GREVE, MAREN AUGUST 2022 PSYCHOLOGICAL SCIENCES

THOUGHT PROBES AS A SOURCE OF MIND WANDERING DEPEND ON ATTENTION CONTROL DEMANDS (46 pp.)

Thesis Advisor: Christopher Was

Mind wandering is a topic of great interest in many areas, but as with all psychological constructs, the interpretation of experimental results might depend on the way it is measured. A common way of measuring mind wandering in experiments is with self-report thought probes, which researchers have done in a variety of tasks. An important question with this methodology is if the probes themselves may be influencing participants' mind wandering. What is missing in the current literature is a comparison of the effect the number of thought probes has on mind wandering. As such, in the three experiments presented here we randomly assigned participants to receive varying amounts of thought probes during two different tasks. In the first experiment, we found that participants who had received fewer probes mind wandered less during an operation span task. In the second experiment, we found that participants who had received fewer probes mind wandered more during a video lecture. In the third experiment, which was conducted as a within-subjects design, we found that participants mind wandered more during the operation span task compared to the video and mind wandered more when they received more probes compared to fewer probes. The results suggest that thought probes interact with attentional control demands to influence mind wandering.

THOUGHT PROBES AS A SOURCE OF MIND WANDERING DEPEND ON ATTENTION CONTROL DEMANDS

A thesis submitted

To Kent State University in partial

Fulfillment of the requirements for the

Degree of Master of Arts

By

Maren Greve

August 2022

© Copyright

All rights reserved

Except for previously published materials

Thesis written by

Maren J. Greve

B.S., Northern Arizona University, 2019

M.A., Kent State University, 2022

Approved by

Christopher Was____, Advisor

Maria Zaragoza____, Chair, Department of Psychological Sciences

Mandy Munro-Stasiuk___________________________________, Dean, College of Arts and Science

LIST OF FIGURES

Acknowledgements

I would like to thank my advisor, Christopher Was, for his help and mentorship in completing these experiments.

Imagine being a Psychology 101 instructor in a large and crowded lecture hall. You would not be surprised to find a few students daydreaming and inattentive. Now, recall the last time you attended a keynote speech at a large conference. Likely, you can imagine yourself thinking about things other than the topic of the talk. Regardless of how important something might be, it feels like the mind cannot help but stray from time to time. Throughout everyday life, this is a common experience (Levinson, Smallwood, & Davidson, 2012; McVay, Kane, & Kwapil, 2009) that impacts task performance (McVay & Kane, 2009, 2012a, 2012b; Unsworth, & McMillan, 2013). Although one might be aware that mind wandering is detrimental to task performance, this knowledge appears to have little effect on the occurrence of mind wandering (Was, Hollis, & Dunlosky, 2019). Mind wandering occurs frequently, we are aware that it occurs, and it is often detrimental, yet we are often unable to prevent it.

Because of the ubiquitous and sometimes uncontrollable nature of mind wandering, psychological research of the phenomenon has spurred a cottage industry. As with all psychological research, the key to mind wandering research is how one measures the phenomenon. In the case of mind wandering, it is possible that measuring it may influence mind wandering itself, thus impacting the results of experiments designed to capture it. The goal of the current investigation was to examine the effects of measuring mind wandering on the occurrence of mind wandering.

Mind wandering is generally defined as thoughts that are irrelevant to one's goals for the current task. In the literature, these off task thoughts are often referred to as task-unrelated thoughts (TUTs). Although research has distinguished between voluntary and involuntary mind wandering (e.g., Robison & Unsworth, 2018; Seli, Wammes, Risko, & Smilek, 2016), the current investigation does not make this distinction, but simply focuses on any off-topic thoughts.

Different explanations exist for why mind wandering occurs. One explanation proposed by Smallwood and Schooler (2006) states that mind wandering takes up attentional (executive control) resources. Because these resources are shared, the demands on executive control induced by mind wandering draw resources away from the task at hand unless one employs proper metacognitive monitoring.

An alternative explanation is one of executive attention. The executive attention explanation posits that working memory deficits lead to a decrease in the ability to maintain goal-oriented thoughts during a task (Engle & Kane, 2004). Put differently, TUTs are the result of a lack of executive control to maintain goal focus. Expanding on the executive control explanation, McVay and Kane (2010; see also 2012a and 2012b) introduced the *control failures x concerns* explanation. This explanation proposed that pre-existing personal concerns (related to previously established but uncompleted goals) can be cued by events or objects in the environment, and once cued, these concerns may alter the contents of thought (McVay & Kane, 2010). McVay and Kane argued that mind wandering represents executive control failures that automatically generate thoughts in response to environmental and mental cues that create interference with which the executive control system must deal. Thus, failure of the executive control system during attention demanding tasks allows for the occurrence of TUTs.

An important aspect to mind wandering research is the way mind wandering is measured. Two popular methods of measuring mind wandering are post hoc self-reports and in vivo probes. Post hoc self-report involves asking participants at the end of the study to estimate or rate the total amount of mind wandering they experienced throughout the study. For example, Schacter &

Szpunar (2015) had participants watch a video lecture and, at the end, rate the extent of their mind wandering over the course of the lecture on a 7-point scale. Another example would be the research of Barron, Riby, Greer, and Smallwood (2011), which used the Dundee Stress State Questionnaire after participants had completed a visual oddball task (a task for examining brain activity associated with processing frequent and infrequent targets). The questionnaire asked participants to identify the type of thoughts they had throughout the task and to rate on a 5-point scale how often they experienced them.

Other studies have employed the use of in vivo mind wandering probes (i.e., thought probes). A thought probe interrupts the participant during a task and requires the participant to report aspects of their thoughts in the few moments prior to the probe. This may entail categorization of the content of their thoughts. For example, Hollis and Was (2016) required participants to watch a 13-minute video lecture. At four points during the lecture, a probe appeared and asked participants what they were thinking about in the 5 seconds prior to the probe. Participants responded by selecting from a choice of seven responses. Responses included "the video," "a memory from the past," and "thinking about using another technology", among others. In another study, participants were probed during a 60-minute lecture and asked whether they were mind wandering at the time the probe occurred (Risko et al., 2012). Others have asked participants to rate the degree to which they were off task. For example, Christoff et al. (2009) used a Likert-type scale from "completely on task" to "completely off task".

Of the two categories of mind wandering measures described, the present study focuses on the use of in vivo probes. We are not concerned with probe phrasing or response options, but rather how the presence of probes might influence the degree of mind wandering during the task in which the probes are inserted. Put differently, our goal was to examine how the presence of probes might influence mind wandering. Studying mind wandering necessarily relies on measuring mind wandering, but the probes themselves may be influencing participants, especially given the repetition of probes during a task. As the participant is probed throughout the task, they may come to expect future probes, which in turn may elicit mind wandering about the probes themselves. Thus, the goal of staying on task with the knowledge of forthcoming thought probes may produce subsequent TUTs about future thought probes and thus influence the amount or degree of mind wandering being reported. In the context of the control failure x concerns framework, "it is the build-up of the interference from thoughts such as current concerns that overwhelm the control system and cause disruptions to conscious focus that may, in turn, cause performance failures" (McVay & Kane, 2010, p. 331). The expectation of future probes and the potential impact on task performance might act as the concern component described in the framework.

Alternatively, more frequent thought probes might lead to less mind wandering. Thought probes may in fact act as breaks in mundane tasks that in turn reduce the amount of mind wandering experienced by participants. The current investigation was designed to examine the effect of the frequency of thought probes on mind wandering during experimental tasks.

