

The influence of thought probes on performance: Does the mind wander more if you ask it?

Elizabeth A. Wiemers & Thomas S. Redick

Psychonomic Bulletin & Review

ISSN 1069-9384

Volume 26

Number 1

Psychon Bull Rev (2019) 26:367-373

DOI 10.3758/s13423-018-1529-3

Psychonomic Bulletin & Review

VOLUME 26, NUMBER 1 ■ FEBRUARY 2019

PB&R

EDITOR

Gregory Hickok, *University of California, Irvine*

ASSOCIATE EDITORS

Jessica Cantlon, *University of Rochester*

Greig de Zubicaray, *University of Queensland*

Stephen D. Goldinger, *Arizona State University*

Antonia Hamilton, *University College London*

Marc Howard, *Boston University*

Ben Newell, *University of New South Wales*

Alison Preston, *University of Texas at Austin*

Liina Pykkänen, *New York University*

Sarah Shomstein, *George Washington University*

Mark Steyvers, *University of California, Irvine*

Eric-Jan Wagenmakers, *University of Amsterdam*

A PSYCHONOMIC SOCIETY PUBLICATION

www.psychonomic.org

ISSN 1069-9384



**PSYCHONOMIC
SOCIETY®**



Springer



Springer

Your article is protected by copyright and all rights are held exclusively by Psychonomic Society, Inc.. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".



The influence of thought probes on performance: Does the mind wander more if you ask it?

Elizabeth A. Wiemers¹ · Thomas S. Redick¹

Published online: 17 September 2018
 © Psychonomic Society, Inc. 2018

Abstract

Mind-wandering has emerged in the past decade as a popular topic in many areas of psychological research. Numerous studies have demonstrated the potential costs and benefits of mind-wandering in relation to ongoing task performance, along with more recent work examining the nature of different types of mind-wandering. A common method of measuring mind-wandering in laboratory research is to embed self-report thought probes at random intervals within an ongoing task. However, a critical issue to determine is whether or not the presence of the thought probes fundamentally alters how an individual typically performs on the task. In the current study, $N = 149$ participants completed a sustained attention to response task (SART) with and without the presence of mind-wandering thought probes. In addition, participants completed operation and symmetry span measures of working memory capacity, as several studies have examined the relationship between individual differences in working memory capacity and mind-wandering using thought probes on the SART. The results indicate that SART performance does not differ whether thought probes are included or not. Individuals higher in working memory capacity produced better SART performance in the conditions with and without thought probes. In addition, individuals in working memory capacity were negatively correlated with mind-wandering frequency. The results indicate that thought probe measurement is a non-reactive method to measure mind-wandering in attention and inhibition tasks.

Keywords Mind-wandering · Working memory · Thought probes

Mind-wandering has emerged in the past decade as a popular topic in many areas of psychological research. Numerous studies have demonstrated the potential costs and benefits of mind-wandering in relation to ongoing task performance (Smallwood and Schooler, 2015), along with more recent work examining the nature of different types of mind-wandering (e.g., Kane et al., 2017). Additionally, interventions have been developed in an effort to reduce mind-wandering and improve cognitive and

academic performance (Mrazek, Franklin, Phillips, Baird, & Schooler, 2013).

A variety of methods have been used to measure mind-wandering. However, the most widely used method for quantifying mind-wandering behavior is the use of self-report thought probes embedded within a select proportion of trials in a cognitive or perceptual task (Weinstein, 2018). For example, a paradigm used frequently in the mind-wandering literature is to intersperse thought probes between certain trials on the sustained-attention-to-response task (SART). Note that the SART acronym is consistent with the mind-wandering literature, but the SART is simply a go/no-go task. The SART involves items of a certain category (e.g., animals), designated as frequent go trials, mapped to a single button press, and items of a different category (e.g., foods), designated as infrequent no-go trials requiring no response. Although the format and response options of the thought probes used in studies

✉ Elizabeth A. Wiemers
 ewiemers@purdue.edu

¹ Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, IN 47907, USA

with the SART vary widely (Weinstein, 2018), the general pattern of results is that individuals who self-report a higher frequency of mind-wandering make more errors and exhibit more variability in response times (RTs) on correct go trials (McVay & Kane, 2009; Unsworth & McMillan, 2014).

