The eyes betray the mind: Pupil size indicates the presence of sticky mind-wandering

Mathanja R. Verkaik, s2198665

Institute of Artificial Intelligence and Cognitive Engineering, University of Groningen, Groningen, The Netherlands

May 2, 2017

Master thesis Behavioural and Cognitive Neurosciences, C-track

Supervised by Dr. Marieke K. van Vugt

Abstract

An important challenge when studying mind-wandering, is the detection of this covert phenomenon. Additionally, a specific property of mind-wandering, which can be referred to as stickiness, has been associated with vulnerability for developing depressive symptoms. The current study seeks to identify performance and pupillary markers of mind-wandering and its stickiness. During a sustained attention to response task (SART), we recorded participants' accuracy, response times and pupil size. To assess the type of thoughts a person was having, thought probes regularly interrupted the task. Mind-wandering and sticky thoughts were found to be associated with behavioural and pupillary patterns of an automatic and stimulus-driven response style, that has been theorised to place minimal demands on cognitive resources. These findings indicate that mind-wandering and its stickiness could potentially be recognised in various contexts, by looking for signals of this specific energy-saving mode. Additionally, our results imply that mind-wandering is a resource-consuming process that leaves little reserves for the performance of other tasks.

Keywords: Mind-wandering, current concerns, pupil size, sustained attention to response task

Contents

The eyes betray the mind: Pupil size indicates the presence of sticky mind-wandering

Mind-wandering is the familiar phenomenon of turning one's attention towards task-unrelated, internally generated thoughts, while performing a certain task. Mind-wandering is hypothesised to occur more frequently in undemanding and unimportant tasks, drawing attentional resources away from ongoing performance (Klinger, 2013; Smallwood & Schooler, 2006; Smallwood & Schooler, 2015). As such, mind-wandering might have important effects on behavioural measures in many of the repetitive and unstimulating laboratory tasks used in cognitive experiments.

Mind-wandering and task performance

An important challenge for studying mind-wandering is the covert nature of this phenomenon, which makes it very difficult to detect. To overcome this problem, it is important to combine information from different sources, such as self-reports, task performance and psychophysiology (Smallwood & Schooler, 2015). On the behavioural level, several correlates of self-reported mind-wandering have been identified. First of all, mind-wandering has been associated with reduced accuracy, both in reading comprehension and in the performance on cognitive tasks (Cheyne, Solman, Carriere, & Smilek, 2009; Franklin, Smallwood, & Schooler, 2011; Foulsham, Farley, & Kingstone, 2013; van Vugt & Broers, 2016; McVay & Kane, 2009). Secondly, mind-wandering is often paired with longer RT in tasks where quick responding is the main objective (Foulsham et al., 2013; Smallwood et al., 2012; Unsworth & Robison, 2016). However, in tasks where accuracy is important as well, mind-wandering is more often linked to shorter, rather than longer, RT (Franklin et al., 2011; Smallwood et al., 2004; McVay & Kane, 2009; Manly, Robertson, Galloway, & Hawkins, 1999; Manly, Davison, Heutink, Galloway, & Robertson, 2000). Additionally, several studies have reported that response time variability is increased during mind-wandering (Bastian & Sackur, 2013; Cheyne et al., 2009; Mrazek, Smallwood, & Schooler, 2012; McVay & Kane, 2009), although van Vugt and Broers (2016) reported a slight decrease. Finally, during mind-wandering, RT has been shown to depend less on stimulus properties that would normally have a large and robust influence (Franklin et al., 2011).

In a sustained attention to response task (SART), a vigilance task developed to measure lapses in sustained attention (Robertson, Manly, Andrade, Baddeley, & Yiend, 1997), decreases in accuracy have been consistently linked to faster responses (Manly et al., 1999; Manly et al., 2000; Smallwood et al., 2004; McVay & Kane, 2009). According to the SART's developers, this effect is caused by the task's ability to invoke a pre-potent response of pressing a button. Once established, that tendency can only be overridden when the participant is actively paying attention (Manly et al., 1999; Manly et al., 2000). This explanation fits well within Braver (2012)'s dual mechanisms of control framework, which states that people have two different control modes at their disposal. In the *proactive* control mode, people evaluate each stimulus in the light of their current task goal before they respond, thus resulting in slow but accurate responses. When people employ a *reactive* strategy, on the other hand, they conserve their mental resources by selecting a default response that is only overridden when an environmental cue tells them to. With no need to check the stimulus before each response, the reactive strategy shortens reaction times, but also reduces accuracy. Because mind-wandering is theorised to draw mental resources away from an ongoing task (Smallwood & Schooler, 2006), mind-wandering might be observable in behaviour by looking for signs of the energy-saving reactive control mode.

Mind-wandering and pupil size

Apart from measuring task performance, psychophysiological parameters can be used to detect mind-wandering. Of these measures, pupil size is of particular interest to mind-wandering research, because it is relatively unintrusive and strongly related to a person's arousal levels (Joshi, Li, Kalwani, & Gold, 2016; Aston-Jones & Cohen, 2005). Pupil size is usually measured on two different temporal scales, called *tonic* and *phasic* pupil size. Tonic, or baseline, pupil size is positively associated with an individual's arousal level and their current tendency to seek novelty. Phasic pupil size, on the other hand, is a transient widening of the pupil in response to task-relevant processes and is theorised to benefit task performance (Murphy, Robertson, Balsters, & O'Connell, 2011). According to the adaptive gain theory (Aston-Jones & Cohen, 2005), tonic and phasic pupil size have an inverted U-shaped relationship, such that phasic pupil responses and task performance are maximal at an intermediate level of tonic pupil size (Figure 1). At low levels of tonic pupil size, people are theorised to be underaroused, which is reflected in small phasic pupil responses, low performance and inactivity. High tonic pupil sizes are also associated with low phasic pupil size and low performance, but this presumably originates from overarousal, which at the behavioural level is marked by high distractability and exploration of new behaviours.

Where mind-wandering lies on the adaptive gain curve, is still open to debate. A consistent finding across several studies, is that phasic pupil responses are reduced during mind-wandering (Mittner et al., 2014; Smallwood et al., 2011; Unsworth & Robison, 2016), indicating that this state lies either at the high or at the low end of the adaptive gain curve. However, it is difficult to determine which one. Some studies have reported increases in tonic pupil size during mind-wandering (Franklin, Broadway, Mrazek, Smallwood, & Schooler, 2013; Smallwood et al., 2011; Smallwood et al., 2012), which would mean that this state lies near the high end of the adaptive gain curve and, therefore, that it is paired with high arousal and a tendency to explore new behaviours. However, decreases in tonic pupil size have been reported as well (Mittner et al., 2014; Unsworth & Robison, 2016; Grandchamp, Braboszcz, & Delorme, 2014),

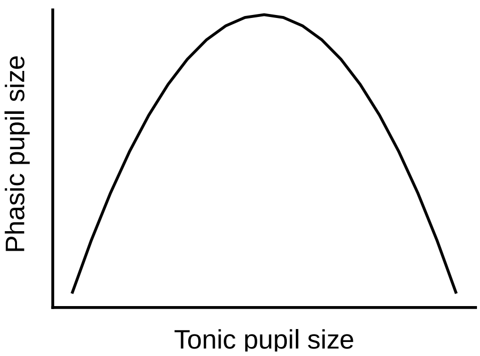

Figure 1. The adaptive gain curve. Phasic pupil responses are largest at intermediate tonic pupil size.

implying that mind-wandering lies at the low end of the adaptive gain curve and is associated with underarousal and inertia. One possible explanation of these divergent findings, is that different types of off-task thought lie at different places on the adaptive gain curve. Support for this interpretation is provided by Unsworth and Robison (2016), who found that tonic pupil size was decreased during inattentiveness and distraction by one's own thoughts, whereas it was increased when people were thinking about their direct environment. As such, inattentiveness and internal distractions may lie towards the low-tonic end of the adaptive gain curve, while external distractions occur at the high-tonic end.

Sticky mind-wandering and current concerns

Within the mind-wandering literature, different subtypes have been distinguished, such as daydreaming, rumination and planning personal goals. Some of these seem to have adaptive consequences, whereas others may be more maladaptive (Mooneyham & Schooler, 2013). For example, Zhiyan and Singer (1997) found that different mind-wandering styles were correlated with different positive and negative personality traits (Mooneyham & Schooler, 2013). Along similar lines, there seems to be a subtype of mind-wandering, that has been associated with vulnerability for developing depressive symptoms (Joormann, Levens, & Gotlib, 2011; Marchetti, Koster, Klinger, & Alloy, 2016; Verplanken, Friborg, Wang, Trafimow, & Woolf, 2007). This subtype is highly habitual and difficult to disengage from and could be referred to as *sticky* mind-wandering (Joormann et al., 2011).

According to Klinger (2013) and Marchetti et al. (2016), sticky mind-wandering develops as a result of thinking about unreached goals. When people commit to a certain goal, they often fill their spare moments evaluating its present state and planning further progress. In principle, such goal-related thoughts are advantageous to the individual. They keep personal goals activated and sensitise cognition to information that could help fulfil them. However, in order to effectively benefit goal pursuit, goal-related mind-wandering needs to be shielded from environmental distractions (Klinger, 2013; Smallwood et al., 2011). This is reflected in situations where we catch ourselves thinking about, for example, the holiday we are planning while we should be paying attention to the presentation that is currently being given. This may be a relatively innocent example, but when significant personal goals remain unattained, concerns about them could become increasingly intrusive, sticking around in people's working memories and pushing out incoming information that might be needed for the completion of daily tasks.

The close involvement of current concerns in (sticky) mind-wandering can also be utilised by researchers, in order to manipulate the frequency and stickiness of mind-wandering. McVay and Kane (2013) and van Vugt and Broers (2016) recently demonstrated this in a SART. Participants in these experiments viewed a stream of words, one at a time, and pressed a button whenever these were spelled in lower-case letters. Interestingly, when three successive words represented one of the participant's current concerns, frequency of off-task thought increased. Additionally, these mind-wandering thoughts were considered more sticky when they occurred. These results thus suggest that the presentation of current concerns can be used in experiments to make especially sticky mind-wandering occur more often. Since people normally only mind-wander 30 to 50 percent of the time (Killingsworth & Gilbert, 2010), implementing current concerns in investigations of sticky mind-wandering can provide a welcome increase in useful observations.

