.com]

viewing (Krasich et al., 2018). Given that slides are usually comprised of textual and graphical information (as in this study), MW may have similar effects on the when decision as in reading and scene viewing. Therefore, we expected that MW would be associated with increased duration for fixations on the slides.

We examined the where question by looking at the dispersion of fixations on the slides. Fixation dispersion measures how fixations are spread across a given area. It can be computed as the root mean square of the Euclidean distance from each fixation to the average position of all fixations (as in Krasich et al., 2018). A larger dispersion suggests that the eyes were moving across the entire area rather than concentrating on a small part of it. Previous studies found inconsistent results regarding dispersion; in a scene-viewing task, MW was associated with increased dispersion (Krasich et al., 2018), whereas in a simulated driving task, MW was associated with decreased dispersion (He, Becic, Lee, & McCarley, 2011). It remains an open question whether fixation dispersion will be associated with MW in the current study.

3 | ANALYZING THE TIME COURSE OF EYE MOVEMENTS

To evaluate whether certain objective patterns are uniquely associated with MW, a typical approach is to compare MW and non-MW data right before the thought probe. However, the range of the analysis window is often chosen somewhat arbitrarily. For example, studies have used a time window of 5 s (Smilek, Carriere, & Cheyne, 2010; Uzzaman & Joordens, 2011), 10 s (Christoff, Gordon, Smallwood, Smith, & Schooler, 2009), 40 words (corresponding to about 10 s; Franklin, Broadway, Mrazek, Smallwood, & Schooler, 2013), 10 sentences (corresponding to about 40 s; Franklin, Mooneyham, Baird, & Schooler, 2014), and so forth. Experience-sampling methods do not indicate the onset of an MW thought (Smallwood, 2013). Thus an arbitrary decision must be made about how much data to be included in the analyses. However, limiting analyses to a single interval may not be the best way to evaluate the rich information contained in time-series data.

In this study, we used growth curve analysis (Mirman, Dixon, & Magnuson, 2008) to model eye movements during a long period that leads up to the self-caught report/the thought probe (50 s or ten 5-s bins). In this way, we could see how MW develops over time across a series of consecutive time bins. The choice of using 5-s bins was arbitrary but consistent with some previous studies on eye movement correlates of MW (Krasich et al., 2018; Smilek et al., 2010; Uzzaman & Joordens, 2011). Moreover, instead of examining differences for each bin (e.g., Krasich et al., 2018), growth curve analysis examines the overall difference between two time courses. Thus this method can be more powerful and less sensitive to the selection of a particular bin size, which can reduce the risk associated with arbitrariness.

In order to begin to understand what happens after MW, we compared eye movements immediately before and after self-caught/probe-caught MW. A self-caught report or the onset of a thought probe usually marks the termination of an MW episode, either

because it has reached awareness or because it is interrupted by the thought probe's presentation. Then participants might exit the MW state, at least temporarily, and redirect themselves back to the task. Thus, comparing what happens before and after these critical moments can offer insights in how learners recover from MW. As an exploratory analysis, we compared eye movement measures during the 15 s before and after the onset of a thought probe or the onset of a self-caught MW report. We used three bins to reduce potential artifacts caused by the process of reporting MW (e.g., looking at the keyboard to respond to the probe and then looking back to the screen).

4 | THE CURRENT STUDY

To summarize, the current studies examined whether MW was associated with temporal changes in eye movements in a video lecture setting. We examined the following eye-movement measures: (a) the allocation of fixations on the instructor (defined as the number of fixations on the instructor divided by the total number of fixations on the instructor and the slides), (b) the duration of fixations on the slides, and (c) the dispersion of fixations on the slides. For each eye movement measure, we examined how it evolved over time during a period before self-caught MW/thought probe onset and whether it abruptly changed immediately after these critical moments.

On the basis of previous literature, we predicted that MW should be associated with increased fixation duration. However, previous literature did not help us to clearly predict fixation dispersion and fixation allocation. Moreover, because the application of growth curve analysis was new in this setting, we could not hypothesize how MW and on-task time series would differ; they could differ in their overall mean, their trends, or both.

To offer evidence of generalizability, we decided to run the same experimental procedure and analysis protocol with two lecture videos. The two videos covered different topics, used different instructors, and had slight differences in their display setting. For clarity, we discuss the first video as Study 1 and the second video as Study 2. Moreover, we used either the self-caught or the probe-caught method to obtain MW and on-task reports, with each participant using only one kind of report method. Self-caught MW only captured MW episodes that eventually reached awareness, whereas probe-caught MW could capture MW with and without the participant's awareness. By including both approaches, we can assess whether our findings depend on the specific ways in which MW is assessed.

5 | STUDY 1

5.1 | Methods

5.1.1 | Participants

Seventy-seven undergraduate students from a large Midwestern U.S. university participated in this study for course credit. We decided

to collect as much data as we could during one semester, taking into account constraints on time and personnel. We discarded data from three participants for technical failures that resulted in incomplete recording and two participants for low tracking ratio, which left a total of 72 participants for data analysis (mean age = 18.86 years, $SD = 0.97$ years, and 61.11% female). Of them, 34 were randomly assigned to the probe-caught condition. All participants reported normal or corrected normal eye sight. The study has been approved by the Institutional Review Board at the authors' university.

5.1.2 | Stimuli and apparatus

The video was a 19-min scripted lecture on International Comparisons in Education that introduced methods used in large-scale international student assessments and how they informed educational improvement efforts (adapted from Samudra, Min, Cortina, & Miller, 2016). The video can be found at <https://osf.io/ptj75/files/>. The video display was divided into a slides window and an instructor window (see Figure 1, left panel). Window size and position remained fixed throughout the lecture. The instructor was a female graduate student who read a script on a teleprompter. The slides contained static texts and figures. The video was played through a 1,280 × 1,024 resolution computer screen at a viewing distance of approximately 60 cm. The slides area subtended approximately 19 × 15 of visual angle, and the instructor area subtended approximately 7 × 4 of visual angle.

Binocular gaze data were recorded by an Eye Tribe tracker at 60 Hz. The Eye Tribe tracker can provide satisfactory gaze tracking quality for fixation-based analysis (Dalmajer, 2014). It was used in several previous studies in a similar context (e.g., Hutt, Hardey, et al., 2017; Hutt, Mills, et al., 2017). The average tracking ratio in this study was 80%, and the average calibration accuracy was 0.62° of visual angle (0.5–1.0 as claimed by the Eye Tribe developers; as cited in Ooms, Dupont, Lapon, & Popelka, 2015). Raw gaze data were fed to the open gaze and mouse analyzer algorithm (Voßkühler, Nordmeier, Kuchinke, & Jacobs, 2008) for an offline detection of fixations. The experiment was implemented by the OpenSesame software (Mathôt, Schreij, & Theeuwes, 2012) with the PyGaze package (Dalmajer, Mathôt, & Stigchel, 2014). Areas of interest (noted by dotted lines in Figure 1, left) were defined prior to data collection.

5.1.3 | Procedure

The definition of MW was communicated to the participants as *any thoughts experienced throughout the lecture that are not related to the content being presented during the lecture*, including cases where *thoughts simply pop into your head and you may choose to think about something other than the lecture content* (as in Lindquist & McLean, 2011). Participants learned this concept and were trained to distinguish MW from task-related interference/elaboration and external distraction (Smallwood & Schooler, 2015). Specifically, participants were simply told not to report MW if they were distracted by external objects and

events in their surroundings. Then participants learned to distinguish between MW and task-related interference/elaboration in seven hypothetical scenarios; for each scenario, participants responded whether they should indicate MW. For example, you find yourself thinking about your plans for tonight should be reported as MW, but you find yourself thinking about how much longer the video will be was task-related interference/elaboration and thus should not be reported. The experimenter explained the correct answer to participants for all incorrect responses. These questions can be found in Data S1.

Participants were randomly assigned to the self-caught condition or the probe-caught condition. In the self-caught condition, participants were instructed to press any key whenever they caught themselves MW. Participants were also asked to rest a finger on the keyboard while watching in order to minimize the delay between MW detection and report. In the probe-caught condition, a beep sound was played at four fixed times in the video (2 min 48 s, 6 min 28 s, 11 min 31 s, and 17 min 23 s), and participants were asked to report whether, at this exact moment, they were MW by pressing the corresponding key (Y for MW and N for on-task).

After learning the instructions, participants took a pretest of the video lecture to measure preexisting knowledge. The pretest consisted of 5 multiple choice questions and required participants to finish within 5 min. Then the eye tracker underwent a 9-point calibration. Next, participants watched the video lecture with eye movements recorded. Participants were told to *watch the video carefully as if you are in a real lecture* and that *you will be tested on the video after watching*. Depending on their assigned group, participants used either the self-caught or the probe-caught method to report MW. The video played continuously despite key presses. When the video was over, participants took a comprehension test that consisted of 18 multiple choice questions and required participants to finish within 15 min. As in Samudra et al. (2016), 12 of the questions were lower level definitional questions on the basis of recalling information from the video, and the other six questions were conceptual and required applying the information from the video to a new context. The post-test questions covered but were not limited to the content presented before the thought probes.¹ Participants also rated their engagement level during the video (*The material covered in this lesson was interesting* and *My attention was fully focused on the video*, from 1 = *strongly agree* to 5 = *strongly disagree*; responses were reverse coded, and the average of the two items was used to index engagement level). Participants had the option to retrospectively disclose their MW thoughts occurred during the video (*What were you thinking while you were mind-wandering? Please give us a few examples.*). These reports were not analyzed in the current study.