The role of thought-probe frequency in mind-wandering research

Despite the widespread usage of self-report thought probes within the SART in mind-wandering research, little work has systematically focused on methodological choices that may influence both mind-wandering behavior and SART performance. For example, the frequency with which thought probes occur within the SART varies across studies, sometimes appearing on as little as 2% (Seli, Risko, & Smilek, 2016) or more frequently at nearly 7% (McVay & Kane, 2009) of all trials. However, a critical issue is to determine whether the presence of the thought probes fundamentally alters how an individual typically performs on the task. One possibility is that a participant made aware of their off-task thoughts via the presence of thought probes may subsequently focus more on the task, knowing they will be periodically asked whether they are on task, producing better overall performance. Alternatively, a participant made aware of their off-task thoughts may subsequently consider off-task thoughts more frequently due to their awareness of them, resulting in poorer performance throughout the task. A third option is that the inclusion of thought probes throughout the SART has no effect on performance compared to versions without thought probes, which is probably the preferred outcome for mind-wandering researchers using this paradigm. To our knowledge, the current study is the first to explicitly examine whether inclusion of thought probes affects SART performance. The research question is relatively simple, but the results have important implications for the mind-wandering literature more broadly.

Two studies are particularly relevant to the current work. First, Seli, Carriere, Levene, and Smilek (2013) explicitly addressed how frequency of thought probes affected ongoing task performance, albeit with the metronome response task as the primary activity instead of the SART. In the metronome response task, participants pressed a button to keep time with a metronome tone that was presented. Across participants as a between-subjects manipulation, mind-wandering thought probes were presented after 0.8–4.2% of all trials. Seli et al. (2013) observed the typical finding for a cognitively-demanding task, specifically that higher self-reported mind-wandering was associated with worse task performance (*viz.*, more variable RTs). More germane to the current work, endorsement of off-task behavior increased as the frequency of thought probes decreased; however, the frequency of thought

probes had no relationship with metronome tone task RT variability. As noted by Seli et al., the observation that individuals subjectively reported mind-wandering more often when thought probes were less frequent could represent an actual increase in off-task thoughts, or it might be evidence of a bias to report increased mind-wandering without an actual increase in its incidence. Because participants in the various thought-probe frequency conditions did not vary in their task performance, Seli et al. concluded the response-bias interpretation was more likely.

Another relevant study is Robison, Miller, and Unsworth (2017), which we were unaware of until after our data collection was complete. In Experiment 1, Robison et al. manipulated thought probe frequency during the SART as a between-subjects variable, with participants receiving thought probes 6.6% or 13.1% of the time. The results indicated that SART performance did not differ between the two conditions of thought-probe frequency. In addition, subjective reports of the proportion of time spent on on-task/off-task behavior did not vary between the two thought-probe conditions.

Although both studies reported minimal effects of different amounts of thought probes during an ongoing task, our goal was to determine if SART performance is similar with and without thought probes. Because the thought probe specifically calls the participants' attention to their own relationship with the task, the presence of any probes during the task could have consequences for their performance that manipulating frequency would not further affect.

Mind-wandering and working memory capacity

Another consideration for investigating the role of thought probes on performance is whether or not all individuals are equally affected by thought probe presence during the task. Working memory capacity (WMC) is a promising candidate as an individual-differences variable that has been studied extensively in the mind-wandering literature. WMC is the ability to control attention in a goal-directed manner, using maintenance and retrieval of relevant information to guide behavior (Engle & Kane, 2004). Speculatively, interruptions caused by thought probes during demanding cognitive tasks could be either beneficial (*e.g.*, provide a break) or detrimental (*e.g.*, require re-engagement with cognitive task), particularly for low-WMC individuals. Although most research indicates that individuals lower in WMC are more likely to report off-task thoughts during ongoing task performance (see below), some studies have suggested a more nuanced relationship between WMC and mind-wandering propensity. For example, high-WMC individuals have reported more mind-wandering, with no performance decrement, during tasks with low cognitive demand (Levinson, Smallwood, & Davidson, 2012; but

see Robison & Unsworth, 2017, for conflicting results). Further, Rummel and Boywitt (2014) posit that cognitive control allows those with higher WMC to adjust their level of mind-wandering based on the task demands.