The current experiment

The present study aimed to identify eye-tracking and performance markers of mind-wandering and its stickiness, by measuring performance and pupil size while people completed a SART. To increase the occurrence of especially sticky mind-wandering, participant's own current concerns, as well as someone else's, were inserted in this task (McVay & Kane, 2013; van Vugt & Broers, 2016). To measure mind-wandering and its stickiness, participants occasionally reported what type of thoughts they were having and how difficult it was to disengage from these.

Based on the discussed findings in previous SART studies, we hypothesised that mind-wandering would be paired with lower accuracy, faster responses and increased response time variability. With respect to pupil size, we predicted that phasic pupil responses would be smaller during mind-wandering than during on-task thought. However, since previous studies on tonic pupil size provided mixed results, we had no predictions about the effects of mind-wandering on this measure. Moreover, since the presentation of people's current concerns should increase mind-wandering frequencies, we expected a decrease in on-task and an increase in off-task thought reports. Additionally, we expected to observe the same patterns in performance and pupil size immediately after the presentation of someone's own concerns as we expected to see during mind-wandering. Furthermore, because sticky mind-wandering may represent a more internally focused and perceptually decoupled form of mind-wandering, it was expected that the described behavioural and eye-tracking markers of mind-wandering, would also be observed when thoughts were rated as more difficult to disengage from. Finally, because the presentation of current concerns has been shown to result in more sticky thoughts, we hypothesised that thoughts right after these concerns would be rated as more difficult to disengage from.

Methods

Participants

34 Native Dutch speakers between 18 and 30 years old (M = 22.7, SD = 2.56; 20 female) participated in this experiment. All of them received a monetary compensation for their time spent on the experiment. Participants were recruited via social media, the Artificial Intelligence study association and through a mailing list for Artificial Intelligence students. By consequence, most participants were university students (n = 25, mean age = 22.2, 14 female), but some studied at a vocational university (n = 6, mean age = 24.2, 5 female) or followed intermediate vocational education (n = 1, age = 25, 0 female). Two participants had not followed any form of higher education (mean age = 23.5, 1 female). All of the participants gave informed consent prior to the

experiment.

Stimuli and apparatus

Questionnaires. Participant's current concerns were collected through an online version of the *Personal Concerns Inventory* (PCI; adapted from Cox & Klinger, 2004). In this questionnaire, participants are asked to report goals and concerns in a number of different areas (e.g. family, finance, hobbies) and rate the importance of those concerns on a scale from one to ten. The *Behavioural Inhibition System / Behavioural Approach System* scales (BIS/BAS; adapted from Carver & White, 1994) and the Habit Index of Negative Thinking (HINT; Verplanken et al., 2007) were used as distractor questionnaires, to lead participant's attention away from the PCI. *Google Forms* was used to administer these questionnaires online.

Task. The SART in the current experiment was based on those used by van Vugt and Broers (2016) and McVay and Kane (2013). It consisted of 720 black-printed words, presented one by one against a grey background. The majority of these words were lower-case *go* stimuli, in response to which participants pressed a button as fast as they could. However, 80 of the presented words were *no-go* stimuli, spelled in all-caps, in response to which participants were instructed to withhold a button press.

As in the experiments by McVay and Kane (2013) and van Vugt and Broers (2016), participant's current concerns (*Own concerns*) were included in the task. These were presented in the shape of three subsequent words that together reflected a participant's reported concern. For each participant, the *Own concerns* of another participant were included in the task as well, serving as a control condition called *Other concerns*.

To asses whether and how stickily participants were mind-wandering, so-called *thought probes* were included in the task. These consisted of two questions that interrupted the SART at predetermined moments (Figure 2). The first of these questions was based on Unsworth and Robison (2016), the second on van Vugt and

What were you just thinking about?

1) I was completely focused on the task.
2) I was evaluating aspects of the task.
(For example my performance or how long it is taking)
3) I was thinking about personal matters.
4) I was distracted by my environment.
(For example sound, temperature, my physical state)
5) I was daydreaming / I was thinking of task-unrelated matters.
6) I was not paying attention, but neither was I thinking about anything specifically.

Press the number that matches your answer to continue.

How difficult was it to disengage from the thought?

1) Very difficult
2) Difficult
3) Neither difficult nor easy
4) Easy
5) Very easy

Press the number that matches your answer to continue.

Figure 2. An English translation of the thought probes.

Broers (2016). In total, the experiment included 16 *OwnConcern* triplets, 16 *OtherConcern* triplets and 48 thought probes.

The stimulus words that did not reflect any concerns were selected from the Dutch subtitle-based word frequency database *SUBTLEX-NL* (Keuleers, Brysbaert, & New, 2016). Selection of words was based on the variable called *Lg10CD*, because it explained the most variance in accuracy and RT on lexical decision tasks (Keuleers et al., 2016). After the most ($>$ 20 per million) and least ($<$ 1000 per million) frequent words were removed, 312 words were selected with a *Lg10CD* value around the mean. Numbers, non-words and high-arousal words were removed and replaced equally by words from both sides of the *Lg10CD* mean.

Materials. The SART was created and run using the open-source application *PsychoPy* (version 1.83.04; Peirce, 2007). Stimuli were presented centrally on an 18-inch (1200 x 1600 pixels) LCD monitor with a refresh rate of 60 Hz. Participants responded by pressing the 'm' key (SART trials) or the number keys (thought probes) on a USB keyboard. Pupil dilation and gaze direction were recorded by an *EyeLink 1000* eye-tracker at a sampling rate of 250 Hz. Data analysis was performed using the *lme4* and *BayesFactor* packages of the *R* statistical software (Bates, Maechler, Bolker, & Walker, 2014; R Core Team, 2016; Bates, Mächler, Bolker, & Walker, 2015; Morey & Rouder, 2015).

Procedure

Questionnaires. Before coming to the lab, participants were asked to fill in the HINT, BIS/BAS and PCI online. Based on the PCI answers, the experimenter chose two of the participant's concerns, translated them into concern triplets and included those in the SART. This process was carried out as follows. First of all, the experimenter looked at the two concerns with the highest importance rating. If any of these concerns were too common or too general, the experimenter moved to the highest-rated concern after that. Whenever two concerns were rated equally important, the most specific or unique one was selected. When two relatively unique and important concerns were selected, these were summarised into three-word, telegraph-style phrases, where the use of participants' own words was avoided. For example, if a participant reported (1), this was translated into (2).

(1) "Er zijn nog wat dingen die ik moet voorbereiden voordat ik kan beginnen met een tussenjaar."
"There are still some things I need to arrange before I can start taking a gap year."

(2) pauze loopbaan prepareren
prepare break career

Laboratory session. The lab session started with setting up the eye-tracker. Participants sat on a non-rolling chair and placed their heads in a chin rest, located approximately 60 cm from the screen. The eye-tracker was calibrated to the participant's dominant eye or the eye that provided the best signal. After camera set-up, the task instructions appeared on the participant's screen. These were followed by a practice session, consisting of ten SART trials (including 1 *no-go* trial) and a thought probe. Finally, the actual experiment began and the eye-tracker started recording.

Participants were allowed to take a break every 2 blocks, which was approximately every 10 to 15 minutes. During this break, participants could take their head out of the chin rest. The computer task lasted for 40 to 60 minutes, while the session as a whole took about 45 to 75 minutes, depending on how easily the eye-tracker could be set up.

Task. The stimulus words were presented in a random order, which differed across participants. The order of the trial types, on the other hand, was fixed and based on that of McVay and Kane (2013). Whenever a concern triplet was presented, it was followed by 4 lower-case *go* trials, one upper-case *no-go* trial, and a thought probe. Figure 3 illustrates this sequence.

An overview of an individual SART trial can be found in Figure 4. Each trial started with a fixation cross, followed by a stimulus word, a mask and another fixation cross. The first fixation cross constituted the intertrial interval (ITI) and stayed on the screen for a randomly chosen duration between 1500 and 2100 ms. Both the stimulus and the mask were displayed for 300 ms. The final fixation cross represented the response interval and remained on the screen for 3000 ms. Responses were recorded from the moment the stimulus appeared on the screen, even during stimulus

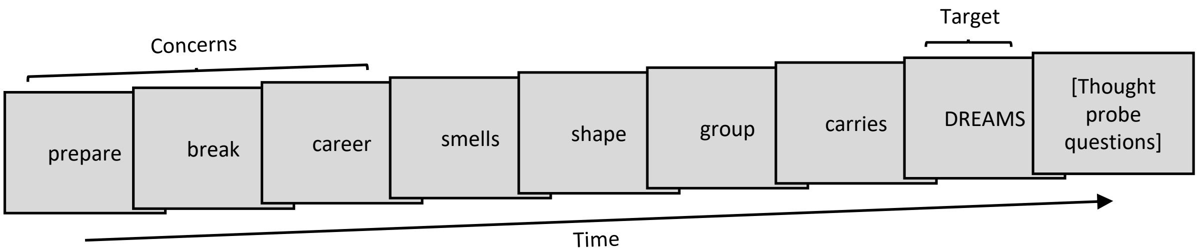

Figure 3. Overview of a task fragment.

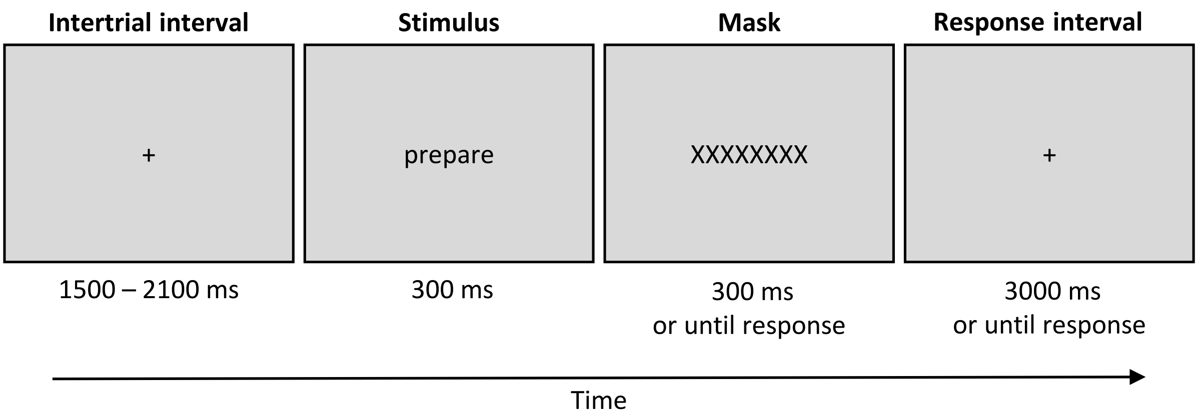

Figure 4. Overview of a trial.

presentation. However, the stimulus remained on the screen for the full 300 ms before proceeding to the next ITI. If the participant answered during the mask or the response interval, the task immediately continued to the next ITI. Thought probes remained on the screen until response.

