CrossMark

Validating a visual version of the metronome response task

Patrick Laflamme[1] · Paul Seli[2] · Daniel Smilek[3]

Abstract

The metronome response task (MRT)—a sustained-attention task that requires participants to produce a response in synchrony with an audible metronome—was recently developed to index response variability in the context of studies on mind wandering. In the present studies, we report on the development and validation of a visual version of the MRT (the visual metronome response task; vMRT), which uses the rhythmic presentation of visual, rather than auditory, stimuli. Participants completed the vMRT (Studies 1 and 2) and the original (auditory-based) MRT (Study 2) while also responding to intermittent thought probes asking them to report the depth of their mind wandering. The results showed that (1) individual differences in response variability during the vMRT are highly reliable; (2) prior to thought probes, response variability increases with increasing depth of mind wandering; (3) response variability is highly consistent between the vMRT and the original MRT; and (4) both response variability and depth of mind wandering increase with increasing time on task. Our results indicate that the original MRT findings are consistent across the visual and auditory modalities, and that the response variability measured in both tasks indexes a non-modality-specific tendency toward behavioral variability. The vMRT will be useful in the place of the MRT in experimental contexts in which researchers' designs require a visual-based primary task.

Keywords Mind wandering · Sustained Attention · MRT · vMRT · Metronome Response Task

Recently, there has been rapidly growing interest in the mental state of "mind wandering"—a phenomenon often defined as task-unrelated thought—with a particular focus on how mind wandering influences ongoing primary-task performance (see Smallwood & Schooler, 2006, 2015, for reviews). The available evidence indicates that mind wandering interferes with performance on numerous tasks, ranging from basic continuous-response tasks in which participants respond to the presentation of frequently presented "go stimuli" and withhold responses to infrequently presented "no-go stimuli" (e.g., Christoff, Gordon, Smallwood, Smith, & Schooler, 2009; McVay & Kane, 2009; Seli, 2016), to more complex tasks such as those assessing

reading comprehension (e.g., Feng, D'Mello, & Graesser, 2013; Unsworth & McMillan, 2013). Notably, mind-wandering rates during some tasks have been shown to increase as time on task progresses, with commensurately increasing costs to performance (Thomson, Seli, Besner, & Smilek, 2014).

As research on mind wandering has progressed, studies have begun to reveal that, in some situations (particularly those requiring constrained responding), mind wandering is associated with behavioral variability (e.g., Seli, Carriere, et al., 2014; Seli, Cheyne, & Smilek, 2013). Specifically, relative to moments of on-task attentiveness, moments of mind wandering have been associated with increased levels of fidgeting (Seli, Carriere, et al., 2014) and increased response time variability (e.g., Seli, Cheyne, & Smilek, 2013). The relation between mind wandering and behavioral variability has also been observed at the level of individual differences: People who report more fidgeting in their daily lives also tend to report higher rates of everyday spontaneous mind wandering (Carriere, Seli, & Smilek, 2013).

The foregoing findings are intriguing at a theoretical level because they could be interpreted as suggesting that mind wandering is associated with a general failure of control, not only over one's stream of consciousness, but also over one's body. Relatedly, these findings could be taken to suggest that

✉ Patrick Laflamme
  patrickj.laflamme@gmail.com

[1] Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada

[2] Department of Psychology, Harvard University, Cambridge, MA, USA

[3] Department of Psychology, University of Waterloo, Waterloo, Ontario, Canada

⚫ Springer

mind wandering is related to an underlying tendency toward experiencing variability, both in terms of one's thoughts and one's actions/behaviors. Indeed, it could be argued that variability in one's thoughts and behaviors is the default state of human beings, and that limiting thought and behavior to a focal task or goal is the more unusual and remarkable ability (Seli, Carriere, et al., 2014). Along these lines, mind wandering and response variability during a constrained task (such as one requiring button presses to a target stimulus) might reflect a retreat to one's natural state of variability.

One task commonly used to assess mind wandering, and one that has been used to examine the link between mind wandering and response variability (McVay & Kane, 2009), is the Sustained Attention to Response Task (SART; Robertson, Manly, Andrade, Baddeley, & Yiend, 1997). In this task, single digits are presented on a screen, one at a time, and participants are instructed to respond (via button press) to each of the digits except for one infrequently occurring target digit, which, historically, is often the digit 3. Studies employing the SART have shown that, relative to periods of on-task performance, periods of mind wandering are associated with increased failures to withhold a response to the infrequent target digit (i.e., errors of commission). More relevant for present purposes, individual differences in everyday inattention (as measured by subjective-report measures) have also been shown to relate to individual differences in response time variability on non-target trials in the SART (e.g., Cheyne, Solman, Carriere, & Smilek, 2009; McVay & Kane, 2009; Seli, Cheyne, Barton, & Smilek, 2012).

Although the SART's measure of response variability has been touted as an index of inattention and mind wandering (Cheyne et al., 2009; McVay & Kane, 2009; Seli, Cheyne, Barton, & Smilek, 2012), unfortunately, this task has some noteworthy limitations that render this measure problematic. One general limitation concerns the instructions used in the SART. Typically, participants are instructed to respond, "as quickly and accurately as possible" (Robertson et al., 1997). As has been highlighted in numerous studies (e.g., Seli, Cheyne, & Smilek, 2012; Seli, Jonker, Cheyne, & Smilek, 2013; Seli, Jonker, Solman, Cheyne, & Smilek, 2013), the weakness of such instructions is that they could be interpreted as emphasizing the importance of either the speed of responding or the accuracy of one's responses (Seli, Cheyne, & Smilek, 2012). Given that the SART inherently includes a speed–accuracy trade-off (Helton, Kern, & Walker, 2009; Seli, Cheyne, Barton, & Smilek, 2012; Seli, Jonker, Solman, et al., 2013), it is unclear whether moments of inattention manifest in reaction times and/or in accuracy scores. To deal with this issue, researchers have applied various statistical methods allowing them to better individuate or model the reaction time data in the context of the SART (e.g., Seli, 2016; Seli, Jonker, Cheyne, & Smilek, 2013).

Perhaps the most substantive problem with the use of the SART as an index of response variability is that the presence of rare target trials may induce perturbations in response variability that are conflated with perturbations in response variability that owe specifically to mind wandering/inattention (that latter of which are often of key interest to researchers; Cheyne et al., 2009; Smallwood et al., 2004). In particular, response time variability might be affected by posttarget processing, such as posterror slowing, which has been well-documented in the SART (e.g., Jonker, Seli, Cheyne, & Smilek, 2013). In addition, target expectation effects have been observed in the SART, as evidenced by participants' tendency to slow down their responses as the time since the appearance of the last target increases and the presentation of the next target becomes more imminent (Cheyne, Carriere, Solman, & Smilek, 2011). Critically, these target-related determinants of response variability might be quite different from the response variability caused by inattention, and it is therefore unclear how statistical methods might tease apart these different types of response variability.