However, studies investigating the relationship between WMC and mind-wandering using the SART with thought probes have shown that individuals with higher WMC show better SART performance (higher d' and lower RT variability) and lower rates of self-reported mind-wandering (McVay & Kane, 2009; Unsworth & McMillan, 2014). Of note, a separate large-sample study produced correlations of similar magnitude between four complex span tasks and SART performance without the inclusion of thought probes (Redick et al., 2016). This latter finding hints that the relationship between WMC and SART performance may be unaffected by the inclusion of thought probes, although a comparison of the correlations within the same sample would provide more direct evidence.

Current research

The current study used a within-subjects manipulation to compare SART performance when thought probes were and were not administered during the task. In addition, we measured individual differences in WMC in order to determine whether the effect of thought-probe presence on SART performance varies as a function of WMC, as it is specifically the low-WMC individuals we expect would be affected.

Method

Participants

Students from Purdue University were offered course credit to participate. In total, 149 students between the ages of 18 and 35 years completed this single-session study, which took around 45 min to complete. The sample size is just short of our target of $N = 150$ because of a computer error, resulting in an incomplete session for one participant.

The final sample size used in the analyses reported below was $N = 137$. Two participants apparently reversed the go/no-go response mapping, resulting in near 0% performance on both trial types. An additional seven participants were excluded based on go accuracy less than 80%, and three more participants were excluded based on no-go accuracy less than 11%¹. The mean age of the final sample was 19.31 (1.51) years, and the participants' self-report responses to a demographics questionnaire indicated the composition of the sample was 45% female and 89% native-English speakers.

Tasks

Operation span (Redick et al., 2012; Unsworth, Schrock, Heitz, & Engle, 2005) In this task, participants alternated viewing a letter and judging whether a proposed answer to an algebraic equation is correct. Between three and seven letters with interleaved equations appeared before participants were asked to serially recall the letters, using the mouse to click on the box next to the letters presented. There were three trials of each list length, with 1 point for each letter recalled in the correct order, resulting in a maximum possible score of 75. Participants were instructed to maintain an accuracy of at least 85% on the interrupting equation questions, and this percentage was provided to them throughout the task to monitor.

Symmetry span (Redick et al., 2012; Unsworth, Redick, Heitz, Broadway, & Engle, 2009) This complex span task was similar in structure to operation span. However, participants recalled locations of a red square presented within a 4×4 grid, and made judgments about whether a black and white figure is symmetrical. List lengths of square locations ranged from 2 to 5, with three trials of each length for a maximum possible recall score of 42.

SART. In this task, Arabic numerals between 0 and 9 were presented centrally on-screen for 300 ms, followed by a 900-ms mask, which together resulted in a 1,200-ms response window. Participants were asked to press the spacebar as quickly as they could whenever the number presented was not 3 and to withhold responding when a 3 appeared. There were ten practice trials with letters instead of numbers before each of the two blocks. Then, for each block, 270 real trials were presented, with the first ten always being go trials. Of the 270 trials, 30 were no-go trials. For the block with thought probes, the participants were warned during the instructions that the question “What were you thinking about just before this screen appeared?” would appear intermittently with four response options (“Blank,” “Off task, anything else,” “On task, not the goal,” “On task, thinking about the goal”). During a practice screen, participants were shown an example of a thought for each possible response and were asked to be as honest as possible. The four possible response options were presented either in the order above or reversed to limit the possible bias of scale direction (Weinstein, 2018). On the thought probe screen, participants clicked on one of the four options to

¹ Of all the main effects and interactions across analyses reported below, only two differed between the final ($N = 137$) versus the entire ($N = 149$) sample. First, with accuracy as the dependent variable, the WMC by trial type interaction was not significant when analyzed with the full sample ($p = .401$). Second, with mean RT as the dependent variable, the WMC by probe presence interaction was significant when reanalyzed with the full sample ($p = .013$). The difference in the results is likely because participants with extremely low accuracy, contributing relatively few correct RTs in the calculation of their mean RT, were excluded from the final sample. This includes the two participants who reversed the appropriate stimulus-response mapping.

