



Examining the effects of probe frequency, response options, and framing within the thought-probe method

Matthew K. Robison¹  · Ashley L. Miller² · Nash Unsworth²

© The Psychonomic Society, Inc. 2019

Abstract

A recent surge of interest in the empirical measurement of mind-wandering has led to an increase in the use of thought-probing to measure attentional states, which has led to large variation in methodologies across studies (Weinstein in *Behavior Research Methods*, 50, 642–661, 2018). Three sources of variation in methodology include the frequency of thought probes during a task, the number of response options provided for each probe, and the way in which various attentional states are framed during the task instructions. Method variation can potentially affect behavioral performance on the tasks in which thought probes are embedded, the experience of various attentional states within those tasks, and/or response biases to the thought probes. Therefore, such variation can be problematic, both pragmatically and theoretically. Across three experiments, we examined how manipulating probe frequency, response options, and framing affected behavioral performance and responses to thought probes. Probe frequency and framing did not affect behavioral performance or probe responses. But, in light of the present results, we argue that thought probes need at least three responses, corresponding to on-task, off-task, and task-related interference. When researchers are specifically investigating mind-wandering, the probe responses should also distinguish between mind-wandering, external distraction, and mind-blanking.

Keywords Mind-wandering · Thought probes · Sustained attention

Psychology has recently experienced a surge of interest in mind-wandering (Callard, Smallwood, Golchert, & Margulies, 2013; Smallwood & Schooler, 2006, 2015). Because mind-wandering is an inherently internal process with no overt associated behavior, researchers have developed a variety of techniques to measure it. These include diaries in which people record instances of mind-wandering during their day-to-day lives (e.g., Unsworth, Brewer, & Spillers, 2012), experience-sampling applications that ask people to report their thoughts periodically through a mobile device (Kane et al., 2007; Killingsworth & Gilbert, 2010), questionnaires (e.g., Broadbent, Cooper, FitzGerald, & Parkes, 1982; Cheyne, Carriere, & Smilek, 2006; Singer & Antrobus, 1970), and retrospective reports following a task (e.g., Antrobus, Singer, & Greenberg, 1966). In the past decade,

however, the method that has been used most commonly is the thought-probe technique (Weinstein, 2018).

Thought probes are screens (or auditory cues) that are inserted at fixed, random, or pseudorandom intervals within a task. These screens ask participants to report the current contents of their thoughts. For example, a thought probe might ask, “What are you thinking about right now?,” and the participant must categorize or verbally describe their thoughts (e.g., focused on the task, mind-wandering, or externally distracted). As compared to other methods, thought probes have a number of advantages. First, they allow researchers to examine moment-to-moment changes in task engagement and attentional state. Second, they do not require participants to be aware that their thoughts have drifted away from the task at hand. In other words, the probes can “catch” participants mind-wandering, even if participants are not aware they are doing so (Schooler, Reichle, & Halpern, 2004). Third, they do not require participants to retrospectively estimate what percentage of time they spent mind-wandering (or experiencing other off-task attentional states) during a task (Schoen, 1970). Thus, thought probes are not subject to memory biases. Fourth, the frequency of such reports can be used as a measure of individual differences and/or to estimate how various task

✉ Matthew K. Robison
matthewkrobison@asu.edu

¹ Department of Psychology, Arizona State University, Tempe, AZ, USA

² Department of Psychology, University of Oregon, Eugene, OR, USA

parameters affect task engagement (e.g., Forster & Lavie, 2009; McVay, Kane, & Kwapil, 2009). For these reasons, the thought-probe technique has become quite common.

As a corollary of such frequent use, a diversity of specific thought-probe methodologies have emerged. Recently, Weinstein (2018) completed a review of studies using thought probes between 2006 and mid-2015. She found 102 published articles comprising 132 experiments that used the technique. Importantly, Weinstein identified 69 variants of the thought-probe technique across these studies. Thus, thought probes are far from standardized. In the present study, we sought to examine whether variation in methodology might lead to substantially different conclusions about behavior, the experience of mind-wandering, and/or the reporting of mind-wandering. To do so, we decided to focus on three sources of variation in use of the thought-probe technique: probe frequency, number of response options, and the framing of mind-wandering.

Because thought probes have been embedded in so many different tasks, one source of variation within the method is the frequency with which probes appear during tasks. This variation might be problematic for several reasons. Primarily, the proportion of time that participants spend mind-wandering or experiencing other thoughts during a given task is a common dependent variable. With regard to replicability, it is important to be able to compare such proportions across studies. If probe frequency affects such proportions, comparability across studies would become problematic. Furthermore, if a study uses a particular manipulation to affect the proportion of mind-wandering, we need to be sure that the measurement is reliable and valid. Probe frequency can affect reliability and validity in at least a few ways. In one way, extremely rare thought probes can reduce variability across individuals and turn what is often assumed to be a continuous variable into an ordinal one. On the other hand, extremely frequent thought probes can be viewed by the participant as annoying or frustrating interruptions. This can potentially lead to acquiescence in which participants start habitually responding to the probes, rather than accurately characterizing their thoughts. Of course, probe frequency can also affect the actual experience of mind-wandering. On the one hand, if the probes are few and far between, participants might be more likely to fall into an off-task attentional state, which would lead to higher proportions of mind-wandering. On the other hand, if the probes are quite frequent, they might work to continually reengage participants, thus preventing mind-wandering from ever occurring. A third possibility is that frequent probing might actually cause participants to mind-wander more, since they are continually interrupted during the task. Experiment 1 addressed the issue of probe frequency by directly manipulating this aspect of the method between two groups of participants.

