**When attention wanders: pupillometric signatures of fluctuations in external attention**

Mahiko Konishi1, Kevin Brown2,3, Luca Battaglini4, and Jonathan Smallwood1

1 Department of Psychology, University of York, Heslington, England, United Kingdom.

2 Departments of Biomedical Engineering, Physics, Chemical and Biomolecular Engineering and Marine Sciences, University of Connecticut, Storrs, CT, United States of America

3 Connecticut Institute for the Brain and Cognitive Sciences, CT, USA

4 Department General Psychology, University of Padua, Padua, Italy

Address for correspondence:

Mahiko Konishi, Department of Psychology, University of York, Heslington, England.

Email: konishi@gmail.com.

**Abstract**

Attention is not always directed to events in the external environment. On occasion our thoughts wander to people and places distant from the here and now. Sometimes, this lack of external attention can compromise ongoing task performance. In the current study we set out to understand the extent to which states of internal and external attention can be determined using pupillometry as an index of ongoing cognition. In two experiments we found that periods of slow responding were associated with elevations in the baseline pupil signal over the three and a half seconds prior to a behavioural response. In the second experiment we found that unlike behavioural lapses, states of off-task thought, particularly those associated with a focus on the past and with an intrusive quality, were associated with reductions in the size of the pupil over the same window prior to the probe. These data show that both states of large and small baseline pupil size are linked to states when attention is not effectively focused on the external environment, although these states have different qualities. More generally, these findings illustrate that subjective and objective markers of task performance may not be equivalent and underscore the importance of developing objective indicators that can allow these different states to be understood.

**Keywords:** mind-wandering; pupillometry; task focus; baseline pupil; self-reports.

1. **Introduction**

Attention is not always focused on the external environment; experiences like mind-wandering and daydreaming illustrate situations when cognition is generated based on our factual knowledge of the world, and episodic memories about the people we know and the places we have visited over the course of our lives (Smallwood, 2013; Smallwood et al., 2016a; Smallwood & Schooler, 2015). Although we now know that these experiences make contributions to our well-being (Killingsworth & Gilbert, 2010; Poerio et al., 2013), can arise either intentionally or spontaneously (Seli et al., 2016), and can compromise ongoing performance (Mcvay & Kane, 2009; Smallwood et al., 2008), the intrinsic nature of these experiences has hindered our capacity to understand their contributions to the human condition.

One barrier to the investigation of self-generated states is a reliance on measures of self-report. Although introspective evidence allows the internal landscape of personal experience to be described, and participants have been shown to be reliable assessors of their task focus (Mittner et al., 2016; Seli et al., 2015), the requirement that participants must explicitly reflect on the contents of their experience makes it possible that results that are generated in this fashion may alter the nature of the experiences that are being investigated (Konishi & Smallwood, 2016). One way to understand, and ultimately overcome, these issues, is through the development of an indirect marker that could be used as a proximal measure for the occurrence of self-generated thoughts. The current study attempts to address this issue using pupillometry as a covert marker for ongoing cognitive processing.

Prior studies have found that when the baseline diameter of pupils is unusually small or large, attention is not always effectively focused on the external environment. For example, momentary lapses in attention, as indexed by slow response times or errors in performance, are preceded by periods of both large and small baseline pupil size (Gilzenrat et al., 2010; Smallwood et al., 2011, 2012; Van Orden et al., 2000). A similar pattern has been observed across studies of mind-wandering, with some finding increased pupil diameter co-occurring with self-reports of off task/mind-wandering episodes (Franklin et al., 2013), while others have found the reverse (Grandchamp et al., 2014). A more recent study (Unsworth & Robison, 2016), which differentiated between types of off task states, found increased baseline pupil size before reports of external distraction and reduced pupil size before both reports of mind wandering episodes and inattentiveness. As it is widely accepted that pupillometry provides an indirect measure of arousal (Aston-Jones & Cohen, 2005; Morad et al., 2000; Murphy et al., 2011; Stanners et al., 1979; Wilhelm et al., 1998; Yoss et al., 1970) it seems plausible that states of optimal focus may be indicated by moderate levels of arousal, with extremely large or small pupils indicating situations when attention is not engaged with the external environment to the same degree (Aston-Jones & Cohen, 2005).