respond. After the response was selected, the next screen prompted participants to press the spacebar to continue. This prompt ensured participants returned their hand to the spacebar, ready to respond to go trials, and gave warning that the trials would begin again. Thought probes followed 12 (40%) of the 30 no-go trials. Undetectable to the participants, the task was programmed in mini-blocks to ensure a distribution of go and no-go trials and thought probes throughout the task. The thought probes were selected to appear randomly with the constraint that they follow two of every five no-go trials. No-go trials were randomly spaced with the constraint that there be three within every 26 trials. This constraint ensured that no-go trials and thought probes were spaced throughout the task without following a predictable pattern.

Procedure

After providing consent, participants completed a demographic questionnaire. Then, all participants completed (in order) operation span, symmetry span, and the SART. Within the SART, the order of the no-probe and probe conditions was counterbalanced across participants.

Analyses

To evaluate the effect of the presence of thought probes and the relationship to WMC, separate analyses of covariance (ANCOVAs) were conducted on SART performance evaluating: (a) accuracy on go and no-go trials, (b) d' , subtracting commission errors on no-go trials (false alarms) from correct responses on go trials (hits), (c) mean RT, and (d) ISD RT. Following the results of Stanislaw and Todorov (1999), we used the loglinear correction for hit and false alarm rates equal to 0 or 1, by adding 0.5 to all hit and false alarm totals and adding 1 to the number of trials. RT analyses were conducted on correct go trials with a RT > 150 ms.

A WMC composite score, calculated by averaging the z-scores from each memory span task, was the between-subjects covariate in each ANCOVA. Task order was the between-subjects factor, and probe presence was the within-subjects factor in each ANCOVA. Trial type (go vs. no-go) was an additional within-subjects factor in the accuracy ANCOVA.

Bayesian model comparisons were also computed, specifically comparing a model with probe presence as a factor against a null model for each outcome variable. These Bayes factors were computed in CRAN R software using the BayesFactor package (Morey & Rouder, 2015). This analysis gives a measure of support for one model over the other, a way to show amount of support for a null result that traditional significance testing does not allow. Finally, Hotelling-Williams tests were used to statistically test whether WMC correlations with SART performance were different between the no-probe and probe conditions.

Results

Descriptive statistics for all variables are shown in Table 1. Inspection of Table 1 shows that SART performance in the probe and no-probe conditions appears to be nearly identical. The scores for operation and symmetry span were consistent with normative data (Redick et al., 2012). In addition, the correlation between operation and symmetry span was significant, $r(135) = .25, p = .003$.

Effect of thought probe presence on SART performance

Full ANCOVA output results are provided in Table 2. For accuracy, there was no main effect of thought-probe condition, and no significant interactions involving thought-probe condition. There was no significant main effect of, nor any interactions with, the between-subjects factor of task order. There was a significant main effect of trial type, because go accuracy was higher than no-go accuracy. There was a significant WMC main effect and a significant WMC by trial type interaction, explored further in the correlational analyses below. Similarly, for d' there were no main effects of nor interactions with thought-probe condition or task order, but a significant main effect of WMC, indicating better performance for individuals higher in WMC. For mean RT, there was no main effect of probe presence, WMC, or task order. There were no significant interactions involving WMC. There was a significant interaction between probe presence and task order. Decomposing this interaction, we compared probe versus no-probe conditions when completed as the first block and as the second block separately. For the first completed task, those in the probe condition had a mean RT of 395 ms, and those who did the no-probe condition first had a mean RT of 401 ms, which were not different from each other, $t(135) = 0.42, p = .676$. For the second completed task, those who did the probe condition second had a mean RT of 412 ms, and those who had the no-probe condition second had a mean RT of 413 ms, which were not different from each other, $t(135) = 0.04, p = .972$. So, despite the significant probe presence by task order interaction for mean RT, the follow-up analyses indicate no difference in performance between the probe and no-probe conditions.

For ISD RT, there were no main effects of, nor interactions with, probe presence or task order. There was a significant main effect of WMC, such that low-WMC individuals' speed in responding to go stimuli was more variable across both probe conditions.