A second source of variation across studies is the number of response options provided. In some cases, thought probes

allow participants to verbally report their thoughts. Such responses are later coded and categorized by the researcher (e.g., Baird, Smallwood, & Schooler, 2011). But in most cases, participants are given several options from which to choose. In her review of the technique, Weinstein (2018) found that a common method was to simply provide a binary response option (e.g., are you mind-wandering? yes/no.) But other studies provide more than just two response options, and these other options can refer to other conscious states, such as task-related interference (i.e., thinking about one's task performance), external distraction, and mind-blanking (e.g., Stawarczyk, Majerus, Maj, Van der Linden, & D'Argembeau, 2011; Ward & Wegner, 2013). Another type of categorization allows participants to report different types of mind-wandering (e.g., intentional/unintentional, past-, present-, or future-focused; Forster & Lavie, 2009; Seli, Risko, Smilek, & Schacter, 2016; Smallwood, Nind, & O'Connor, 2009). Another common method allows participants to report their thoughts on a scale from *completely focused on the task* to *completely disengaged* (e.g., Christoff, Gordon, Smallwood, Smith, & Schooler, 2009). Variation in the number of response options can be important for several reasons. On one hand, with the binary response option, participants may be forced to sort their thoughts into one of two categories, even when that thought does not fall neatly into one category or the other. For example, if a participant is experiencing task-related interference, should he or she report being on-task or mind-wandering? What if a participant is distracted by a noise outside the laboratory? With a binary response option, we may be missing out on meaningful categorizations of thoughts. On the other hand, providing participants with a bevy of response options may overwhelm them. In that case, instead of accurately characterizing their thoughts, participants may fall into a response mode in which they continually choose the same response option, just to avoid the extra work of self-reflection. In Experiment 2, we manipulated the number of response options across conditions to examine this issue.

A third source of variation with the thought-probe technique is the way in which mind-wandering and the probes themselves are framed. For example, some probes ask, "What are you thinking about?" (e.g., Forster & Lavie, 2009). Others specifically ask, "Are you mind-wandering?" (e.g., Krawietz, Tamplin, & Radvansky, 2012). In one study, Weinstein, De Lima, and van der Zee (2018) specifically manipulated the phrasing of thought probes to see how it biased responding. In both conditions, participants read a text and were periodically probed by an auditory signal. In one condition, the thought probe stated, "At the time of the ding, my mind was on something other than the text." In this condition, the response options were (1) Yes, I was thinking about something else, and (2) No, I was thinking about the text. In the other condition, the thought probe stated, "At the time of the

ding, my mind was on the text.” In this condition, the response options read (1) Yes, I was thinking about the text, and (2) No, I was thinking about something else. Participants received ten such probes at pseudorandom intervals over a 20-min task. When the thought probe gave an on-task response as the default option, participants reported mind-wandering on an average of 2.25 probes. When the thought probe gave a mind-wandering response as the default option, participants reported mind-wandering on an average of 3.4 probes, resulting in a significant difference across the framing conditions (Weinstein et al., 2018). Thus, minor task discrepancies can lead to significant changes in response tendencies.

In addition to the way in which the actual probe is framed, researchers might characterize mind-wandering (either intentionally or unintentionally) in different ways during the introduction to the task. Often, researchers don’t specifically state what they tell the participants about mind-wandering. If participants come to the lab with a particular preconceived notion about mind-wandering (e.g., that it is an embarrassing occurrence that should be avoided), they might be hesitant to admit to it when asked. During instructions, researchers might also accidentally prime participants to think about mind-wandering in a particular way. In our own work, we tell participants that it is perfectly normal to mind-wander, daydream, or become distracted during tasks like the one they are about to complete. And when asked to characterize their thoughts, they should be as honest and accurate as possible. But even with this instruction, participants might be loath to report mind-wandering, due to a social desirability bias (Weinstein, 2018). If mind-wandering is framed differently—for example, as a “bad” thing that people should avoid—participants might report fewer instances of mind-wandering. This could be due to actual reductions in mind-wandering, because participants devote more attention to the task in order to resist mind-wandering, or it could simply reduce their willingness to report mind-wandering, rather than actually increasing task engagement. Thus, depending on how mind-wandering is framed, in addition to how the probe question is framed, participants might alter their behavior, task engagement, and/or response biases. In Experiment 3, we investigated this issue by manipulating how we framed mind-wandering across conditions.

Altogether, in the present study we sought to investigate whether variation in thought-probe techniques might lead to variation in behavior, mind-wandering, and response biases. In all three experiments, we used the semantic version of the Sustained Attention to Response Task (SART; McVay & Kane, 2009; Robertson, Manly, Andrade, Baddeley, & Yiend, 1997). We chose the SART for several reasons. First, the SART has been used rather extensively to measure mind-wandering and other attentional states. Therefore, the findings of the present study would be relevant to prior (and future) work. Second, the SART gives several behavioral indicators of attentional focus: omission errors (failure to respond to

frequent nontargets), commission errors (failing to withhold a response to rare targets), mean response time, and response time variability. So, in addition to examining how our various experimental manipulations affected thought-probe responses, we could use behavioral performance to corroborate these responses. For example, if changes in thought-probe responses across conditions were consistent with changes in behavioral performance, we could surmise that the experimental manipulations were indeed affecting the experience of mind-wandering (and other attentional states). If, however, the experimental manipulations affected the thought-probe responses without any corroborating changes in behavioral performance, the manipulations might only be affecting response biases.

Experiment 1

The goal of Experiment 1 was to determine whether probe frequency has any effect on behavioral performance or probe responses during a relatively common laboratory task. To examine this issue, we gave two groups of participants nearly identical versions of the SART. The only difference between conditions was that we doubled the number of thought probes in one condition. Probe frequency could affect behavioral performance and probe responses in a number of ways. First, increasing probe frequency might work to keep individuals more on-task, since the probes would more frequently catch individuals mind-wandering, and perhaps move them back into an on-task state. If this were the case, we should observe an increase in on-task reports, a decrease in mind-wandering (or other off-task reports), and improvements to behavioral performance, in the forms of fewer errors and decreased response time variability. Increasing probe frequency could also have the opposite effect: By disrupting the flow of the task, frequent probes might cause participants to be hyperaware of their own thoughts, leading to worse behavioral performance and more off-task thoughts. Furthermore, we might observe changes in the probe responses with no associated change in behavioral performance. This would indicate that participants were changing their response biases without any change in task engagement. Finally, probe frequency might have no effect on either behavior or response biases. If this were the case, we should see no substantial differences in behavioral or subjective indicators of task engagement across conditions.