The experiment consisted of eight blocks, each containing an equal number of targets, concerns and thought probes. Each block included only one of the two *OwnConcerns* and one of the two *OtherConcerns*. Which ones exactly, alternated between blocks. In order to conceal the repeated concern words, each of the blocks was composed of two identical sequences of 45 SART trials and 3 thought probes.

Data analysis

All data were analysed using linear mixed effects models. Each model contained one dependent and one independent variable, as well as random intercepts for *Subject*. The presence of an overall relationship was investigated using likelihood ratio tests and Bayes factors (BF). Both methods compared alternative models with the independent variable of interest to null models that did not include this variable. The Bayes factors were calculated by dividing the likelihood of the alternative model by the likelihood of the null model (alternative / null). As such, high BF values provided evidence for the alternative model and low values for the null model, whereas Bayes factors around 1 provided no evidence in either direction. After model comparison, the precise relationship between the dependent and independent variable was evaluated through significance testing of the regression coefficients, where $|t| \leq 2$ was considered significant.

Behavioural measures. During the SART, we collected behavioural, pupillometric and categorical variables. The three performance variables used in the analyses were *no-go* accuracy, RT and RT variability. *No-go* accuracy was a binomial variable (1 = correct, 0 = incorrect), describing participants' performance on the *no-go* trial immediately preceding each thought probe. As can be seen from Figure 4, this was the only *no-go* trial in the interval between each concern manipulation and subsequent thought probe. RT was measured as the response time in ms on all trials between each concern manipulation and subsequent thought probe. To measure the RT variability, we used the so-called reaction time coefficient of variability (RTCV), which in the past has shown to reflect mind-wandering (van Vugt & Broers, 2016). The RTCV is calculated by taking the standard deviation of a series of RTs and dividing it by the mean of these same RTs. In the current experiment, this series of RTs consisted of any RTs that were recorded between each concern manipulation and subsequent thought probe. Thought probes with less than three RTs in the interval of interest (n = 2), were excluded from analysis.

Pupillometric measures. Before the eye-tracking data could be analysed, they needed to be preprocessed. First of all, any sample-to-sample jumps larger than 0.05 standard deviations (SD) were marked as artefacts and removed from the dataset. Subsequently, all of these missing data points - as well as a 100-ms window around them - were linearly interpolated. The final dataset contained between 2.6 and 28.1 percent of these interpolated data points per subject (M = 14.4). After this artefact correction, the data were down-sampled to 100 Hz, using the median of each 10-ms interval, and divided into stimulus-locked epochs of -500 to 2000 ms. In order to eliminate individual differences, these segmented data were then z-transformed within participants. Next, the baseline pupil diameter was calculated as the average diameter on the 500 ms before each stimulus. This baseline was both used as the dependent measure in the analyses of tonic pupil size and for baseline correction of the phasic pupil data. After baseline correction, phasic pupil size was calculated as the maximum pupil diameter in the 2000 ms post-stimulus.

Categorical variables. Apart from these continuouse measures, the three categorical variables *Concern condition*, *Thought category* and *Stickiness level* were included in the analyses. *Concern condition* indicated the type of concern that was presented just before the trials of interest. This could either be an *Own concern*, an *Other concern* or *No concern. Thought category* and *Stickiness level* represented the answer given on the first and second thought probe question, respectively. In the order presented in Figure 2, the thought categories were called *On-task*, *Task interference*, *Concerns*, *External distraction*, *Mind-wandering* and *Blank*. The stickiness levels were labelled *Very sticky*, *Sticky*, *Neutral*, *Non-sticky* and *Very non-sticky*.

Because we could not control the number of times each response alternative was chosen, some categories ended up under-represented in the data. In fact, four of the thought categories were selected so infrequently (mean frequency ≤ 5.2; see Appendix Figure 1), that we decided to combine them into larger categories. Because the thought probe category *Concerns* is actually a special case of mind-wandering and because there were no major differences between the categories *Concerns* and *Mind-wandering*, these two response alternatives were merged into the overarching category *Internal distraction*. Likewise, the two infrequently chosen, off-task alternatives *Blank* and *External distraction* were combined to make the overarching category *Other distraction*. For the *Stickiness levels*, too, binning was necessary to compensate for infrequently selected response alternatives (mean frequency ≤ 4.5; see Appendix Figure 2). Consequently, the low-frequency option *Very non-sticky* was combined with *Non-sticky* to make a larger category that was also called *Non-sticky*. Similarly, the infrequently chosen alternative *Very sticky* was merged with *Sticky* to make the overarching category *Sticky*. This aggregation of response alternatives resulted in more data points per category and thus higher statistical power. For this reason, we used the aggregated categories for statistical analysis.

Results

Behavioural results

Thought probes and stickiness reports. First of all, we wanted to study the effect of the presented concerns on the types of thoughts people reported. Therefore, we compared the number of times participants chose each thought category in each of the the three concern conditions (Figure 5). The concern conditions had a significant effect on the number of times the *On-task* option was selected ($\chi^2(2) = 12.80$, $p = 0.002$, $BF = 16.63$). *On-task* frequency was higher after *No concerns* ($b = 1.59$, $SE = 0.49$, $t = 3.25$) and *Other concerns* ($b = 1.59$, $SE = 0.49$, $t = 3.25$) than after *Own concerns*,

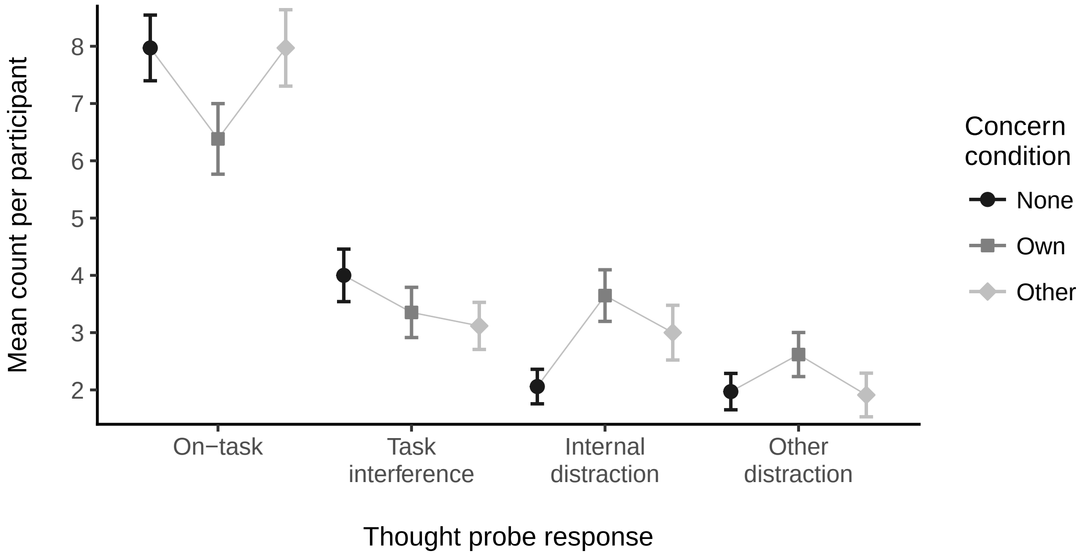

Figure 5. Number of times each thought category was chosen in each concern condition. Error bars represent standard errors of the mean.

indicating less on-task thinking just after a participant's own concerns had been presented. Additionally, the presentation of concerns influenced the number of times participants were distracted, both by their own thoughts ($\chi^2(2) = 14.54$, $p < 0.001$, $BF = 35.25$) and by other things ($\chi^2(2) = 6.41$, $p = 0.04$, $BF = 1.17$). Specifically, people reported less *Internal distraction* after *No concerns* than after a participant's *Own concerns*, pointing towards higher mind-wandering frequencies after the presentation of concerns. Similarly, less *Other distraction* was reported after the presentation of *No concerns* ($b = -0.65$, $SE = 0.30$, $t = -2.14$) and *Other concerns* ($b = -0.71$, $SE = 0.30$, $t = -2.34$), than after *Own concerns*. This also suggests that people were more regularly distracted when their own concerns were presented, compared to when when no concerns or someone else's concerns were presented. For *Task interference*, on the other hand, we found no evidence of a relationship with the presented concerns ($\chi^2(2) = 4.94$, $p = 0.08$, $BF = 0.65$). However, we also found no evidence of the absence of a relationship. Hence, the current data are indecisive about whether the frequency of *Task interference* was influenced by the presentation of concerns or not.

Subsequently, we investigated the effect of the concerns manipulation on the

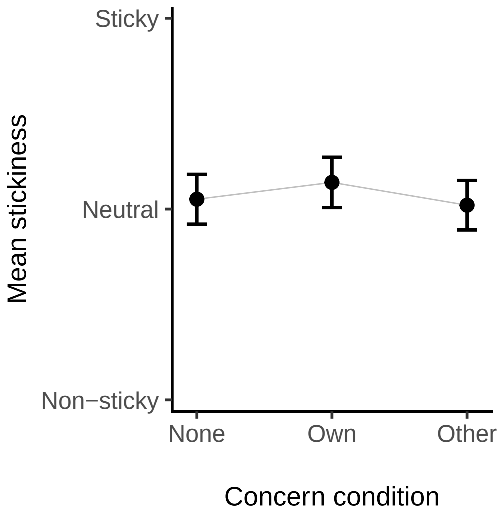

Figure 6. Mean stickiness in each concern condition. Error bars represent standard errors of the mean.

stickiness of participants' ongoing thoughts, by comparing mean stickiness across the concern conditions (Figure 6). The presence of an overall effect of concern condition on stickiness was debatable ($\chi^2(2) = 9.29$, $p = 0.01$, $BF = 0.68$). Even so, it was found that stickiness was slightly lower after the presentation of *No concerns* ($b = -0.09$, $SE = 0.04$, $t = -2.17$) and *Other concerns* ($b = -0.12$, $SE = 0.04$, $t = -2.94$), than after someone's *Own concerns*. This indicated that people's thoughts were a bit more difficult to disengage from when their own concerns were presented, compared to when someone else's or no concerns were presented.