Here, we focus on one particular task that was developed to overcome the aforementioned shortcomings of the SART: the metronome response task (MRT; Seli, Cheyne, & Smilek, 2013). In the standard version of the MRT, participants are required to respond to an auditory tone that is presented once every 1,300 ms. More specifically, participants are instructed to respond (via button press) synchronously with the onset of each tone so that they produce a button press at the exact moment at which each tone is presented (Seli, Cheyne, & Smilek, 2013). The primary measure of interest yielded by the MRT is the variability in participants' responses to the metronome tone. Also of interest in the MRT are participants' reports of mind wandering, which have typically been obtained via the intermittent presentation of "thought probes" that require participants to report on the content of their thoughts (e.g., "on task" or "mind wandering") just prior to the onset of each probe.

Numerous studies have now reliably shown that, as participants' minds wander away from the MRT, the variability in their responses to the metronome tones tends to increase (e.g., Seli, Carriere, et al., 2013; Seli, Carriere, et al., 2014; Seli, Cheyne, & Smilek, 2013; Seli, Cheyne, Xu, Purdon, & Smilek, 2015; Seli, Jonker, Cheyne, Cortes, & Smilek, 2015). The relation between mind wandering and response variability has been demonstrated in two complementary ways. First, the variability in responses on trials immediately preceding thought probes has been found increase as a function of participants' self-reported depth of mind wandering (e.g., Seli, Carriere, et al., 2014). Second, at the level of individual differences, correlational analyses have shown that people who report higher rates of mind wandering also tend to exhibit greater response variability throughout the MRT (e.g., Seli, Carriere, et al., 2013; Seli, Cheyne, et al., 2015).

The general conclusion regarding response variability and its co-occurrence with reported mind wandering in the MRT is that it does not simply reflect a modality-specific tendency to synchronize responding with an auditory tone, but that it instead reflects a more general tendency toward behavioral variability that generalizes across modalities. However, to date, the MRT has not been validated across modalities. This lack of cross-modality validation of the MRT is problematic because responding in synchrony to an auditory tone might be a rather unique behavior. Indeed, given the similarity between MRT responses and the common behavior of tapping along to the beat of a song, it is possible that participants can produce a stable pattern of responses to the MRT tones independently of their level of attention to the task. If synchronizing with an auditory tone can be done automatically, then the MRT might not accurately estimate the relation between mind wandering and performance variability. However, tapping in synchrony to a visual metronome is much less common, and as such, a task requiring such tapping should help to mitigate this problem. For this reason, a visual version of the MRT may help to increase the task's sensitivity to states of mind wandering.

In the present article, we report on the development and validation of a visual version of the MRT. Developing and validating a visual version of the MRT (a vMRT) is important for two primary reasons: First, it allows for the assessment of the generality of the original MRT findings across auditory and visual modalities. Second, the ability to administer a visual task that does not suffer from the shortcomings of the SART allows for greater experimental flexibility. For instance, accessibility to a visual version of the MRT would allow researchers to examine MRT performance in the context of auditory distractions, which are common in everyday life, and hence, a visual version of the MRT could permit more ecologically valid research on mind wandering.

To adapt the MRT to a visual form, we replaced each metronome tone with a gray square, which was presented on a computer monitor. To confirm that mind wandering was linked with increased response variability in the vMRT—as in the standard version of the MRT (see Seli, Cheyne, et al., 2013)—throughout the vMRT, we intermittently presented thought probes that required participants to report on the depth of their mind wandering (as in Seli, Carriere, et al., 2014).

Study 1

Method

Participants Forty-two participants (mean age = 20.4, 28 female) were recruited from the undergraduate Research Experiences Group (REG) at the University of Waterloo. As per the conditions of recruitment, all participants reported that they had normal or corrected-to-normal vision, had normal or

corrected-to-normal hearing, and could read and write fluently in English. In this validation study, up to five participants were tested in the same room at a time. As in Seli, Cheyne, et al. (2013), we first identified participants whose rates of omissions (i.e., failures to respond on a given trial) were greater than 10%, which indicates a failure to comply with the task instructions. These participants' data were then removed from all subsequent analyses. In total, seven participants' data were excluded for this reason, leaving data from 35 participants for the subsequent analyses.

vMRT stimuli and procedures The vMRT was designed to closely match the presentation parameters of the standard MRT (see Seli, Cheyne, et al., 2013). The vMRT consisted of 900 trials, and each trial began with the presentation of a blank screen for 650 ms, followed by a gray square for 150 ms, followed by another blank screen for 500 ms. From the participants' perspective, the onsets of the gray squares were separated by 1,300 ms (see Fig. 1). Critically, the visual stimuli were presented for a longer period of time than are the auditory stimuli in the standard MRT. This was done in order to ensure that the stimulus would synchronize with the frame rate of the monitor being used (60 Hz), with the stimulus being presented for exactly nine frames. As a result, the tempo of the metronome remained the same as in the original MRT, but the interstimulus interval was reduced by 75 ms. The gray box measured 1.5 cm × 1.5 cm and was located in the center of the screen. The square was set to RGB values of 126, 126, 126 (i.e., gray), and the background RGB values were set to 0, 0, 0 (i.e., black). Participants were instructed to "press the spacebar in synchrony with the flashing box so that you press the spacebar exactly when each box is presented."

Thought probes To assess participants' depth of mind wandering, 18 thought probes were pseudorandomly presented throughout the vMRT. One thought probe was presented in each block of 50 trials, with the constraint that no two thought probes were presented within ten trials of each other (Seli, Carriere, et al., 2013). When a thought probe was presented, the vMRT temporarily stopped and participants were instructed to select the degree to which they were focused on the vMRT or were thinking about task-unrelated concerns. The response options were as follows: (1) "completely on task," (2) "mostly on task," (3) "equally on task and thinking about unrelated concerns," (4) "mostly thinking about unrelated concerns," and (5) "completely thinking about unrelated concerns" (e.g., Mrazek, Franklin, Phillips, Baird, & Schooler, 2013; Seli, Carriere, et al., 2014). After a response had been provided to each probe, the vMRT resumed. Participants were instructed that being "on task" meant they were thinking about things related to the task (e.g., their performance on the task, the gray box, or their response), whereas thinking about unrelated concerns meant that they were thinking about things that
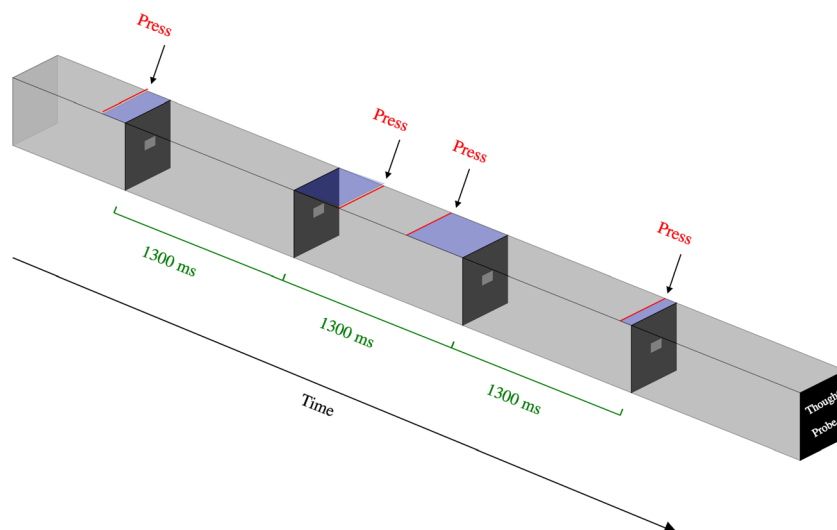
Fig. 1 Visual depiction of the vMRT. Participants are instructed to press the spacebar in synchrony with the onset of each gray box on the screen. Each gray box remains on the screen for 150 ms, after which time the screen turns black. In total, 1,300 ms elapses between stimulus onsets. Periodically throughout the task, a thought probe is presented, asking participants to report their depth of mind wandering immediately prior to the onset of the probe. The relative response time (RRT) represents the time between the trial response and stimulus onset. If the response preceded the onset of the stimulus, the associated RRT was negative.

were not related to the task at all (e.g., plans with friends, an upcoming test, plans for dinner, etc.).