Ours is not the first investigation of the use of thought probes in mind wandering studies. In recent years, researchers have emphasized the importance of probe usage and impact. Some studies have focused on the framing of probes (Weinstein, De Lima & van der Zee, 2018), whereas others focused on the merits of different types of probes (Varao-Sousa & Kingstone, 2019). Weinstein (2018) provides a review of probe usage, including the framing, probe type, and response options. However, research regarding the *number* of probes inserted in a task is of particular relevance, but the findings thus far are inconclusive. For example, Robison, Miller &

Unsworth (2019) found no effect of probe rate on mind wandering. In this experiment there were two groups, one with the double the amount of mind wandering probes as the other. There was no significant difference between the groups for amount of mind wandering reported or task performance on a sustained attention to response task (SART). Wiemers and Redick (2019) also found that high working memory capacity participants performed equally as well on the SART whether or not thought probes were inserted, suggesting that thought probes are a non-reactive measure of TUTs.

Contrary to the null results found by Robison et al. (2019) and Wiemers and Redick (2019), others have found evidence of more mind wandering for participants who received fewer probes. For example, Schubert, Frischkorn, & Rummel (2019) also split participants into two groups (one with double the amount of probes) and required participants to complete a SART. They found mind wandering was more likely for those who received fewer probes. Additionally, Seli, Carriere, Levene & Smilek (2013) varied the number of probes and the time between probes. Participants received between 5 and 25 probes presented across 600 trials of the Metronome Response Task (MRT). In the MRT, participants tap along to a steady beat and occasional probes measure attentional states by asking participants if they were on-task or mind-wandering at probe-onset. Seli et al. found a positive correlation between time between probes and the proportion of mind wandering reported during the task. In other words, the use of less frequent probes was related to more mind wandering.

In the studies just described, in which less frequent thought probes resulted in reports of more mind wandering, the sustained attention tasks are thought to measure the ability of participants to continue detecting environmental stimuli over time (Helton et al., 2009) and developed on the grounds of mindlessness theory (Manly et al., 1999, Robertson et al., 1997).

Mindlessness theory proposes that participants' withdrawal of effortful attention from the boring, repetitive task and their subsequent pre-occupation with their own thoughts results in reduce attention that may contribute to poor performance. As the task progresses, participants are more likely to engage in the task in an automatic, unconscious manner. As conscious effort fades and the participant disengages from the task, the participant becomes preoccupied with TUTs and daydreams (Giambra, 1995; Smallwood et al., 2004). Thus, errors of sustained attention are attributed to mind wandering.

Although some propose that errors on the SART are due to mind wandering, other researchers give an alternative interpretation of errors of sustained attention. The SART asks participants to complete trials quickly and accurately. Because it is impossible to be both maximally fast and accurate, participants must choose how to approach the task and balance the demand for speed and accuracy. Completing the task more quickly results in less accuracy and vice versa. This is known as the speed-accuracy trade-off. Dang, Figueroa, & Helton (2018) proposed that commission errors reflect strategy choice in response to the speed-accuracy trade-off. They found a large between-subjects correlation (r=-0.846) and within-subjects correlation (r=-0.767) between correct response times and commission errors. Dang and colleagues interpret these correlations as evidence that making commission errors is in fact fulfilling the task instructions to respond to trials quickly or accurately. This would be a strategy choice on the part of the participant and reflect attention (rather than inattention).

Although this is a valid alternative explanation of performance on the SART, many researchers have used SART to investigate mind wandering (McVay & Kane, 2009; Robison, Miller & Unsworth, 2019; Wiemers & Redick, 2019; Schubert, Frischkorn, & Rummel, 2019). Furthermore, Kane et al. argue that mind wandering rates are reliable across tasks and cite a number of studies that used latent variable analysis of multiple tasks with thought probes that reported mind wandering rates across multiple tasks converged into a single latent factor (e.g., Hollis & Was, 2016; Robinson & Unsworth, 2017, 2018). Germane to the current study, Kane et al. report on TUT-performance correlations in experimental tasks in which participants were not provided with overt feedback regarding performance. For example, Hollis and Was (2016) probed participants during a video lecture and then later measured comprehension for the video lecture content. TUTs have also been demonstrated to be associated with performance on tasks following errors not detected by participants (Kane et al.). Thus, we agree with Kane et al. that TUT reports do indicate inattention even without explicit feedback of performance. However, our interest is not in performance per se. Rather, as we stated, the goal of the current study was to determine if the number of thought probes administered during a given task would lead to more or less mind wandering subsequent to the probes.

Head & Helton (2016) suggested that response to thought probes might be made in response to performance feedback. Put differently, participants may justify their low performance by reporting that they were mind wandering when presented with a thought probe after performance feedback. We did our best to eliminate this possibility in the present investigation. Participants were not given any feedback on their performance before they were asked to respond to the thought probes. In Experiment 1, using a complex span task, probes appeared after the completion of a given trial, but always before the feedback screen. Participants only had their own self-judgment as a means of performance evaluation. Similarly in Experiment 2, using a video lecture, all probes appeared during the video but before they were given a posttest. Thus, in the current experiments, performance should not influence thought probe response beyond their own self-evaluation of performance, due to a lack of explicit feedback.

The Current Investigation

Similar to previous studies of the effect of thought probes on mind wandering, the present studies also seek to examine the relationship between the degree of mind wandering and the number of thought probes administered during a task. However, our studies differ in two important ways. First, we addressed this question with four conditions of increasing probes (e.g., 1, 3, 5, and 6 probes), as opposed to the two conditions used in some other studies. Second, we examined the impact of the varying number probes in tasks that differ from the SART in the amount of effortful attention required during the task. In our first experiment, we used the automated operation span task (OSPAN; Unsworth, Heitz, Schrock, & Engle, 2005) as the focus task, and in our second experiment we used a video lecture as the focus task.

We chose these two tasks because one of the important findings in the mind wandering literature is that participants tend to report more TUTs when completing easy compared to more difficult tasks (e.g., Giamba, 1989; Teasdale et al., 1995). Put differently, participants typically report more TUTs when completing tasks requiring less effortful attention. Although the focus of our investigation is not the influence of the amount of effortful attention required, we wanted to use two different types of tasks that have been employed in previous research that vary in the amount of effortful attention required, as the thought probes may impact the amount and degree of TUTs depending on the effortful attention that successful completion of the task requires.

Our first experiment employed the OSPAN task, as this is a common task used to measure working memory, but also in mind wandering research. For example, Mrzarek et al. (2012) added three thought probes at unpredictable intervals to three complex span tasks: the OSPAN, RSPAN, and SSPAN. Although not the focus of their study, comparison of participants who received thought probes to those in another study that completed the span tasks without thought probes resulted in no significant difference in performance on the tasks between the two groups. However, Mrzarek et al. did not examine whether the two groups differed in the amount or degree of TUTs reported during the tasks. A video lecture was chosen for the second study as having participants watch prerecorded lectures is also a common procedure used in the mind wandering literature (e.g., Hollis & Was, 2016; Risko et al., 2012; Wammes & Smilek, 2017; Was et al., 2019).

The differences in effortful attention required to complete the OSPAN task and watching a video lecture is quite large. The OSPAN task requires participants to solve a series of mathematical operations (e.g., 7 x 2 – 4 = ?) while maintaining a series of to-be-recalled letters. Thus, each trial in the OSPAN task requires effortful attention for successful completion. Watching a 13-minute video lecture on the other hand, does not require constant effortful attention and is likely to lead to participants to behave in a more automatic, unconscious manner. Hence, the differences in effortful attention will allow us to examine how thought probes might influence mind wandering under unique conditions.

There are two possible outcomes of our experiments. The first possibility is that administration of more thought probes will reduce the degree of mind wandering reported during the task. Seli et al. (2018) found that more intentional mind wandering occurred during an easy task (choice response time task) compared to a more difficult working memory task. However, they also found that unintentional mind wandering was greater during the working memory task. Multiple thought probes administered during the task might cue participants to stay on task. Put differently, when participants receive a thought probe(s) early in the task, it may cue them to stay on task and thus report less mind wandering later in the task.