The null ANCOVA results for the probe-presence manipulation were supported by Bayesian analyses, showing evidence against a model with probe presence as a factor. That is, for all dependent variables, there was "moderate" evidence, values between 0.14 and 0.22, for the null versus a model with

Table 1 Descriptive statistics

	M	SD	Skew	Kurtosis	Correlation with WMC
No probe					
Go Accuracy	98.02	3.53	-2.85	8.71	.267*
No-go Accuracy	57.79	22.07	0.05	-0.81	.163 ⁺
<i>d'</i>	2.57	1.04	-0.01	-0.32	.300*
Mean RT	406.78	102.31	1.24	1.71	.023
ISD RT	119.35	47.15	0.97	1.07	-.287*
Probe					
Go Accuracy	98.39	2.14	-3.25	12.65	.306*
No-go Accuracy	59.93	21.30	0.05	-0.94	.254*
<i>d'</i>	2.70	1.00	-0.08	-0.57	.348*
Mean RT	403.74	94.26	1.15	1.00	.108
ISD RT	117.56	43.19	1.45	3.19	-.319*
On task/on goal	5.72	3.46	0.29	-1.01	.262*
On task/not goal	3.05	2.22	0.68	0.14	.010
Off task/else	1.82	1.93	0.98	0.20	-.252*
Blank	1.42	2.18	2.29	7.06	-.204*

Note. Go accuracy and no-go accuracy are percent correct; Mean RT and ISD RT are ms; *Indicates $p < .05$, ⁺ indicates $p = .056$

probe presence as a factor, using the criteria determined by Wagenmakers et al. (2018). These Bayes factors are presented in Table 3.

Individual differences in WMC and SART performance

The relationships between individual differences in WMC and SART performance in the probe and no-probe conditions were assessed by correlating the *z*-score WMC composite with the various SART dependent variables (Table 1). Note that the correlations between go accuracy and WMC should be interpreted with caution, because the high go accuracy led to skewness and/or kurtosis values classified as “extreme” by Kline (1998; skewness $> |3.0|$ and kurtosis $> |8.0|$). Inspection of Table 1 clearly indicates that individuals with lower WMC were less accurate on no-go trials and produced more variable correct RTs on go trials, but there was no relationship between WMC and mean RT on go trials – these results are consistent with previous studies. Critically, Hotelling-Williams *t*-tests confirmed that none of the correlations significantly differed as a function of the probe versus no-probe conditions (all *t*'s < 1.68 , all *p*'s $> .096$). Thus, the inclusion of the thought probes did not affect the relationship between individual differences in WMC and SART performance.

Finally, as an exploratory analysis into the content of the mind-wandering responses in the thought-probe present condition, the relationship between individual differences in WMC and mind-wandering was evaluated by correlating the *z*-score composite WMC variable with each of the thought probe response options (Table 1). Overall, WMC was related

inversely to mind-wandering proneness. The “On task, thinking about the goal” response was positively correlated with WMC, the intermediate “On task, not the goal” response was not correlated, and both off-task/mind-wandering responses (“Blank” and “Off task, anything else”) were negatively correlated with WMC.

Discussion

The current project addressed an important question for mind-wandering researchers, namely whether or not including thought probes within a cognitive task is a non-reactive method to quantify mind-wandering phenomenology. Our results showed no differences in SART performance, or WMC correlations with multiple dependent variables from the SART, as a function of the presence or absence of thought probes. Further, we found moderate evidence for the null for all outcome measures using Bayes factors. This is the first explicit evidence that inclusion of thought probes does not fundamentally alter the cognitive processes involved in performing the SART.