5.1.4 | Data analysis

Data analysis was conducted in R (Version 3.5.1; R Core Team, 2018).² Fixations greater than 2,000 ms or shorter than 80 ms were discarded (5.80% of data). We chose a relatively high upper bound because previous studies showed that MW tended to produce long fixations (Bixler &

D'Mello, 2016; Foulsham et al., 2013; Krasich et al., 2018; Reichle et al., 2010). We obtained fixations that occurred during the 50 s before and 15 s after the onset of a thought probe or a self-caught key press. These fixations were then put into thirteen 5-s bins (10 before and 3 after) on the basis of each fixation's onset time. For example, if a fixation started at bin n and ended at bin $n + 1$, it was put into bin n .

In the self-caught condition, the interval between two reports could be shorter than 65 s. Thus the 15 s after the current self-report could overlap with the 50 s before the next report. We addressed this problem by doing the following: (a) if the participant pressed key twice within 5 s, we deleted the first key press (1.86% of self-caught reports) and used the second one as the endpoint of that MW episode; and (b) we prioritized data selection for the current report over the next report in case there was an overlap. Thus we ensured that each MW episode was associated with a unique portion of data. The probe-caught condition did not have this issue because the intervals between thought probes were all greater than 65 s.

We conducted a series of growth curve analysis to examine how eye movements changed in the period leading up to MW reports. Growth curve analysis uses orthogonal polynomials to identify components in the curvilinear time course of longitudinal data. We only used the first and the second order polynomials because they can sufficiently capture the effects specified in the hypotheses.³ Data for self-caught and probe-caught MW were modeled separately. For each model, the outcome variable was the participant-level average of fixation allocation, duration, or dispersion. To construct a reference level for self-caught MW, we used the average of 1,000 epochs of 65 s randomly selected from each participant's full eye movement data. All models included the two time terms, attention (MW or not), and their interactions as fixed effects. Random effects included time terms for each participant, as well as time terms for each observed participant-attention combination. The correlation between random parameters were constrained to be 0 to help reach convergence. Approximated p values were obtained using the Satterthwaite's method from the lmerTest package (Kuznetsova et al., 2017). See Appendix A for full model results.

To examine how eye movements changed abruptly before and after MW reports, we used repeated-measure analyses of variance (ANOVAs) to compare the last three time bins before with the first

three time bins after probe presentation or self-caught report. Effect sizes are reported in generalized eta-squared or η_G^2 (Bakeman, 2005).

5.2 | Results

5.2.1 | MW, engagement, and comprehension

As an initial indicator of the validity of participants' MW monitoring, we correlated participants' level of MW with their self-rated engagement during the video and their lecture comprehension (see Table 1). MW level in the probe-caught condition was defined as the percentage of probe-caught MW; MW level in the self-caught condition was defined as the total length of the video divided by the number of reports, which was the estimated average interval in seconds between two self-reports (thus a larger value indicates fewer cases of reported MW).⁴ We found a significant relationship between MW level and engagement in both conditions. MW level was also significantly correlated with posttest score. This relationship held true even after controlling for pretest score (self-caught: partial $r = .41$, $p = .012$; and probe-caught: partial $r = -.45$, $p = .008$). Thus participants' in-the-moment assessments of MW were connected to their overall self-reports of engagement and to what they learned from the lecture.

5.2.2 | Fixation allocation

The time course of fixation allocation is shown in Figure 2, left panel. For self-caught MW, there was a significant linear trend during a period of 50 s before self-report, $b = 5.15$, standard error (SE) = 1.49, $t = 3.45$, $p = .001$. This indicates that, during self-caught MW, the relative frequency of fixations on the teacher area increased over time. Importantly, the interaction between the linear term and attention was significant, $b = -5.48$, SE = 1.97, $t = -2.78$, $p = .008$. This indicates that the time course of self-caught MW evolved differently from the baseline. Moreover, a significant effect of attention on the intercept term, $b = -1.52$, SE = 0.55, $t = -2.75$, $p = .009$, indicates a global difference between the two curves.

TABLE 1 The average level of mind wandering and its relationship with engagement and comprehension

Report method	Study	MW level	SD	MW-engagement	MW-posttest
Probe-caught	Study 1	0.32	0.22	-0.35*	-0.42*
Self-caught	Study 1	215.65	199.60	0.38*	0.37*
Probe-caught	Study 2	0.32	0.23	-0.59***	-0.08
Self-caught	Study 2	201.05	171.46	0.49**	0.04

Note: Mind wandering (MW) level in the probe-caught condition was defined as the percentage of probe-caught MW; MW level in the self-caught condition was defined as the total length of the video divided by the number of reports, which gives an estimation of the average interval in seconds between two self-reports (thus a larger value indicates fewer cases of reported MW). MW-engagement is the correlation between MW level and engagement level; MW-comprehension is the correlation between MW level and posttest score.

Abbreviation: MW, mind wandering.

*** $p < .001$; ** $p < .01$; * $p < .05$.

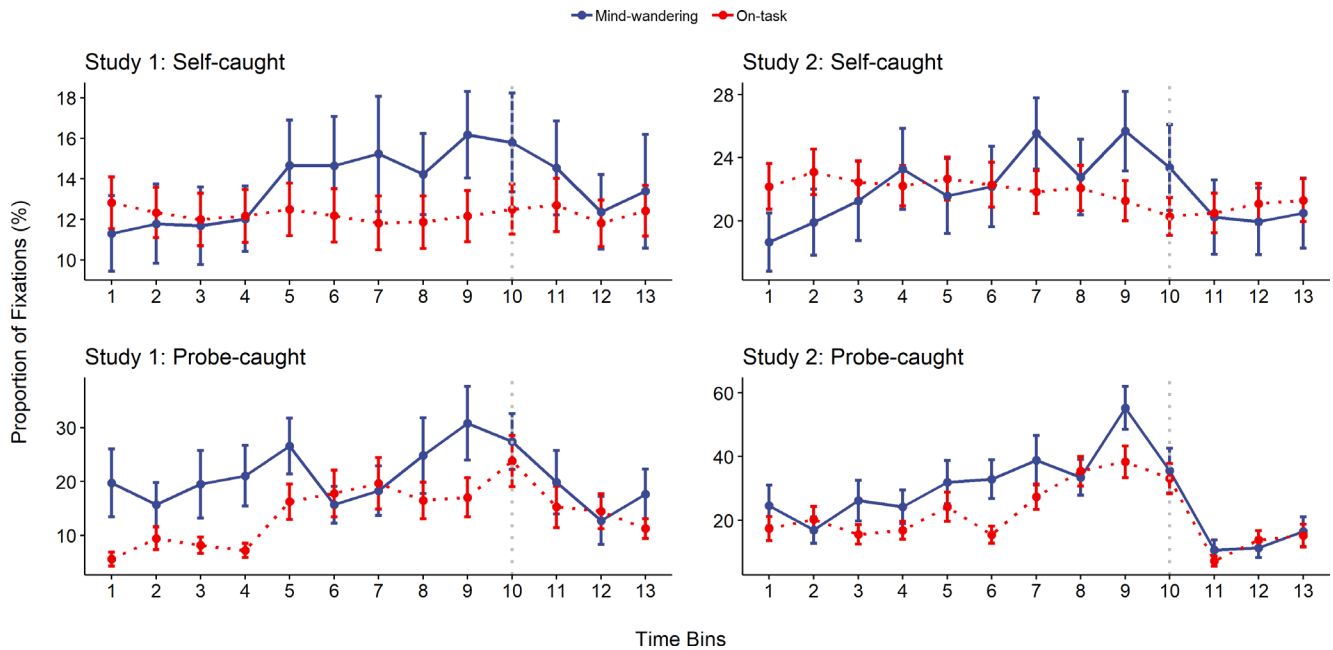


FIGURE 2 The proportion of fixations on the instructor. The y axis represents fixation allocation, defined as the number of fixations on the instructor divided by the total number of fixations on the instructor and the slides. Each number on the x axis represents a 5-s bin. The dots show the average across participants, and error bars show ± 1 standard error. The blue solid line represents mind wandering, and the red dotted line represents the baseline (in the self-caught condition) or on-task episodes (in the probe-caught condition). The dotted vertical line marks the last bin before self-caught report or probe onset. A colored version of this figure is available in the online version of this paper [Colour figure can be viewed at wileyonlinelibrary.com]

For probe-caught MW, there was a marginally significant linear trend during a period of 50 s before probe onset, $b = 9.99$, $SE = 4.93$, $t = 2.03$, $p = .047$. The linear trend did not significantly interact with attention, $p = .356$, which suggests that probe-caught MW and on-task episodes followed similar trends. However, we observed a significant effect of attention on the intercept term, $b = -8.32$, $SE = 3.68$, $t = -2.26$, $p = .030$. This indicates that probe-caught MW, compared with on-task episodes, was associated with a larger proportion of fixations allocated to the teacher area.

We compared the last three time bins before with the first three time bins after a probe or self-caught episode (see Figure 2). Time point (before or after), attention (MW or not), and their interaction term were submitted to a two-way repeated measure ANOVA. For the self-caught condition, the main effect of time was not significant, $p = .197$; the main effect of attention was marginally significant, $F(1, 37) = 3.99$, $p = .053$, $\eta_G^2 = 0.01$; and the interaction between time and attention was not significant, $p = .123$. Therefore, the difference between MW and the baseline seemed to persist after self-caught reports.

For the probe-caught condition, we observed a significant main effect of time, $F(1, 26) = 7.82$, $p = .010$, $\eta_G^2 = 0.04$. The main effect of attention was not significant, $p = .223$; and the interaction between attention and time was also not significant, $p = .182$.