Multiple probes might also interrupt a boring task and thus prohibit a participant's preoccupation with their own thoughts. Therefore, the tendency for conscious effort to fade and the participant to disengage from the task might be disrupted by thought probes, allowing for the participant to exhibit greater attention control.

An alternative possible outcome is that multiple thought probes act as a distraction. As participants are administered more probes during the task, they may come to anticipate future probes, which in turn may elicit mind wandering about the probes themselves. Thus, the goal of staying on task with the knowledge of forthcoming thought probes may produce subsequent TUTs about future thought probes and thus influence the amount or degree of mind wandering being reported. Experiment 1 was designed to tests these competing hypotheses.

Experiment 1

Methods

Participants

One hundred and thirty-eight undergraduates enrolled in a general psychology course at a large Mid-Western state university participated in the study for course credit. Sample size was determined using an estimated effect size of Cohen's $f = .30$, power of .80, and alpha of .05. It was determined that a sample of 128 participants would provide the necessary power. The study was approved by the university's institutional review board and all ethical standards of the American Psychological Association were followed.

Apparatus

Participants performed all measures on PCs with SVGA monitors and standard keyboards. Programming of all tasks was completed with E-Prime 2.0 software (Schneider, Eschman, & Zuccolotto, 2002). E-Prime controlled the stimulus presentation, timing, and data collection.

Design and Procedures

Participants were tested in groups of between one and six individuals in a room with barriers separating six computer carrels. Task instructions were presented via computer, and a research assistant was available if participants had questions.

To test the impact of multiple thought probes, we adapted the automated OSPAN task (Unsworth et al., 2005). In the OSPAN task, participants were presented a mathematical operation with a question mark at the end (e.g., "(4 × 2) + 3 = ?"). After mentally solving the problem, participants click the mouse. The question mark was then replaced with a proposed solution, and participants were required to indicate whether the solution was true or false by clicking the correct response with the mouse. Participants were then presented a letter from the English alphabet that they were instructed to remember for later recall. Following the final operation and letter pair of the trial, participants were presented with a screen displaying 12 possible letters next to check boxes. Participants used the mouse to click, in serial order, the letters they were asked to recall.

Participants first completed a practice block of the OSPAN prompting them to solve 15 mathematical equations. Participants next completed a practice block of the primary task: the letter memory part of the test. Letters were shown sequentially for 1000 ms, and at the end of the trial, participants were shown the entire pool of 12 letters plus a blank space and were asked to select the letters in the order they were presented by clicking the check box next to the letter. Participants could select the blank space for any letters they did not remember.

Participants then completed a final practice block that included both tasks (eight trials each with two letters and two mathematical equations). At the end of each trial, participants were asked to recall the letters and their positions. At the end of the practice, participants were shown their mean response time and accuracy on the mathematical equations.

In the experimental block, participants completed 18 trials with sets of three to seven letters from a set of 12 that were presented for 250ms per letter. Participants were instructed to obtain at least 85% accuracy on the secondary task (the equations) and were informed that on some trials they would be asked to report whether they were paying attention to the task. In addition to the trial-by-trial feedback on the secondary task, the sixth and ninth trials were followed by feedback on the cumulative accuracy rates. Trials in which participant took longer than $M + 2.5$ *SD* (person-specific, based on the second practice round) were counted as errors. Span scores were calculated as the total number of items recalled in correct serial order across all trials (Conway et al., 2005). Thought probes occurred immediately following the recall portion of the experimental trials but before the response time and accuracy feedback. At intervals specified by the researcher but unknown to the participants, response screens were followed by thought probes which asked participants to indicate the extent that their attention was on task-unrelated thoughts or on-task using a 5 point Likert-type scale (1 - *completely on-task,* 2 - *mostly on-task,* 3 - *both on the task and unrelated concerns,* 4 - *mostly on unrelated concerns,* 5 - *completely on unrelated concerns*). This thought probe procedure provides the opportunity to assess mind wandering during the crucial moment of task processing. After answering the thought probe, participants were instructed that they would begin a new trial.

As our interest is whether thought probes influence mind wandering during subsequent task attention, participants were assigned to receive different numbers of thought probes. Table 1 presents the distribution of thought probes during the OSPAN task. Participants received either 1, 3, 5, or 6 thought probes. The thought probes replaced screens at trials 5, 8, 10, 13, 16, 18. In the condition that only received 1 probe, the probe replaced response screen 18. The 3-probe condition received the probes at response screens 13, 16, and 18. This pattern was followed for the other conditions. We used this pattern because our intent was to determine if thought probes influenced participants responses to subsequent thought probes. We chose 6 as the maximum number of probes during the tasks as this approximates the number of thought probes often reported in the extant literature for a task of this length (e.g., Feng et al, 2013; Hollis & Was, 2016; Was et al, 2019; Szpunar et al, 2013).

Table 1

Distribution of Probes by Condition in Experiment 1

Condition	n	Trial 5	Trial 8	Trial 10	Trial 13	Trial 16	Trial 18
1-Probe	38						X
3-Probe	33				X	X	X
5-Probe	33		X	X	X	X	X
6-Probe	34	X	X	X	X	X	X

Results

Preliminary analysis

To ensure that participants were complying with the task instructions and engaging in both the storage and processing tasks for the OSPAN task, we checked performance on the processing component of the task. High performance on the processing component of the task suggested that participants were complying with the task instructions (*M* = 92%, *SD* = 6).

Table 2 presents the percentage for each thought probe response. The response percentages indicate that participants were more likely to respond 1-completely on task (35.5% of responses) or 2-mostly on task (31.9%) than they were to respond with the other options. Only thirty-two percent of participants responded with 3, 4, or 5. Regarding the degree of mind wandering reported, on the last (common across conditions) thought probe, the mean response was 2.20 (*SD* = 1.20).

Table 2

Response Rate on Final Mind Wandering Probe in Experiment 1

Probe Response	Frequency	Percent
1 (Completely on Task)	49	35.5
2 (Mostly on Task)	44	31.9
3 (On Both the Task and Unrelated Concerns)	21	15.2
4 (Mostly on Unrelated Concerns)	17	12.3
5 (Completely on Unrelated Concerns)	7	5.1
Total	138	100

Figure 1 depicts the comparison of the means of the four conditions on the final probe. Our original analytic plan was to conduct an analysis of variance comparing the four conditions on the final probe response. However, due to the disproportionate use of the response categories and the non-normality of the final probe response data (Kolmogorov-Smirnov = .24, $p < .001$), we revised our analytic strategy. [i] First, we bifurcated the final probe response into on-task (responses 1 and 2) and TUTs (responses 4 and 5). We eliminated response 3 because, conceptually, the response suggests one is equally on-task and off-task. We then conducted a binary logistic regression by predicting final probe response (i.e., dependent variable) by probe condition. We coded TUTs as 0 and on-task as 1. A logistic regression model was used to determine if condition (number of probes) had an impact on mind wandering at the final probe. We selected the 1-probe condition as the referent group in the analysis, as our interest is whether more probes would lead to more TUTs and compared the 1-probe condition to each of the other conditions.

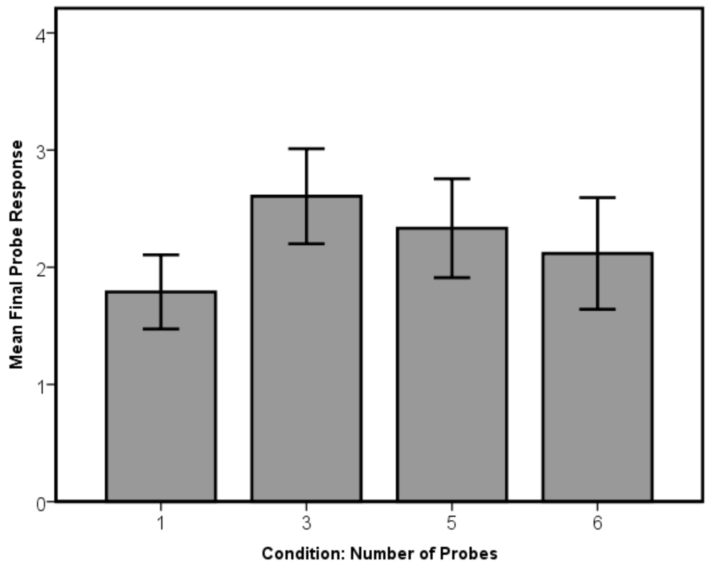

Figure 1. Mean Degree of Mind Wandering Response on Final Probe in Experiment 1.