In order to evaluate whether the types of thoughts people had were associated with different levels of reported stickiness, mean stickiness was compared across the different thought categories. Indeed, there was a strong association between the a thought's category and its stickiness ($\chi^2(3) = 215.59$, $p < 0.001$, $BF > 1000$). As visible in Figure 7, stickiness ratings were lower when people were on-task, relative to when they were internally distracted ($b = -0.49$, $SE = 0.05$, $t = -10.55$). This indicates that on-task thought is perceived as easier to disengage from than internally generated distractions.

Performance measures.

NoGo accuracy. The first performance measure to be analysed was *NoGo accuracy*, which reflects how effectively people can withhold a response when they

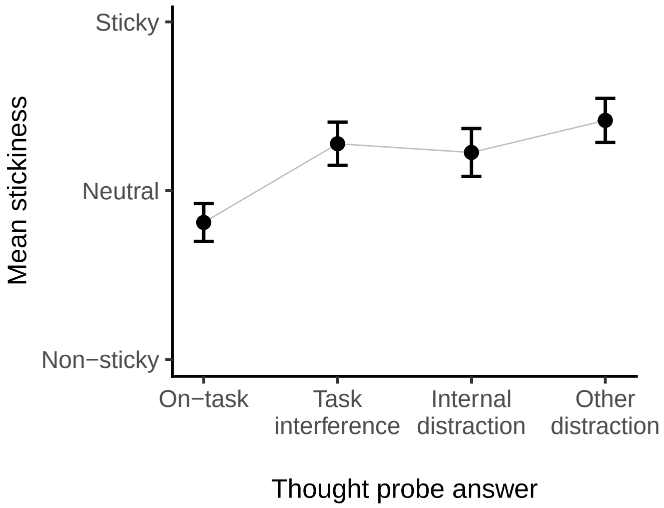

Figure 7. Mean stickiness of each thought category. Error bars represent standard errors of the mean.

encounter an infrequent *no-go* stimulus. To assess whether this accuracy measure was associated with the category of their ongoing thoughts, *no-go* accuracy was compared across the different thought categories. A strong relationship between *no-go* accuracy and thought category was present ($\chi^2(3) = 229.94$, $p < 0.001$, $BF > 1000$). As one might expect, *NoGo accuracy* was higher for the thought category *On-task* than for *Internal distraction* ($b = 0.31$, $SE = 0.03$, $t = 10.36$), indicating that people are better at withholding a response when they are concentrating on the task. No differences in accuracy were found for the other thought categories.

To determine whether and how *no-go* accuracy was affected by the concerns manipulation, we compared its mean value in each of the concern conditions (Figure 8). Even though an overall influence of concern condition on *no-go* accuracy was disputable ($\chi^2(2) = 10.86$, $p = 0.004$, $BF = 1.49$), a small effect was found. Accuracy was slightly higher after *No concerns* than after someone's *Own concerns* ($b = 0.08$, $SE = 0.03$, $t = 3.07$). No such differences were found between *Other concerns* and *Own concerns*. Together, these results indicate that the presentation of any concerns results in less accurate performance on a SART, irrespective of whose concerns they are.

Finally, it was examined whether people make more mistakes when they have more difficulty disengaging from their ongoing thoughts. To test this prediction, we

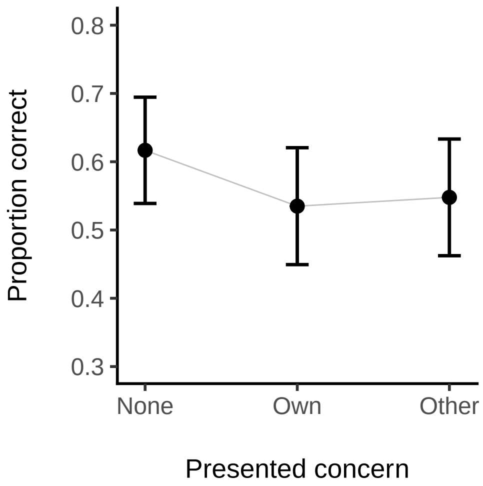

Figure 8. *No-go* accuracy by concern condition. Error bars represent standard errors of the mean.

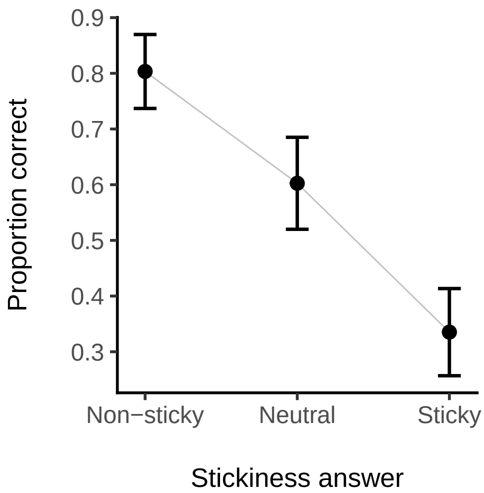

Figure 9. *No-go* accuracy by stickiness level. Error bars represent standard errors of the mean.

compared participants' mean accuracy across the three stickiness categories (Figure 9). There appeared to be a strong relationship between *no-go* accuracy and stickiness ($\chi^2(2) = 188.57$, $p < 0.001$, $BF > 1000$). We found a significant linear relationship, such that *no-go* accuracy decreased as stickiness increased ($b = -0.30$, $SE = 0.02$, $t = -13.91$). This finding indicates that people indeed seem to make more mistakes when their thoughts are more difficult to disengage from.

Response time. To find out whether RT differed across different types of thoughts, we compared mean RT across the different thought categories. Although it is disputable whether any real effects were present ($\chi^2(3) = 11.13$, $p = 0.01$, $BF = 0.06$), RT was found to be marginally longer when people reported being *On-task* ($b = 10.41$, $SE = 3.88$, $t = 2.68$) than when they were internally distracted. Possibly, this could reflect a shift in participant's speed-accuracy trade-off, such that participants emphasised accuracy, rather than speed, when they were concentrating on the task. This might also explain the complex pattern of RT across the concern conditions (see below).

Next, we compared mean RT was across concern conditions, to examine the effect of the concerns manipulation on RT, (Figure 10). There was a substantial effect of concern condition on RT ($\chi^2(2) = 24.61$, $p < 0.001$, $BF = 352.82$), but the nature of this relationship was unexpected. On the one hand, RTs were found to be shorter after the presentation of *Other concerns*, than after a person's *Own concerns* ($b = -8.20$, $SE = 3.15$, $t = -2.61$), indicating that responses were slower when the presented concerns were the participant's own. Since slower responses are regarded as worse performance, this result supports the hypothesis that the presentation of *Own concerns* results in decreased performance on the SART. However, on the other hand, RTs were

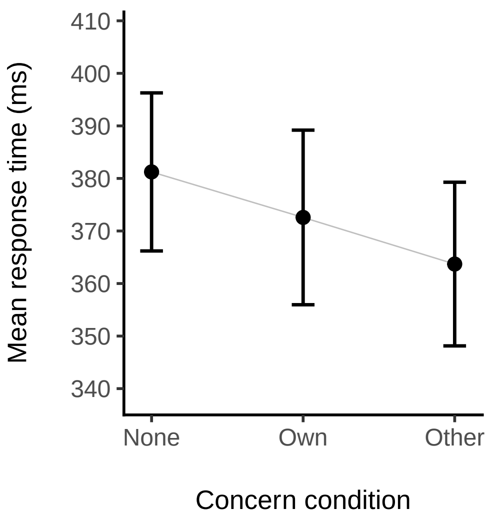

Figure 10. Mean response time by concern condition. Error bars represent standard errors of the mean.

found to be longer after *No concerns* than after *Own concerns*, pointing towards faster responses after the presentation of *Own concerns*. Consequently, this finding seems to contradict our hypothesis of a performance decrease after the presentation of *Own concerns*. It seems, then, that a complex interaction is at work, where the presence of any concerns results in faster responses, but RTs are slowed down again when the concerns are the person's own.

The last RT analysis examined whether this measure could reflect participants' reported stickiness. To answer this question, RT was compared across the different stickiness categories (Figure 11). Although the presence of a general association was debatable ($\chi^2(2) = 8.15$, $p = 0.02$, $BF = 0.13$), a significant linear trend was found ($b = -7.98$, $SE = 2.80$, $t = -2.85$). Specifically, it was found that participants responded faster when their thoughts were stickier. This effect, although weak, may also point towards a shift in people's speed-accuracy trade-off, such that SART performance becomes faster but less accurate when ongoing thoughts are more difficult to disengage from.

Response time variability. Previous studies have suggested that response time variability is a more sensitive measure of mind-wandering than plain RT. Therefore, we decided to include the response time coefficient of variation (RTCV) in

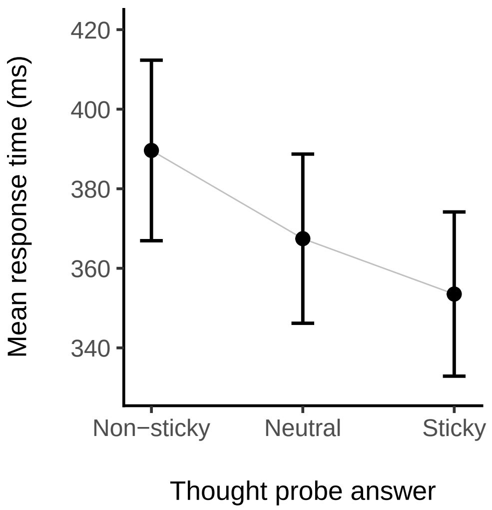

Figure 11. Mean response time by stickiness level. Error bars represent standard errors of the mean.

the analyses as well. To investigate whether RT variability might reflect participants' choice of a thought category, RTCV was analysed across the different thought categories. There appeared to be no relationship between RT variability and self-reported thought content ($\chi^2(3) = 5.84$, $p = 0.12$, $BF = 0.05$).