Measures Rhythmic response times (RRTs) were calculated as the relative time difference (in milliseconds) between the moment at which the response was recorded and the moment of stimulus onset (Seli, Cheyne, & Smilek, 2013). Because participants' responses could precede or follow stimulus onset, the RRT value was negative if a participant responded prior to the stimulus onset, and positive if a participant responded following the stimulus onset. Figure 1 shows a few example trials and the corresponding time periods represented by the RRT. Since variability in response times is the main measure of attention yielded by the MRT, three measures of the variability of RRTs were calculated for the vMRT. Our first measure of variability, overall mean RRT variability, was computed using a moving window of the current and preceding four trials across all trials throughout the task (see Seli, Jonker, Cheyne, & Smilek, 2013).[1] Our second measure of variability, odd/even RRT variability, was obtained by separately computing the variance of RRTs on all nonoverlapping even and odd five-trial windows throughout the task (i.e., the overall mean RRT variance across trials 5–9, 15–19, 25–29, etc., and across trials 10–14, 20–24, 30–34, etc., respectively). Our third measure of variability was computed as the variance in RRTs produced on the five trials preceding each of the five thought-probe responses. As in previous work (e.g., Seli, Carriere, et al., 2013), the variance data were highly positively skewed, so we adjusted each variance measure using a

natural-logarithm transform. All analyses were performed using the R statistical language (R Core Team, 2015).

Results

In Study 1, we conducted four primary analyses. The first analysis focused on establishing the reliability of the vMRT. The second explored whether vMRT response variability on the trials immediately preceding thought probes increased linearly as a function of participants' depth of mind wandering (see Seli Cheyne, & Smilek, 2013, and Seli, Carriere, et al., 2014, for similar analyses in the context of the MRT). Third, a time-course analysis was performed in order to assess the changes in depth of mind wandering and MRT task performance as the task progressed (Thomson et al., 2014). Finally, at the level of individual differences, we explored the relation between the average response variability across the entire vMRT and the average depth of mind wandering (see Seli, Carriere, et al., 2013, and Seli, Cheyne, et al., 2015, for similar analyses in the context of the original MRT).

Moment-to-moment reliability As in Seli, Carriere, et al. (2014), to estimate the reliability of the RRT variance measure, we conducted a correlational analysis examining the relation between the log transformed RRT variance on all the nonoverlapping even and odd five-trial windows throughout the task. This analysis yielded a strong significant positive correlation coefficient, $r(33) = .96$, $p < .001$, indicating good reliability of the vMRT variance measure.

Split-half reliability In addition to the moment-to-moment measure of reliability, we examined the reliability of mean

---

[1] We used the moving-window analysis because it retained the temporal component of the variance across the task.

overall RRT variance between the first and second halves of the task. The goal of this secondary reliability analysis was to provide an overall estimate of the reliability of changes in response variability during the task, across participants. This reliability score was quantified with a Pearson product–moment correlation analysis examining the relation of the mean overall log-transformed RRT variance scores between the first and second halves. As we observed in the moment-to-moment measure of reliability, we found a very strong, positive relationship between the mean overall log-transformed RRT variance of the first and second halves of the task, $r(33) = .86, p < .001$.

Performance prior to thought probes Next we sought to determine whether the vMRT variance on the five trials immediately preceding thought probes varied as a function of each of the five possible probe reports (i.e., depth of mind wandering). To explore this possibility, we conducted a linear mixed-effects analysis with depth of mind wandering (1–5) as a fixed factor and participant as a random factor, which allowed both the intercept and the effect of depth of mind wandering to vary by participant. Importantly, this analysis allowed for the inclusion of data from participants who did not report at least one instance of each level of mind-wandering depth (i.e., 1, 2, 3, 4, and 5) across the 18 thought probes. In addition, this analysis permitted the inclusion of each observation for each participant, which thereby provided an estimate of within-subjects variability. To evaluate the significance of a term within a linear mixed-effects model, we compared the performance of the complete model, with all effects, with the performance of a model will all but the effect of interest (Magezi, 2015). The depth-of-mind-wandering measure contributed significantly to the model [estimate = 0.21; $\chi^2(1) = 12.132, p < .001$], indicating that vMRT variance increased as a function of increasing depth of mind wandering (see Fig. 2).

Performance and mind wandering over time on task To further test the convergent validity of the vMRT, we sought to test for decrements in task performance as a function of time on task. If the vMRT assesses inattentiveness, then we would expect to see a performance decrement over time and for this decrement to be associated with an increase in mind-wandering rates (as reported by Thomson et al., 2014). We examined this possibility by splitting the task into six blocks, each with 150 vMRT trials and three thought probes, with Block 1 corresponding to the first 150 trials, Block 2 corresponding to the second 150 trials, and so on. A linear mixed-effects model was fit to the data to assess the linear effect of block on the reported depth of mind wandering. The slope associated with a block and the intercept were permitted to vary by participant. We found a significant positive effect of block on depth of mind wandering [estimate = 0.27; $\chi^2(1) = 56.20, p < .001$], as is shown in Fig. 3. A similar model was fit
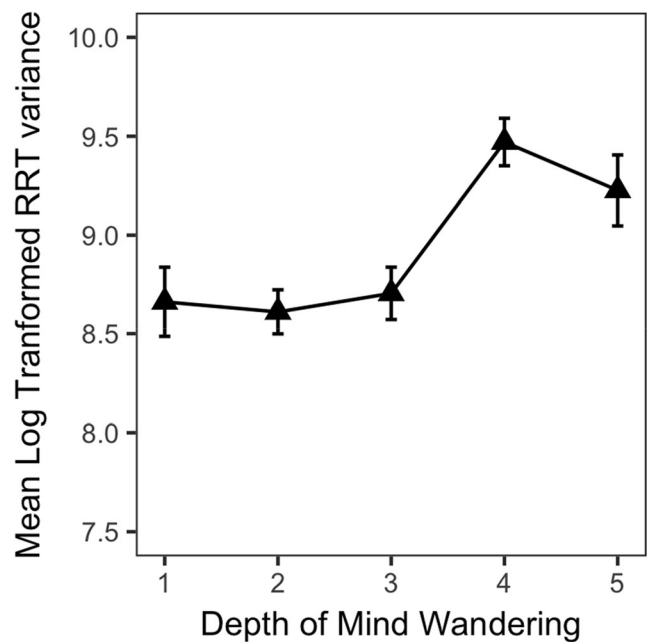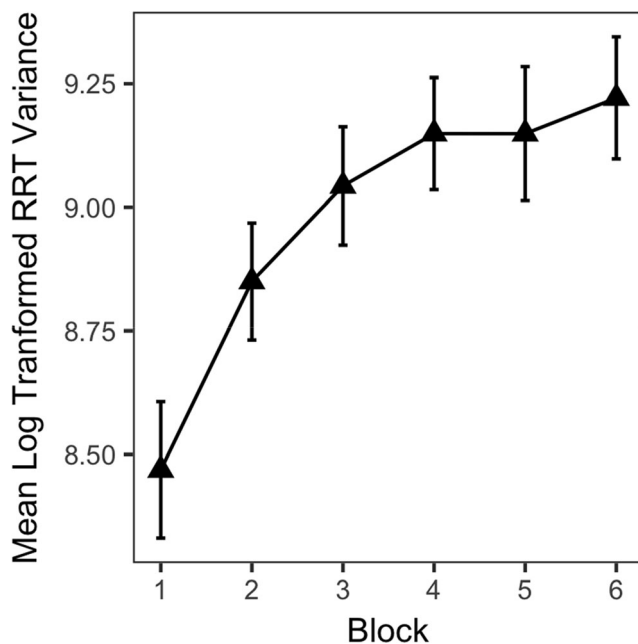