5.2.3 | Fixation duration

We hypothesized that MW would increase the duration of fixations on the slides area during MW. However, it was unclear whether fixations on the instructor would be similarly affected. We found that

fixations on the instructor were substantially longer than those on the slides (540.13 vs. 276.77 ms). Thus, the two areas needed to be analyzed separately. However, because the instructor area received limited attention in general (12.35% vs. 83.76% of fixations), the SEs for the parameter estimates of the instructor-only growth curve became very large, rendering such models severely underpowered. Therefore, we only report results for the slides area (see supporting information for the instructor area's results).

The time course of fixation duration is shown in Figure 3, left panel. For self-caught MW, we observed a significant linear term during 50 s before self-report, $b = 36.37$, $SE = 6.31$, $t = 5.76$, $p < .001$. This indicates that fixation duration increased over time during self-caught MW. Importantly, this linear trend significantly differed from the baseline, $b = -37.92$, $SE = 8.93$, $t = -4.25$, $p < .001$. The quadratic term for self-caught MW was also significant, $b = 20.45$, $SE = 8.17$, $t = 2.50$, $p = .014$. This quadratic trend was marginally different from the baseline, $b = -20.78$, $SE = 11.56$, $t = -1.80$, $p = .076$.

For probe-caught MW, we also observed a significant linear term during the same period, $b = 87.72$, $SE = 20.30$, $t = 4.32$, $p < .001$. Similarly, we observed a significant interaction between the linear term and attention, $b = -90.68$, $SE = 26.78$, $t = -3.39$, $p = .001$, indicating a significantly weaker linear trend during on-task episodes.

We then contrasted mean fixation duration before and after MW. Time point (before or after), attention (MW or not), and their interaction term were submitted to a two-way repeated measure ANOVA. In the self-caught condition, we found a significant main effect of time, $F(1, 37) = 12.98$, $p = .001$, $\eta_G^2 = 0.02$; and a significant main effect of attention, $F(1, 37) = 4.94$, $p = .032$, $\eta_G^2 = 0.01$. The interaction term was also significant, $F(1, 37) = 11.29$, $p = .002$,

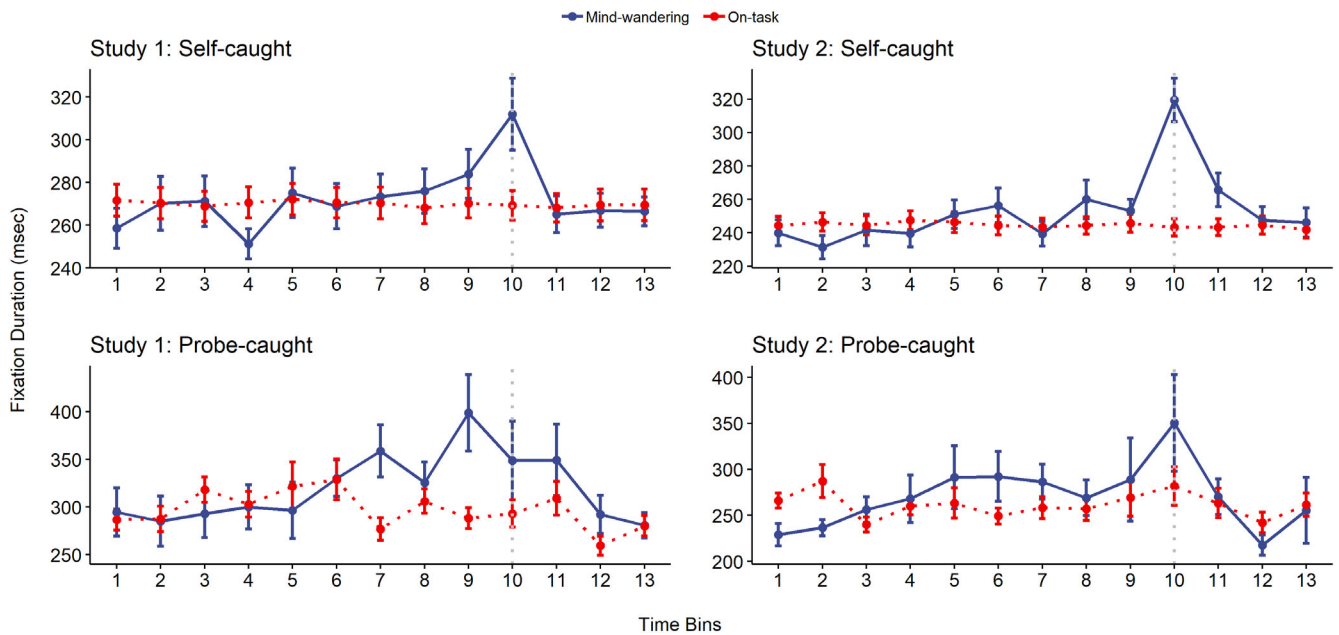


FIGURE 3 The mean duration of fixations on the slides. The y axis represents the mean fixation duration. Each number on the x axis represents a 5-s bin. The dots show the average across participants, and error bars show ± 1 standard error. The blue solid line represents mind wandering, and the red dotted line represents the baseline (in the self-caught condition) or on-task episodes (in the probe-caught condition). The dotted vertical line marks the last bin before self-caught report or probe onset. A colored version of this figure is available in the online version of this paper [Colour figure can be viewed at wileyonlinelibrary.com]

$\eta_G^2 = 0.02$. Post hoc comparisons revealed that fixation duration significantly decreased (difference: 24 ms) after reports of self-caught MW, $t = 4.89$, $p < .001$; but this was not the case for the baseline (difference: 1 ms), $p = .724$.

In the probe-caught condition, we found a significant main effect of time, $F(1, 25) = 4.98$, $p = .035$, $\eta_G^2 = 0.04$; and a significant main effect of attention, $F(1, 25) = 5.14$, $p = .032$, $\eta_G^2 = 0.04$. However, the interaction was not significant, $p = .264$.

5.2.4 | Fixation dispersion

Fixation dispersion was normalized to a 0 to 1 scale by dividing dispersion on the slides area by the maximum possible dispersion. A smaller value means that participants limited their fixations to a smaller portion of the slides. The time course of fixation dispersion on the slides area is shown in Figure 4, left panel. For self-caught MW, we observed a significant linear term during 50 s before self-report, $b = -0.02$, $SE = 0.01$, $t = -3.82$, $p < .001$, indicating that fixation dispersion decreased over time during self-caught MW. Importantly, this linear trend significantly differed from the baseline, $b = 0.02$, $SE = 0.01$, $t = 3.01$, $p = .005$. Moreover, there was also an effect of attention on the intercept term, $b = 0.01$, $SE = .003$, $t = 2.68$, $p = .011$. This indicates that fixation dispersion was in general smaller during self-caught MW than in the baseline.

For probe-caught MW, the linear term for the same period was also significant, $b = -0.06$, $SE = 0.02$, $t = -3.17$, $p = .002$, indicating decreased dispersion during probe-caught MW. The interaction between attention and the linear term was not significant, $p = .113$. However, there was a significant effect of attention on the intercept term, $b = 0.03$, $SE = 0.01$,

$t = 2.46$, $p = .020$, indicating that probe-caught MW episodes in general had lower dispersion than did on-task episodes.

As with the two previous measures, we conducted an abrupt change analysis for fixation dispersion using repeated-measure ANOVA. In the self-caught condition, we found a significant main effect of time, $F(1, 37) = 11.87$, $p = .001$, $\eta_G^2 = 0.02$; and a significant main effect of attention, $F(1, 37) = 7.84$, $p = .008$, $\eta_G^2 = 0.01$. The interaction between time and attention was also significant, $F(1, 37) = 14.18$, $p = .001$, $\eta_G^2 = 0.03$. Post hoc analysis showed that fixation dispersion significantly increased (difference: .022) after self-caught MW, $t = 5.21$, $p < .001$. But this was not the case for the baseline (difference: .001), $p = .733$.

In the probe-caught condition, the main effect of time was not significant, $p = .273$. The main effect of attention was significant, $F(1, 25) = 7.07$, $p = .013$, $\eta_G^2 = 0.06$. Importantly, there was a significant interaction between time and attention, $F(1, 25) = 7.05$, $p = .014$, $\eta_G^2 = 0.05$. Post hoc analysis showed that fixation dispersion significantly increased (difference: .042) after the presentation of thought probes reported as MW, $t = 2.69$, $p = .010$; but this was not the case for on-task reports (difference: .018), $p = .263$.