Method

We report how we determined our sample size, all measures and manipulations, and all data exclusions (when necessary).

Participants and procedure

A total of 73 participants from the human subject pool at the University of Oregon completed the experiment in exchange for partial course credit. After completing informed consent and demographic forms, participants completed either the standard semantic SART or the double-probe semantic SART, based on random assignment. The session lasted 30 min. Due to the semantic nature of the task, we excluded ten participants who indicated that English was their nonnative language. These participants were replaced with new participants. Our final sample included 63 participants (29 in the standard condition and 34 in the double-probe condition).¹ Our target sample was 60 participants (~30 in each condition). We specified this target sample size based on what we figured a typical between-subjects design would comprise. However, this was not based on a power analysis, since we did not know what effect size to anticipate. With this sample size, we would have detected an effect size (d) of 0.72 with a power of .80. We stopped data collection on the day we reached this target sample.

Task

Semantic SART Participants completed a semantic version of the SART (Kane et al., 2016; McVay & Kane, 2012; McVay et al., 2009).² Participants were instructed to respond as quickly as possible to relatively frequent nontargets (animals) and to withhold their response to nonfrequent targets (vegetables). On each trial, a word appeared in lowercase 18-point Times New Roman black font on a white background for 300 ms. A 900-ms mask (12 capitalized Xs) immediately followed the presentation of each word. Participants made responses by pressing the spacebar. The task was separated into five blocks of 135 trials (675 total trials), which were separated further into 45-trial mini-blocks. Within each mini-block, 40 nontargets and five targets were presented. The separations of blocks and mini-blocks were invisible to participants. After the initial instructions, participants completed ten practice trials in which the categories were boys' and girls' names. Before beginning the 675 experimental trials, participants completed ten unanalyzed buffer trials.

In the standard condition, thought probes appeared after two targets and one nontarget in each mini-block, resulting in 45 total probes across the entire task. In the double-probe condition, thought probes appeared after four targets and two nontargets in each mini-block, resulting in 90 probes across the entire task. The difference in the frequency of thought

probes was the only difference between the conditions. We analyzed several dependent measures from the SART: mean response time (RT), intrasubject variability in RT (the coefficient of variation: RT CV), failures to withhold responses to targets (commission errors), and failures to respond to nontargets (omission errors).

Thought probes During the instructions, the participants were told, "We are also interested in finding out how often your mind wanders and you are distracted during tasks like this. In order to examine this, the computer will periodically ask you what you were just thinking about. It is normal to mind-wander, 'zone-out,' or be distracted now and then during a task like this. Please try your best to honestly assess your thoughts." Participants were then given an example thought-probe screen that said:

You will periodically see a screen like this:

What are you thinking about?

1. I am totally focused on the current task.
2. I am thinking about my performance on the task or how long it is taking.
3. I am distracted by sights/sounds or by physical sensations (hungry/thirsty).
4. I am thinking about things unrelated to the task/my mind is wandering.
5. My mind is blank.

Participants were given specific instructions regarding each thought-probe response. Specifically, they were told,

When you are asked to characterize your current conscious experience, please choose 1) I am totally focused on the current task, if you were thinking about the words you saw, pressing the space bar, or if you were totally focused on the task and nothing else. Please choose 2) I am thinking about my performance on the task or how long it is taking, if you were thinking about how well you are doing on the task, how many you are getting right, how long the task is taking, or any frustrations with the task. Please choose 3) I am distracted by sights/sounds or by physical sensations (hungry/thirsty), if you are distracted by noises inside or outside the room, sights in the room, or any other distracting stimuli in the current environment. Please choose 4) I am thinking about things unrelated to the task/my mind is wandering, if you were mind-wandering/day-dreaming and thinking about something other than the task or the current environment. Please choose 5) My mind is blank, if you were not focused on the task and you were not thinking about anything specific, but just zoned out.

¹ The unequal numbers of participants in the two conditions resulted from a disproportionate number of nonnative English speakers having been assigned to the standard condition during random assignment.

² We thank Mike Kane for sending us these task materials.

Table 1 Descriptive statistics by condition for Experiment 1

Measure	Condition		<i>t</i>	<i>p</i>	BF ₀₁
	Standard (<i>N</i> = 29)	Double-Probe (<i>N</i> = 34)			
Mean RT	504 (89)	475 (94)	1.24	.22	2.04
RT CV	.34 (.11)	.37 (.18)	0.85	.40	2.85
Commission errors	33.41 (13.80)	36.23 (14.19)	− 0.80	.43	3.28
Omission errors	30.10 (30.43)	41.24 (60.55)*	− 0.90	.37	3.35
On-task	.40 (.23)	.36 (.21)	0.63	.53	3.69
Task-related interference	.26 (.12)	.28 (.19)	− 0.59	.56	3.77
External distraction	.08 (.10)	.09 (.07)	− 0.35	.73	3.61
Mind-wandering	.19 (.16)	.18 (.12)	0.26	.80	2.97
Mind-blanking	.08 (.11)	.09 (.13)	− 0.41	.68	2.76

RT = response time, CV = coefficient of variation, BF₀₁ = Bayes factor in favor of null hypothesis. For each dependent variable, means are listed with standard deviations in parentheses. * The numerically larger mean and standard deviation in this condition is due to two participants with relatively large numbers of errors. Removing these participants did not change the pattern of results. The *p* values listed do not include a correction for multiple comparisons. However, because none of the tests were significant, the correction would not have affected our interpretation of the results

The final instruction screen told participants, “When you see a screen like this, please respond based on what you were thinking just before the screen appeared. Do not try to reconstruct what you were thinking during the preceding words on the screen, and please select the category that best describes your thoughts as accurately as you can. Remember that it is quite normal to have any of these kinds of thoughts during an ongoing task.” When the thought probes appeared, participants indicated their subjective state by pressing the key corresponding to the option that best characterized their thoughts. Response 1 was scored as *on-task*, Response 2 as *task-related interference*, Response 3 as *external distraction*, Response 4 as *mind-wandering*, and Response 5 as *mind-blanking*. Participants advanced through these instructions at their own pace. After the instructions, participants were told to ask the experimenter if they had any questions about the task. If they did not have any questions, they proceeded through the task.