Fig. 2 Mean log-transformed RRT variance for each of the five levels of depth of mind wandering, with Level 1 corresponding to "completely on task" responses and Level 5 corresponding to "completely off task" responses. Error bars depict one standard error of the mean.

in order to assess the linear effect of block on the measured log-transformed RRT variance. The slope associated with a block and the intercept were again permitted to vary by participant. There was again a significant positive effect of block on the log-transformed RRT variance [estimate = 0.14; $\chi^2(1) = 42.00, p < .001$], shown in Fig. 4.
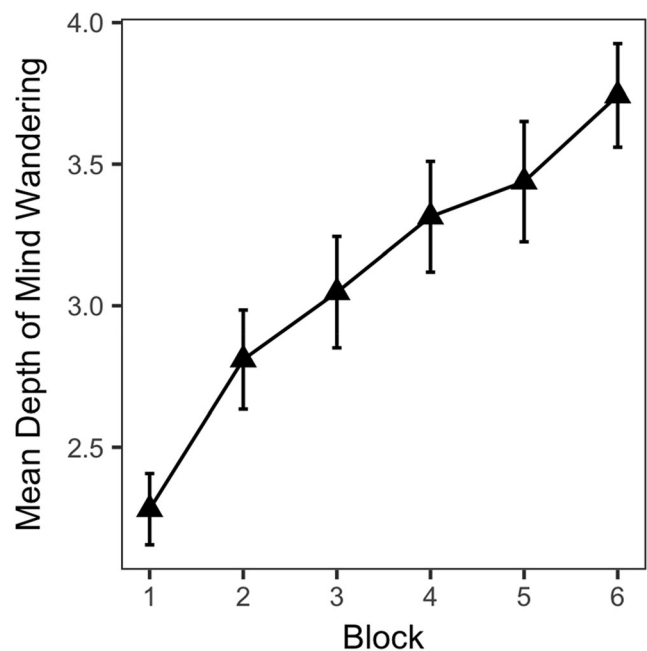
Fig. 3 Mean depth of mind wandering, with Level 1 corresponding to "completely on task" responses and Level 5 corresponding to "completely off task" responses, for each of the six task blocks. Error bars depict one standard error of the mean.
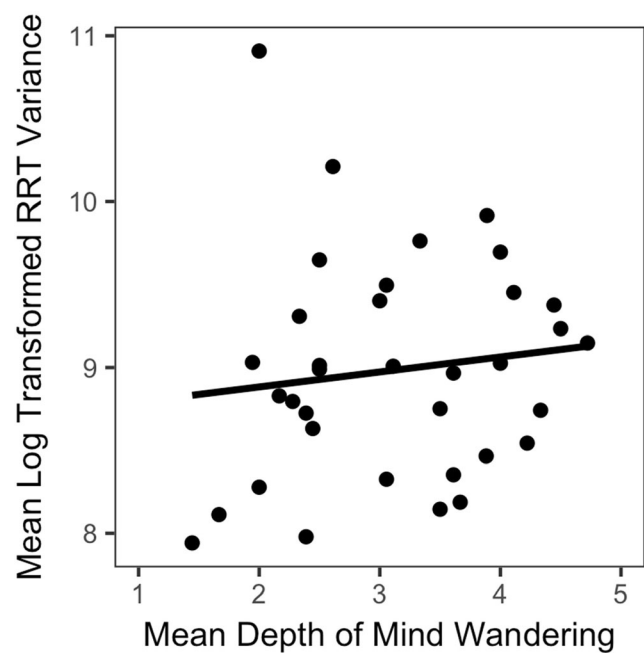
Fig. 4 Mean log-transformed RRT variance for each of the six task blocks. Error bars depict one standard error of the mean.

Fig. 5 Scatterplot showing the relation between mean depth of mind wandering over 18 thought probes and the mean log-transformed RRT variance over the entire duration of the task.

Individual differences (correlational) analysis Finally, we examined the relation between depth of mind wandering and vMRT performance by conducting a Pearson product–moment correlation, entering in each participant's mean overall variance across the entire task and their average depth of mind wandering. This analysis revealed that these measures were significantly positively correlated, $r(33) = .266$, $p = .002$ (see Fig. 5). Thus, as in previous work employing the standard MRT (e.g., Seli, Carriere, et al., 2013), participants who reported higher rates of mind wandering tended to produce greater response variability.

Discussion

The primary purpose of Study 1 was to verify that the measures associated with the MRT extended beyond the auditory domain. This was achieved by testing for similarities between the MRT and vMRT in terms of their behavioral outcomes and their relationships to mind wandering. Our results suggest that the vMRT has four key similarities with the MRT. First, the vMRT shows very high moment-to-moment reliability; a property that is mirrored in the MRT (Seli, Cheyne, et al., 2013). This reliability measure suggests that two separate measurements of behavioral variability within close temporal proximity are very similar in nature, and the strong relationship between the two measurements implies that the measure of behavioral variability is highly reliable.

Second, periods of greater self-reported depth of mind wandering were associated with greater response variability. This finding suggests that, as has been observed with the SART and MRT, behavioral variability is reliably associated

with mind wandering in the vMRT (Cheyne et al., 2009; McVay & Kane, 2009; Seli, Cheyne, Barton, & Smilek, 2012). The consistency of this effect across modalities and tasks provides strong evidence that self-reported mind wandering and behavioral variability are linked at a level that is not modality- nor task-specific.

Third, the results showed that both depth of mind wandering and response variability increased as a function of time on task. Time on task has been associated with increased mind wandering and decreased performance on a variety of tasks (e.g., Thomson et al., 2014). That both mind wandering and response variability follow this trend supports the notion that fluctuations in the vMRT response variability metric do indeed reflect fluctuations in depth of mind wandering, thus lending further support for the validity of the behavioral measure.