It has been suggested that understanding the relationship between self-generated thought and other aspects of neurocognitive functioning depends upon taking account of both the content of the experiences and the task context within which they take place (Smallwood & Andrews-Hanna, 2013). For example, studies have found that when experience is focused on events from the past, this is often associated with lower levels of happiness (Poerio et al., 2013; Ruby et al., 2013; Smallwood & O’Connor, 2011). By contrast, thoughts about the future have been linked to reductions in levels of social stress (Engert et al., 2014) and may contribute to the processes through which people consolidate personal goals (Medea et al., 2016). Neurocognitive investigations have also highlighted differences between these classes of experiences. Self-generated thoughts about the past were linked to higher connectivity between lateral temporal lobe regions and the hippocampus, reflecting the heightened role of episodic memory when we retrospect on the past, and relatively greater decoupling between medial prefrontal cortex and medial visual cortex than for individuals who tend to think more about the future (Smallwood et al., 2016a). Together, these observations support the content-regulation hypothesis that the *content* of self-generated thought in part determines its relationship to other neurocognitive measures.

As well as taking into account the content of self-generated thought, it is important to consider the *context* in which self-generated thought occurs. Studies have shown that although executive control measures tend to predict lower levels of off-task thought when tasks are complex (Mcvay & Kane, 2009; Unsworth & Robison, 2016) the relationship can often reverse when tasks are less demanding (Bernhardt et al., 2014; Kane et al., 2007; Levinson et al., 2012; Rummel & Boywitt, 2014; Smallwood et al., 2013). Indeed, task demands modulate different types of off-task thought, such as intentional vs. unintentional mind wandering (Seli et al., 2016; Seli et al., 2016b), and also the relationship between mind wandering and task performance (Thomson et al., 2014). More generally, self-generated thought is more common during tasks that lack complex demands (Ruby et al., 2013; Smallwood et al., 2009; Teasdale et al., 1993), presumably because there is a greater availability of cognitive resources to devote to self-generated thought. Together these lines of evidence suggest that understanding the context in which self-generated occurs can be important in understanding its neurocognitive basis.

The current study aims to elucidate the relationship between baseline pupil size and the extent to which attention is deployed to the external environment. We measured baseline pupil size in the context of a paradigm in which we manipulated the degree of external task focus by means of the addition of a working memory load (see Figure 1). We have previously used this paradigm to vary the amount of attention that participants devote to an ongoing task, a manipulation that is reflected in the speed and effectiveness with which decisions are made, as well as in changes in reports of task focus. In a prior study we acquired functional magnetic resonance imaging data during this task and found that performance of the easy task is accompanied by greater engagement of regions of the default mode network (Konishi et al., 2015) – a neural system important in self-generated thought (Allen et al., 2013; Christoff et al., 2009; Mason et al., 2007; Stawarczyk et al., 2011).

Using this paradigm, we conducted two experiments on healthy participants in which we measured baseline pupil size while they performed alternating blocks of the 0-back and 1-back versions of the task. We acquired two different indicators of the focus of attention. In Experiment 1 we acquired measures of behavioural task performance, and in Experiment 2 we also measured the content of ongoing thought using Multi-Dimensional Experience Sampling (MDES; Karapanagiotidis et al., 2016; Medea et al., 2016; Smallwood et al., 2016b; Smallwood & Schooler, 2015). We measured both subjective and objective indicators of attention to explore whether they had the same signature in terms of baseline pupil size. Our motivation for only measuring subjective indicators of attention in the second experiment was to address the concern that the act of introspecting on experience would alter the nature of any pupil-behaviour relationships observed in the first experiment. Using these two experiments we aimed to identify the relationship between pupil size and both objective and subjective indicators of attention.