5.3 | Summary of Study 1

Our results show temporal changes in eye movement patterns during a period of 50 s leading to MW reports: we observed an increase in the allocation of fixations to the instructor during MW. For the fixations on the slides, the average duration increased but the dispersion decreased. The growth curves during MW distinguished themselves from those

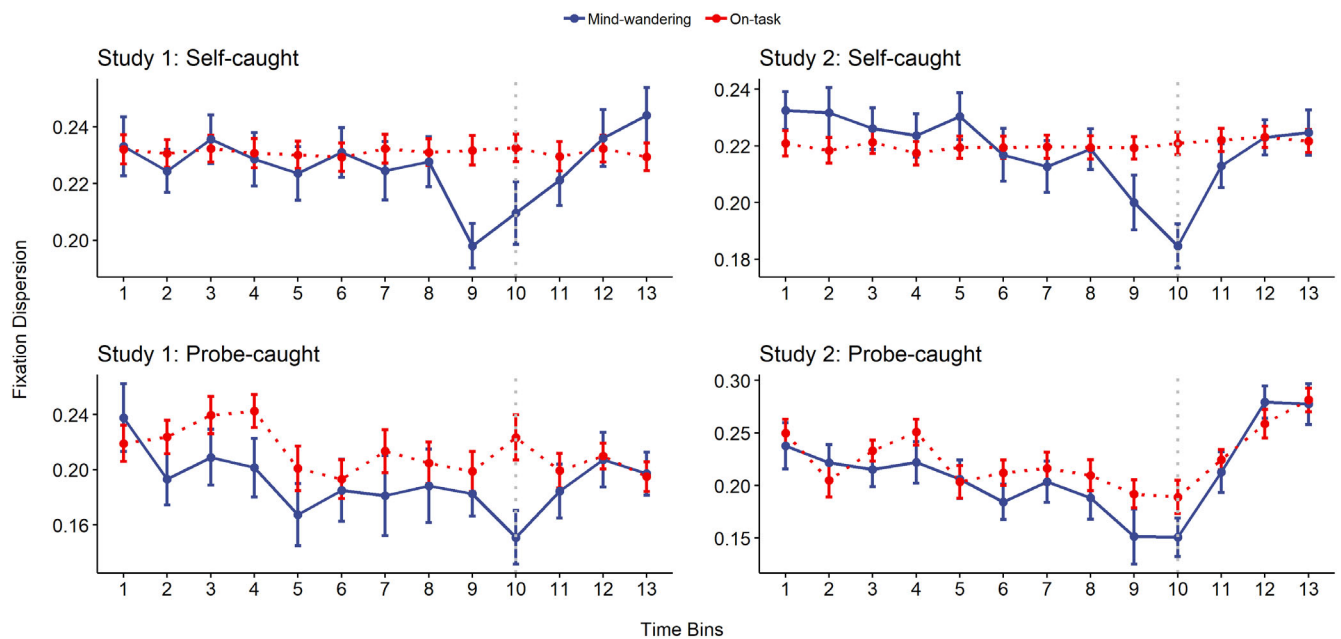


FIGURE 4 The dispersion of fixations on the slides. The Y axis represents fixation dispersion. Each number on the X axis represents a 5-s bin. The dots show the average across participants, and error bars show ± 1 standard error. The blue solid line represents mind wandering, and the red dotted line represents the baseline (in the self-caught condition) or on-task episodes (in the probe-caught condition). The dotted vertical line marks the last bin before self-caught report or probe onset. A colored version of this figure is available in the online version of this paper [Colour figure can be viewed at wileyonlinelibrary.com]

during non-MW episodes through differences in the overall curve height or in the time trends.

To further illustrate the results, we constructed fixation heat maps using the probe-caught data (Figure 5, upper panel). Fixations during the entire 50 s were pooled together across participants and thought probes and were overlaid on a canonical display. Fixations were also weighted by their duration. A Gaussian kernel with a width of 70 pixels and a standard deviation of 10 pixels was used to smooth the heat maps. Visual attention appears to be bias towards the instructor area after weighting fixations by duration, a pattern consistent with previous research (Kizilcec, Papadopoulos, & Sritanyaratana, 2014; van Wermeskerken & Gog, 2017; Wang & Antonenko, 2017). Importantly, the heat map shows a reduction of coverage of the slides area during MW, which potentially contributed to the increased proportion of fixations on the teacher and the decreased dispersion of fixations on the slides.

Preliminary evidence showed that some eye movement patterns associated with MW diminished after MW was reported. Specifically, the dispersion of fixations on the slides area increased abruptly after MW reports but not after reports of on-task. We also found an abrupt reduction in the duration of fixations on the slides area after MW reports, but it is not clear whether a similar change occurred when participants were on-task.

6 | STUDY 2

In Study 2, a different sample of participants watched a different video lecture. The experiment followed the same procedure, and we conducted

the same set of analyses. We wanted to make sure our results were not bound to a particular instructor, lecture content, or visual layout.

6.1 | Methods

6.1.1 | Participants

Seventy-five undergraduate students from a large Midwestern U.S. university participated in this study for course credit. We decided to collect as much data as we could during one semester, given constraints on time and personnel. We discarded data from two participants for incomplete recording and two participants for low tracking ratio. This left 71 participants for data analysis (mean age = 18.85 years, $SD = 0.92$ years, and 57.75% female), and 34 were randomly assigned to the probe-caught condition. All participants reported normal or corrected normal eye sight.

6.1.2 | Stimuli and apparatus

The video used in Study 2 was a 22-min scripted lecture on the subject *Introduction to Genetics*, which reviewed some past and current views on how genes and environment shaped human behaviors (adapted from Min, 2018). The video can be found at <https://osf.io/ptj75/files/>. The video display was divided into a slides window and an instructor window (see Figure 1, right panel). Window size and position remained fixed throughout the lecture. The instructor was a different female graduate student who read a script presented on a teleprompter.

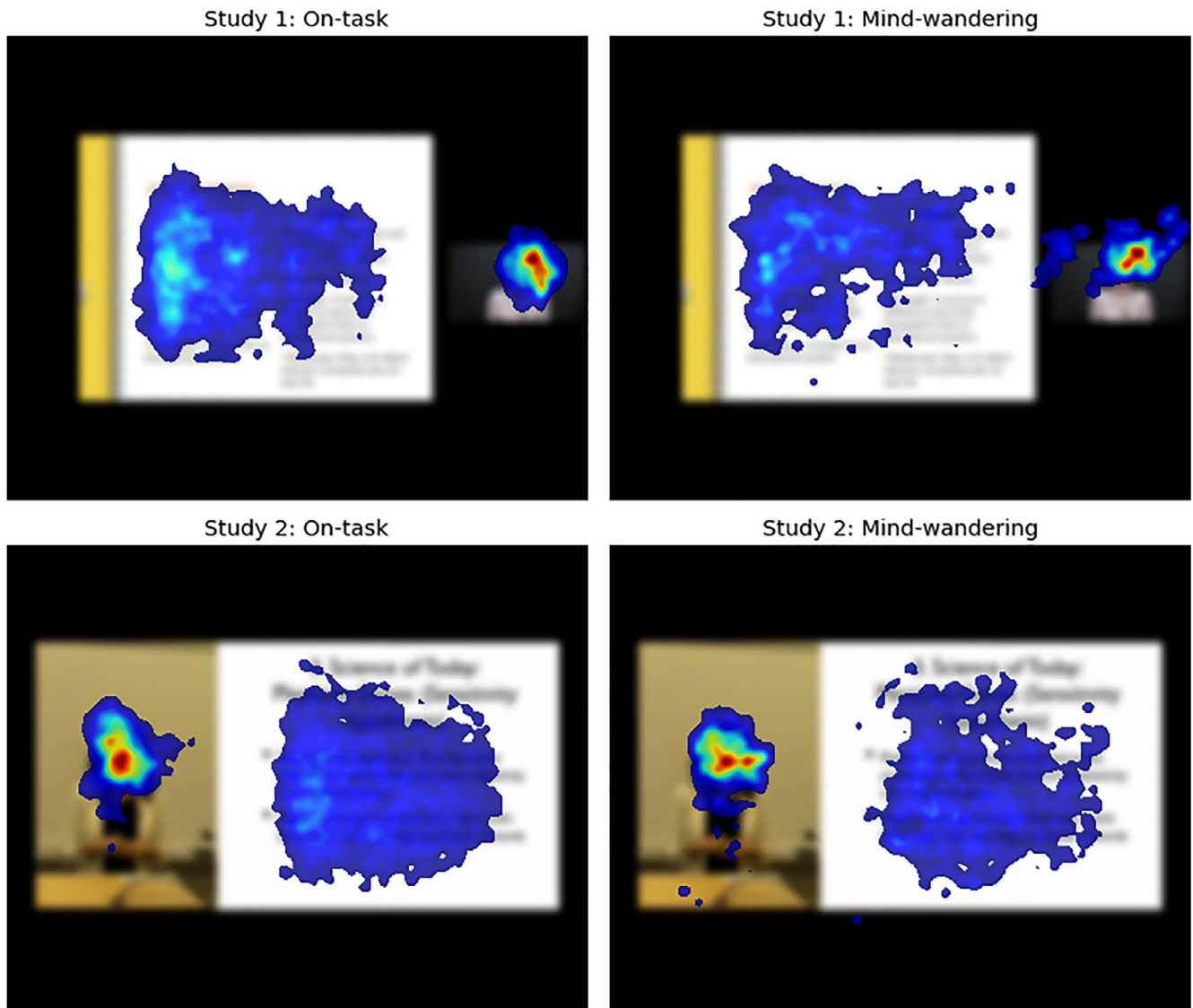


FIGURE 5 Fixation heat maps during 50 s prior to the presentation of thought probes, overlaid on a canonical depiction of the display. Fixations during the entire 50 s were pooled across participants and thought probes. Fixations were also weighted by their duration. A Gaussian kernel with a width of 70 pixels and a standard deviation of 10 pixels was used to smooth the heat maps [Colour figure can be viewed at wileyonlinelibrary.com]

Gaze data in the probe-caught condition were collected using the Eye Tribe tracker at 60 Hz. However, we used a Tobii T60 tracker (60 Hz) for the self-caught condition due to a connection issue with the Eye Tribe server. The video was played at approximately 60 cm from the participant, with the slides area subtending approximately 18×15 of visual angle, and the instructor area subtending approximately 9×15 of visual angle. The average tracking ratio was 82%, and the calibration accuracy was 0.64° of visual angle. Raw gaze data from both trackers were fed to a single algorithm (i.e., open gaze and mouse analyzer; Voßkühler et al., 2008) for offline detection of fixations.

6.1.3 | Procedure

The procedure was the same as in Study 1. Participants completed a pretest, a video lecture with eye tracking, a posttest, and an exit survey.