Results and discussion

Means and standard deviations for each dependent variable of interest in each condition are listed in Table 1. We summed reports of on-task, task-related interference, external distraction, mind-wandering, and mind-blanking for each participant, then divided by 45 for the standard-probe condition and by 90 for the double-probe condition, and averaged across the participants in each condition. This gave estimates of the proportion of time participants in each condition reported being in each of the five attentional states. We analyzed differences between the conditions in the behavioral performance and thought-probe responses with *t* tests and Bayes factors. Because all of the *t* tests were nonsignificant, we report the

Bayes factor in favor of the null hypothesis (BF₀₁).³ This statistic can be interpreted as the ratio of evidence in favor of the null hypothesis (no difference) over the alternative hypothesis. Thus, a BF₀₁ of 2 would mean that the null hypothesis is twice as likely to be true, given the data. For these analyses we assumed equal prior likelihoods (i.e., the null and alternative hypotheses were assumed to be equally likely to occur).

We did not observe any evidence that the frequency of the probes during the task had any substantial effect on behavioral performance (measured by mean RT, RT CV, commission errors, and omission errors) or on subjective states measured by the thought probes. The behavioral performance was about equal, and the various subjective states were reported with about equal proportions in the two conditions. The *t* tests across conditions did not reveal any significant effects of the manipulation of probe frequency, and in all cases the null hypothesis (no difference) was at least twice as likely, given the data, as indicated by the Bayes factors. Therefore, in a task that is often employed to measure mind-wandering (i.e., SART), doubling the frequency with which thought probes appeared did not have any impact on either objective or subjective indicators of mind-wandering.

Experiment 2

We had several goals in our next experiment. The primary goal was to determine whether adding additional probe

³ In later sections, we will denote the Bayes factor of evidence in favor of the alternative hypothesis over the null as BF₁₀.

responses had any effect on the behavioral performance or thought-probe responses. A secondary goal was to examine how often participants attribute task-related interference to on- and off-task states. To do so, we administered the same SART as in Experiment 1 (with the standard probe frequency). We manipulated the number of possible probe responses between subjects. In one condition, the only two responses were *on-task* and *mind-wandering*. In a second condition, the response options included *on-task*, *task-related interference*, and *mind-wandering*. In a third condition, the response options included *on-task*, *task-related interference*, *mind-wandering*, *external distraction*, and *mind-blanking*.

Method

Participants and procedure

A total of 95 participants from the human subject pool at the University of Oregon completed the study in exchange for partial course credit. After completing informed consent and demographic forms, participants were randomly assigned to one of the three conditions. They then completed the SART, which took about 25 min. Five participants reported speaking English as a second language, so we excluded these participants and replaced them with new participants. We stopped data collection once we had reached 30 participants in each condition (after excluding the second-language English speakers). One participant from the two-option condition and one participant from the five-option condition were eliminated for not responding on the majority of trials, resulting in a final sample of 88 participants. With the final sample size, we could have detected an effect size (f) of 0.33 with a power of .80.

Task

Semantic SART See Experiment 1.

Thought probes The number of thought-probe responses was the only difference across conditions. The probe screens appeared with equal frequency in all three conditions. In one condition, the probe screen gave two options: (1) *I am totally focused on the current task*, and (2) *I am thinking about things unrelated to the task/my mind is wandering*. In a second condition, the probe screen gave three options: (1) *I am totally focused on the current task*, (2) *I am thinking about things unrelated to the task/my mind is wandering*, and (3) *I am thinking about my performance on the task or how long it is taking*. In a third condition, the probe screen gave five options: (1) *I am totally focused on the current task*, (2) *I am thinking about things unrelated to the task/my mind is wandering*, (3) *I am thinking about my performance on the task or how long it is taking*, (4) *I am distracted by sights/sounds or by physical*

sensations (hungry/thirsty), and (5) *My mind is blank*. Response 1 was scored as *on-task*, Response 2 as *mind-wandering*, Response 3 as *task-related interference*, Response 4 as *external distraction*, and Response 5 as *mind-blanking*. Thought probes appeared after two targets and one nontarget in each mini-block, resulting in 45 total probes across the entire task. The probe instructions across conditions were identical to those in Experiment 1, with the exception that participants did not receive specific instructions about probe responses that were not included in their respective condition.

Results and discussion

Descriptive statistics for each condition are listed in Table 2. In our first set of analyses we examined whether any of the behavioral performance measures differed across conditions. To do so, we submitted the mean RTs, RT CVs, commission errors, and omission errors to analyses of variance (ANOVAs) with condition as the between-subjects factor. None of the behavioral measures differed across conditions (mean RT: $F < 1$, $BF_{01} = 9.34$; RT CV: $F < 1$, $BF_{01} = 4.60$; omission errors: $F < 1$, $BF_{01} = 8.97$; commission errors: $F < 1$, $BF_{01} = 7.37$).⁴ Thus, the manipulation of response options did not have any effect on behavioral performance.

Our next set of analyses focused on the thought-probe responses. First, we compared on-task and mind-wandering reports between the two-option and three-option conditions. This comparison allowed us to determine how often participants attributed task-related interference to on-task thoughts and mind-wandering, respectively. The participants in the two-option condition reported being on task slightly yet significantly more often than the participants in the three-option condition [$t(57) = 4.55$, $p < .001$, $BF_{10} = 669.80$]. But the participants in the two-option condition also reported mind-wandering significantly more often than the participants in the three-option condition [$t(57) = 2.04$, $p = .047$, $BF_{10} = 1.47$]. Importantly, the participants in the three-option condition reported task-related interference 29% of the time, or just as frequently as mind-wandering. So, if we assume that the participants in these two conditions were performing the task about equally, as indicated by the equivalent behavioral measures, we can estimate that about 1/3 of task-related interference is attributed to mind-wandering, and about 2/3 is attributed to on-task thoughts. It is likely that some participants considered task-related interference as being on-task, whereas others considered it off-task. Without a within-subjects comparison, however, we cannot make firm conclusions about this possibility. Finally, comparing the three-option and five-

⁴ Again, we did not include a correction for multiple comparisons in these analyses, but because none of the tests were significant, our interpretation of the results is the same.