1. **Method**

2.1 Task Paradigm

The task used in both studies was programmed using PsychoPy2 (Peirce, 2007, 2008). The task featured a 0-back and a 1-back condition that continuously switched from one another throughout the experimental session (see Figure 1). Our paradigm is the same used in Konishi et al. (2015). In both conditions participants saw different pairs of shapes (Non-Targets, NT) appearing on the screen divided by a vertical line. The shape pairs could be: a circle and a square, a circle and a triangle, or a square and a triangle for a total of 6 possible pairs (two different left/right configurations for each). The pairs never had shapes of the same kind (e.g. a square and a square). In both tasks, a block of NT was followed by a target requiring participants to make a manual response. The target was a small stimulus presented in either blue or red. In the 0-back condition the target was flanked by one of two shapes and participants had to indicate by pressing the appropriate button which shape matched the target shape. In the 1-back condition, the target was flanked by two question marks and participants had to respond using the left and right arrow keys, depending on which side the target shape was on the previous trial.

Each block lasted between 40 to 120 seconds before switching to the other condition; the change of condition was signalled by a message (“SWITCH”) that remained on screen for 4 seconds. The number of NTs preceding the targets varied between 2 and 6, the number of targets per block varied between 2 and 5 and the total number of blocks was 8 for each condition. The total number of targets was 15 to 20 per condition. The blue and red colours used in the two conditions were matched for luminance as displayed on an LCD monitor, and the colour/condition pairing was counterbalanced across participants. The starting order of conditions was counterbalanced across participants and a single run of the task lasted ~15 minutes for Experiment 1 and ~20 minutes for Experiment 2 (due to the addition of thought probes). Each participant completed two runs for study 1, and two to six runs in the span of three different days (two sessions a day) for study 2. Presentation rate of the stimuli was jittered in the following way: fixation crosses ranged from 1.8-2.2 seconds (average 2s) in steps of 0.05s, Non-targets were varied from 1.3–1.7 seconds (average 1.5s) in steps of 0.05s. Target durations ranged from 2.1-2.5 seconds (average 2.3s) in steps of 0.05s and lasted for the full duration, regardless of participant response.



2.2 Multi-Dimensional Experience Sampling (Experiment 2)

In order to sample participants’ ongoing experiences in Experiment 2, we used a probe-caught, experience sampling method (Kahneman et al., 2004; Smallwood & Schooler, 2006). The task was designed so that there was a 20% chance of a thought probe being presented in place of a Target in a condition block. The experience sampling protocol consisted of a series of ten questions (shown in Table 1), the first of which always prompted participants to rate their focus level (“Just before this question appeared, were you focused on the task or were you thinking about something else?”) on a continuous slider scale from “completely off task” to “completely on task”. Before the experiment began, participants were instructed on the meaning of being completely on task (e.g. “I’m focused and only thinking about the computer task”) and completely off task (e.g. “I was thinking about something unrelated, like a past vacation or what I’ll have for dinner”). The other nine questions were split into five questions regarding the content, and four questions regarding the form of the thoughts experienced by the participants. Participants answered using a slider scale that always had “not at all” and “completely” at the extremes, apart from one question regarding the thoughts’ valence, for which the scale went from “negative” to “positive”. These questions have been used in previous investigations (Medea et al., 2016; Smallwood et al., 2016a).

**Table 1**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Type | Question | Dimension | Left Extreme | Right Extreme |
| Content | My thoughts were focused on the task I was performing. | Focus | Not at all | Completely |
| Content | My thoughts involved other people. | Other | Not at all | Completely |
| Content | The content of my thoughts was: | Emotion | Negative | Positive |
| Content | My thoughts involved past events. | Past | Not at all | Completely |
| Content | My thoughts involved myself. | Self | Not at all | Completely |
| Content | My thoughts involved future events. | Future | Not at all | Completely |
| Form | My thoughts were in the form of words. | Words | Not at all | Completely |
| Form | My thoughts were in the form of images. | Images | Not at all | Completely |
| Form | My thoughts were vague and non-specific. | Vague | Not at all | Completely |
| Form | My thoughts were intrusive. | Intrusive | Not at all | Completely |