Results of the logistic regression suggest that participants in each of the multiple-probe conditions were more likely to report a greater degree of mind wandering at the final probe than the 1-probe condition. Although Nagelkerke's R^2 of .09 indicated a small relationship between prediction and grouping and the model was deemed to not fit the data well according to the omnibus test, $\chi^2(3) = 7.05$, $p = .070$, prediction success overall was 79.5% and the Wald criterion from Model 1 (see Table 3 for the regression weights and Wald statistics) indicated that participants in the 1-probe condition were less likely to report mind wandering at the final probe than participants in the other conditions. We conducted a second logistic regression in which we compared the 1-probed condition to the mean of all other conditions. This analysis was significant, $\chi^2(1) = 7.04$, $p = .008$, Nagelkerke's $R^2 = .09$, $B = -1.71$, Wald z = 4.89, $p = .027$, Exp(B) = .18, 95% CIs [.18; .82] indicating that participants in the 1-probe condition were less likely to report mind wandering that participants in the other conditions combined.

Table 3

Logistic Regression Estimates in Experiment 1

Source		**Wald**	**OR**	**CI OR**	P
Intercept	2.74 (.73)***	14.11	.07		
1-Probe		4.91			
3-Probe	-1.74 (.85)*	4.17	.175	[1.07, 30.40]	.041
5-Probe	-1.69 (.85)*	3.94	.184	[1.02, 28.79]	.047
6-Probe	-1.69 (.84)*	4.05		[1.05, 27.81]	.044

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$

We also conducted repeated logistic regression in which we used the 6-probe condition as the referent group. Only the 1-probe condition differed from the 6-probe condition, B = -1.65, SE =.837, Wald = 4.05, p = .044. All other comparisons were nonsignificant (ps > .90). We also conducted a repeated contrast logistic regression in which each condition was compared to the condition with the next larger number of probes. In this regression, the only significant comparison was between the 1-probe condition and the 2-probe condition, B = -1.74, SE =.85, Wald = 4.17, p = .041. All other comparisons were nonsignificant (ps > .90).

As a last test of the amount of mind wandering reported at the final condition, we conducted a planned contrast comparing the 1-probe condition to all other conditions combined on the degree of mind wandering reported at the final probe using the full scale of responses. The result suggests that participants in the 1-probe condition reported a lesser degree of mind wandering than the other conditions, $F(3, 134) = 3.09$, $MSe = 4.22$, $p = .03$.

It was also important to determine if more thought probes lead to a reduction of speed and accuracy in the working memory task. Table 4 presents the means and standard deviations of OSPAN scores across conditions. As reaction time and accuracy during the OSPAN task were correlated, $r = -.28$, $p = .014$, we conducted a MANOVA with operation accuracy and operation reaction time (RT) as dependent variables with a planned contrast comparing the 1-probe condition to the mean of the other conditions using the 1-probe condition as the referent. The MANOVA effect was not significant, Pillais' Trace = .03, $F(2, 135) = 2.32$, $p = .10$. The univariate F tests indicated no significant differences in accuracy between the 1-probe condition and the other conditions, $F < 1$, but a significant difference in reaction time, $F(2, 135) = 4.68$, $p = .032$, suggesting that participants in the 1-probe condition exhibited faster reaction times than participants in the other conditions.

Table 4

Mean OSPAN Scores and Proportion of Participants Reporting TUTs at Final Probe by Probe Condition in Experiment 1

Condition	N	Mean (SD)	Proportion Reporting TUTs
1-probe	38	37.24 (15.96)	.06
3-probe	33	31.03 (17.01)	.21
5-probe	33	34.97 (16.38)	.21
6-probe	34	35.59 (16.05)	.24

Last, to examine the relationship between final probe response and performance on the OSPAN task, we conducted two correlations: 1) A point bi-serial correlations between OSPAN performance and the bifurcated final self-reports of mind wandering and 2) a Pearson correlation between OSPAN performance and the full scale final self-reports of mind wandering. The first correlation was not significant, $r = -.14$, $p = .13$, but the second correlation was significant, $r = -.19$, $p = .03$. These results suggest that performance on the OSPAN was negatively related to the degree of mind wandering at the final probe.

Discussion

Even though the effects are not large, the results of Experiment 1 suggest that more thought probes led to more mind wandering at the final probe during the OSPAN task. The logistic regression analysis indicated that participants in each of the multiple thought-probe conditions reported more TUTs than those in the single probe condition. Although the differences in mind wandering reported at the final probe are small, participants reporting a greater degree of mind wandering on the final probe performed more poorly on the OSPAN task than those reporting a lesser degree of mind wandering.

Although the results of Experiment 1 suggest that more thought probes led to more mind wandering at the final probe, they are certainly not conclusive. Only participants in the 1-probe condition reported a greater degree of mind wandering at final probe than the other conditions. If more thought probes lead to a greater degree of mind wandering, than we might expect to see differences between the other conditions as the number of thought probes increased. Also, the task in which the thought probes were embedded requires effortful attentional control for successful completion. This effortful control is likely to result in less mind wandering. Indeed, as the responses to the thought probes suggest, participants were on task the majority (67%) of the time. Therefore, it is quite likely that the nature of the task produced a minimum of mind wandering, thus hampering our ability to determine if more thought probes impact the amount of mind wandering.

In regard to the performance-mind wandering correlation, self-report of mind wandering on the last probe might have been influenced by performance feedback during the entire task. Put differently, although each thought probe preceded trial feedback, the accumulation of feedback throughout the task might have influenced participants' responses to the final probe. Therefore, although we did not expect individual trial feedback to influence probe responses, the buildup of feedback might have still influenced participants' reports of their mind wandering, as suggested by Head and Helton (2016).

Experiment 2 was designed to test the effects of multiple thought probes during a task requiring less control of attention effortful control of attention and without performance feedback during the task. A good number of studies regarding mind wandering and performance use educational materials, such as video lectures, as the stimuli (e.g., Hollis & Was, 2016; Risko et al., 2012) that require less effortful control of attention and do not provide feedback during learning. Therefore, we adapted the materials used by Hollis and Was (2016) to examine the effects of multiple thought probes while participants watched a lecture video.

Experiment 2

Method

Participants

Two hundred and eighty-five undergraduate students enrolled at the same university participated in Experiment 2. Sample size for Experiment 2 was also determined using an estimated effect size of Cohen's f = .30, power of .80, and alpha of .05. It was determined that a sample of 128 participants would be sufficient. This study was posted on the Psychology Department's subject pool website, near the end of the semester and students self-enrolled in the study through the site. Thus, the stopping rule for data collections was the end of semester closing for participant enrollment.

Design and procedure

Experiment 2 was conducted online. After enrolling in the experiment, participants were directed to a Qualtrics survey that handled the randomization of participants to the different conditions. All participants watched a 13-minute video lecture about the history of public relations. They watched this lecture on Edpuzzle (https://edpuzzle.com), an online video editing and sharing platform. Before the video started, a message gave participants additional instructions. They were asked to make the screen full screen, to refrain from taking notes, and to not alter the timeline of the video in any way (e.g. pausing, speeding up, etc.). They were also told there would be a test at the end.