Finally, at the level of individual differences, participants who reported greater depths of mind wandering also tended to produce greater response variability. This, too, is in agreement with previous findings, with similar results being observed in the original MRT (Seli et al., 2013). The relatively small magnitude of the effect in this study is not altogether surprising. Such crude measures of mind wandering and response variability would be unlikely to be strongly associated, since many additional variables that are not associated with mind wandering may influence average response variability across individuals. For instance, individual differences in experience with holding a rhythm may account for some variance in response variability, with individuals with greater experience with rhythms having a lower overall response variability. In

any case, the small relation between response variability and depth of mind wandering across individuals adds to the previously mentioned findings pointing to the similarity between the MRT and vMRT.

Overall, the results of Study 1 suggest a high degree of similarity among the behavioral outcomes of the MRT and vMRT. This, in turn, suggests that the two tasks measure a common construct. However, to verify this conclusion, direct comparison of the two tasks was required.

Study 2

Study 2 was designed to extend the results of Study 1 in two ways. First, to confirm our initial findings, we attempted to directly replicate the results of Study 1 with a larger sample size. Second, we sought to directly compare the results from the vMRT to those from the original (auditory) MRT by having participants complete each of these tasks (within subjects). As in Study 1, participants again responded to periodically presented thought probes assessing their depth of mind wandering.

Method

Participants One hundred twenty one participants between the ages of 18 and 35 years (mean age = 29.83, *SD* = 4.23, 57 female) were recruited via Amazon's Mechanical Turk. Of the 121 participants, 102 reported that they had normal or corrected-to-normal vision, five reported not having corrected-to-normal vision, and 14 participants declined to answer the question. In addition, 99 participants reported that they had normal or corrected-to-normal hearing, four reported not having corrected-to-normal hearing, and 17 participants declined to answer the question. Except for one case (outlined below), the results of Study 2 did not vary as a function of whether we excluded data from participants who (1) reported non-normal vision/hearing or (2) did not respond to these questions; as such, all subsequent analyses included the full sample of participants.

Participants were assigned to one of two counterbalance conditions. In one condition, participants completed the auditory MRT prior to completing the vMRT. In the other condition, participants completed the vMRT prior to completing the auditory MRT. As in Study 1 and in previous work (e.g., Seli, Cheyne, & Smilek, 2013), the data from participants who failed to respond to >10% of the trials in either or both tasks were removed from subsequent analyses. With such a criterion for exclusion, 26 participants were removed from the subsequent analyses.

vMRT and MRT All details of the vMRT were identical to those reported in Study 1. The MRT was identical to the vMRT, with the following exceptions: (1) a metronome tone replaced the presentation of the gray box in the center of the screen (the

metronome tone consisted of an audible click, and was obtained from the audio recording software Ableton Live 9), and (2) as in the original MRT, the duration of the tone was 75 ms, as compared to the 150-ms duration of the gray box (this resulted in a slightly shorter poststimulus interval in the visual version of the task; the overall trial length, however, remained the same). Finally, all measures were computed in the same manner as in Study 1.

Results

Moment-to-moment reliability As in Study 1, we conducted a Pearson product–moment correlation analysis examining the mean overall variance for nonoverlapping even and odd windows separately for the vMRT and the MRT. In the vMRT, we found a strong positive relation between these variance measures, $r(93) = .93, p < .001$, and the same was true in the original MRT, $r(93) = .97, p < .001$. Importantly, these results suggest a high degree of moment-to-moment reliability in both tasks.

Split-half reliability In addition to the moment-to-moment analysis, we also performed a secondary analysis of reliability by comparing the mean overall variances across the first and second halves of each task. This analysis was also performed separately for the vMRT and MRT. In the vMRT, there was a strong positive relation between mean overall log-transformed RRT variances across the first and second halves of the task, $r(93) = .78, p < .001$. In the MRT, there was likewise a strong positive relationship between the mean overall log-transformed RRT variances in the first and second halves of the task, $r(93) = .85, p < .001$. This further supports the idea that RRT variance is a highly reliable measure.

Cross-modal reliability To directly compare MRT and vMRT performance, we conducted a Pearson product–moment correlation analysis examining the relation between participants' mean overall variance scores across the two tasks. This analysis yielded a strong positive relation between the mean overall variances in the MRT and the vMRT, $r(93) = .72, p < .001$. Critically, this finding further validates the measures of the MRT and vMRT by showing that individual differences in participants' response variabilities are consistent across tasks that involve different modalities. These findings thus support the notion that the response variability measured in both tasks is not modality-specific, but rather, that both tasks measure a general tendency toward behavioral variability.

Performance prior to thought probes A within-subjects analysis was conducted (on both the MRT and vMRT data) to test whether the log-transformed RRT variances differed as a function of depth of mind wandering in each of the two tasks (see Fig. 6). As in Study 1, a linear mixed-effects model was fit to the data from each task. In the models, depth of mind
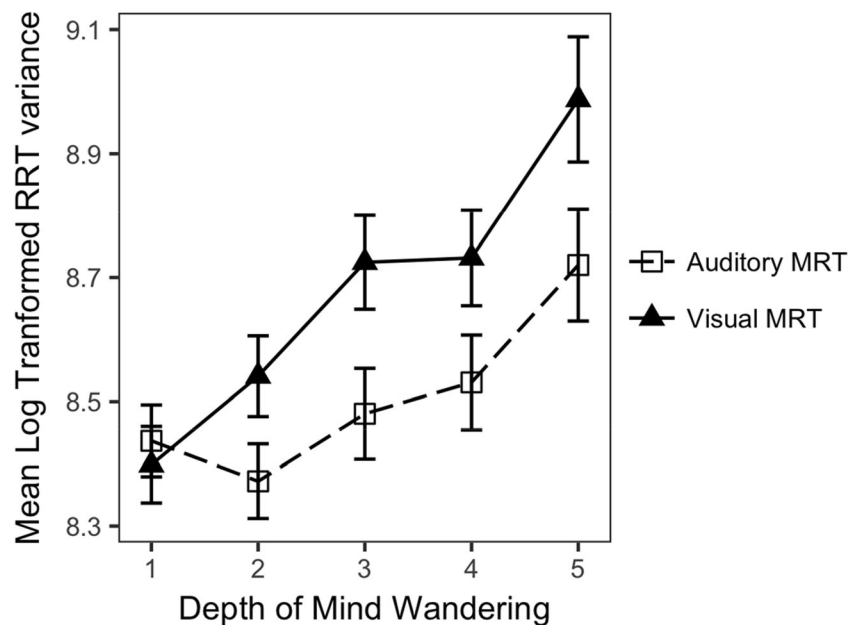
Fig. 6 Mean log-transformed RRT variances for each of the five levels of depth of mind wandering, with Level 1 corresponding to "completely on task" responses and Level 5 corresponding to "completely off task"

responses. Data for both the MRT and the vMRT are shown. Error bars depict one standard error of the mean.

wandering, as measured by the thought-probe responses, ranging from 1 ("completely on task") to 5 ("completely mind wandering"), and task modality (visual, auditory) were entered as fixed effects, and participant was entered as a random effect. No significant interaction between depth of mind wandering and task modality emerged [estimate = 0.04; $\chi^2(1) = 1.42, p = .23$]. The depth-of-mind-wandering term contributed significantly to the model, suggesting that response variance increased as depth of mind wandering increased [estimate = 0.155; $\chi^2(1) = 60.05, p < .001$]. Finally, task modality also contributed significantly to the model, suggesting that there were significantly higher levels of variance in the vMRT than in the MRT [estimate = 0.03; $\chi^2(1) = 4.47, p = .03$].