We recorded a total of 848 thought probes. Following Medea et al. (2016), Ruby et al. (2013a), Ruby et al. (2013b), Smallwood et al. (2016), we decomposed these data at the trial level using exploratory factor analysis with varimax rotation and selected three components for both the content questions and three components for the form questions which explained ~80 % of the variance in both cases. These are presented in Figure 2. For the content of thoughts, this resulted in: 1) a Past/Off Task/Others component, weighting on thinking about the past and about other people, while being off task; 2) a Future/Self component, weighting on thinking about the future and on one’s self, and 3) a Positive/On Task component, weighting on having positively valenced thoughts, while being on task. For the form of thoughts, the analysis identified: 1) an Images component, weighting on thinking in images and also not thinking in words, 2) an Intrusive component, weighting on reporting one’s thoughts as being intrusive, and 3) a Vague component, weighting on having vague thoughts. These are similar to the solutions produced in prior investigations using this approach. For the purpose of our analysis, we projected these solutions back onto the trial level data.



2.3 Participants

Forty-two participants (18-28 years, mean age 19.4; 8 males) completed Experiment 1; 9 participants were removed from the analyses due to excessive amount of missing pupil data (as defined below in the pre-processing section) and one was removed as target accuracy was at chance; after filtering of participants, 32 participants (64 total sessions) were used in the analyses. Forty-two participants (18-39 years, mean age 21.5; 11 males) completed Experiment 2; five participants were removed from the analyses due to excessive amount of missing pupil data, one participant was removed due to abnormally slow reaction times; thirty-six participants (216 total sessions) were finally used in the analyses.

2.4 Apparatus and Setup

Pupil size was recorded using an EyeLink 1000 Desktop Mount (SR Research Ltd., Mississauga, ON, Canada), at a rate of 250hz, from the right eye only of participants. Pupil diameter was measured in arbitrary units as recorded by the eye-tracker. The study was conducted in a small, dark room with no windows, resulting in ambient light levels below 0.1 cd/m2. Visual stimuli were presented on an 18-inch LCD monitor located 60 cm from the chinrest (with forehead support) that participants used, and the eye-tracker was placed right below the computer screen. Presentation of all stimuli was controlled with PsychoPy2 v.1.81.03 together with the ioHub Python package (<http://www.isolver-solutions.com/iohubdocs/>) to interface with the eye-tracker.

2.5 Procedure

Participants were provided an information sheet for the experiment at least 24 hours ahead of the testing day; on the day, participants were welcomed in the lab and signed a consent form. They were then introduced to the task paradigm and the eye-tracker setup, after which they completed a practice trial of the task, which included full instructions. A randomized target order 9-point calibration routine was performed and a separate validation was performed using the EyeLink 1000 software. Once calibration was completed, the experiment began. All investigation was conducted according to the principles expressed in the Declaration of Helsinki; the study and the process for gaining informed consent was approved by the University of York Department of Psychology’s Ethics Committee, and by the Ministry of Defence Research Ethics Committees.

2.6 Pre-Processing of Eye-Tracking Data

The following pre-processing steps were taken in order to remove possible artefacts before data analysis, for each run, for each participant. If pupil data was missing in time periods longer than 1 second it was discarded, and it was linearly interpolated for periods shorter than 1 second. If more than 25% of the data was missing (e.g. for excessive blinking, drowsiness, falling asleep), the run was discarded and not used in subsequent analysis. Pupil data was z-scored and pupil measures with absolute values larger than 3.5 or smaller than -3.5 were discarded. The pupil time series was also median filtered (order 5) to remove spikes, and low-pass filtered with a 10hz cut-off and then downsampled to 80ms (12.5hz).