Upon enrolling in the study, participants were randomly assigned to receive either 1, 2, 4, or 6 probes during the video. The number of probes in Experiment 2 differs from that of Experiment 1. The reason for this change was to evenly distribute the number of probes throughout the video. The probes were given at evenly spaced points, with all groups receiving the last probe at the same point in the video (see Table 5). The probes paused the video and asked, "Just now, where was your attention focused?". Participants were given a selection of responses to choose from (1 - *completely on the task*, 2 - *mostly on the task*, 3 - *somewhat on the task*, 4 - *somewhat on unrelated concerns*, 5 - *mostly on unrelated concerns, 6 - completely on unrelated concerns*). We changed the thought probe response options in order to force a choice between being on vs. being off task. After entering their selection, the lecture continued. At the end of the video, participants completed a 10-item posttest on the content of the lecture. The probes and posttest were edited into the video using the Edpuzzle interface.

Table 5

Distribution of Probes by Condition in Experiment 2

Condition	**0:00 min**	**3:00 min**	**5:00 min**	**7:00 min**	**9:00 min**	**11:00 min**	**13:00 min**	**13:08 min**
1-Probe							X	Posttest
2-Probe						X	X	Posttest
4-Probe				X	X	X	X	Posttest
6-Probe		X	X	X	X	X	X	Posttest

Results

Preliminary analysis

Before statistical tests were conducted, data from participants who 1) re-watched any part of the video, 2) accessed the video more than once, or 3) started but did not complete the video were eliminated from further analyses. In total, 83 participants were excluded for a total sample size of 202 participants.

As with Experiment 1, we planned to conduct an ANOVA comparing conditions on the final thought probe (see Figure 2). However, there was again an imbalance of responses to the final thought probe (see Table 6 for the responses rates). Seventy percent of responses to the final probe were responses 1-3. Because we again experienced a lack of range and a lack of normality in the response options (final probe response $M = 3.0$, $SD = 1.4$, Kolmogorov-Smirnov = .22, $p < .01$), we bifurcated the final thought probe into a binary variable. We recoded the first three response options (completely on the task, mostly on the task, somewhat on the task) as 1 (on task) and the last three options (somewhat on unrelated concerns, mostly on unrelated concerns, completely on unrelated concerns) as 0 (TUTs).

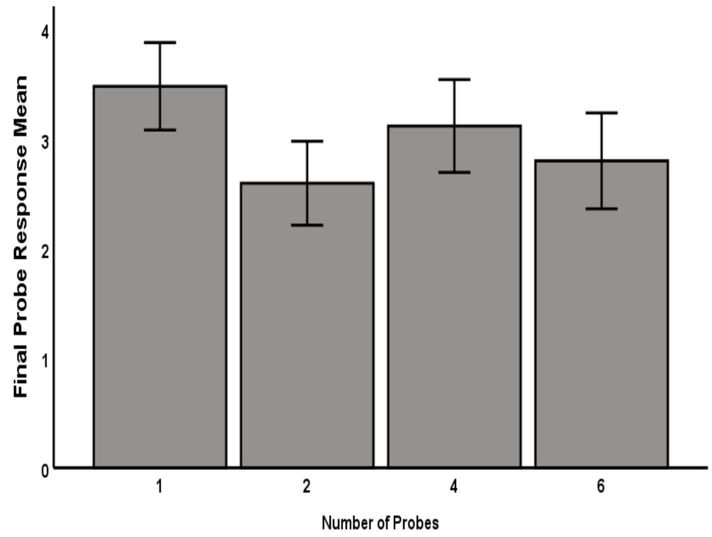

Figure 2. Mean Degree of Mind Wandering Response on Final Probe in Experiment 2.

Table 6

Response Rate on Final Mind Probe in Experiment 2

Probe Response	Frequency	Percent
1 (Completely on Task)	25	12.3
2 (Mostly on Task)	71	35.0
3 (Somewhat on Task)	47	23.2
4 (Somewhat on Unrelated Concerns)	17	8.4
5 (Moslty on Unrelated Concerns)	26	12.8
6 (Completely on Unrelated Concerns	17	8.4
Total	203	100

As with Experiment 1, we ran a series of binary logistic regressions. Results of the first logistic regression, in which the 1-probe condition was used as the referent group, suggest that participants in the 1-probe conditions were more likely to report mind wandering at the final probe than those in all other probe conditions. The Wald criterion estimates indicated that participants in the 1-probe condition were more likely to report mind wandering at the final probed than participants in the other conditions (see Table 7 for the regression weights and Wald statistics). The omnibus test of model fit was significant, $\chi^2(3) = 12.07$, $p = .007$, and Nagelkerke's R^2 of .08 indicated a small relationship between prediction and grouping. Prediction success overall was 70.4%.

Table 7

Logistic Regression Estimates for Experiment 2

Condition		**Wald**	**Odds Ratio**	***P***
1-Probe		11.678		.009
2-Probe	1.484 (.463)	10.266	4.410	.001
4-Probe	.892 (.423)	4.454	2.440	.035
6-Probe	.966 (.429)	5.069	2.627	.024

Results of the second repeated logistic regression, in which we used the 6-probe condition as the referent group suggest that only the 1-probe condition differed from the 6-probe condition, $B = .97$, $SE = .43$, Wald = 5.07, $p = .024$. All other comparisons were nonsignificant (ps > .30).

We also ran a repeated contrast logistic regression in which each condition was compared to the condition with the next larger number of probes. In this regression, the only significant comparison was between the 1-probe condition and the 2-probe condition, $B = 1.48$, $SE = .46$, Wald = 10.27, $p > .001$. All other comparisons were nonsignificant (*p*s > .22).

Next, we wanted to examine task performance. To this end, we conducted a one-way ANOVA with a planned contrast comparing the 1-probe condition to the mean of all other conditions. The results were significant, $F(1, 202) = 5.70$, $MSe = 4.88$, $p = .018$, indicating that participants in the 1-probe condition performed better on the video posttest than participants in the other three conditions. Table 8 contains means and standard deviations for video posttest scores by condition.

Table 8

Correct Posttest Responses by Condition

Condition	n	Mean	SD
1-probe	50	5.52	2.31
2-probe	51	5.10	2.54
3-probe	52	4.25	1.82
5-probe	50	5.32	2.01

Last, to examine the relationship between final probe response and performance on the video posttest we conducted two correlations: 1) A point bi-serial correlations between posttest performance and the bifurcated final self-reports of mind wandering and 2) a Pearson correlation between posttest performance and the full scale final self-reports of mind wandering. The first correlation was not significant, $r = -.11$, $p = .14$, but the second correlation was significant, $r = -.18$, $p = .01$. These results suggest, similar to those of Experiment 1, that performance on the task was negatively related to the degree of self-reported mind wandering at the final probe.

Discussion

The results of Experiment 2 are in stark contrast to the results of Experiment 1: Individuals in the 1-probe condition were *more* likely to report mind wandering compared to the other conditions at the final probe shared by all conditions. The most immediate explanation of the contrasting results is that the attention control demands of the two tasks in the two experiments lend themselves to varying degrees of mind wandering. Regarding the reports of a lesser degrees of mind wandering at the final probe during the video, one explanation is that thought probes may have acted as a cue to focus attention.

The results of our first two experiments indicate that participants in the 1-probe condition in Experiment 1 were *less* likely to report mind wander compared to the other conditions at the final probe shared by all conditions. Conversely, participants in the 1-probe condition in Experiment 2 were *more* likely to report mind wander at the final probe compared to the other conditions. We propose that the attentional control demands of the tasks used in the experiments may differentially impact the degrees of mind wandering reported, and that thought probes interact differently with those demands.

Task demands have been found to play a role in the degree to which participants mind wander in experimental tasks (Rummel & Boywitt, 2014). As Rummel and Boywitt reported, when they controlled task demands, participants reported higher mean TUT rates when task demands were low as compared to high task demands. Robinson, Miller, & Unsworth (2020) also found that task demands were related to the amount of mind wandering reported and that individual differences (e.g., working memory and attention control) mediated that relationship. Further examples include Seli et al. (2018), who found that participants experienced more unintentional mind wandering in an easy task (choice response time task) compared to a difficult task (working memory task), and Smallwood, McSpadden, & Schooler (2007) who found an increase in off-task reports during a go/no-go task for an easier, high target probability version of the task as compared to a more difficult, low target probability version. It is reasonable to assume the task demands of attention and difficulty described here that influenced these studies also influenced ours.