Table 2 Descriptive statistics by condition for Experiment 2

Measure	Condition		
	Two options ($N = 29$)	Three options ($N = 30$)	Five options ($N = 29$)
Mean RT	475 (91)	472 (100)	481 (105)
RT CV	.32 (.11)	.36 (.17)	.39 (.26)
Commission errors	34.90 (11.38)	37.46 (16.88)	38.27 (17.69)
Omission errors	34.55 (28.31)	30.90 (22.53)	31.93 (35.48)
On-task	.62 (.18)	.42 (.16)	.43 (.26)
Mind-wandering	.38 (.18)	.29 (.15)	.19 (.14)
Task-related interference	—	.29 (.13)	.21 (.16)
External distraction	—	—	.07 (.09)
Mind-blanking	—	—	.09 (.16)

RT = response time, CV = coefficient of variation. For each dependent variable, means are listed with standard deviations in parentheses

option conditions revealed roughly equal proportions of on-task thoughts [$t(57) = -0.15$, $p = .88$, $BF_{01} = 3.75$], but slightly different proportions of mind-wandering [$t(57) = 2.63$, $p = .01$, $BF_{10} = 4.37$] and task-related interference [$t(57) = 2.01$, $p = .049$, $BF_{10} = 1.41$]. Although less common than on-task reports, task-related interference, and mind-wandering, external distraction and mind-blanking collectively accounted for a significant proportion of the thought-probe responses in the five-option condition.

On the basis of these findings, we argue that it is important to distinguish mind-wandering from other types of off-task thoughts and from task-related interference. We are not the first to make this point, as Stawarczyk et al. (2011) made similar arguments. In our own work (Robison, Gath, & Unsworth, 2017; Robison & Unsworth, 2018; Unsworth & Robison, 2016), we typically include a mind-blanking option, since this attentional state is not task-related, but it is also not technically mind-wandering. Rather, it lacks any internal or external focus (Ward & Wegner, 2013). We also have evidence that, in some situations, mind-wandering and mind-blanking associate with different pretrial and task-evoked pupil diameters, which indicates that these two states might reflect different states of arousal and attentional orientation (Unsworth & Robison, 2018). Although the number of response options did not affect performance on the SART in any systematic way, the thought-probe responses indicate that it is important for researchers to distinguish mind-wandering from other forms of task-related and task-unrelated thought. If researchers are interested in the behavioral, neural, contextual, and dispositional correlates of mind-wandering per se, they need to account for other types of off-task thought. If researchers are interested in catching the occurrence of *any* type of off-task thought, they should be careful to instruct participants that thoughts like external distraction and mind-blanking should fall in the off-task thought category. Because task-related interference is neither task-unrelated

nor specifically task-focused, this option should also be included. This will prevent researchers from confounding these types of thoughts with on-task and various other off-task thoughts.

Experiment 3

The goal of Experiment 3 was to determine whether the framing of mind-wandering and of other off-task thought has any effect on either behavioral performance or thought-probe responding. To do so, we gave participants the same task as in Experiments 1 and 2. During the instructions, off-task thought was framed positively, negatively, or with neutral language (see the [Method](#) section below for the verbatim instructions). Where Weinstein et al. (2018) tested the framing of the thought-probe question, we sought to investigate a different framing effect. We attempted to manipulate attitudes toward experiencing and reporting mind-wandering. If participants come to the laboratory with the preconceived notion that mind-wandering is a bad (or good) thing, this might alter their willingness to experience (or report) mind-wandering and other off-task states. If framing off-task thought in a negative manner significantly affects the degree to which participants engage in off-task thoughts, we should observe a decrease in off-task reports and improvements in behavioral performance, relative to the neutral-framing condition. It is also possible that participants would be less willing to admit to off-task thoughts following a negative framing. If this were the case, we should observe a decrease in off-task reports but no change in behavioral performance. Similarly, if framing off-task thought in a positive manner encourages participants to engage in such thoughts, we should see an increase in off-task reports and a decrease in behavioral performance. If the positive framing simply makes participants more willing to report off-task thoughts, we should then see an increase in off-task

reports but no associated changes in behavioral performance. A final caveat to Experiment 3 is the inclusion of separate response options for unintentional and intentional mind-wandering. As previous research has demonstrated the importance of distinguishing between these two types of mind-wandering in various contexts (e.g., Robison & Unsworth, 2018; Seli, Cheyne, Xu, Purdon, & Smilek, 2015; Seli, Risko, Purdon, & Smilek, 2017; Seli, Risko, & Smilek, 2016), we decided to do so in Experiment 3. It is possible that framing mind-wandering in a positive or negative light would affect intentional mind-wandering but have no effect on other types of off-task thought. Specifically, positive framing might lead individuals to engage in (or be more willing to report) intentional mind-wandering, whereas negative framing might lead individuals to resist engaging in (or reporting) intentional mind-wandering.

Method

Participants and procedure

A total of 92 participants from the University of Oregon human subject pool completed the study in exchange for partial course credit. After completing informed consent and demographic forms, participants were randomly assigned to one of three conditions. They then completed the SART, which took about 25 min. We stopped collecting data when we had reached a sample size of 30 participants in each condition. Two participants' ages exceeded our traditional age range (18–35). Given the fact that age does appear to affect mind-wandering rates (Giambra, 1977–78; Jackson & Balota, 2012; Jackson, Weinstein, & Balota, 2013; Krawietz et al., 2012; McVay, Meier, Tournon, & Kane, 2013), we decided to replace these two participants with two new participants. Preliminary examination of the data also revealed two participants who did not make a response on the majority of trials. These participants were excluded from the final analyses, leaving a final sample of 88 participants. With the final sample size, we could have detected an effect (f) of 0.32 with a power of .80.