2.7 Analysis

Pupil data was analysed using linear mixed models as implemented in R through the package *lme4* (Bates et al., 2015). For both experiments we ran models to investigate how baseline pupil size predicted behaviour (target RTs), and for Experiment 2 we also ran models to investigate how pupil size predicted on task focus and mind wandering reports (PCA components) in the two task conditions. We selected a time window of interest comprising the non-target (NT) and fixation cross immediately preceding either a correctly reported target, in case of behaviour, or a thought probe, for the analysis of the self-reports. This resulted in an average time window of 3.5 seconds. In all our models the average pupil diameter in that time window, and task condition, were included as fixed effect predictors; on task reports, PCA components extracted from the mind wandering reports (Experiment 2), and reaction times (Experiment 1 and 2; log-transformed in order to normalise their distribution) were the predicted variables in our models. Participants and sessions were modelled as random factors, in a nested fashion; sessions were also modelled as a fixed factor, in order to investigate possible learning effects. We then compared models that had only task condition or only baseline pupil size to a null model through a Likelihood Ratio Test, to see if it would improve the model’s fit. Next, we compared models that included both predictors to models that included only one of them, to see if the additional predictor would improve model fit. Finally, we compared models that included the interaction between task condition and baseline pupil size, to models only having the two predictors but with no interaction. Fixed effects on the predicted variables were plotted using the *effects* (Fox, 2003) and *ggplot2* (Wickham, 2009) R packages.

1. **Results**

3.1 Task condition effects on behaviour and pupil size (Experiment 1 & 2)

We analysed participants’ accuracy and response time data for both of our studies using paired sample t-tests. We replicated the task-condition effects found in Konishi et al. (2015). Participants were both slower, and less accurate, to targets in the 1-back task relative to the 0-back (Experiment 1 RTs: *t* (35) = 10.00, p<.001, Experiment 1 accuracy: *t*(35) = 3.93, p < .001; Experiment 2 RTs: *t* (31) = 7.99, p<.001, Experiment 2 accuracy: *t*(31) = 5.08, p<.001). In Experiment 2 we also replicate the effects of differential task focus for the two conditions, so that participants reported to be more off task in the 0-back condition than in the 1-back (*t*(35)=4.80, p<.001). Finally we found that average pupil diameter (PD) was larger throughout the 1-back task relative to the 0-back, in both studies (Experiment 1 PD: *t* (35) =2.02, p = .051; Experiment 2 PD: *t*(31) = 3.03, p = .005). All results are shown in Figure 1 (bottom panel).

3.2 Effects of condition, session, and baseline pupil size on target reaction times (Experiments 1 & 2)

Task condition significantly predicted RTs in both Experiment 1 (χ2(1) = 258.15, *p* < .0001) and Experiment 2 (χ2(1) = 322.17, *p* < .0001), as the 1-back condition resulted in longer RTs. Moreover, adding baseline pupil to the model significantly improved the fit in both Experiment 1 (χ2 (1) = 6.70, *p* = .009) and Experiment 2 (χ2(1)= 22.36, *p* < .0001), with larger baseline pupils predicting longer RTs. Adding an interaction factor between task condition and baseline pupil did not improve model fit in either Experiment 1 (χ2 (1) = 0.47, *p* = .494) or Experiment 2 (χ2 (1) = 2.02, *p* = .155). Finally, adding the session factor also significantly improved the fit of the model in both Experiment 1 (χ2 (1) = 7.14, *p* = .007) and Experiment 2 (χ2 (1) = 25.48, *p* < .0001). These analyses therefore show that slower responding was preceded by larger pupils in both the 0-back and the 1-back task and that RT got shorter across the sessions. All results are shown in Figure 3.

3.3 Effects of condition and baseline pupil size on reports of task focus (Experiment 2)

Our next analysis examined the relation between baseline pupil size and the response to the task focus questions. Task condition significantly predicted task focus in Experiment 2 (χ2 (1) = 44.38, *p* < .0001), as participants reported being more on task in the 1-back condition. Adding baseline pupils to the model significantly improved the fit (χ2 (1) = 21.68, *p* < .0001), with larger baseline pupils predicting more on task focus reports. Adding an interaction factor between task condition and baseline pupil slightly improved model fit, albeit not significantly (χ2 (1) = 2.79, *p* = .094). To understand this trend we compared the effect of the pupil signal in each task separately. This subsequent analysis confirmed differences in pupil size for reports of task focus exist in the 0-back task (χ2 (1) = 21.59, *p* < .0001), that were not clear in the 1-back task (χ2 (1) = 2.02, *p* = .155). All results are shown in Figure 3.