In Experiment 1, the demands of the OSPAN task involve a great degree of attentional control as participants are continuously required to make a judgment about the operations while maintaining the memory load across trials. In other words, the OSPAN is a highly active and attention control demanding task, with processing requirements at every single trial. Because of this, administration of thought probes may have been distracting or disruptive to participants, in turn allowing for TUTs possibly related to concerns regarding future thought probes. By contrast, the video lecture in Experiment 2 is a passive activity in which participants are only required to watch a video and not engage in effortful processing for the duration of the task, thus the task requires less attention control. This lack of attention control demands would likely promote mind wandering. The interruption provided by the thought probes disrupts the monotony of the task and thus, thought probes might have helped participants maintain attention control during the task.

Although we feel the evidence presented suggests that thought probes interact with the attention control demands of the task in which they are embedded, we also believe it is necessary to address an issue regarding the specificity of the results. In Experiments 1 and 2, only the 1-

probe condition differed from the other conditions. This suggests that there is something idiosyncratic about the 1-probe condition. However, when participants were only administered 1 thought probe, this probe always occurred at the end of the task. This design choice was made because if attention control demands are related to the amount or degree of mind wandering experienced during a task, the condition in which only 1-probe occurred requires that probe to be at the end of the task. In future studies, instead of a 1-probe condition, post hoc self-report of the number of TUTs, percent of time mind wandering, or scale report of the degree of mind wandering could be used in conjunction with several different probe conditions and a no probe condition. We do feel that because the two experimental tasks produce opposite results in terms of mind wandering at the end of the task, our conclusions are justified.

There are alternative explanations for our results. Head & Helton (2016) suggested that the response to thought probes might be made in response to performance. Participants may explain low performance by concluding that they were mind wandering and indicate such in a probe. We do not think that this is the case in Experiment 2 as all probes appeared during the video and before they were given the posttest. Thus, explicit feedback on performance could not have been used as an indicator of mind wandering, and performance did not influence thought probe response beyond their own self-evaluation of performance during the video portion of the task. However, it is also possible that the probes during the video simply acted as a cue to refocus attention. It may also be the case that participants were attempting to be compliant. That is, the probes acted as reminders that the "goal" of watching the video was to pay attention and to perform well on the posttest. Therefore, we do not think self-assessment of performance influenced thought probe responses, but responses may have not been flawless indicators of mind wandering.

Performance feedback may have influenced responses to the final probe during the OSPAN. In the OSPAN task, probes appeared after the completion of a given trial, but always before the feedback screen. At the first thought probe, participants could only use their own self-judgment to evaluate their task performance. Thus, participants were not likely to use trial performance to determine if they were mind wandering. However, it is possible that over the course of the task, participants administered a greater number of probes might have used cumulative performance feedback in a retrospective way as an indicator of mind wandering. If so, this may explain why more probes during the OSPAN task led to more reports of mind wandering at the final probe. Put differently, the ongoing feedback during the OSPAN may lead participants to attribute their poor performance to lapses in attention.

Overall, we speculate based on the present two experiments that mind wandering is a multi-faceted phenomenon, affected by both the measurement of mind wandering, the task in which one is engaged, and individual factors. We recognize that our study design limits our ability to draw robust conclusions regarding task demands and the effect that thought probes have on subsequent mind wandering. To investigate the potential effect of task demands on the occurrence of mind wandering, we conducted Experiment 3.

Experiment 3

Methods

Participants

Eighty-seven undergraduate students from the same state university participated in this study for course credit. A power analysis with an effect size of Cohen's *f* = .15, power of .80, and alpha of .05 suggested a sample size of 62 participants. Experiment 3 was conducted completely in the laboratory. Participants self-enrolled in the study via the Psychology Department's subject pool website.

Design and procedure

Because we suspected the number of probes might interact with the attentional demands of different tasks, participants were again randomly assigned to receive different numbers of thought probes. We reduced the number of conditions to two: a 1-probe condition and a 6-probe condition. We eliminated the middle conditions because did not anticipate a difference between these conditions. We also changed the scaling of the thought probes to a 1 (completely off task) to 10 (completely on task) point scale, due to the lack of range in the responses we experienced before.

Additionally, in order to address the conflicting results found in Experiments 1 and 2, Experiment 3 was conducted as a within-subjects design in order to determine the effect of type of task on reports of mind wandering. To that end, all participants completed the same two tasks that were used in Experiments 1 (OSPAN task) and Experiment 2 (video lecture), the order of which was counterbalanced. The procedure of the OSPAN remained identical to Experiment 1.

The procedure of the video changed slightly due to the shift from online back to in-person study. The video was programmed using E-Prime 3.0 on laboratory desktop computers. Instructions, probes, and posttest items were inserted into the video. The instructions were simplified because we no longer needed to tell participants to make the video full screen or to avoid skipping ahead or going back.

Upon arriving at the lab, participants were assigned to either the 1-probe or 6-probe condition. A research assistant started up the experiment on the computer for them and clarified any confusion. After reading the instructions, participants clicked the computer mouse to continue to the first task. After completing the first task, participants were instructed to press a button on the keyboard if that had been their first task or inform the research assistant if that had been their second task, at which point the research assistant would terminate the program.

Results

Table 9 provides the descriptive statistics (means and standard deviations) for response to the final probe in the OSPAN, response to the final probe in the video lecture, OSPAN total score, and video posttest score. Table 10 shows the distribution of responses to the final thought probe for both tasks. Final, response to the final thought probe was not normally distributed for in both the OSPAN (Kolmogorov-Smirnov = .245, $p < .001$) and the video (Kolmogorov-Smirnov = .151, $p = .001$). However, participants used the whole scale for the video, so we did not bifurcate probe response as we did in the previous experiments.

Table 9

Descriptive Statistics in Experiment 3

	Mean	SD
OSPAN Final Probe	8.64	1.60
Video Final Probe	6.34	2.48
OSPAN Total score	56.28	11.91
Video Posttest	5.53	1.55

Table 10

Response Rate on Final Mind Wandering Probe in Experiment 3

	OSPAN		Video	
Probe Response	Frequency	Percent	Frequency	Percent
1 (Completely Off Task)			2	3.1
2			5	7.8
3			3	4.7
4	3	4.7	5	7.8
5	1	1.6	7	10.9
6	3	4.7	7	10.9
7	4	6.3	11	17.2
8	11	17.2	9	14.1
9	18	28.1	11	17.2
10 (Completely on Task)	24	37.5	4	6.3
Total	64	100	64	100

Preliminary analysis

Before conducting our main analyses, we excluded data from participants who 1) obtained an accuracy of .79 or lower on the processing part of the OSPAN or 2) obtained an accuracy of .24 or lower on the posttest of the video. Following these criteria, we excluded 23 participants, resulting in a total sample size of 64 participants. We also checked our data for homogeneity of variance: For response to the final thought probe in the video only, Levene's test was significant, $F(3, 60) = 2.802$, $p = .047$. Therefore, we report Greenhouse-Geisser values below.

Main analysis

For our main analysis, we conducted a 2x2x2 repeated measures analysis of variance. The within-subjects variable (task type) contained two levels (OSPAN and video). The first between-subjects variable (number of probes) contained two levels (1-probe and 6-probe). The second between-subjects variable (order of tasks) contained two levels (OSPAN first and video first). There was no significant main effect of order, nor was there any significant interaction with order. Therefore in further analyses, we ignore the order of the tasks and focus instead solely on task type and the number of probes in each task.