Task

Semantic SART See Experiment 1.

Thought probes The participants in all three conditions received the same thought probes with the same frequency. In each 45-trial mini-block, participants received two thought probes following target trials and three following nontarget trials, resulting in 75 total probes. In all conditions, the thought probe presented a screen that said the following:

What are you thinking about?

1. I am totally focused on the current task.

2. I am thinking about my performance on the task or how long it is taking.
3. I am distracted by sights/sounds or by physical sensations (hungry/thirsty).
4. I am intentionally thinking about things unrelated to the task.
5. I am unintentionally thinking about things unrelated to the task.
6. My mind is blank.

Participants were given specific instructions regarding each probe response. For Responses 1, 2, 3, and 6, these instructions were identical to those in Experiment 1. For Response 4 (intentional mind-wandering), the instructions said, "Please choose 4) I am intentionally thinking about things unrelated to the task, if you were deliberately mind-wandering/day-dreaming and thinking about something other than the task or the current environment." For Response 5 (unintentional mind-wandering), the instructions said, "Please choose 5) I am unintentionally thinking about things unrelated to the task, if you were mind-wandering/day dreaming, but were not intending to do so."

The only difference between the conditions was how off-task thought was framed. In the neutral condition, the instructions said, "It is perfectly normal for your thoughts to drift off-task every now and then. When you are asked to report your thoughts, please do so honestly." In the positive condition, the instructions said, "It is perfectly normal for your thoughts to drift off-task every now and then. Previous research has shown that off-task thoughts can actually help people plan, problem solve, and be more creative. When you are asked to report your thoughts, please do so honestly." In the negative condition, the instructions said, "It is perfectly normal for your thoughts to drift off-task every now and then. However, previous research as shown that off-task thoughts are bad for task performance. Do your best to resist all types of off-task thoughts so you can maximize your performance. But when you are asked to do so, please try to honestly assess your thoughts."

Results and discussion

Table 3 shows descriptive statistics for all dependent measures. All the measures appeared consistent across conditions, but we analyzed each dependent measure as a function of condition. Additionally, we provide the BF_{01} in addition to the relevant F statistic from the ANOVA. Neither mean RT ($F < 1$, $BF_{01} = 9.54$) nor RT CV ($F < 1$, $BF_{01} = 5.36$) differed across conditions. Commission errors ($F < 1$, $BF_{01} = 9.65$) and omission errors ($F < 1$, $BF_{01} = 4.69$) were also roughly equal across conditions. Finally, the proportions of time that participants reported being on-task ($F < 1$, $BF_{01} = 6.53$) or experiencing task-related interference ($F = 2.18$, $BF_{01} =$

Table 3 Descriptive statistics by condition for Experiment 3

Measure	Framing condition		
	Neutral ($N = 29$)	Positive ($N = 30$)	Negative ($N = 29$)
Mean RT	499 (102)	501 (113)	507 (115)
RT CV	.31 (.09)	.30 (.13)	.34 (.13)
Commission errors	19.65 (10.82)	19.07 (9.69)	19.17 (9.09)
Omission errors	34.55 (31.74)	25.27 (33.18)	25.62 (21.00)
On-task	.40 (.26)	.46 (.26)	.45 (.24)
Task-related interference	.39 (.21)	.29 (.19)	.29 (.11)
External distraction	.05 (.05)	.07 (.11)	.07 (.12)
Intentional MW	.04 (.09)	.02 (.04)	.04 (.08)
Unintentional MW	.10 (.14)	.09 (.07)	.08 (.11)
Mind-blanking	.04 (.06)	.08 (.13)	.07 (.11)

RT = response time, CV = coefficient of variation, MW = mind-wandering. For each dependent variable, means are listed with standard deviations in parentheses

1.79), external distraction ($F < 1$, $BF_{01} = 6.61$), intentional mind-wandering ($F < 1$, $BF_{01} = 6.37$), unintentional mind-wandering ($F < 1$, $BF_{01} = 6.97$), and mind-blanking ($F < 2$, $BF_{01} = 4.15$), did not differ as a function of condition.⁵

Collectively, we did not find any evidence that framing mind-wandering in a specific manner had any appreciable effect on the task performance or probe responses. Furthermore, most comparisons across conditions yielded Bayes factors in favor of the null hypothesis that there were no differences. Thus, at least with the present task, telling participants that mind-wandering is a normal thought (neutral frame), a beneficial thought (positive frame), or a harmful thought (negative frame) did not significantly alter their task focus or the degree to which they were willing to admit to falling into various attentional states throughout the task.

General discussion

Given the recent expansion of interest in mind-wandering in the past decade, it is perhaps unsurprising that there is so much variability in the precise methods people have used to measure mind-wandering. Even within the thought-probe technique, there has been a large amount of variation across studies in the nature of the thought probes (Weinstein, 2018). In the present study, we sought to examine whether this variability amounts to any appreciable differences in typical dependent variables.

In Experiment 1 we examined variation in probe frequency. We found no evidence that probe frequency affected either behavioral performance or the proportion of time participants

reported being in various attentional states. So, at least with the current task, probe frequency did not significantly affect any dependent measures. This finding is at odds with at least one other study. Seli, Carriere, Levene, and Smilek (2013) found that including fewer probes led participants to report more instances of mind-wandering. However, they did not find a relationship between probe frequency and behavioral performance (RT variability). Therefore, they offered the possibility that probe frequency may not be affecting the actual occurrence of mind-wandering, but rather the willingness to report it. In the present study, we saw neither behavioral nor thought-probe response differences, so we are reasonably confident that probe frequency did not affect the experience of mind-wandering. Right now, it is unclear why Seli et al. (2013) found more mind-wandering reports with fewer probes and we did not. Although we did not observe any significant effects, probe frequency is still an issue researchers should consider when designing a task. Our recommendation is to use a probe frequency that closely aligns with prior work that has used the same or a similar task. If the task is entirely novel, researchers should base their probe frequency on the temporal duration of the task so that it matches reasonably with the average probe-to-probe interval in comparable tasks.