Regardless of probe condition, we observed a familiar pattern of WMC and SART relationships – compared to high-WMC individuals, low-WMC individuals made more errors and were more variable in their RTs on correct trials, although they were not slower with regard to mean RT (Redick et al., 2016; Stawarczyk, Majerus, Catale, & D'Argembeau, 2014; Unsworth & McMillan, 2014). Individuals lower in WMC were more likely to exhibit lapses of attention that characterize their performance in a variety of situations that have been

Table 2 ANCOVAs

		F	p	η^2
Accuracy	Main Effects			
	Probe presence	2.425	.122	.018
	Trial type	624.007	<.001	.823
	Order	0.013	.908	.000
	WMC	9.308	<.001	.065
	Interactions			
	Probe \times Trial Type	1.456	.230	.011
	Probe \times Order	0.728	.395	.005
	Probe \times WMC	1.346	.248	.010
	Probe \times Trial \times WMC	1.635	.203	.012
d'	Main Effects			
	Probe presence	3.155	.077	.023
	Order	0.001	.980	.000
	WMC	19.351	<.001	.126
	Interactions			
	Probe \times Order	0.087	.768	.001
	Probe \times WMC	0.278	.599	.002
Mean RT	Main Effects			
	Probe Presence	0.413	.521	.003
	Order	0.029	.864	.000
	WMC	0.567	.453	.004
	Interactions			
	Probe \times Order	8.037	.005	.057
	Probe \times WMC	1.851	.176	.014
ISD RT	Main Effects			
	Probe Presence	0.300	.585	.002
	Order	0.627	.430	.005
	WMC	17.327	<.001	.115
	Interactions			
	Probe \times Order	1.225	.270	.009
	Probe \times WMC	0.026	.872	.000

demonstrated across decades of research (Christopher & Redick, 2016).

Table 3 Bayes factors

Outcome measure	Bayes factor
Go accuracy	0.20
No-go accuracy	0.18
d'	0.22
Mean RT	0.14
ISD RT	0.14

Note. Bayes factors < 1 show support for the null. Factors between 0.10 and 0.30 are considered “moderate” evidence according to Wagenmakers et al. (2018)

Interestingly, WMC was significantly related to self-reported mind-wandering, such that individuals lower in WMC more often reported mind-wandering and less often reported on-task/on-goal thoughts. This result echoes previous WMC-SART findings (McVay & Kane, 2009; Unsworth & McMillan, 2014), and is also consistent with goal-maintenance and attention control theories (Engle & Kane, 2004), which posit that individuals higher in WMC are better able to maintain the goal of a task and control attention appropriately.

The current results may not be applicable to mind-wandering in other contexts, such as during reading of text passages (Feng, D'Mello, & Graesser, 2013). In addition, Kane et al. (2017) showed a distinction in the relationship between WMC and mind-wandering measured in the laboratory with thought probes versus outside of the lab with event-sampling methodology. Kane et al. (2017) reported that WMC negatively predicted mind-wandering likelihood during performance of tasks in the laboratory, but WMC predicted more mind-wandering in daily activities outside of the laboratory when individuals reported less of an effort to concentrate on the task at hand. Future work studying the role of thought probes in these additional contexts will likely provide a more complete story about the reactivity of thought probes, and whether they affect task performance and relationships with constructs of interest.

In conclusion, SART performance does not differ based on the presence or absence of thought probes. Individuals higher in WMC produced better SART performance with and without thought probes. Finally, individual differences in WMC were negatively correlated with mind-wandering frequency. The results indicate that thought probe measurement is a non-reactive method to evaluate mind-wandering in attention and inhibition tasks.

Acknowledgements While collecting the data and writing the manuscript, EAW and TSR were supported by the National Science Foundation (Award # 1632403) and the National Institutes of Health (Award # 2R01AA013650-11A1). We thank Kandice Campbell, Evelyn Cover, Alexis Hernandez, Ashlyn Hines-Lanham, and Lauren Lahey for their assistance with data collection. This project was pre-registered by the authors at <https://aspredicted.org> on 10 April 2017.

References

- Christopher, E. A., & Redick, T. S. (2016). Working memory. In V. Zeigler-Hill & T. Shackelford (Eds.), *Encyclopedia of Personality and Individual Differences*. Springer. doi: <https://doi.org/10.1007/978-3-319-28099-81039-1>
- Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *Psychology of Learning and Motivation*, 44, 145–199. doi:[https://doi.org/10.1016/S0079-7421\(03\)44005-X](https://doi.org/10.1016/S0079-7421(03)44005-X)