We then gauged the influence of the concerns manipulation on RT variability, by comparing mean RTCV across concern conditions. A small overall effect of *Concern condition* may have been present ($\chi^2(2) = 8.07$, $p = 0.02$, $BF = 0.40$), but no specific difference seemed to drive this effect. However, looking at Figure 12, RTCV does seem to display the same pattern as RT (Figure 10). As such, it is possible that the same complex interaction visible in RT also applies to RTCV, but that the pattern for RTCV did not reach statistical significance due to a lower number of data points (1 per thought probe for RTCV, compared to 4 for RT).

Last of all, RTCV was compared across stickiness levels, to see whether RT variability and stickiness were linearly related. We found that RTCV held no linear relationship with stickiness ($\chi^2(2) = 1.63$, $p = 0.44$, $BF = 0.03$). Hence, there seems to be no association between RT variability and the degree to which reported thoughts are difficult to disengage from.

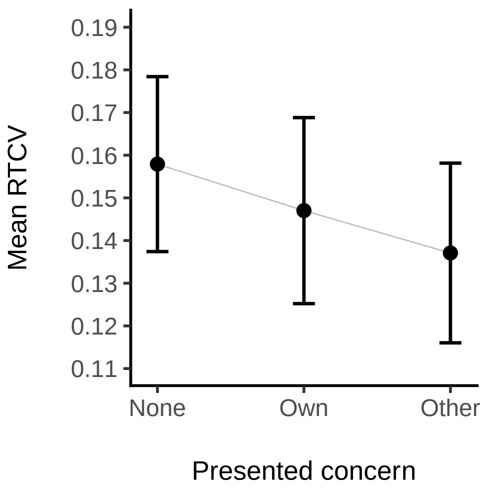

Figure 12. RTCV by concern condition. Error bars represent standard errors of the mean.

Eye-tracking results

Tonic pupil size. We then examined whether mind-wandering and its stickiness were related to pupil size. The first pupillary measure we looked at was the tonic pre-stimulus pupil size, which is thought to indirectly reflect arousal. We first investigated whether participants' tonic pupil size was related to self-reported thought contents. To this end, baseline pupil diameter was compared across the different thought probe responses (Figure 13). Indeed, there was a strong association $\chi^2(3) = 63.85$, $p < 0.001$, $BF > 1000$). Tonic pupil size was larger when participants chose the options *On-task* ($b = 0.15$, $SE = 0.03$, $t = 4.85$) and *Task interference* ($b = 0.20$, $SE = 0.03$, $t = 5.72$), compared to when they chose *Internal distraction*. Consequently, tonic pupil size seems to be lower when people are mind-wandering than when they are thinking about or concentrating on the task.

To investigate how tonic pupil size was influenced by the presentation of concerns, baseline pupil diameter was analysed across concern conditions. Baseline pupil diameter was strongly affected by the displayed concerns ($\chi^2(2) = 213.39$, $p < 0.001$, $BF > 1000$); Figure 14). Tonic pupil size was higher after the presentation of *No concerns* ($b = 0.38$, $SE = 0.03$, $t = 14.69$) and *Other concerns* ($b = 0.17$, $SE = 0.03$, $t = 6.58$) than after a participant's *Own concerns*. These findings imply that the

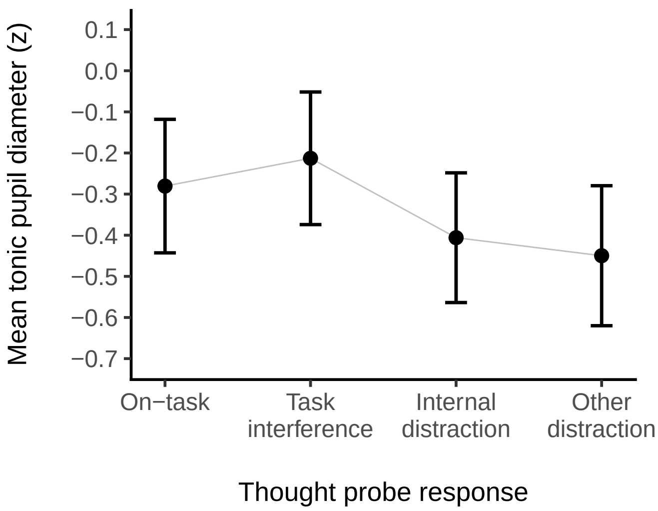

Figure 13. Mean baseline pupil diameter by thought category. Based on the trials between each concern manipulation and subsequent thought probe.

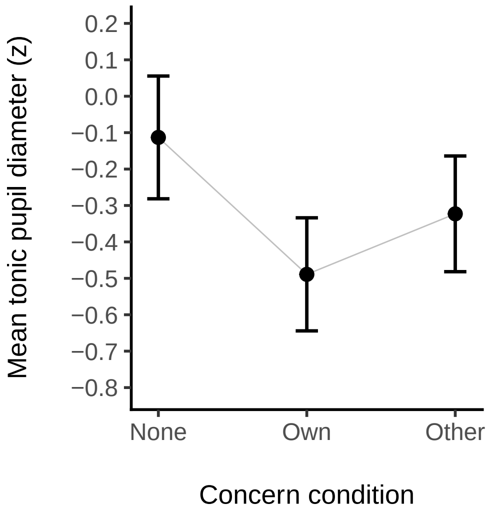

Figure 14. Mean baseline pupil diameter by concern condition. Based on the trials between each concern manipulation and subsequent thought probe.

presentation of a person's own concerns leads to a reduction in tonic pupil size.

The final analysis of tonic pupil size investigated whether this measure was related to the stickiness of participants' ongoing thoughts. Baseline pupil diameter was compared across the different stickiness levels and a strong relationship was found ($\chi^2(2) = 47.37$, $p < 0.001$, $BF > 1000$; Figure 15). A significant linear trend was present in the data, such that higher stickiness ratings were accompanied by lower baseline pupil diameter ($b = -0.15$, $SE = 0.022$, $t = -6.98$). The level of experienced

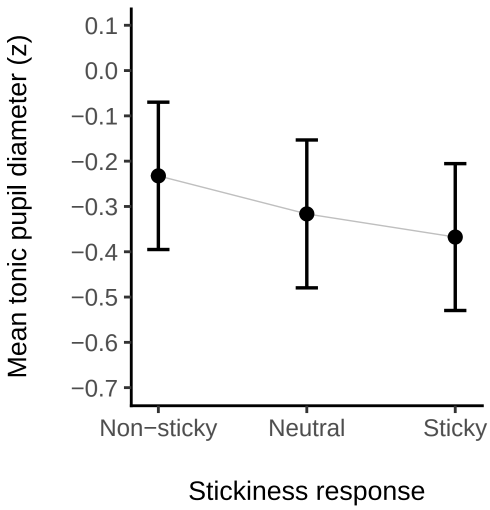

Figure 15. Mean baseline pupil diameter by stickiness level. Based on the trials between each concern manipulation and subsequent thought probe.

difficulty when disengaging from ongoing thoughts thus seems to be reflected in people's tonic pupil size. Since this measure is theorised to reflect arousal, the present finding may indicate that people experience more difficulty disengaging from their current thoughts when they are under-aroused.

Phasic pupil size. The next eye-tracking measure to be analysed was the phasic, stimulus-evoked pupil response. Phasic pupil size is theorised to reflect processing of task-relevant stimuli, where the amplitude of the evoked response represents the amount of cognitive effort spent on this process. Because the implications of the stimuli were very different on *go* than on *no-go* trials, phasic pupil size was expected to behave differently during these two types of trials. For this reason, the results for phasic pupil size will be considered separately for *go* and *no-go* trials.

Go trials. First, we look at the stimulus-evoked pupil response during *go* trials. Since the majority of trials were *go* trials, these can be regarded as the baseline of task performance. We started by examining whether people's pupillary response to a stimulus might give an indication of what type of thoughts they are having. To answer this question, peak phasic pupil diameter was compared across the different thought categories. We found that there was no relationship between people's phasic pupil size and the type of their subsequent thoughts ($\chi^2(3) = 7.02$, $p = 0.07$, $BF = 0.01$). These results imply that the mental effort spent on processing *go* stimuli holds no relationship with the kind of thoughts a person is having. As such, phasic pupil size on *go* trials seems incapable of discriminating between mind-wandering and on-task thinking.

To study the influence of the presented concerns on phasic pupil size, peak phasic pupil diameter was compared between the different concern conditions (Figure 16). A large effect was found ($\chi^2(2) = 28.04$, $p < 0.001$, $BF > 1000$), such that phasic pupil diameter was lower when no concern was presented, than when someone's own concerns were presented ($b = -0.07$, $SE = 0.02$, $t = -4.61$). No such differences were found between *Own concerns* and *Other concerns* ($b < 0.001$, $SE = 0.01$, $t = 0.06$). As such, it seems that participants' phasic pupil response to *go* stimuli increased after the presentation of concerns, irrespective of whether these concerns were their own or

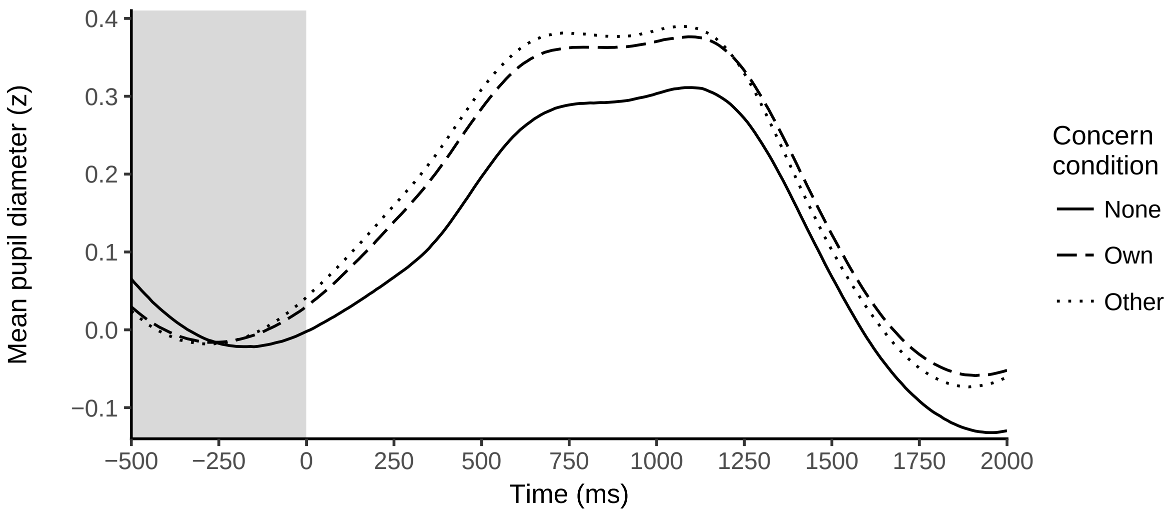

Figure 16. Mean phasic pupil diameter by concern condition. Based on the *go* trials between each concern manipulation and subsequent thought probe. The shaded area was used as the baseline.

someone else's. Since phasic pupil size reflects cognitive effort spent on stimulus processing, it seems that the presentation of a concern triplet makes people spend more effort on the processing of subsequent *go* stimuli. Perhaps this could be attributed to a more reactive - rather than a proactive - response style, invoked by the presented concerns.