The second issue we addressed was the number of responses available on thought probes. Our recommendation is that if researchers are interested in *any* off-task thought, they should provide response options for on-task thoughts, task-related interference, and all other thoughts. If, however, researchers are specifically interested in *mind-wandering*, they need to be careful to allow participants to report other off-task states (e.g., external distraction, mind-blanking). In this way, the experience of other off-task thoughts is not confounded with occurrences of mind-wandering per se.

⁵ Although we did not explicitly correct for multiple comparisons, none of the tests were significant, and thus our interpretation was unaffected.

In Experiment 3, we examined the effect on behavioral performance and thought-probe responses of how mind-wandering is framed. We did not find any evidence that the framing differences across conditions had any effect on behavioral performance or thought-probe responses. Thus, it did not appear that framing off-task thoughts as neutral, positive, or negative affected the experience of or the willingness to report various off-task thoughts. One possible limitation is that our participants moved through the instructions at their own pace, and we did not collect data on reading times for each instruction screen. Therefore, it is possible that participants read through the framing instructions without fully incorporating this information. One aspect worthy of note in Experiment 3 is that the overall mind-wandering rates (combining intentional and unintentional mind-wandering) were lower than those in Experiments 1 and 2 [$t(178) = 2.87, p = .001$]. Therefore, something about alerting people to the intentionality of their thoughts might reduce mind-wandering or participants' willingness to report it. This finding will require future research.

Overall, the present study suggests that the typical findings from thought-probing techniques may be rather robust to mild variability in task parameters. If anything, the findings are quite comforting, in that the results from studies using varying methods can be compared with a reasonable amount of confidence. If we had observed drastic changes in behavioral performance or thought-probe responses with the relatively minor adjustments we made to the thought-probe parameters, this might have put in doubt the ability to compare findings across methods. But given the consistency of behavioral performance and probe responses within and across experiments, perhaps many of these methodological alterations produce only minor or negligible effects on probe response tendencies. However, we do not endorse careless manipulations of method. In any study, researchers should be sure to tether their method to prior work. When possible, they should use the exact task parameters used in prior studies, in order to maximize comparability and replicability. When the experimental question requires alteration of the method, such alterations should be justified given the research question. This, of course, is simply part of good science, but it deserves repeating here.

We should note that in the present study we used just one task, which lasted about 25 min. We chose the SART because it is used relatively often, and we used the same task in every experiment in order to maintain consistency. However, in many studies, especially individual-difference investigations, thought probes are included in several tasks, and sessions last much longer than 25 min. Therefore, these results might not generalize to individual-difference investigations that examine thought-probe responses across many tasks. Future research may be necessary in order to examine how manipulations to thought probes such as those in the present study might affect

the patterns of interindividual relationships. In these situations, the probe options should remain consistent across tasks.

If we have one general recommendation, it is to be as consistent as possible with one's method. Furthermore, the specifics of the probing technique, including the number of response options (as well as verbatim wording), the spacing and frequency of probes, task duration, probe framing, and the specific task instructions given to participants, should be outlined when researchers report their method. Some labs have already started putting stimulus materials for mind-wandering studies, which include the actual tasks used, in open-access repositories such as the Open Science Framework (e.g., Kane et al., 2017). Practices like these will make such work more replicable and generalizable in future studies.

Author note This research was supported by Office of Naval Research Grant No. N00014-15-1-2790. All data are publicly available at the Open Science Framework (<https://osf.io/uww2d>) or upon request from the first author.

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

References

- Antrobus, J. S., Singer, J. L., & Greenberg, S. (1966). Studies in the stream of consciousness: Experimental enhancement and suppression of spontaneous cognitive processes. *Perceptual and Motor Skills*, 23, 399–417. <https://doi.org/10.2466/pms.1966.23.2.399>
- Baird, B., Smallwood, J., & Schooler, J. W. (2011). Back to the future: Autobiographical planning and the functionality of mind-wandering. *Consciousness and Cognition*, 20, 1604–1611.
- Broadbent, D. E., Cooper, P. F., FitzGerald, P., & Parkes, K. R. (1982). The Cognitive Failures Questionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology*, 21, 1–16.
- Callard, F., Smallwood, J., Golchert, J., & Margulies, D. S. (2013). The era of the wandering mind? Twenty-first century research on self-generated mental activity. *Frontiers in Psychology*, 4, 891. <https://doi.org/10.3389/fpsyg.2013.00891>
- Cheyne, J. A., Carriere, J. S. A., & Smilek, D. (2006). Absent-mindedness: Lapses of conscious awareness and everyday cognitive failures. *Consciousness and Cognition*, 3, 578–592.
- Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., & Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proceedings of the National Academy of Sciences*, 106, 8719–8724. <https://doi.org/10.1073/pnas.0900234106>
- Forster, S., & Lavie, N. (2009). Harnessing the wandering mind: The role of perceptual load. *Cognition*, 111, 345–355. <https://doi.org/10.1016/j.cognition.2009.02.006>
- Giambra, L. M. (1977–78). Adult male daydreaming across the lifespan: A replication, further analyses, and tentative norms based upon retrospective reports. *International Journal of Aging and Human Development*, 8, 197–228.
- Jackson, J. D., & Balota, D. A. (2012). Mind-wandering in younger and older adults: Converging evidence from the Sustained Attention to Response Task and reading for comprehension. *Psychology and Aging*, 27, 106–119.