- Feng, S., D'Mello, S., & Graesser, A. C. (2013). Mind wandering while reading easy and difficult texts. *Psychonomic Bulletin & Review*, 20(3), 586–592. doi:<https://doi.org/10.3758/s13423-012-0367-y>
- Kane, M., Gross, G., Chun, C., Smeekens, B., Meier, M., Silvia, P., & Kwapil, T. (2017). For whom the mind wanders, and when, varies across laboratory and daily-life settings. *Psychological Science*, 28(9), 1271–1289. doi: <https://doi.org/10.1177/0956797617706086>
- Kline, R. B. (1998). *Principles and practice of structural equation modeling*. Guilford Press.
- 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. doi: <https://doi.org/10.1177/0956797611431465>
- 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. doi: <https://doi.org/10.1037/a0014104>
- Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes Factors for Common Designs. R package version 0.9.12-2. <https://CRAN.R-project.org/package=BayesFactor>
- Mrazek, M. D., Franklin, M. S., Phillips, D. T., Baird, B., & Schooler, J. W. (2013). Mindfulness training improves working memory capacity and GRE performance while reducing mind wandering. *Psychological Science*, 24(5), 776–781. doi: <https://doi.org/10.1177/0956797612459659>
- Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, 28(3), 164–171. doi: <https://doi.org/10.1027/1015-5759/a000123>
- Redick, T. S., Shipstead, Z., Meier, M. E., Montroy, J. J., Hicks, K. L., Unsworth, N., Kane, M. J., . . . Engle, R. W. (2016). Cognitive predictors of a common multitasking ability: Contributions from working memory, attention control, and fluid intelligence. *Journal of Experimental Psychology: General*, 145(11), 1473–1492. doi: <https://doi.org/10.1037/xge0000219>
- Robison, M. K., Miller, A. L., & Unsworth, N. (2017). *Examining the effects of probe frequency, response options, and framing within the thought-probe method*. Manuscript submitted for publication.
- Robison, M. K., & Unsworth, N. (2017). Working memory capacity and mind-wandering during low-demand cognitive tasks. *Consciousness and Cognition*, 52, 47–54. doi: <https://doi.org/10.1016/j.concog.2017.04.012>
- Rummel, J., & Boywitt, C. D. (2014). Controlling the stream of thought: Working memory capacity predicts adjustment of mind-wandering to situational demands. *Psychonomic Bulletin and Review*, 21, 1309–1315. doi: <https://doi.org/10.3758/s13423-013-0580-3>
- Seli, P. S., Carriere, J., Levene, M., & Smilek, D. (2013). How few and far between? Examining the effects of probe rate on self-reported mind wandering. *Frontiers in Psychology*, 4, 430. doi: <https://doi.org/10.3389/fpsyg.2013.00430>
- 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. doi: <https://doi.org/10.1177/0956797616634068>
- Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering: Empirically navigating the stream of consciousness. *Annual Review of Psychology*, 66, 487–518. doi: <https://doi.org/10.1146/annurev-psych-010814-015331>
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31(1), 137–149. doi: <https://doi.org/10.3758/BF03207704>
- Stawarczyk, D., Majerus, S., Catale, C., & D'Argembeau, A. (2014). Relationships between mind-wandering and attentional control abilities in young adults and adolescents. *Acta Psychologica*, 148, 25–36. doi:<https://doi.org/10.1016/j.actpsy.2014.01.007>
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, Instruments, & Computers*, 37(3), 498–505. doi: <https://doi.org/10.3758/BF03192720>
- Unsworth, N., & McMillan, B. D. (2014). Similarities and differences between mind-wandering and external distraction: A latent variable analysis of lapses of attention and their relation to cognitive abilities. *Acta Psychologica*, 150, 14–25. doi: <https://doi.org/10.1016/j.actpsy.2014.04.001>
- Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, 17(6), 635–654. doi: <https://doi.org/10.1080/09658210902998047>
- Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., . . . Morey, R. D. (2018). Bayesian inference for psychology: Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, 25(1), 58–76. doi: <https://doi.org/10.3758/s13423-017-1323-7>
- Weinstein, Y. (2018). Mind-wandering, how do I measure thee with probes? Let me count the ways. *Behavior Research Methods*, 50, 642–661. doi: <https://doi.org/10.3758/s13428-017-0891-9>