With respect to the last finding, we note that the main effect of task modality was marginal when we excluded the data from participants who (1) indicated nonnormal vision/ hearing or (2) did not respond to the questions pertaining to their vision and hearing [estimate = 0.05; $\chi^2(1) = 3.39, p = .07$] (all other analyses led to similar results, irrespective of whether these participants' data were included). It is difficult to determine why the removal of these participants' data produced a marginal effect of task modality. On the one hand, it is possible that factors surrounding nonnormality of vision and/ or hearing contributed to this effect. Perhaps more likely, though, is that the removal of these participants' data simply led to a reduction in statistical power.

Performance and mind wandering over time on task As in Study 1, linear mixed-effects models were used to assess the time course of reported depth of mind wandering and response

variance. To this end, we divided each of the two tasks into six blocks, each of which consisted of 150 trials and three thought probes. A model was fit to the data, with block number and task modality as fixed effects and with participant as a random effect, predicting depth of mind wandering (see Fig. 7). The slopes of block number and task modality, as well as the intercepts, were all allowed to vary by participant. There was no interaction between task modality and block [estimate = 0.04; $\chi^2(1) = 0.11, p = .74$], indicating that removing the interaction term from the model did not significantly reduce its predictive performance relative to the model including the interaction term as well as the two main effect terms. We also observed no main effect of task modality [estimate = − 0.01; $\chi^2(1) = 0.64, p = .42$]. There was, however, a main effect of block number, suggesting that when we collapsed across the MRT and vMRT, depth of mind wandering increased along with time on task [estimate = 0.22; $\chi^2(1) = 37.48, p < .001$]. These results suggest that the depth of mind wandering was equivalent across tasks and that it increased with time on task in similar fashions in both tasks.

We also fit models predicting the log-transformed RRT variance (see Fig. 8), with the same fixed and random effects. When predicting mean overall variance, the interaction between block number and task modality was not significant [estimate = 0.04; $\chi^2(1) = 1.23, p = .27$]. Also, the task modality term did not contribute significantly to the model, suggesting a nonsignificant main effect of task modality [estimate = − 0.01; $\chi^2(1) = 1.93, p = .16$]. However, the block number term did contribute significantly to the model, indicating that response variance increased significantly as time on task increased [estimate = 0.14; $\chi^2(1) = $
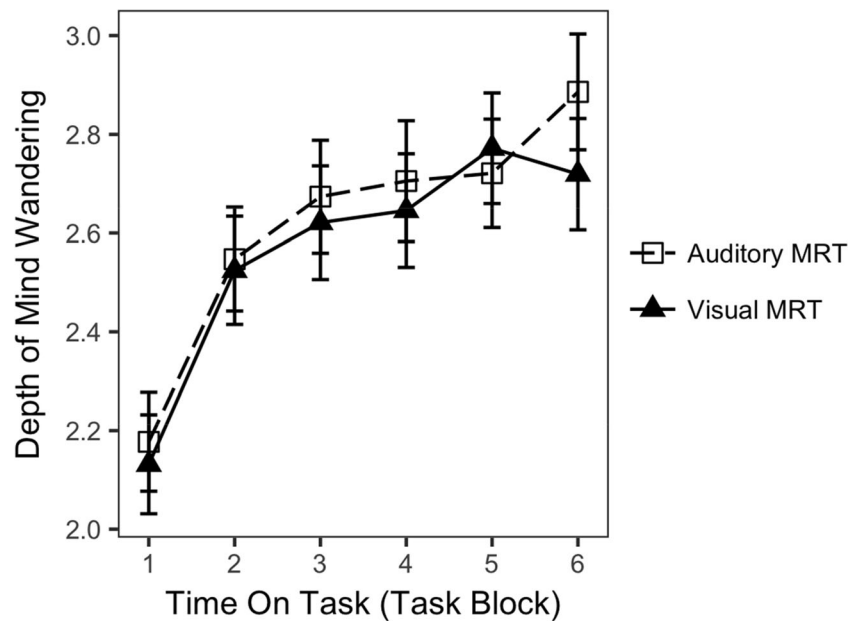
Fig. 7 Mean depth of mind wandering, with Level 1 corresponding to "completely on task" responses and Level 5 corresponding to "completely off task" responses, for each of the six task blocks. Data for both the MRT and the vMRT are shown. Error bars depict one standard error of the mean.

79.16, $p < .001$]. Therefore, like depth of mind wandering, response variability was equivalent across tasks and increased with time on task in a similar fashion in each task.

Individual differences (correlational) analysis Finally, we assessed the relation between the response variability of the tasks and depth of mind wandering at the level of individual differences. This was done by conducting a correlational analysis examining participants' mean depth of mind wandering and their mean overall response variance. The analysis was performed separately for each task. In the MRT, the relation between the mean overall log-transformed RRT variance and mean depth of mind wandering was not significant, $r(93) = -.08$, $p = .39$, as is shown in Fig. 9. Likewise, in the vMRT we found a non-significant relation between the mean overall log-transformed RRT variance and mean depth of mind wandering, $r(93) = -.04$, $p = .69$ (see Fig. 10)
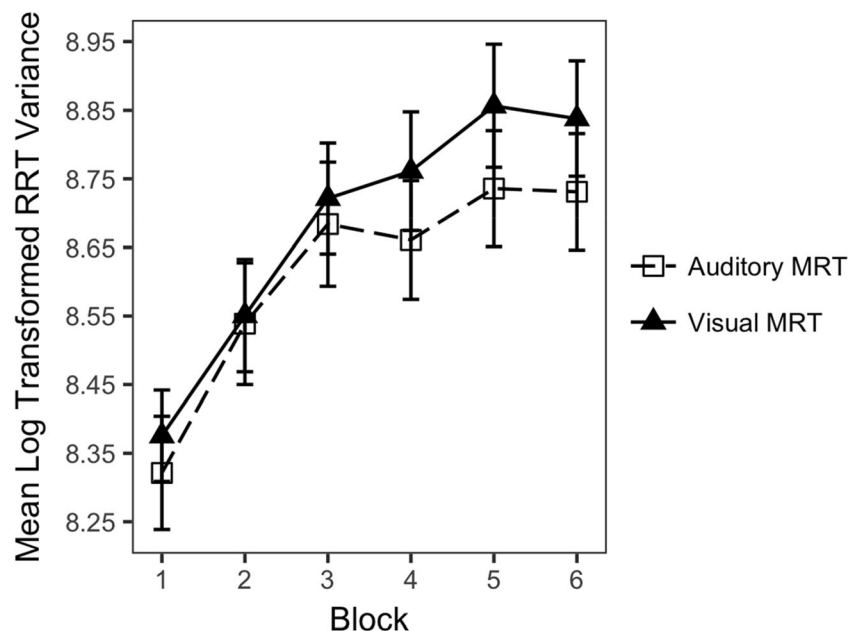
Fig. 8 Mean log-transformed RRT variances for each of the six task blocks. Data for both the MRT and the vMRT are shown. Error bars depict one standard error of the mean.
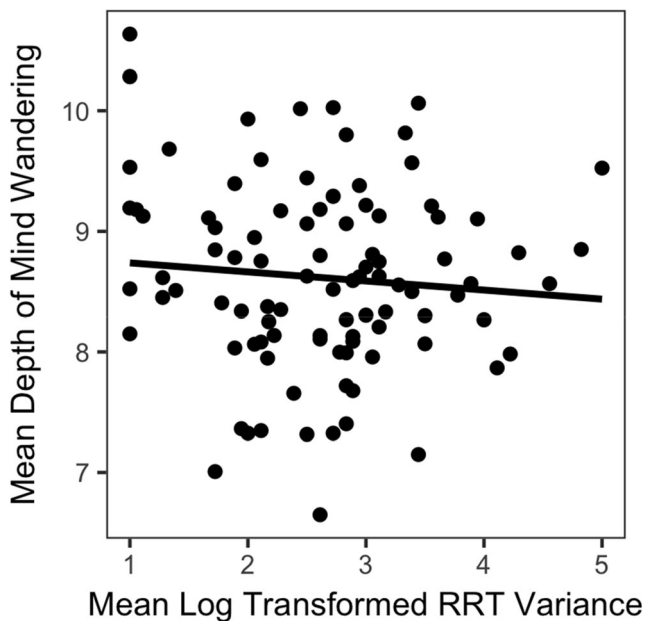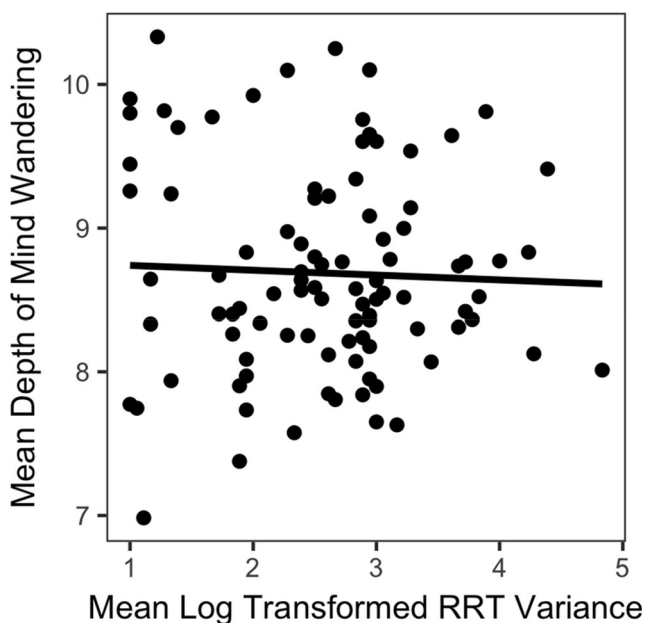
Fig. 9 Scatterplot showing the relationship between the mean depth of mind wandering over 18 thought probes and the mean log-transformed RRT variance over the entire duration of the task in the MRT.