Finally, it was assessed whether the persistence of ongoing thoughts might be negatively related to the amount of resources spent on *go* stimulus processing. To investigate this question, phasic pupil diameter was compared across the different stickiness levels. No significant effects were found ($\chi^2(2) = 2.57$, $p = 0.28$, $BF = 0.009$), indicating that the mental effort spent on processing *go* stimuli is unrelated to the stickiness of a person's ongoing thoughts. Consequently, people's phasic pupil response on *go* trials cannot be used as an indicator of a thought's stickiness.

No-go trials. Subsequently, we discuss the results regarding phasic pupil size on *no-go* trials. Because *no-go* stimuli were presented less frequently and required the opposite response of *go* trials, it might be expected that they require more cognitive effort to process correctly. As a result, phasic pupil responses to *no-go* stimuli may be more sensitive to mind-wandering than those to *go* stimuli. The first relationship to be

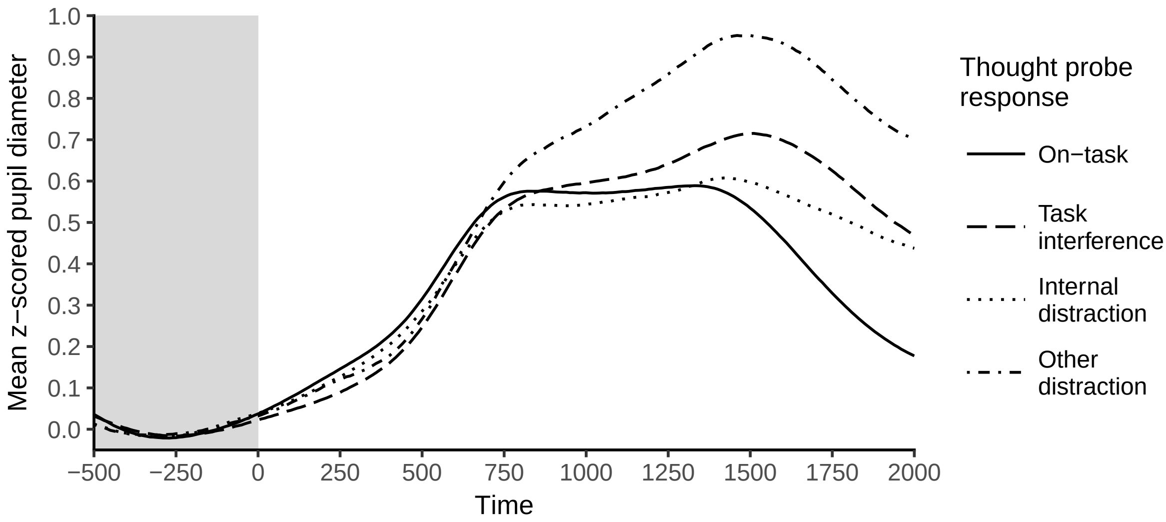

Figure 17. Mean phasic pupil diameter by thought probe answer. Based on the last *no-go* trial before each thought probe. The shaded area was used as the baseline.

investigated, was that between processing intensity of *no-go* stimuli and the type of thoughts people had. To test whether such a relationship might exist, peak phasic pupil diameter was compared across the different thought categories (Figure 17). Indeed, there was a large association between phasic pupil size and subsequent thought categories ($\chi^2(3) = 28.55$, $p < 0.001$, $BF < 1000$). Peak pupil size was significantly higher for the *Other distraction* category than for *Internal distraction* ($b = 0.18$, $SE = 0.06$, $t = 3.00$), but no differences were found between mind-wandering and on-task thinking ($b = -0.08$, $SE = 0.05$, $t = -1.58$). Hence, it appears that the phasic pupil response on *no-go* SART trials can discriminate between mind-wandering and other types of distraction, but not between mind-wandering and on-task thought. Moreover, this ability to discriminate between mind-wandering and other distractions seems to arise from increased cognitive effort spent on the processing of *no-go* stimuli during other distractions than mind-wandering. Perhaps, internally generated distractions are more "calculated" than other distractions, resulting in a larger surprise when they detect a *no-go* stimulus.

We then examined whether exerted cognitive effort was influenced by the presented concerns, by comparing peak evoked pupil diameter across concern conditions (Figure 18). There was indeed a strong effect of the presented concerns on phasic pupil

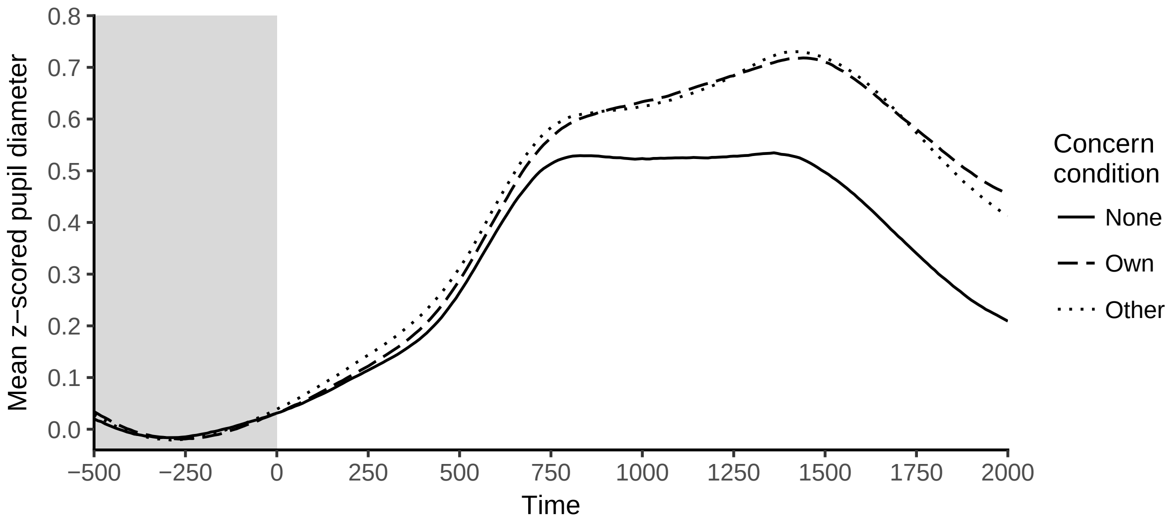

Figure 18. Mean phasic pupil diameter by concern condition. Based on the last *no-go* trial before each thought probe. The shaded area was used as the baseline.

diameter ($\chi^2(2) = 26.84$, $p < 0.001$, $BF > 1000$). As was the case for *go* trials, phasic pupil size was lower when no concerns were presented than when someone's own concerns were presented ($b = -0.17$, $SE = 0.04$, $t = -4.25$). No differences were found between *Own* and *Other concerns* ($b = 0.02$, $SE = 0.04$, $t = 0.47$). The presentation of concerns - either the participant's own or someone else's - thus results in increased processing of *no-go* stimuli. Since mind-wandering frequencies were higher after concern triplets, the found increase in *no-go* stimulus processing may be the result of participants being more surprised by such a stimulus when they were not paying attention to the task. Conversely, participants might have been more prepared for a *no-go* trial when they were on-task.

Last of all, we studied the relationship between *no-go* processing intensity and the stickiness of people's ongoing thoughts. Peak phasic pupil diameter was compared across stickiness alternatives and a large effect was found ($\chi^2(2) = 34.6$, $p < 0.001$, $BF < 1000$; Figure 19). This effect was driven by a strong linear trend, such that participants' peak evoked pupil response increased with increasing stickiness ($b = 0.20$, $SE = 0.03$, $t = 5.91$). These results indicate that people invest more mental effort in processing *no-go* stimuli when their current thoughts are difficult to disengage from. Perhaps, this means that people need to invest more cognitive effort in stimulus

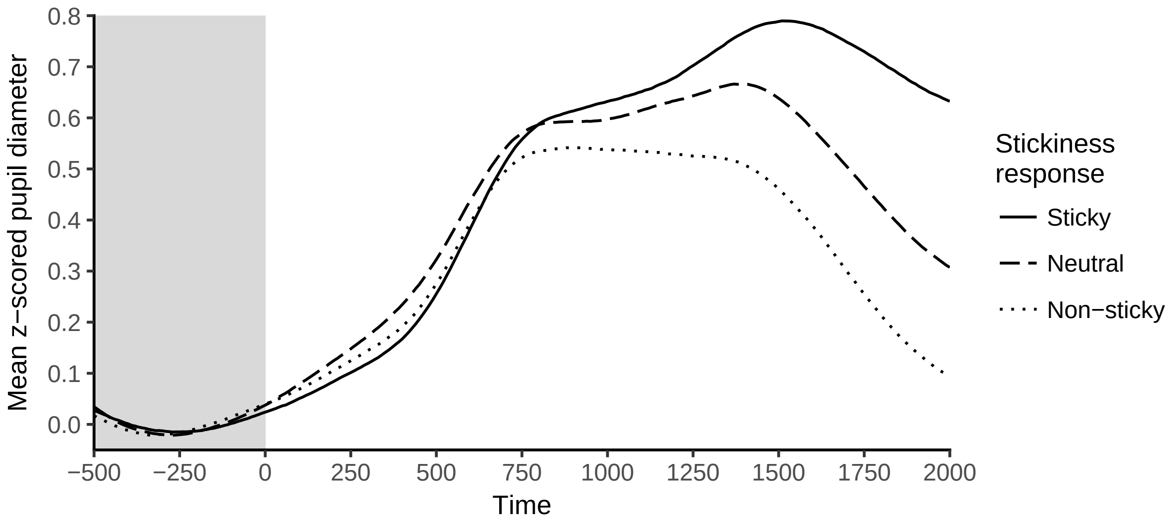

Figure 19. Mean phasic pupil diameter by stickiness level. Based on the last *no-go* trial before each thought probe. The shaded area was used as the baseline.

processing, in order to overcome the difficulty they experience when trying to disengage from their ongoing thoughts. Additionally, these results imply that phasic pupil size on a *no-go* trial could potentially provide a nice indication of an ongoing thought's stickiness.