Results suggested there was no interaction between task type and number of probes, $F(1, 60) = .047$, $p = .829$, $\eta^2_p = .001$. However, the analysis indicated that there was a significant main effect of task type, for which participants were more on task during the OSPAN and mind wandered more during the video, $F(1, 60) = 39.996$, $p < .001$, $\eta^2_p = .400$, and a significant main effect of number of probes, for which participants were more on task with 6 probes and mind wandered more with 1 probe, $F(1, 60) = 4.256$, $p = .043$, $\eta^2_p = .066$.

Discussion

We sought to resolve the conflicting pattern found in the first two experiments by again administering varying number of probes to participants in different tasks, this time as a within-subjects design. Of particular relevance to our research question, we did not find a significant interaction between task type and number of probes. Rather, we found that these variables were only impactful individually.

Contrary to our predictions, the number of probes participants received in a given task impacted their mind wandering, with those in the 1-probe condition reporting more mind wandering, and those in the 6-probe condition reporting more on-task thoughts. We originally thought that the 6-probe condition would help prevent mind wandering during a monotonous task like the video lecture and be a distraction during the demanding OSPAN task. However, it may be the case that, although these tasks differ in processing requirements, participants benefited from receiving multiple probes to refocus their attention in both tasks.

Alternatively, the effect of number of probes may also be due to social desirability. Participants may not have wanted to report being off-task, especially in the laboratory environment. In the 1-probe condition, participants only received the probe at the very end of the task and may have felt less pressure to respond in a performative way, compared to participants who received 6 probes, and thus, were potentially more aware of the variables of interest to the researcher. Participants receiving 6 probes were likely to be aware that the researchers were interested in their mind wandering during the task and may have used the probes to indicate that they were following directions and complying with the task.

In line with our predictions, we found a main effect of task type, such that participants mind wandered more during the video and were more on task during the OSPAN. We still contribute this to the attentional demands required of participants in these two different tasks. Further evidence for this effect comes from the following between experiment analysis.

Between Experiment Analysis

Before discussing the results of the three experiments, we present the results of a between experiment analysis of the degree of mind wandering reported at the final thought probe. We conducted an independent samples *t*-test with 1000 bootstrapped samples with the full scale report of mind wandering at the final probe as the dependent variable. Participants in Experiment 2 (the video) reported a greater degree of mind wandering (M = 3.00, SD = 1.48) at the final probe than participants in Experiment 1 (M = 2.05, SD = 1.25). This significant result indicates that participants in Experiment 2 self-reported a greater deal of mind wandering at the final probe than participants in Experiment 1, $t(318) = 5.80$, $Mean_{\text{Difference}} = .94$, $p < .001$, 95% CI [.64, 1.24], $d = .67$. This suggests to us that the type of task influenced participants' mind wandering.

General Discussion

In this series of experiments, we were interested in the impact of varying numbers of thought probes on the occurrence of mind wandering in different tasks. We investigated this by giving participants between 1 and 6 probes in an OSPAN task in Experiment 1, and a video lecture in Experiment 2. Finally, we gave participants either 1 or 6 probes in both tasks in Experiment 3.

Across these 3 experiments, we found an inconsistent relationship between number of thought probes and mind wandering. In Experiment 1, we found that more probes led to more mind wandering during the OSPAN task. However, recall the overall model did not provide an adequate fit to the data. Additionally, the model in which we compared the 1-probe condition to each next condition was not significant; only the model comparing the 1-probe condition to the mean of all other conditions was significant. Given these weak results, it could be that the results of Experiment 1 are inconclusive.

In Experiment 2, we found that more probes led to less mind wandering during a video lecture. In Experiment 3, participants in the 6-probe condition experienced significantly less mind wandering than the 1-probe condition, and this held across both tasks. Thus, Experiments 2 and 3 are consistent with each other regarding the effect of number of probes. Supporting this, we ran an analysis on the OSPAN portion of Experiment 3. We compared the 1- and 6-probe conditions and found a moderate effect, $t(62) = 2.17$, $p = .034$, $d = .55$, suggesting that participants mind wandered more with 1 probe and were more on task with 6 probes.

We did find support for task demands affecting mind wandering. In Experiment 3, we found a main effect of task type: participants were significantly more on task during the OSPAN and mind wandered more during the video. This suggested to us that the attentional demands related to each task might explain the direction of the results.

In Experiment 3, we also anticipated an interaction between the type of task a participant is engaged in and the number of probes the participant receives during that task because Experiments 1 and 2 appeared to show trends in opposite directions. Interestingly, although both main effects were significant in Experiment 3, we found no support for an interaction between task type and number of thought probes. This surprised us, given the conflicting results found in Experiments 1 and 2. This lack in consistency may be due to differences in the experiments. Experiments 1 and 2 each consisted of a single task, whereas Experiment 3 required participants to complete two tasks in a row. This additional time on task may have produced enough additional fatigue to affect participants' mind wandering, which some other research has demonstrated in a working memory task (Krimsky, Forster, Llabre, & Jha, 2017) and in a video lecture (Risko, et al., 2012). If this is the case, then it would make sense that we observed no order effects in our data. Put differently, it may be that extra time on task are imposing an increased burden, regardless of the type of task. Alternatively, the type of task could indeed matter, and that extra time may be compounded by the specific demands of the second task. This is a testable hypothesis that could be investigated with a modified version of the Experiment 3 design: comparing three conditions of either two working memory tasks, two video lectures, or one of each task type. Because the length would be comparable, differences should be due to the differences in the characteristics of the tasks.

How might the control failures x concerns framework explain our results? The control failure x concerns frameworks posits that a lapse in attention can allow pre-existing personal concerns to steal away attention, allowing TUTs. In Experiment 1, more probes led to more mind wandering. This may be because probes reminded participants of personal concerns (e.g. concerns over task performance). In Experiment 2, we found that more probes led to less mind wandering, so while probes might serve as a reminder for personal concerns, this in turn might cue participants to stay on task. Experiment 3 reflected Experiment 2: more probes led to less mind wandering in both tasks. Thus, Experiment 1 is at odds with Experiments 2 and 3. From a control failures perspective, we might expect a more demanding task to lead to more personal concerns being triggered and thus more mind wandering, but that is not what we observed in Experiment 3. In fact, participants were more on task in the OSPAN and mind wandered more in the video, reinforcing the idea that task demands are a greater influence than any concerns that may have stemmed from the probes.

Revisiting Head & Helton's argument, they might predict that low performing participants would have reported more mind wandering. However, we ran four correlations between task performance and response to the thought probe in Experiment 3, and only one was significant. The correlation between OSPAN score and mind wandering at the final probe was significant for the 1-probe condition, $r = .42$, $p = .034$, but not for the 6-probe condition, $r = -.039$, $p = .82$. The correlation between video posttest score and mind wandering at the final probe was not significant for either the 1-probe condition, $r = .288$, $p = .154$, or the 6-probe condition, $r = .122$, $p = .467$. The lack of significance could be because task performance overall for both tasks was rather low.

Implications and future directions

This research has real world implications for education. The need to pay attention is universally applicable, but especially in education. Relevant to our use of video lectures in the current experiments, online classes have become more and more prevalent in the current day and age, and so has the concern with students paying attention in online classes. Inserting some type of task interruption into online lectures could potentially be an easy way for instructors to check in with students and give an opportunity to refocus those who are mind wandering. This could take the form of thought probes by themselves or could involve combining thought probes with other components of the lecture (e.g. placing a thought probe before an important part of the lecture like a key concept). This kind of break in the task as a method of reducing mind wandering could have wide application to educational settings.