Discussion

The primary goals of Study 2 were to (1) attempt to replicate the findings from Study 1 with a larger sample, and (2) direct-ly compare the results from the vMRT to the MRT by having participants complete both tasks.

As in Study 1, we first assessed the reliability of both tasks using three separate reliability analyses: a moment-to-

Fig. 10 Scatterplot showing the relationship between the mean depth of mind wandering over 18 thought probes and the mean log-transformed RRT variance over the entire duration of the task in the vMRT.

moment reliability measure, a split-half reliability measure, and a between-task reliability measure. With respect to moment-to-moment reliability, both the vMRT and the orig-inal MRT showed a high degree of reliability. Similarly, in the split-half reliability analysis, both tasks consistently showed high degrees of reliability. Finally, the overall reli-ability in response variability across tasks (i.e., between modalities) was likewise relatively large. Taken together, these measures of reliability suggest that the MRT and vMRT are highly reliable across participants, over a wide range of time scales.

Second, we investigated the task performance prior to thought probes and its relation to reported depth of mind wan-dering. Here we found that, in both the MRT and vMRT, response variability increased as a function of depth of mind wandering. Critically, this finding suggests that behavioral variability is reliably associated with mind wandering in both the MRT and the vMRT. Moreover, the consistency of this effect across modalities and tasks provides further evidence that self-reported mind wandering and behavioral variability are linked at a level that is not modality- or task-specific. Third, we examined depth of mind wandering and response variability as a function of time on task. As in previous re-search (Thomson et al., 2014), both depth of mind wandering and response variability increased as a function of time on task, suggesting that these two measures are tightly coupled with one another.

Finally, at the level of individual differences, we observed a nonsignificant relationship between mean depth of mind wan-dering and mean overall log-transformed RRT variance in both the MRT and the vMRT.

Conclusions

In the present studies, we developed and validated a visual version of the MRT (the vMRT). Across both studies, we observed various noteworthy findings. First, the vMRT showed high levels of reliability. This was measured in two separate ways: (1) the response variances for even and odd nonoverlapping windows were compared, and (2) the mean overall variances were examined for the first and second halves of each task. Both methods confirmed that there was high reliability in the vMRT's measure of response variability. Second, we observed a strong positive correlation between response variability in the vMRT and response variability in the original auditory version of the MRT (Study 2), which indicates that, within individuals, response variability is high-ly consistent across modalities (i.e., tasks). Third, at the within-participant level, we found that response variability in the vMRT systematically increased as a function of depth of mind wandering. Importantly, this finding indicates that re-sponse variability, as measured in both tasks, can be

considered a behavioral index of mind wandering. Fourth, the results indicated that response variability in the vMRT and depth of mind wandering both increased over time on task, which in turn provides further evidence to suggest that response variability and mind wandering co-occur reliably. Taken together, these findings indicate that both the vMRT and the original MRT index behavioral variability that is (1) not modality-specific and (2) related to fluctuations in task inattention, as indexed by mind wandering.

It is important to note that one finding did not generalize across our two studies. Namely, in Study 1 we observed a positive correlation between the mean depth of mind wandering and mean overall variance at the individual-differences level. However, this relation was not replicated in the vMRT in Study 2, nor was it found when examining variability and depth of mind wandering in the MRT (Study 2), despite the fact that previous research has reported such a relation when examining MRT performance (e.g., Seli, Cheyne, et al., 2015). However, as was recently noted by Thomson, Seli, Besner, and Smilek (2014), individual-difference correlations between rates of mind wandering and performance on continuous-performance tasks, including the SART and MRT, are seemingly tenuous and do not always obtain significance. This is likely due to substantial differences in individuals' baseline propensity to mind-wander and their baseline ability to perform well on such tasks. For this reason, we suggest that the MRT and vMRT would be most effective when used primarily for within-participant comparisons and group-level analyses.

To our knowledge, three tasks have typically been used to examine the relation between response variability and mind wandering: the SART, the MRT, and reading tasks during which eyetracking technology is used (e.g., Reichle, Reineberg, & Schooler, 2010). We have discussed the reasons we believe that the MRT and vMRT provide measures of behavioral variability that are superior to those yielded by the SART. However, the general conclusions about the relation between behavioral variability and mind wandering seem at odds with findings from studies that have employed eyetracking technology. Indeed, in Reichle, Reineberg, and Schooler's study, participants who were asked to read a passage and report mind wandering had lower variability in their fixation durations during episodes of mind wandering than during episodes of on-task thought. Unlike the findings in Studies 1 and 2, as well as previous findings in the literature (e.g., Seli, Carriere, et al., 2013; Seli, Cheyne, et al., 2015), this finding points to a decrease in behavioral variability during mind wandering. This likely occurs because attentive reading leads eye fixation durations to closely track variability in word familiarity, with familiar words leading to shorter fixations than do unfamiliar words. However, when an individual mind-wanders and stops paying attention to the text, his or her eye fixations likely become insensitive to word familiarity, thus yielding lower variability in fixation durations.