Discussion

The current experiment investigated the 'stickiness' dimension of ongoing on- and off-task thought, as well as its influence on performance and pupil size. To make the possibility of observing a sticky type of concern-related mind-wandering more likely, we used a current concerns paradigm, which has been shown to increase this type of off-task thought (van Vugt & Broers, 2016; McVay & Kane, 2013). The results of this study are summarised in Table 1. We found that episodes of internal distraction were associated with decreased *no-go* accuracy, faster responses and reduced tonic pupil size. Additionally, we showed that these observed mind-wandering markers, along with self-reported mind-wandering frequency, were present more strongly after participants had been presented with their personal goals and concerns. Finally, like van Vugt and Broers (2016), we demonstrated, that internal distractions were perceived as more difficult to disengage from than on-task thinking. This stickiness of ongoing thoughts

Table 1

Overview of results

Dependent measure	Thought type: On–task	Task interference	Internal distraction	Other distraction	Concern condition: None	Own	Other	Linear relationship stickiness
Thought type (freq.)								
On–task					↑	↓	↑	
Task interference					-	-	-	
Internal distraction					↓	↑	-	
Other distraction					↓	↑	↓	
NoGo accuracy	↑	-	↓	-	↑	↓	-	↓
RT	↑	-	↓	-	↑	↕	↓	↑
RTCV	-	-	-	-	-	-	-	-
Tonic pupil diameter	↑	↑	↓	-	↑	↓	↑	↓
Phasic pupil diameter								
Go trials	-	-	-	-	↓	↑	-	-
NoGo trials	-	-	↓	↑	↓	↑	-	↑
Stickiness	↓	-	↑	-	↓	↑	↓	

↑ Measure increased/positive relationship. ↓ Measure decreased/negative relationship.
↕ Measure both increased and decreased, depending on comparison. - No effects found.

was also increased after people encountered their own concerns and was associated with the decreases in accuracy and tonic pupil diameter that were also found for mind-wandering episodes.

On-task versus off-task thinking

The reductions in accuracy and response time during mind-wandering are consistent with the theory that mind-wandering, in combination with the SART's repetitiveness, leads people to respond in an automatic and stimulus-driven manner (Manly et al., 1999; Manly et al., 2000; Braver, 2012). Additionally, since the assumed function of this so-called reactive control mode is to conserve cognitive resources, these findings also support Smallwood's notion that mind-wandering draws mental resources away from an ongoing task (Smallwood & Schooler, 2006). Thus, it seems that mind-wandering uses limited mental resources and therefore can be detected during task performance, by looking for signs of a reactive control strategy.

Discriminating even more strongly between on- and off-task thought than accuracy and RT, was the measure of tonic pupil size. Specifically, we found that baseline pupil diameter for any type of distraction was lower than for any type of thought about the task. This finding supports Unsworth and Robison (2016)'s theory that the mind-wandering state lies towards the low end of the adaptive gain curve, which is marked by underarousal and inertia. By consequence, the current finding conflicts with Smallwood's idea that mind-wandering is a state of exploration and high arousal that lies towards the high end of the adaptive gain curve. (Franklin et al., 2013; Smallwood et al., 2011; Smallwood et al., 2012). Contrary to our hypotheses, this decreased tonic pupil size was not accompanied by a reduced phasic pupil size.

Current concerns

As can be seen in Table 1, people reported more distraction and less on-task thought after their personal goals and concerns had been presented. They also displayed the decreases in accuracy, RT and tonic pupil size, found to characterise mind-wandering, right after the presentation of people's own concerns. Taken together, these results indicate that the current concerns manipulation successfully invoked increased off-task thought. However, it did not always matter whether the presented concerns were the participant's own or someone else's. Even though we attempted to select concerns that were as unique as possible, some overlap between concerns probably still existed.

The concerns manipulation may also shed some light on the absence of a difference in phasic pupil size between mind-wandering and on-task thought. Assuming that mind-wandering frequency is the primary difference between the trials after own versus no concerns, an increase of the stimulus-evoked pupil response might represent an additional marker of mind-wandering. If so, this finding would be consistent with the notion that people tend to use a reactive control strategy during mind-wandering. Because reactive control involves minimal mental preparation for upcoming stimuli, a lot of cognitive work still needs to be carried out at the moment of stimulus presentation. Since phasic pupil size reflects the amount of processing during stimulus presentation, an increased stimulus-evoked pupil response could also be a marker of a reactive control strategy.

If increased phasic pupil size is indeed an additional marker of off-task thought, the finding that phasic pupil size was increased when people's minds were blank and when they were distracted by their environment, but not when they were mind-wandering, may represent a complex interaction. In that case, any form of distraction might result in a reactive response style and increased phasic pupil size, while specifically internal distractions result in sensory decoupling and reduced phasic pupil size, thus bringing it back to the level of on-task thought.

Sticky thoughts

The increased stickiness for internal distraction is consistent with the view that mind-wandering is a less flexible, more engrossing type of thinking that could be a precursor of rumination (Marchetti et al., 2016; Joormann et al., 2011). This interpretation is further supported by the decrease in accuracy and increase in phasic pupil size observed for sticky thought. These two findings are both consistent with the use of a reactive strategy, which in turn provides further evidence that sticky thoughts occupy resources that otherwise could have been used for task performance. However, we did not find the decrease in RT that usually accompanies a reactive strategy in a SART. Instead, we found a puzzling lengthening of response times. Perhaps, sticky thoughts are even more disruptive than 'ordinary' mind-wandering, such that stimuli are not detected on time. This would leave even a reactive strategy useless for producing fast reaction times.

The finding that stickiness of people's thoughts was amplified by presenting their current goals and concerns, supports Klinger's theory that sticky thoughts are usually concerned with planning and evaluating personal goals (Klinger, 2013; Marchetti et al., 2016; McVay & Kane, 2013). Because these thoughts are both functionally and emotionally important to the individual, they tend to linger in the person's working memory. As such, it seems that the concerns manipulation not only stimulates mind-wandering by increasing its frequency, but also by increasing its persistence.

Finally, the observed decrease in tonic pupil diameter with increasing stickiness indicates that the mental state in which sticky thoughts occur, is similar to the one in which mind-wandering takes place. That is, both states are associated with reduced tonic pupil size and thus seem to lie towards the low end of the adaptive gain curve. This further supports Unsworth and Robison (2016)'s theory that the mind-wandering state is one of underarousal and inactivity, rather than one of overarousal and distractability (Franklin et al., 2013; Smallwood et al., 2011; Smallwood et al., 2012).

Implications for future studies

Although most results for accuracy, RT and pupil size neatly fit into the theory that mind-wandering evokes a reactive control mode, we did not find the expected increase in RT variability that is usually also associated with distraction and decreased SART performance (Manly et al., 2000; Bastian & Sackur, 2013; Cheyne et al., 2009; Mrazek et al., 2012). This probably cannot be explained by a lack of statistical power, since most Bayes factors were quite low ($\leq \frac{1}{20}$), implying that we did have enough power, but simply found more evidence for the absence of any differences in RT variability. One possible explanation might be that a window of 3-5 trials is not enough to reliably estimate RT variability, since most previous studies used larger intervals. This poses a trade-off for future mind-wandering studies to take into account: larger intervals render more reliable RTCVs, but the responses given during this interval are further removed from their corresponding thought probe.

In the present study, we came across the problem of overlapping own and other concerns, as mentioned before by McVay and Kane (2013). The current results show that this problem might not be as severe as initially feared, since we did find differences between the two different types of concerns in RT, tonic pupil size, stickiness and two thought type frequencies. However, for mind-wandering frequency, no-go accuracy and phasic pupil size, it remains unclear whether the effects of the presented concerns should be attributed to their content or their physical properties. For example, meaningfulness and presentation frequency of words have been found to affect response times (Jolsvai, McCauley, & Christiansen, 2013; Craik & Tulving, 1975). Since the concern triplets were more meaningful and were presented more frequently than the random filler words, these factors may have impacted results for the concern triplets. Consequently, it would be advisable for future concern studies to use *control triplets*, like McVay and Kane (2013) did. These control triplets did not contain any concerns, but they did have a clear meaning and were presented as frequently as the triplets with participants' own concerns. Ideally, such control concerns would closely match the concern triplets in all respects, except for the personal meaning to the participant.

To our knowledge, the present study is the first to combine thought probes and pupil size measurements in a SART. This has resulted in some interesting theoretical discoveries, but it has some very practical implications as well. First of all, we found that the stimulus-evoked pupil response in a SART is more informative on *no-go* than that on *go* trials. Perhaps, this originates from the same ceiling effect that has led to the convention of reporting accuracy for no-go trials only. Moreover, our results show that tonic pupil size is a powerful way to discriminate between on- and off-task thought. Because pupil size can be measured continuously during task performance, this provides a first step towards the online detection of mind-wandering. Additionally, since phasic pupil size is apparently capable of discriminating between mind-wandering on the one hand and external distractions and "zone-outs" on the other, this measure could potentially be used to distinguish mind-wandering from other types of covert distraction during task performance. In order to enable the real-life use of both tonic and phasic pupil size as detectors of mind-wandering, mechanisms need to be developed that can identify the presently found mind-wandering markers on a trial-by-trial basis. Fortunately, there are machine-learning algorithms, called classifiers, that are very suitable for this job and that are already beginning to be used for it (Franklin et al., 2011; Mittner et al., 2014).

Taken together, most findings in the present study point towards the use of a more reactive response style during episodes of mind-wandering. However, the findings reported here could be specific to the SART and may not be generalisable to other tasks and real-life situations. Therefore, this theory needs to be tested in more different tasks and contexts. This would help us understand when mind-wandering happens and how it can be recognised. One way to test whether mind-wandering indeed leads to a reactive response style, would be to study it in a task that was specifically designed to identify reactive versus proactive strategies. A good example of such a task is the AX continuous performance task (Braver, 2012). If thought probes and perhaps pupil size measurements would be included in such a task, we could say with more confidence whether mind-wandering leads to a reactive control mode in other contexts as well.