During the video, participants were asked to either self-catch MW or respond to four sound beeps during the lecture (3 min 19 s, 10 min 48 s, 15 min 01 s, and 21 min 15 s) to report MW or on-task. The pretest consisted of five multiple choice questions, and the time limit was 5 min. The posttest consisted of 14 multiple choice questions, and the time limit was 15 min. Four of the questions were lower level definitional questions, and the other nine were higher level conceptual questions. As in Study 1, the posttest questions covered but were not exclusive to the content presented before the thought probes. Both pretest and posttest questions are available at <https://osf.io/ptj75/files/>.

6.1.4 | Data analysis

Data analysis proceeded the same way as in Study 1. The deletion of extreme values resulted in 4.25% of data loss. The deletion of close

self-caught reports (within 5 s of a previous report) resulted in a 2.71% loss of total self-caught reports.

6.2 | Results

6.2.1 | MW, engagement, and comprehension

As in Study 1, we correlated participants' level of MW with participants' engagement during the video and their lecture comprehension (see Table 1). Again, we found a significant relationship between MW level and self-rated engagement. However, the relationship between MW level and comprehension score was not significant, even after controlling for pretest score (self-caught: partial $r = 0.05$, $p = .772$; and probe-caught: partial $r = -0.07$, $p = .690$).

6.2.2 | Fixation allocation

The right panel of Figure 2 shows fixation allocation over time. For self-caught MW, there was a significant linear trend during 50 s before self-report, $b = 5.38$, $SE = 1.94$, $t = 2.78$, $p = .007$. This linear trend differed from the baseline, $b = -7.21$, $SE = 2.68$, $t = -2.69$, $p = .011$, indicating a different evolving pattern during self-caught MW compared with the baseline.

For probe-caught MW, the linear term during 50 s before the probe onset was also significant, $b = 25.18$, $SE = 6.42$, $t = 3.92$, $p < .001$. Attention did not significantly interact with the linear term, $p = .629$. However, there was a significant effect of attention on the intercept term, $b = -6.93$, $SE = 2.75$, $t = -2.52$, $p = .017$, indicating that the proportion of fixations on the teacher overall was higher during MW than during on-task episodes.

As in Study 1, we conducted an abrupt change analysis using repeated-measure ANOVA. In the self-caught condition, the main effect of time was significant, $F(1, 36) = 6.94$, $p = .012$, $\eta_G^2 = 0.01$. The main effect of attention was not significant, $p = .366$, but there was a significant interaction between time and attention, $F(1, 36) = 7.09$, $p = .012$, $\eta_G^2 = .007$. Post hoc comparisons revealed that attention on the teacher significantly decreased (difference: 3.71%) after reports of self-caught MW, $t = 3.77$, $p < .001$; but this was not the case for the baseline (difference: .10%), $p = .92$.

In the probe-caught condition, we observed a significant main effect of time, $F(1, 26) = 41.40$, $p < .001$, $\eta_G^2 = 0.27$. This indicates that attention on the slides was increased in general after responding to thought probes. On the other hand, we did not find a significant effect of attention, $p = .138$, or a significant interaction between time and attention, $p = .205$.

6.2.3 | Fixation duration

Fixations on the instructor were substantially longer than those on the slides (534.91 vs. 248.90 ms), but the majority of the fixations

were on the slides area (77.96% vs. 19.97% of fixations). As in Study 1, we report results for the slides area.

The right panel of Figure 3 shows the time course of fixation duration. For self-caught MW, we observed a significant linear trend during 50 s prior to self-report, $b = 53.12$, $SE = 5.83$, $t = 9.10$, $p < .001$. This linear trend significantly differed from the baseline, $b = -54.77$, $SE = 8.19$, $t = -6.68$, $p < .001$. The quadratic term was also significant, $b = 31.06$, $SE = 6.39$, $t = 4.86$, $p < .001$; and it significantly differed from the baseline, $b = -32.31$, $SE = 8.88$, $t = -3.64$, $p = .001$. Overall, baseline fixations were shorter than those in the self-caught MW state, $b = -8.19$, $SE = 2.45$, $t = -3.34$, $p = .001$.

For probe-caught MW, we also found a significant linear trend during the same period, $b = 97.98$, $SE = 26.78$, $t = 3.66$, $p = .001$. There was also a significant interaction between the linear term and attention, $b = -84.95$, $SE = 27.17$, $t = -3.13$, $p = .004$, indicating that the linear trend was significantly reduced during on-task episodes.

An abrupt change analysis for the self-caught condition showed a significant main effect of time, $F(1, 36) = 12.85$, $p = .001$, $\eta_G^2 = 0.03$. There was also a significant main effect of attention, $F(1, 36) = 37.03$, $p < .001$, $\eta_G^2 = 0.07$. Importantly, the interaction term was also significant, $F(1, 36) = 10.84$, $p = .002$, $\eta_G^2 = 0.02$. Post hoc comparisons showed that fixation duration significantly decreased (difference: 24 ms) after reports of self-caught MW, $t = 4.89$, $p < .001$; but this was not the case for the baseline (difference: 1 ms), $p = .845$.

An abrupt change analysis for the probe-caught condition did not find a significant main effect of time, $p = .112$, or a significant main effect of attention, $p = .296$. There was a marginally significant interaction between time and attention, $F(1, 26) = 3.60$, $p = .069$, $\eta_G^2 = 0.02$. Post hoc comparisons showed that fixation duration significantly decreased (difference: 60 ms) after MW responses, $t = 2.45$, $p = .018$. A similar abrupt change was not found after on-task responses (difference: 5 ms), $p = .831$.

6.2.4 | Fixation dispersion

The right panel of Figure 4 shows the time course of fixation dispersion. For self-caught MW, there was a significant linear trend during the 50 s prior to self-report, $b = -0.04$, $SE = 0.01$, $t = -6.68$, $p < .001$. Thus fixations on the slides became less spread over time. Moreover, this linear trend significantly differed from baseline, $b = 0.04$, $SE = 0.01$, $t = 4.75$, $p < .001$. There was also a significant quadratic trend associated with self-caught MW, $b = -0.02$, $SE = .005$, $t = -3.09$, $p = .002$; and this quadratic trend significantly differed from the baseline, $b = 0.02$, $SE = 0.01$, $t = 2.41$, $p = .016$.

For probe-caught MW, the linear term during the same period was also significant, $b = -0.08$, $SE = 0.02$, $t = -4.86$, $p < .001$. Attention did not significantly interact with the linear trend, $p = .151$. But there was a significant effect of attention on the intercept term, $b = 0.02$, $SE = 0.01$, $t = 2.21$, $p = .036$, indicating a larger dispersion overall during on-task episodes than during MW episodes.

Again, we conducted a repeated-measures ANOVA to examine fixation dispersion before and after MW. In the self-caught condition,

we found a significant main effect of time, $F(1, 36) = 16.91, p < .001, \eta_G^2 = 0.03$; and a significant main effect of attention, $F(1, 36) = 8.02, p = .008, \eta_G^2 = 0.03$. Importantly, the interaction term was also significant, $F(1, 36) = 9.11, p = .005, \eta_G^2 = 0.02$. Post hoc comparisons showed that fixation dispersion significantly decreased (difference: .019) after reports of self-caught MW, $t = 5.07, p < .001$; but this was not the case for the baseline (difference: .004), $t = .29, p = .447$.

In the probe-caught condition, we found a significant main effect of time, $F(1, 26) = 71.55, p < .001, \eta_G^2 = 0.24$. The main effect of attention was not significant, $p = .268$. The interaction between time and attention was also not significant, $p = .160$.

6.3 | Summary of Study 2

In Study 2, results from the growth curve analysis were consistent with those obtained in Study 1, despite changes in the instructor, visual layout, and content of the video lecture. Specifically, we found, during MW, an increased proportion of fixations on the teacher, increased duration, and reduced dispersion for fixations on the slides. We also constructed a fixation heat map on the basis of the probe-caught data with the same procedure (see Figure 5, lower panel). Similar to Study 1, a reduction in the coverage of the slides during MW potentially contributed to the change in fixation allocation as well as the reduced dispersion for fixations on the slides.

Similar to Study 1, both self-caught and probe-caught MW were correlated with the offline rating of engagement. Different from Study 1, we did not find a significant relationship between MW frequency and comprehension. There was a lower average score on the comprehension test (37% in Study 2 vs. 60% in Study 1), so it may be that differences in difficulty of the content or the test might play into this difference in results.

For fixation duration and dispersion, the self-caught data across the two studies consistently showed an abrupt change right after reports of MW. However, some inconsistencies existed in the probe-caught condition between the two studies. For example, in Study 2, we found that fixation duration significantly decreased after probe-caught MW but not after on-task episodes. This interaction was not statistically significant in the probe-caught condition of Study 1. In Study 2, the dispersion of fixation on the slides seemed to increase in general after probe presentation, regardless of participants' responses. In Study 1, however, this was only the case for probes responded as MW.

7 | GENERAL DISCUSSION

The current study examined the relationship between MW and eye movements during video lectures. A similar set of eye movement patterns were found in two video lectures. First, MW, compared with on-task episodes, was associated with a larger proportion of fixations on the instructor. Second, during MW, fixations on the slides tended to become longer in duration and less spread in location. These patterns

were observed during both self-caught and probe-caught MW, which significantly differed from those observed during non-MW episodes.