- Jackson, J. D., Weinstein, Y., & Balota, D. A. (2013). Can mind-wandering be timeless? Atemporal focus and aging in mind-wandering paradigms. *Frontiers in Psychology, 4*, 742. <https://doi.org/10.3389/fpsyg.2013.00742>
- Kane, M. J., Brown, L. E., Little, J. C., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007). For whom the mind wanders, and when: An experience-sampling study of working memory and executive control in daily life. *Psychological Science, 18*, 614–621.
- Kane, M. J., Meier, M. E., Smeekens, B. A., Gross, G. M., Chun, C. A., Silvia, P. J., & Kwapil, T. R. (2016). Individual differences in executive control of attention, memory, and thought, and their associations with schizotypy. *Journal of Experimental Psychology: General, 145*, 1017–1048.
- Killingsworth, M. A., & Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science, 330*, 932. <https://doi.org/10.1126/science.1192439>
- Krawietz, S. A., Tamplin, A. K., & Radvansky, G. A. (2012). Aging and mind wandering during text comprehension. *Psychology and Aging, 27*, 951–958.
- McVay, J. C., & Kane, M. J. (2009). Conducting the train of thought: Working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 196–204. <https://doi.org/10.1037/a0014104>
- McVay, J. C., & Kane, M. J. (2012). Drifting from slow to “D’oh!”: Working memory capacity and mind wandering predict extreme reaction times and executive control errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 525–549. <https://doi.org/10.1037/a0025896>
- 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*, 857–863. <https://doi.org/10.3758/PBR.16.5.857>
- McVay, J. C., Meier, M. E., Touron, D. R., & Kane, M. J. (2013). Aging ebbs the flow of thought: Adult age differences in mind wandering, executive control, and self-evaluation. *Acta Psychologica, 142*, 136–147.
- Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). “Oops!”: Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia, 35*, 747–758. [https://doi.org/10.1016/S0028-3932\(97\)00015-8](https://doi.org/10.1016/S0028-3932(97)00015-8)
- Robison, M. K., Gath, K. I., & Unsworth, N. (2017). The neurotic wandering mind: An individual differences investigation of mind-wandering, neuroticism, and executive control. *Quarterly Journal of Experimental Psychology, 70*, 649–663.
- Robison, M. K., & Unsworth, N. (2018). Cognitive and contextual predictors of spontaneous and deliberate mind-wandering. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*, 85–98.
- Schoen, J. R. (1970). Use of consciousness sampling to study teaching methods. *Journal of Educational Research, 63*, 287–390.
- Schooler, J. W., Reichle, E. D., & Halpern, D. V. (2004). Zoning-out while reading: Evidence for dissociations between experience and metaconsciousness. In D. T. Levin (Ed.), *Thinking and seeing: Visual metacognition in adults and children* (pp. 203–226). Cambridge, MA: MIT Press.
- Seli, P., Carriere, J. S. A., Levene, M., & Smilek, D. (2013). How few and far between? Examining the effects of probe rate on self-reported mind wandering. *Frontiers in Psychology, 4*, 430. <https://doi.org/10.3389/fpsyg.2013.00430>
- Seli, P., Cheyne, J. A., Xu, M., Purdon, C., & Smilek, D. (2015). Motivation, intentionality, and mind wandering: Implications for assessments of task-unrelated thought. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 1417–1425. <https://doi.org/10.1037/xlm0000116>
- Seli, P., Risko, E. F., Purdon, C., & Smilek, D. (2017). Intrusive thoughts: Linking spontaneous mind wandering and OCD symptomatology. *Psychological Research, 81*, 392–398. <https://doi.org/10.1007/s00426-016-0756-3>
- Seli, P., Risko, E. F., & Smilek, D. (2016). On the necessity of distinguishing between unintentional and intentional mind wandering. *Psychological Science, 27*, 685–691. <https://doi.org/10.1177/0956797616634068>
- Seli, P., Risko, E. F., Smilek, D., & Schacter, D. L. (2016). Mind-wandering with and without intention. *Trends in Cognitive Sciences, 20*, 605–617.
- Singer, J. L., & Antrobus, J. S. (1970). *Imaginal processes inventory*. Princeton, NJ: Educational Testing Service.
- Smallwood, J., Nind, L., & O’Connor, R. C. (2009). When is your head at? An exploration of the factors associated with the temporal focus of the wandering mind. *Consciousness and Cognition, 18*, 118–125. <https://doi.org/10.1016/j.concog.2008.11.004>
- Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin, 132*, 946–958. <https://doi.org/10.1037/0033-2909.132.6.946>
- Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering: Empirically navigating the stream of consciousness. *Annual Review of Psychology, 66*, 487–518. <https://doi.org/10.1146/annurev-psych-010814-015331>
- Stawarczyk, D., Majerus, S., Maj, M., Van der Linden, M., & D’Argembeau, A. (2011). Mind-wandering: Phenomenology and function as assessed with a novel experience sampling method. *Acta Psychologica, 136*, 370–381.
- Unsworth, N., Brewer, G. A., & Spillers, G. J. (2012). Variation in cognitive failures: An individual differences investigation of everyday attention and memory failures. *Journal of Memory and Language, 67*, 1–16.
- Unsworth, N., & Robison, M. K. (2016). Pupillary correlates of lapses of sustained attention. *Cognitive, Affective, & Behavioral Neuroscience, 16*, 601–615. <https://doi.org/10.3758/s13415-016-0417-4>
- Unsworth, N., & Robison, M. K. (2018). Tracking arousal state and mind wandering with pupillometry. *Cognitive, Affective, & Behavioral Neuroscience, 18*, 638–664. <https://doi.org/10.3758/s13415-018-0594-4>
- Ward, A. F., & Wegner, D. M. (2013). Mind-blanking: When the mind goes away. *Frontiers in Psychology, 4*, 650. <https://doi.org/10.3389/fpsyg.2013.00650>
- Weinstein, Y. (2018). Mind-wandering, how do I measure thee with probes? Let me count the ways. *Behavior Research Methods, 50*, 642–661. <https://doi.org/10.3758/s13428-017-0891-9>
- Weinstein, Y., De Lima, H. J., & van der Zee, T. (2018). Are you mind-wandering, or is your mind on task? The effect of probe framing on mind-wandering reports. *Psychonomic Bulletin & Review, 25*, 754–760. <https://doi.org/10.3758/s13423-017-1322-8>