3.4 Effects of condition and baseline pupil size on mind wandering reports (Experiment 2)

Having demonstrated a relationship between the pupil signal and on and off task thought, we next examined how this related to the different types of mind-wandering as described by the decomposition analysis. Task condition significantly predicted self-reports regarding the Off task/Past (χ2 (1) = 7.04, *p* = .007), Future/Self (χ2 (1) = 12.52, *p* < .001), Images (χ2 (1) = 25.97, *p* < .0001), and Intrusive (χ2 (1) = 4.82, *p* = .028) components: in the 0-back condition, compared to the 1-back, participants reported being more off task and thinking more about the past, thinking more about the future, thinking more in images than in words, and described having more intrusive thoughts. Adding baseline pupils to the model only improved fit for the Past/Off task (χ2(1) = 8.17, *p* = .004) and Intrusive component (χ2(1) = 15.94, *p* < .0001). There were no interactions between task condition and baseline pupils for any of the PCA components. All results are shown in Figure 3. These results suggest that the same relationship between pupils and ongoing experience that was linked to off task thought, was associated with thoughts regarding the past and with an intrusive nature.

1. **Discussion**

In two experiments we find evidence of significantly larger pupils predicting poor response times in a simple working memory task. In both datasets, longer reaction times were preceded by larger pupil diameter. A link between larger baseline pupils and slower responding has been observed before in different paradigms, such as Unsworth & Robison (2016) using a sustained attention task; Gilzenrat et al. (2010) in an oddball task; Bradshaw (1968) in a reaction-time task; and Smallwood et al. (2011) in a 1-back task similar to ours. Our data, therefore, adds to a growing body of evidence that unusually large pupils are a signature that external information is not being processed correctly.

We also found evidence of a link between baseline pupil size and ongoing experiential states. In our data, off task states were associated with significantly smaller pupil diameter than was observed preceding on-task reports replicating prior studies (Grandchamp et al., 2014; Mittner et al., 2014; Unsworth & Robison, 2016). Our analysis of the content of thoughts suggests that reduced pupil size is a marker for experiential states that are focused on the past, or that are particularly intrusive in nature. Prior studies using the same decomposition procedure has identified that past thoughts are most strongly linked to unhappiness (Ruby et al., 2013). A recent fMRI experiment (Smallwood et al., 2016a) demonstrated that past thoughts depended on coupling between the hippocampus and the lateral temporal lobe and decoupling between the medial prefrontal cortex and the occipital cortex, extending into the lingual gyrus. As neural activation in the lingual gyrus is a correlate of large pupils (Kuchinsky et al., 2016; Murphy et al., 2014) it seems plausible that the association between small pupils when thinking about events from the past may describe a pattern of stronger perceptual decoupling than is engaged more than in other types of self-generated thought. This hypothesis could be tested by measuring external attention directly using event related potentials (Baird et al., 2014). More generally, it seems possible that a combination of intrusive thoughts, as well as a tendency to focus on the past may constitute the experiential correlates of a state of rumination (Poerio et al., 2013; Smallwood & O’Connor, 2011). Moving forward, this study suggests that reductions in the size of pupils might be a useful marker in clinical research.