Although we found compelling evidence of an effect of task type and number of probes during a task, replication is needed to ensure these effects hold across different time points, and even across difference tasks. For example, an interesting conceptual replication would be to repeat Experiment 3, but with a more difficult educational task, such as a reading comprehension task, and an easier cognitive task, such as the SART. A replication of the main effect of task attentional demands from Experiment 3 would indicate that the nature of the task demands is driving the results, and a replication of the main effect of number of probes would suggest that probes can help redirect participants to the task at hand. Future education-focused research could also evaluate thought probes versus other types of questions inserted into a task in their ability to deter mind wandering. For example, future research could evaluate the efficacy of thought probes versus quiz questions as a means of minimizing mind wandering during class.

Conclusions

Although the results of Experiment 1 are inconclusive, Experiments 2 and 3 present robust evidence that more thought probes (compared to fewer thought probes) lead to less mind wandering in a task. We also found that greater task demands (compared to less attention demanding tasks) lead to less mind wandering in a task. This research suggests to us that mind wandering is a nuanced phenomenon, and we hope researchers consider the number of thought probes and the type and characteristics of the task in which they are used.

Open Practices Statement. Data presented in this manuscript are available on the Open Science Framework at DOI 10.17605/OSF.IO/DNS45

References

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.

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. doi:10.1073/pnas.0900234106

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786. https://doi.org/10.3758/bf03196772

Dang, J. S., Figueroa, I. J., & Helton, W. S. (2018). You are measuring the decision to be fast, not inattention: the Sustained Attention to Response Task does not measure sustained attention. Experimental brain research, 236(8), 2255-2262.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), The psychology of learning and motivation (pp. 145–199). New York, NY: Academic Press.

Feng, S., D'Mello, S., & Graesser, A. C. (2013). Mind wandering while reading easy and difficult texts. Psychonomic bulletin & review, 20(3), 586-592.

Giambra, L. M. (1989). Task-unrelated-thought frequency as a function of age: A laboratory study. *Psychology & Aging, 4*, 136–143.

Giambra, L. M. (1995). A laboratory method for investigating influences on switching attention to task-unrelated imagery and thought. *Consciousness and Cognition*, *4*(1), 1-21.

Head, J., & Helton, W. S. (2018). The troubling science of neurophenomenology. *Experimental brain research*, *236*(9), 2463–2467. https://doi.org/10.1007/s00221-016-4623-7

Helton, W. S., Kern, R. P., & Walker, D. R. (2009). Conscious thought and the sustained attention to response task. *Consciousness and Cognition*, *18*(3), 600-607.

Hollis, R. B., & Was, C. A. (2016). Mind wandering, control failures, and social media distractions in online learning. *Learning and Instruction, 42*, 104–112. https://doi.org/10.1016/j.learninstruc.2016.01.007

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

Levinson, D. B., Smallwood, J., Davidson, R. J. (2012). The persistence of thought: evidence for a role of working memory in the maintenance of task-unrelated thinking. *Psychological Science*, *23*(4), 375-380. https://doi.org/10.1177/0956797611431465

Manly, T., Robertson, I. H., Galloway, M., & Hawkins, K. (1999). The absent mind:: further investigations of sustained attention to response. *Neuropsychologia*, *37*(6), 661-670.

McVay, J.C., & Kane, M.J. (2009). Conducting the Train of Thought: Working Memory Capacity, Goal Neglect, and Mind Wandering in an Executive-Control Task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(1), 196-204. https://doi.org/10.1037/a0014104.

McVay J.C., Kane M.J. (2010) Adrift in the Stream of Thought: The Effects of Mind Wandering on Executive Control and Working Memory Capacity. In: Gruszka A., Matthews G., Szymura B. (eds) Handbook of Individual Differences in Cognition. The Springer Series on Human Exceptionality. Springer, New York, NY. https://doi.org/10.1007/978-1-4419-1210-7_19

McVay, J. C., & Kane, M. J. (2012a). 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.

McVay, J. C., & Kane, M. J. (2012b). 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

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

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

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.proxy.library.kent.edu/10.1002/acp.1814

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, *35*(6), 747-758.

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

Robison, M. K., Miller, A. L., & Unsworth, N. (2020). A multi-faceted approach to understanding individual differences in mind-wandering. *Cognition*, *198*, 104078.

Robison, M. K., & Unsworth, N. (2017). Working memory capacity and mind-wandering during low-demand cognitive tasks. *Consciousness and cognition*, *52*, 47–54. https://doi. org/10.1016/j.concog.2017.04.012

Robison, M.K., & Unsworth, N. (2018). Cognitive and Contextual Correlates of Spontaneous and Deliberate Mind-Wandering. Journal of *Experimental Psychology: Learning, Memory, and Cognition*, *44*, 85–98.

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*(5), 1309-1315.

Schacter, D. L., & Szpunar, K. K. (2015). Enhancing attention and memory during video-recorded lectures. Scholarship of Teaching and Learning in Psychology, 1(1), 60–71. https://doi-org.proxy.library.kent.edu/10.1037/stl0000011

Schubert, A.-L., Frischkorn, G. T., & Rummel, J. (2019). The validity of the online thought-probing procedure of mind wandering is not threatened by variations of probe rate and probe framing. *Psychological Research*. https://doi-org.proxy.library.kent.edu /10.1007/s00426-019-01194-2

Seli, P., Carriere, J. S. A., Levene, M., & Smilek, D. (2013). How few and far between?

Examining the effects of probe rate on self-reported mind wandering. *Frontiers in Psychology, 4*. https://doi-org.proxy.library.kent.edu/10.3389/fpsyg.203.00430

Seli, P., Konishi, M., Risko, E. F., & Smilek, D. (2018). The role of task difficulty in theoretical accounts of mind wandering. *Consciousness and Cognition: An International Journal, 65*, 255–262. https://doi-org.proxy.library.kent.edu/10.1016/j.concog.2018.08.005

Seli, P., Wammes, J. D., Risko, E. F., & Smilek, D. (2016). On the relation between motivation and retention in educational contexts: The role of intentional and unintentional mind wandering. *Psychonomic Bulletin & Review*, *23*(4), 1280–1287. https://doi.org/10.3758/s13423-015-0979-0

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*(4), 657-690.

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

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

Szpunar, K. K., Khan, N. Y., & 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.

Teasdale, J. D., Dritschel, B. H., Taylor, M. J., Proctor, L., Lloyd, C. A., Nimmo-Smith, I., & Baddleley, A. D. (1995). Stimulus-independent-thought depends on central executive resources. *Memory & Cognition, 23,* 551–559.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. https://doi.org/10.3758/bf03192720

Unsworth, N., & McMillan, B.D. (2013). Mind wandering and reading comprehension: examining the roles of working memory capacity, interest, motivation, and topic experience. *Journal of experimental psychology*. *Learning, memory, and cognition, 39*(3), 832-842 .

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.proxy.library.kent.edu/10.3758/s13428-018-1073-0

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.

Was, C.unsworthA., Hollis, R.B., & Dunlosky, J. (2019). Do students understand the detrimental effects of mind wandering during online learning? *Computers in Education,, 135*, 113-122.

Weinstein, Y. (2018). Mind-wandering, how do I measure thee with probes? Let me count the ways. Behavior Research Methods, 50(2), 642–661. https://doi-org.proxy.library.kent.edu/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*(2), 754–760. https://doi-org.proxy.library.kent.edu/10.3758/s13423-017-1322-8

Wiemers, E. A., & Redick, T. S. (2019). The influence of thought probes on performance: Does the mind wander more if you ask it?. *Psychonomic Bulletin & Review*, *26*(1), 367-373.

---

[i] We conducted the ANOVA and found an overall effect of condition, $F(3, 134) = 3.09$, $MSe = 1.37$, $p = .029$, suggesting that there was a significant difference in the degree of mind wandering reported between conditions on the final thought probe. Post hoc analyses using Tukey's HSD, Bonferroni, and Scheffe all indicated that significant differences were only between the 1-probe and 3-probe conditions (p = .02, .02, and .40 respectively).

For example, two prior studies of the effects of probe frequency on mind wandering found that mind wandering was associated with fewer or less frequent probes (Schubert et al., 2019; Seli et al., 2013).