We obtained preliminary evidence that some eye movement patterns associated with MW tended to revert to normal immediately after reports of MW. This may suggest that the state of MW was terminated either because it reached awareness (self-caught) or because it was interrupted by the sudden onset of a thought probe (probe caught), leading to abrupt changes in eye movement behaviors. However, as noted before, the results were not entirely consistent between the two studies, especially for the probe-caught conditions. The sudden onset of a thought probe is naturally distracting and may interfere with both on-task and off-task processes. Moreover, analysis using the probe-caught data might have limited statistical power because it requires that probes occur at the same time participants are MW. In this sense, the self-caught method might be more beneficial in terms of statistical power because it does not limit the number of reports. Although the current results on how participants respond to realizing they are MW remain inconclusive, recovery from MW is an important topic in its own right, and we believe that contrasting eye movements before and after MW is a fruitful way to further understand the phenomenon.

In the sections below, we connect findings from the growth curve analysis to existing theories and discuss limitations that should be addressed in future research.

7.1 | Looking at the instructor

There are several possible explanations for the change in fixation allocation during MW, for example, a loss of top-down control during MW. As a result, the eyes were spontaneously drawn to salient objects. Wilson et al. (2018) argue that visual features of the instructor contain what Harp and Mayer (1998) termed *seductive details*, features that are salient but irrelevant to the task. The instructor's visual presence may bring in seductive details such as the instructor's physique, movements, and clothes. Although these features are often not central to learning, they may be more appealing than the static presentation on the slides. Although the learner may exert top-down control to avoid these seductive details when they are on-task, they might nevertheless lose control of their eyes when the mind is wandering.

Another possibility is that this pattern reflects a balance between on-task and off-task processing. It has been suggested that MW and task-relevant processing draw from the same limited pool of attentional resources (Thomson, Besner, & Smilek, 2015). Understanding texts and graphs can be effortful and must be supported by limited attentional resources, but simply looking at the instructor speaking may be less effortful and thus permit the mind to wander. If participants do not want to fully engage in the lecture (either because they have understood it or because they are not motivated to learn), they might avoid looking at the slides so that they can think about things that they deemed more important. This possibility can be tested by examining the intentionality of MW (Seli, Cheyne, Xu, Purdon, &

Smilek, 2015; Seli, Risko, & Smilek, 2016). Unintentional MW occurs when the learner loses control over thoughts despite their best intention to focus on the task, whereas intentional MW occurs when the learner decides to disengage from the task. Thus intentional MW is a more deliberate and controlled process than unintentional MW (Seli, Risko, Smilek, & Schacter, 2016). It may be meaningful in future research to examine to what extent the shift of fixation allocation is an intended behavior.

It is also important to consider how a third variable, such as the learner's comprehension, could be associated with both the observed eye movement patterns and MW. This is particularly important given the correlational relation between self-reports and eye movements. For example, comprehension decreases when learners do not know where to look (so they look more at the teacher and explore the slides less). And because the learner cannot follow the lecture content, they might start to mind wander. However, this notion is complicated by the absence of a relationship between MW and comprehension in Study 2.

7.2 | Looking at the slides

We found that MW was associated with increased fixation duration, a finding consistent with some previous studies (Bixler & D'Mello, 2016; Faber et al., 2018; Foulsham et al., 2013; Krasich et al., 2018; Reichle et al., 2010). Overall, the increased duration suggests that visual processing became less efficient, possibly because participants had to spend longer time processing information at one location before moving on to the next. Recent research (e.g., Tatler, Brockmole, & Carpenter, 2017) showed that the when and the where aspects of eye movement control are more connected with each other than previously assumed. Therefore, another possibility is that changes in where to move the eyes affected when to move the eyes. For example, if during MW participants preferred to look at certain stimuli on the slides that take longer to process (e.g., difficult words, longer words, etc.), the overall fixation duration could increase. Finally, it is also possible that participants were simply gazing at the slides blankly and not processing any information at all. Internally oriented cognition is often associated with decreases in external processing (Barron, Riby, Greer, & Smallwood, 2011). The fact that the eyes temporarily stop at one location may not necessarily indicate that the corresponding perceptual information is processed.

The reduced fixation dispersion across the slides area suggests that, during MW, learners were not actively moving their eyes to process information presented on the slides. Together with the fixation duration results, they show that information processing on the slides area was impaired during MW. However, this pattern is not consistent with two recent studies that reported greater dispersion of fixations during MW (Hutt, Mills, et al., 2017; Krasich et al., 2018). This discrepancy might be attributed to task-specific viewing strategies. For example, in the scene-viewing task, participants were asked to view scenes with complex visual details (urban scenes) for a later memory test (Krasich et al., 2018). A part of the memory test required the

participant to identify small vignettes extracted from the scenes. In this context, participants might tend to sample a few distant locations during MW as a strategy for quick and rough processing of the scene. In the current study, however, the processing demand oriented more towards comprehension and less towards remembering perceptual details. Therefore the sampling strategy that might have appeared in Krasich et al. (2018) could be absent in the current study. On the other hand, the current results appear to be consistent with those found in a simulated driving task, where the authors claimed that the driver's focus was narrower during MW (He et al., 2011). It is possible that the dispersion pattern and how it is affected by MW are task specific and more research is needed to make generalizations. Moreover, fixation dispersion may be a rather coarse measure that does not fully capture the underlying mechanism of attention allocation during MW.

7.3 | Limitations and future directions

One advantage of the measures examined in this study is that they can be easily calculated without heavily depending on specific events or details of the video. Features that depend less on the context may be more useful when future research tries to build MW detection algorithms that are generalizable across tasks. Indeed, existing work suggests that using global and local features combined do not improve prediction performance than using global features alone (e.g., Bixler & D'Mello, 2016; Faber et al., 2018; Hutt, Hardey, et al., 2017). However, a potential downside of using generic measures is that they may be less diagnostic of specific cognitive processes during lecture viewing. Using more fine-grained eye movement measures can shed light on the mechanisms of MW. For example, during MW, learners might fail to examine some critical information presented on the slides, which may predict worse memory of such information in the posttest. More generally, a lack of synchrony between the instructor's speech and the learner's eye movements may predict worse comprehension. Examining these questions may help to answer why MW does not always lead to worse learning outcomes (Varao-Sousa & Kingstone, 2015, 2019). It might be the case that the learner was MW at non-critical points of the lecture, which would weaken the relationship between attention and performance. The current study was not designed to answer these questions, but they are nevertheless crucial to advance our understanding of the complex relationship between attention and learning.

It is also worth questioning, given the wide variety in video production formats, to what extent the current findings can be generalized to other lecture styles. We varied the instructor, content, and arrangement of the instructor and slide frames, but this does not begin to exhaust the range of online video-based course formats. Many video lectures are simply classroom recordings, as they are relatively convenient and low cost for production. In another popular format, the teacher appears adjacent to or within the presentation, creating two areas of interest (as in the current study). It is also possible to integrate the instructor's image into the presentation (as often done with a green screen in televised weather forecasts) to allow for

more interactions between the instructor and the content (Bhat, Chinprutthiwong, & Perry, 2015). There are also lectures that show the instructor's image only at selected times (e.g., intermixed with slide presentations). Moreover, virtual agents replace real human instructors in some intelligent tutor systems (Hutt et al., 2016; Li, Kizilcec, Bailenson, & Ju, 2016). Finally, videos in some online learning platforms (e.g., Khan Academy) do not show the instructor at all. The format we used was the most commonly reported in a recent study (Santos-Espino, Afonso-Suárez, & Guerra-Artal, 2016) that surveyed 115 massive open online courses (MOOC) lectures, with 34 (30%) of them using a pairing of a talking head with slides. The question of how video format affects attention and learning in MOOCs is largely open (Kizilcec et al., 2014; Mayer & Moreno, 2003; van Gog, Verveer, & Verwee, 2014; van Wermeskerken & Gog, 2017; Wilson et al., 2018), but eye-tracking measures of student behavior in this context will play a role in understanding the relations between format and student experience of instruction.

Studying the relation between MW and learning is complicated by the otherwise good news that students are usually not in the MW state. In addition to increasing the sample size, there are a number of ways to increase the likelihood of MW, such as including an induction phase (Kopp, D'Mello, & Mills, 2015), extending the task length (Thomson et al., 2015), and inserting more thought probes to obtain more MW episodes, although all of these can alter the learning experience in ways that may make it seem less natural. Because we hope that the methods used here will be extended to other contexts, all data, stimuli, and script are freely available at <https://osf.io/ptj75>.

Overall, the current research showed that MW was associated with temporal changes in when and where the eyes moved during video lectures. These results contribute to a growing body of research uncovering objective signatures of MW in educational settings and may facilitate efforts in finding ways to identify MW to promote learners' attention. Although many questions remain unanswered, we hope that our work piques interests for more thorough studies on this topic in the near future.

DATA AVAILABILITY STATEMENT

All data, material, and scripts are available at <https://osf.io/ptj75/>.

ORCID

Han Zhang  <https://orcid.org/0000-0001-6087-0428>

Xin Sun  <https://orcid.org/0000-0001-5500-0620>

ENDNOTES

¹ Because the current paper primarily concerns eye movements, we did not distinguish the type of posttest questions (definitional/conceptual or probed/nonprobed) in data analysis. Both pretest and posttest questions are available at <https://osf.io/ptj75/files/>.

² We, furthermore, used the R packages *afex* (Version 0.21.2; Singmann, Bolker, Westfall, & Aust, 2018), *aspace* (Version 3.2; Bui, Buliung, & Rummel, 2012), *dplyr* (Version 0.7.8; Wickham, François, Henry, & Müller, 2018), *emmeans* (Version 1.1.3; Lenth, 2018), *ggpubr* (Version 0.1.7; Kassambara, 2018), *gridExtra* (Version 2.3; Auguie, 2017), *kableExtra* (Version 0.9.0; Zhu, 2018), *lme4* (Version 1.1.17; Bates,

Mächler, Bolker, & Walker, 2015), *lmerTest* (Version 3.0.1; Kuznetsova, Brockhoff, & Christensen, 2017), *papaja* (Version 0.1.0.9842; Aust & Barth, 2018), *ppcor* (Version 1.1; Kim, 2015), *stargazer* (Version 5.2.2; Hlavac, 2018), and *tidyr* (Version 0.8.1; Wickham & Henry, 2018).