Conclusion

In the current experiment, we found that mind-wandering was marked by reductions in *no-go* accuracy, response times and tonic pupil size. These overt markers can be used for the online detection of mind-wandering in future SART studies and imply that off-task thought may be recognised in other contexts as well, by looking for signals of a reactive control strategy. We also found that presenting people's current goals and concerns results in similar decreases in accuracy, RT and tonic pupil size as were found during mind-wandering. Together with self-reports of increased mind-wandering and decreased on-task thought, these observations suggest that ongoing goals and concerns play an important role in the emergence of mind-wandering and can be used to increase mind-wandering frequencies. Finally, the current study indicates that mind-wandering thoughts and thoughts after presented concerns are more difficult to disengage from than on-task thought. This so-called 'stickiness' dimension of thoughts was accompanied by lower accuracy and larger phasic pupil size, which are both markers of a reactive control mode. As such, it seems that mind-wandering, current concerns and sticky thoughts are all marked by a reactive control mode, implying that they all require cognitive resources that then cannot be used for proactive task performance.

References

Aston-Jones, G. & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci. 28*, 403–450.

Bastian, M. & Sackur, J. (2013). Mind wandering at the fingertips: Automatic parsing of subjective states based on response time variability. *Frontiers in Psychology*, *4*, 573.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Bates, D., Maechler, M., Bolker, B., Walker, S., et al. (2014). Lme4: linear mixed-effects models using eigen and s4. *R package version*, *1*(7).

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in cognitive sciences*, *16*(2), 106–113.

Carver, C. S. & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the bis/bas scales. *Journal of personality and social psychology*, *67*(2), 319.

Cheyne, J. A., Solman, G. J., Carriere, J. S., & Smilek, D. (2009). Anatomy of an error: A bidirectional state model of task engagement/disengagement and attention-related errors. *Cognition*, *111*(1), 98–113.

Cox, W. M. & Klinger, E. (2004). *Handbook of motivational counseling: concepts, approaches, and assessment.* John Wiley & Sons.

Craik, F. I. & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of experimental Psychology: general*, *104*(3), 268.

Foulsham, T., Farley, J., & Kingstone, A. (2013). Mind wandering in sentence reading: Decoupling the link between mind and eye. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *67*(1), 51.

Franklin, M. S., Broadway, J. M., Mrazek, M. D., Smallwood, J., & Schooler, J. W. (2013). Window to the wandering mind: Pupillometry of spontaneous thought while reading. *The Quarterly Journal of Experimental Psychology*, *66*(12), 2289–2294.

Franklin, M. S., Smallwood, J., & Schooler, J. W. (2011). Catching the mind in flight: Using behavioral indices to detect mindless reading in real time. *Psychonomic Bulletin & Review*, *18*(5), 992–997.

Grandchamp, R., Braboszcz, C., & Delorme, A. (2014). Oculometric variations during mind wandering. *Frontiers in psychology*, *5*, 31.

Jolsvai, H., McCauley, S. M., & Christiansen, M. H. (2013). Meaning overrides frequency in idiomatic and compositional multiword chunks. In *Cogsci*.

Joormann, J., Levens, S. M., & Gotlib, I. H. (2011). Sticky thoughts: Depression and rumination are associated with difficulties manipulating emotional material in working memory. *Psychological science*, *22*(8), 979–983.

Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron*, *89*(1), 221–234.

Keuleers, E., Brysbaert, M., & New, B. (2016). Subtlex-nl: a new measure for dutch word frequency based on film subtitles. *Behavior research methods*, *42*(3), 643–650.

Killingsworth, M. A. & Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science*, *330*(6006), 932–932.

Klinger, E. (2013). Goal commitments and the content of thoughts and dreams: Basic principles. *Frontiers in Psychology*, *4*(415), 10–3389.

Manly, T., Davison, B., Heutink, J., Galloway, M., & Robertson, I. H. (2000). Not enough time or not enough attention? speed, error and self-maintained control in the sustained attention to response test (sart). *Clinical Neuropsychological Assessment*, *3*(10).

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

Marchetti, I., Koster, E., Klinger, E., & Alloy, L. B. (2016). Spontaneous thought and vulnerability to mood disorders: The dark side of the wandering mind. *Clinical Psychological Science*.

McVay, J. C. & Kane, M. J. (2009). Conducting the train of thought: working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 196.

McVay, J. C. & Kane, M. J. (2013). Dispatching the wandering mind? Toward a laboratory method for cuing "spontaneous" off-task thought. *Frontiers in Psychology*, *4*, 570.

Mittner, M., Boekel, W., Tucker, A. M., Turner, B. M., Heathcote, A., & Forstmann, B. U. (2014). When the brain takes a break: A model-based analysis of mind wandering. *The Journal of Neuroscience*, *34*(49), 16286–16295.

Mooneyham, B. W. & Schooler, J. W. (2013). The costs and benefits of mind-wandering: A review. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *67*(1), 11.

Morey, R. D. & Rouder, J. N. (2015). *Bayesfactor: computation of bayes factors for common designs*. R package version 0.9.12-2. Retrieved from https://CRAN.R-project.org/package=BayesFactor

Mrazek, M. D., Smallwood, J., & Schooler, J. W. (2012). Mindfulness and mind-wandering: Finding convergence through opposing constructs. *Emotion*, *12*(3), 442.

Murphy, P. R., Robertson, I. H., Balsters, J. H., & O'Connell, R. G. (2011). Pupillometry and p3 index the locus coeruleus–noradrenergic arousal function in humans. *Psychophysiology*, *48*(11), 1532–1543.

Peirce, J. W. (2007). Psychopy: Psychophysics software in python. *Journal of neuroscience methods*, *162*(1), 8–13.

R Core Team. (2016). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/

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

Smallwood, J., Brown, K. S., Baird, B., Mrazek, M. D., Franklin, M. S., & Schooler, J. W. (2012). Insulation for daydreams. *PloS one*, *7*(4), e33706.

Smallwood, J., Brown, K. S., Tipper, C., Giesbrecht, B., Franklin, M. S., Mrazek, M. D., … Schooler, J. W. (2011). Pupillometric evidence for the decoupling of attention from perceptual input during offline thought. *PloS one*, *6*(3), e18298.

Smallwood, J., Davies, J. B., Heim, D., Finnigan, F., Sudberry, M., O'Connor, R., & Obonsawin, M. (2004). Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness and cognition*, *13*(4), 657–690.

Smallwood, J. & Schooler, J. W. (2006). The restless mind. *Psychological bulletin*, *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.

Unsworth, N. & Robison, M. K. (2016). Pupillary correlates of lapses of sustained attention. *Cognitive, Affective, & Behavioral Neuroscience*, 1–15.

van Vugt, M. K. & Broers, N. (2016). Self-reported stickiness of mind-wandering affects task performance. *Frontiers in psychology*, *7*(732). doi:10.3389/fpsyg.2016.00732

Verplanken, B., Friborg, O., Wang, C. E., Trafimow, D., & Woolf, K. (2007). Mental habits: Metacognitive reflection on negative self-thinking. *Journal of Personality and Social Psychology*, *92*(3), 526.

Appendix

Frequencies of the original thought categories and stickiness levels

The Figures below display for each of the thought probe answer alternatives how often it was selected per participant. The depicted alternatives are the original ones, before they were combined into larger categories. Figure 1 shows the frequencies for the type of thoughts people had, whereas Figure 2 shows the frequencies for the different stickiness levels.

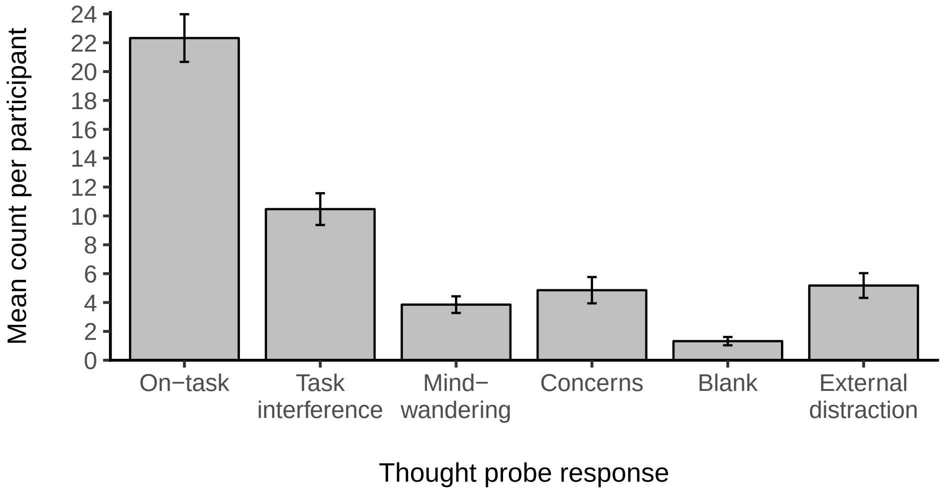

Figure 1. Mean count per participant of the non-binned thought probe alternatives. Error bars represent standard errors of the mean.

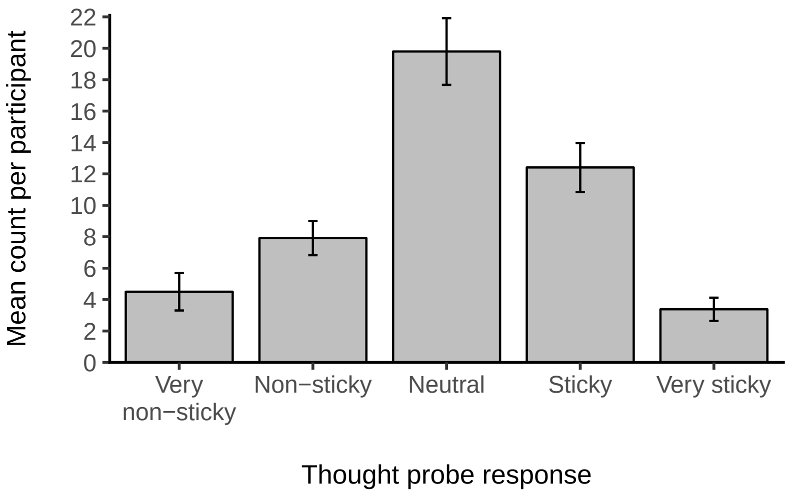

Figure 2. Mean count per participant of the non-binned stickiness alternatives. Error bars represent standard errors of the mean.