With respect to comparing the suitability of eyetracking to that of the MRT and vMRT as an indicator of mind wandering, we suggest that the MRT/vMRT can easily be employed in situations in which the logistics of eyetracking would make it difficult to implement.

Overall, the results of our two studies suggest that the original MRT findings (Seli, Carriere, et al., 2013) are consistent across visual and auditory modalities. Moreover, these results suggest that the vMRT can be used in the place of the MRT in experimental contexts in which researchers' designs require a visual-based primary task, acting as a replacement for the standard SART (Robertson et al., 1997). As was argued by Seli, Carriere, et al. (2013), the primary benefit of the MRT—and by extension, the vMRT—is that it does not suffer the same methodological issues as other tasks that include infrequent target trials (Helton et al., 2009; Seli, Cheyne, & Smilek, 2012). The absence of critical target trials, combined with the explicit instructions for participants to respond rhythmically, allow the MRT and the vMRT to provide a purer measure of behavioral variability.

References

Carriere, J. S., Seli, P., & Smilek, D. (2013). Wandering in both mind and body: Individual differences in mind wandering and inattention predict fidgeting. *Canadian Journal of Experimental Psychology*, *67*, 19. doi:https://doi.org/10.1037/a0031438

Cheyne, J. A., Carriere, J. S., Solman, G. J., & Smilek, D. (2011). Challenge and error: Critical events and attention-related errors. *Cognition*, *121*, 437–446. doi:https://doi.org/10.1016/j.cognition.2011.07.010

Cheyne, J. A., Solman, G. J. F., Carriere, J. S. A., & Smilek, D. (2009). Anatomy of an error: A bidirectional state model of task engagement/disengagement and attention-related errors. *Cognition*, *111*, 98–113. doi:https://doi.org/10.1016/j.cognition.2008.12.009

Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., & Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proceedings of the National Academy of Sciences*, *106*, 8719–8724. doi:https://doi.org/10.1073/pnas.0900234106

Feng, S., D'Mello, S., & Graesser, A. C. (2013). Mind wandering while reading easy and difficult texts. *Psychonomic Bulletin & Review*, *20*, 586–592. doi:https://doi.org/10.3758/s13423-012-0367-y

Helton, W. S., Kern, R. P., & Walker, D. R. (2009). Conscious thought and the sustained attention to response task. *Consciousness and Cognition*, *18*, 600–607. doi:https://doi.org/10.1016/j.concog.2009.06.002

Jonker, T. R., Seli, P., Cheyne, J. A., & Smilek, D. (2013). Performance reactivity in a continuous-performance task: Implications for understanding post-error behavior. *Consciousness and Cognition*, *22*, 1468–1476. doi:https://doi.org/10.1016/j.concog.2013.10.005

Magezi, D. A. (2015). Linear mixed-effects models for within-participant psychology experiments: An introductory tutorial and free,

graphical user interface (LMMgui). *Frontiers in Psychology*, *6*, 2. doi:https://doi.org/10.3389/fpsyg.2015.00002

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. doi:https://doi.org/10.1037/a0014104

Mrazek, M. D., Franklin, M. S., Phillips, D. T., Baird, B., & Schooler, J. W. (2013). Mindfulness training improves working memory capacity and GRE performance whereas reducing mind wandering. *Psychological Science*, *24*, 776–781. doi:https://doi.org/10.1177/0956797612459659

R Core Team. (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from www.R-Project.org/

Reichle, E. D., Reineberg, A. E., & Schooler, J. W. (2010). Eye movements during mindless reading. *Psychological Science*, *21*, 1300–1310. doi:https://doi.org/10.1177/0956797610378686

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). "Oops!": Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, *35*, 747–758. doi:https://doi.org/10.1016/S0028-3932(97)00015-8

Seli, P. (2016). The Attention-Lapse and Motor Decoupling accounts of SART performance are not mutually exclusive. *Consciousness and Cognition*, *41*, 189–198. doi:https://doi.org/10.1016/j.concog.2016.02.017

Seli, P., Carriere, J. S., 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., Carriere, J. S. A., Thomson, D. R., Cheyne, J. A., Martens, K. A. E., & Smilek, D. (2014). Restless mind, restless body. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 660–668. doi:https://doi.org/10.1037/a0035260

Seli, P., Cheyne, J. A., Barton, K. R., & Smilek, D. (2012). Consistency of sustained attention across modalities: Comparing visual and auditory versions of the SART. *Canadian Journal of Experimental Psychology*, *66*, 44–50. doi:https://doi.org/10.1037/a0025111

Seli, P., Cheyne, J. A., & Smilek, D. (2012). Attention failures versus misplaced diligence: Separating attention lapses from speed–accuracy trade-offs. *Consciousness and Cognition*, *21*, 277–291. doi: https://doi.org/10.1016/j.concog.2011.09.017

Seli, P., Cheyne, J. A., & Smilek, D. (2013). Wandering minds and wavering rhythms: Linking mind wandering and behavioral variability. *Journal of Experimental Psychology. Human Perception and Performance*, *39*, 1–5. doi:https://doi.org/10.1037/a0030954

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. doi:https://doi.org/10.1037/xlm0000116

Seli, P., Jonker, T. R., Cheyne, J. A., Cortes, K., & Smilek, D. (2015). Can research participants comment authoritatively on the validity of their self-reports of mind wandering and task engagement? *Journal of Experimental Psychology: Human Perception and Performance*, *41*, 703–709. doi:https://doi.org/10.1037/xhp0000029

Seli, P., Jonker, T. R., Cheyne, J. A., & Smilek, D. (2013). Enhancing SART validity by statistically controlling speed–accuracy trade-offs. *Frontiers in Psychology*, *4*, 265. doi:https://doi.org/10.3389/fpsyg.2013.00265

Seli, P., Jonker, T. R., Solman, G. J. F., Cheyne, J. A., & Smilek, D. (2013). A methodological note on evaluating performance in a sustained-attention-to-response task. *Behavior Research Methods*, *45*, 355–363. doi:https://doi.org/10.3758/s13428-012-0266-1

Smallwood, J., Davies, J. B., Heim, D., Finnigan, F., Sudberry, M., O'Connor, R., & Obonsawin, M. (2004). Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness and Cognition*, *13*, 657–690. doi:https://doi.org/10.1016/j.concog.2004.06.003

Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin*, *132*, 946–958. doi: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. doi:https://doi.org/10.1146/annurev-psych-010814-015331

Thomson, D. R., Seli, P., Besner, D., & Smilek, D. (2014). On the link between mind wandering and task performance over time. *Consciousness and Cognition*, *27*, 14–26. doi:10.1016/j.concog.2014.04.001

Unsworth, N., & McMillan, B. D. (2013). Mind wandering and reading comprehension: Examining the roles of working memory capacity, interest, motivation, and topic experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 832–842. doi: https://doi.org/10.1037/a0029669