³ We suggest the reader be cautious when interpreting the quadratic effects due to the risk of overfitting.

⁴ The number of reports is affected by the total length of the lecture, which is different between Studies 1 and 2. We decided to use average time because it was comparable across lectures.

REFERENCES

- Auguie, B. (2017). GridExtra: Miscellaneous functions for "grid" graphics. Retrieved from <https://CRAN.R-project.org/package=gridExtra>
- Aust, F., & Barth, M. (2018). papaja: Create APA manuscripts with R markdown. Retrieved from <https://github.com/crsh/papaja>
- Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, 37(3), 379–384. <https://doi.org/10.3758/BF03192707>
- Barron, E., Riby, L. M., Greer, J., & Smallwood, J. (2011). Absorbed in thought: The effect of mind wandering on the processing of relevant and irrelevant events. *Psychological Science*, 22(5), 596–601. <https://doi.org/10.1177/0956797611404083>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Beserra, V., Nussbaum, M., & Oteo, M. (2019). On-task and off-task behavior in the classroom: A study on mathematics learning with educational video games. *Journal of Educational Computing Research*, 56(8), 1361–1383. <https://doi.org/10.1177/0735633117744346>
- Bhat, S., Chinprutthiwong, P., & Perry, M. (2015). Seeing the instructor in two video styles: Preferences and patterns. *International Educational Data Mining Society*, Paper presented at the International Conference on Educational Data Mining (EDM) (8th, Madrid, Spain, Jun 26–29, 2015), pp. 305–312.
- 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. <https://doi.org/10.1007/s11257-015-9167-1>
- Bloom, B. S. (1953). Thought-processes in lectures and discussions. *The Journal of General Education*, 7(3), 160–169. Retrieved from <http://www.jstor.org/stable/27795429>
- Bui, R., Buliung, R. N., & Rummel, T. K. (2012). Aspace: A collection of functions for estimating centographic statistics and computational geometries for spatial point patterns. Retrieved from <https://CRAN.R-project.org/package=aspace>
- 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(21), 8719–8724. <https://doi.org/10.1073/pnas.0900234106>
- Core Team, R. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Dalmajer, E. S. (2014). Is the low-cost Eye Tribe eye tracker any good for research? *PeerJ PrePrints*, 2:e585v1. <https://doi.org/10.7287/peerj.preprints.585v1>
- Dalmajer, E. S., Mathôt, S., & Stigchel, S. d. (2014). PyGaze: An open-source, cross-platform toolbox for minimal-effort programming of eye-tracking experiments. *Behavior Research Methods*, 46(4), 913–921.
- Faber, M., Bixler, R., & D'Mello, S. K. (2018). An automated behavioral measure of mind wandering during computerized reading. *Behavior Research Methods*, 50(1), 134–150. <https://doi.org/10.3758/s13428-017-0857-y>

- Farley, J., Risko, E. F., & Kingstone, A. (2013). Everyday attention and lecture retention: The effects of time, fidgeting, and mind wandering. *Frontiers in Psychology*, 4, 619. <https://doi.org/10.3389/fpsyg.2013.00619>
- 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>
- 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. <https://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. <https://doi.org/10.3758/s13423-013-0468-2>
- Gullberg, M., & Holmqvist, K. (2006). What speakers do and what addressees look at: Visual attention to gestures in human interaction live and on video. *Pragmatics & Cognition*, 14(1), 53–82.
- Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*, 90(3), 414–434. <https://doi.org/10.1037/0022-0663.90.3.414>
- He, J., Becic, E., Lee, Y. C., & McCarley, J. S. (2011). Mind wandering behind the wheel: Performance and oculomotor correlates. *Human Factors*, 53(1), 13–21. <https://doi.org/10.1177/0018720810391530>
- Hlavac, M. (2018). Stargazer: Well-formatted regression and summary statistics tables. Bratislava, Slovakia: Central European Labour Studies Institute (CELSI). Retrieved from <https://CRAN.R-project.org/package=stargazer>
- Hutt, S., Hardey, J., Bixler, R., Stewart, A., Risko, E. F., & D'Mello, S. (2017). Gaze-based detection of mind wandering during lecture viewing. In *EDM, Proceedings of the 10th International Conference on Educational Data Mining*, pp. 226–231.
- Hutt, S., Mills, C., Bosch, N., Krasich, K., Brockmole, J., & D'Mello, S. (2017). "Out of the fr-eye-ing pan": Towards gaze-based models of attention during learning with technology in the classroom. In *Proceedings of the 25th conference on user Modeling, adaptation and personalization - UMAP '17* (pp. 94–103). Bratislava, Slovakia: ACM Press. <https://doi.org/10.1145/3079628.3079669>
- Hutt, S., Mills, C., White, S., Donnelly, P. J., & D'Mello, S. K. (2016). *The eyes have it: Gaze-based detection of mind wandering during learning with an intelligent tutoring system*. In *International Educational Data Mining Society, Proceedings of the 9th International Conference on Educational Data Mining*, pp. 86–93.
- Kane, M. J., Smeekens, B. A., von Bastian, C. C., Lurquin, J. H., Carruth, N. P., & Miyake, A. (2017). A combined experimental and individual-differences investigation into mind wandering during a video lecture. *Journal of Experimental Psychology: General*, 146(11), 1649–1674. <https://doi.org/10.1037/xge0000362>
- Kassambara, A. (2018). *Ggpubr: 'Ggplot2' based publication ready plots*. Retrieved from <https://CRAN.R-project.org/package=ggpubr>
- Kim, S. (2015). *Ppcor: Partial and semi-partial (part) correlation*. Retrieved from <https://CRAN.R-project.org/package=ppcor>
- Kizilcec, R., Papadopoulos, K., & Sritanyaratana, L. (2014). Showing face in video instruction: Effects on information retention, visual attention, and affect. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*. <https://doi.org/10.1145/2556288.2557207>
- Kopp, K., D'Mello, S., & Mills, C. (2015). Influencing the occurrence of mind wandering while reading. *Consciousness and Cognition*, 34, 52–62. <https://doi.org/10.1016/j.concog.2015.03.003>
- Krasich, K., McManus, R., Hutt, S., Faber, M., D'Mello, S. K., & Brockmole, J. R. (2018). Gaze-based signatures of mind wandering during real-world scene processing. *Journal of Experimental Psychology: General*, 147(8), 1111–1124. <https://doi.org/10.1037/xge0000411>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lenth, R. (2018). *Emmeans: Estimated marginal means, aka least-squares means*. Retrieved from <https://CRAN.R-project.org/package=emmeans>
- Li, J., Kizilcec, R., Bailenson, J., & Ju, W. (2016). Social robots and virtual agents as lecturers for video instruction. *Computers in Human Behavior*, 55, 1222–1230. <https://doi.org/10.1016/j.chb.2015.04.005>
- Lindquist, S. I., & McLean, J. P. (2011). Daydreaming and its correlates in an educational environment. *Learning and Individual Differences*, 21(2), 158–167. <https://doi.org/10.1016/j.lindif.2010.12.006>
- 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.
- Mayer, R. E. (2005). Principles of multimedia learning based on social cues: Personalization, voice, and image principles. In R. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning (Cambridge Handbooks in Psychology)*, pp. 201–212. Cambridge: Cambridge University Press.
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38, 43–52. https://doi.org/10.1207/S15326985EP3801_6
- McVay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology: General*, 141(2), 302–320. <https://doi.org/10.1037/a0025250>
- Min, I. (2018). *Nurturing cultures-A cross-cultural perspective on the intersection of parenting and teaching* (Doctoral dissertation). the University of Michigan.
- Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, 59, 475–494. <https://doi.org/10.1016/j.jml.2007.11.006>
- Ooms, K., Dupont, L., Lapon, L., & Popelka, S. (2015). Accuracy and precision of fixation locations recorded with the low-cost Eye Tribe tracker in different experimental setups. *Journal of Eye Movement Research*, 8(1), 1–24.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422. <https://doi.org/10.1037/0033-2909.124.3.372>
- Reichle, E. D., Reineberg, A. E., & Schooler, J. W. (2010). Eye movements during mindless reading. *Psychological Science: A Journal of the American Psychological Society / APS*, 21(9), 1300–1310. <https://doi.org/10.1177/0956797610378686>
- Risko, E. F., Anderson, N., Sarwal, A., Engelhardt, M., & Kingstone, A. (2012). Everyday attention: Variation in mind wandering and memory in a lecture. *Applied Cognitive Psychology*, 26(2), 234–242. <https://doi.org/10.1002/acp.1814>
- Samudra, P. G., Min, I., Cortina, K. S., & Miller, K. F. (2016). No second chance to make a first impression: The "thin-slice" effect on instructor ratings and learning outcomes in higher education. *Journal of Educational Measurement*, 53, 313–331. <https://doi.org/10.1111/jedm.12116>
- Santos-Espino, J. M., Afonso-Suárez, M. D., & Guerra-Artal, C. (2016). Speakers and boards: A survey of instructional video styles in MOOCs. *Technical Communication*, 63(2), 101–115.
- Schacter, D. L., & Szpunar, K. (2015). Enhancing attention and memory during video-recorded lectures. *Scholarship of Teaching and Learning in Psychology*, 1(1), 60–71. <https://doi.org/10.1037/stl0000011>
- Schoen, J. R. (1970). Use of consciousness sampling to study teaching methods. *Journal of Educational Research*, 63(9), 387–390. <https://doi.org/10.1080/00220671.1970.10884042>
- Seli, P., Cheyne, J. A., Xu, M., Purdon, C., & Smilek, D. (2015). Motivation, intentionality, and mind wandering: Implications for assessments of

- task-unrelated thought. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(5), 1417–1425. <https://doi.org/10.1037/xlm0000116>
- Seli, P., Risko, E. F., & Smilek, D. (2016). On the necessity of distinguishing between unintentional and intentional mind wandering. *Psychological Science*, 27(5), 685–691. <https://doi.org/10.1177/0956797616634068>
- Seli, P., Risko, E. F., Smilek, D., & Schacter, D. L. (2016). Mind-wandering with and without intention. *Trends in Cognitive Sciences*, 20(8), 605–617. <https://doi.org/10.1016/j.tics.2016.05.010>
- Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2018). *Afex: Analysis of factorial experiments*. Retrieved from <https://CRAN.R-project.org/package=afex>
- Smallwood, J. (2011). Mind-wandering while reading: Attentional decoupling, mindless reading and the cascade model of inattention. *Language and Linguistics Compass*, 5(2), 63–77. <https://doi.org/10.1111/j.1749-818X.2010.00263.x>
- 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. <https://doi.org/10.1037/a0030010>
- Smallwood, J., Fishman, D. J., & Schooler, J. W. (2007). Counting the cost of an absent mind: Mind wandering as an underrecognized influence on educational performance. *Psychonomic Bulletin & Review*, 14(2), 230–236. <https://doi.org/10.3758/BF03194057>
- Smallwood, J., & Schooler, J. (2015). The science of mind wandering: Empirically navigating the stream of consciousness. *Annual Review of Psychology*, 66(1), 487–518. <https://doi.org/10.1146/annurev-psych-010814-015331>
- Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin*, 132(6), 946–958. <https://doi.org/10.1037/0033-2909.132.6.946>
- Smilek, D., Carriere, J. S., & Cheyne, J. A. (2010). Out of mind, out of sight: Eye blinking as indicator and embodiment of mind wandering. *Psychological Science*, 21(6), 786–789. Retrieved from <http://www.jstor.org/stable/41062288>
- Steindorf, L., & Rummel, J. (2019). Do your eyes give you away? A validation study of eye-movement measures used as indicators for mindless reading. *Behavior Research Methods*, 1–15. <https://doi.org/10.3758/s13428-019-01214-4>
- Szpunar, K., Khan, N., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences*, 110(16), 6313–6317. <https://doi.org/10.1073/pnas.1221764110>
- Tatler, B. W., Brockmole, J. R., & Carpenter, R. H. S. (2017). LATEST: A model of saccadic decisions in space and time. *Psychological Review*, 124(3), 267–300. <https://doi.org/10.1037/rev0000054>
- Thomson, D. R., Besner, D., & Smilek, D. (2015). A resource-control account of sustained attention: Evidence from mind-wandering and vigilance paradigms. *Perspectives on Psychological Science*, 10, 82–96. <https://doi.org/10.1177/1745691614556681>
- 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(4), 1882–1886. <https://doi.org/10.1016/j.concog.2011.09.010>
- van Gog, T., Verveer, I., & Verveer, L. (2014). Learning from video modeling examples: Effects of seeing the human model's face. *Computers & Education*, 72, 323–327. <https://doi.org/10.1016/j.compedu.2013.12.004>
- van Wermeskerken, M., & Gog, T. v. (2017). Seeing the instructor's face and gaze in demonstration video examples affects attention allocation but not learning. *Computers & Education*, 113, 98–107. <https://doi.org/10.1016/j.compedu.2017.05.013>
- Varao-Sousa, T. L., & Kingstone, A. (2015). Memory for lectures: How lecture format impacts the learning experience. *PLoS One*, 10(11), e0141587.
- Varao-Sousa, T. L., & Kingstone, A. (2019). Are mind wandering rates an artifact of the probe-caught method? Using self-caught mind wandering in the classroom to test, and reject, this possibility. *Behavior Research Methods*, 51(1), 235–242. <https://doi.org/10.3758/s13428-018-1073-0>
- Voßkübler, A., Nordmeier, V., Kuchinke, L., & Jacobs, A. M. (2008). OGAMA (open gaze and mouse analyzer): Open-source software designed to analyze eye and mouse movements in slideshow study designs. *Behavior Research Methods*, 40(4), 1150–1162. <https://doi.org/10.3758/BRM.40.4.1150>
- Wammes, J. D., & Smilek, D. (2017). Examining the influence of lecture format on degree of mind wandering. *Journal of Applied Research in Memory and Cognition*, 6(2), 174–184. <https://doi.org/10.1016/j.jarmac.2017.01.015>
- Wang, J., & Antonenko, P. D. (2017). Instructor presence in instructional video: Effects on visual attention, recall, and perceived learning. *Computers in Human Behavior*, 71, 79–89. <https://doi.org/10.1016/j.chb.2017.01.049>
- Wickham, H., François, R., Henry, L., & Müller, K. (2018). *Dplyr: A grammar of data manipulation*. Retrieved from <https://CRAN.R-project.org/package=dplyr>
- Wickham, H., & Henry, L. (2018). *Tidyr: Easily tidy data with 'spread()' and 'gather()' functions*. Retrieved from <https://CRAN.R-project.org/package=tidyr>
- Wilson, K. E., Martinez, M., Mills, C., D'Mello, S., Smilek, D., & Risko, E. F. (2018). Instructor presence effect: Liking does not always lead to learning. *Computers and Education*, 122, 205–220. <https://doi.org/10.1016/j.compedu.2018.03.011>
- Young, M. S., Robinson, S., & Alberts, P. (2009). Students pay attention!: Combating the vigilance decrement to improve learning during lectures. *Active Learning in Higher Education*, 10(1), 41–55. <https://doi.org/10.1177/1469787408100194>
- Zhu, H. (2018). *KableExtra: Construct complex table with 'kable' and pipe. Syntax*. Retrieved from <https://CRAN.R-project.org/package=kableExtra>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Zhang H, Miller KF, Sun X, Cortina KS. Wandering eyes: Eye movements during mind wandering in video lectures. *Appl Cognit Psychol*. 2020;34: 449–464. <https://doi.org/10.1002/acp.3632>

APPENDIX A

This appendix included fixed effect estimates of the growth curve analysis models mentioned in the main text.

TABLE A1 Growth curve models for the relative fixation proportion of the teacher area

	Study 1: probe-caught	Study 1: self-caught	Study 2: probe-caught	Study 2: self-caught
On-task	-8.32* (3.68)	-1.52** (0.55)	-6.93* (2.75)	-0.39 (0.84)
Linear term	9.99* (4.93)	5.15*** (1.49)	25.18*** (6.42)	5.38** (1.94)
Quadratic term	3.59 (4.09)	-0.56 (1.14)	-1.27 (5.87)	-1.93 (1.88)
On-task: linear	6.12 (6.58)	-5.48** (1.97)	-4.24 (8.73)	-7.21** (2.68)
On-task: quadratic	-4.67 (5.19)	1.12 (1.52)	7.56 (7.37)	0.80 (2.49)
Constant	22.29*** (2.83)	13.76*** (1.44)	31.50*** (2.92)	22.42*** (1.39)

Note: Reference level for all models is mind wandering; standard errors are in parenthesis.

*** $p < .001$; ** $p < .01$; * $p < .05$.

TABLE A2 Growth curve models for the mean of fixation durations (*slides area*)

	Study 1: probe-caught	Study 1: self-caught	Study 2: probe-caught	Study 2: self-caught
On-task	-18.18 (12.62)	-3.86 (3.33)	-19.62 (13.22)	-8.19*** (2.45)
Linear term	87.72*** (20.30)	36.37*** (6.31)	97.98*** (26.78)	53.12*** (5.83)
Quadratic term	3.68 (23.39)	20.45* (8.17)	4.28 (17.02)	31.06*** (6.39)
On-task: linear	-90.68*** (26.78)	-37.92*** (8.93)	-84.95** (27.17)	-54.77*** (8.19)
On-task: quadratic	-34.56 (29.85)	-20.78 (11.56)	22.74 (22.38)	-32.31*** (8.88)
Constant	318.08*** (11.86)	274.04*** (7.59)	282.90*** (13.11)	253.21*** (5.46)

Note: Reference level for all models is mind wandering; standard errors are in parenthesis.

*** $p < .001$; ** $p < .01$; * $p < .05$.

TABLE A3 Growth curve models for fixation dispersion (*slides area*)

	Study 1: probe-caught	Study 1: self-caught	Study 2: probe-caught	Study 2: self-caught
On-task	0.03* (0.01)	0.01** (0.003)	0.02* (0.01)	0.002 (0.003)
Linear term	-0.06** (0.02)	-0.02*** (0.01)	-0.08*** (0.02)	-0.04*** (0.01)
Quadratic term	0.01 (0.02)	-0.01 (0.01)	-0.01 (0.02)	-0.02** (0.005)
On-task: linear	0.04 (0.02)	0.02** (0.01)	0.03 (0.02)	0.04*** (0.01)
On-task: quadratic	0.001 (0.02)	0.01 (0.01)	0.01 (0.02)	0.02* (0.01)
Constant	0.19*** (0.01)	0.22*** (0.005)	0.19*** (0.01)	0.22*** (0.004)

Note: Reference level for all models is mind wandering; standard errors are in parenthesis.

*** $p < .001$; ** $p < .01$; * $p < .05$.