Our data also highlights the theoretical value of taking into account both the content of experience, and the context in which it occurs, in studies of mind-wandering (Smallwood & Andrews-Hanna, 2013). We found that pupil size was a more reliable predictor of attentional state in the context of the less demanding 0-back task than in the 1- back task. This may reflect the fact that in the less demanding condition of our experiment there is more freedom for attention to fluctuate for intrinsic reasons, and that these are being reflected in the pupil signal. We also observed that the pupil signal was sensitive to the content of experience. Our findings suggest that a focus on the past, and/or experiences with an intrusive aspect during the mind-wandering state, are associated with periods when the size of pupils is maximal. Other aspects of mind-wandering such as focus on the future, by contrast, were not associated with fluctuations in pupil size, but were instead modulated by the level of external demands in the concurrent task. It thus seems likely that the pupil signal does not capture information on all aspects of the mind-wandering state, a complex heterogeneous state whose behaviour changes in a complex fashion across different task conditions. Given these data it seems that studies of mind-wandering reviewed in the introduction have failed to reveal a consistent association with the pupil signal because they have routinely failed to account for this contextual and experiential complexity. Moving forward, it is a priority for studies to take account of different experimental and situational influences in their experimental design.

In conclusion, our study suggests that there are two different physiological states that relate to reductions in the extent to which attention is devoted to the external environment, which differ on their psychological qualities. Assuming that spontaneous changes in baseline pupil size are related to intrinsic variation in arousal (Aston-Jones & Cohen, 2005), our data suggests that states of very high and very low arousal may induce situations when attention is not focused on external processing. However, our data suggests that these different states may not have identical psychological features. In both studies, slow responding to targets was preceded by unusually large pupils whereas smaller pupils were predictive of off-task thoughts, and in particular of thoughts that are related to the past and that are intrusive in nature. In the future, it will be important to distinguish these two states of less effective external attention on other metrics, such as their neural correlates or their associations to personality. More generally, as our study shows that different patterns of baseline pupil size discriminate between different attentional states, it underscores that it is dangerous to make generalisations directly from behaviour to experience because, at least in the context of our paradigm, these metrics do not agree (see Konishi & Smallwood, (2016) for a consideration of this issue).

5. 1 Limitations

Our analysis used a short a priori defined window of ~3.5 seconds as determined by the jittering that our paradigm employed. This relatively short analysis window means our data is unable to determine whether there are longer term trends in the pupil signal that may relate to ongoing experiential or behavioural states. Second, in our paradigm we did not ask participants to distinguish between intentional and unintentional mind wandering. These have been shown to be states that differ in the context in which they arise, and in their relationship with task performance (Seli et al., 2016; Seli et al., 2016b; Thomson et al., 2014); furthermore, a recent study has shown these states to have dissociable relations to cortical thickness and functional connectivity of the brain (Golchert et al., 2016). It is thus possible that intentional and unintentional off-task thought would have differential pupillometry signatures, something that future studies should inquire. Finally, one motivation for conducting this experiment was to identify whether there was information about internal states that can be derived from the pupil signal, but with the ultimate aim of inferring mental states in the absence of introspection; while our study demonstrates that non-optimal behaviour has a similar property under conditions in which participants are, and are not, required to introspect on the contents of experience, our data does not show that introspection does not interfere with ongoing cognition. To assess this question. it will be necessary to develop independent methods of assessing information processing linked to the pupil signal, such as can be achieved by using fMRI or EEG, and understand if these metrics change when participants are asked to introspect on their own experiences.

**Acknowledgements**

This work was funded by the Defence Science and Technology Laboratory (Dstl). JS is also supported by the Economic Research Council (WANDERINGMINDS – [646927](http://www.sciencedirect.com/science/article/pii/S1053811916303743#gts0015)). The authors also wish to thank Susan Sidey, Jacolien van Rij for her help with a previous draft of the paper, Lorenzo Rossi for helping with data processing steps, and Gianmarco Altoè for general comments on the data.

**References**

Allen, M., Smallwood, J., Christensen, J., Gramm, D., Rasmussen, B., Jensen, C. G., … Lutz, A. (2013). The balanced mind: the variability of task-unrelated thoughts predicts error monitoring. *Frontiers in Human Neuroscience*, *7*(November), 743. https://doi.org/10.3389/fnhum.2013.00743

Aston-Jones, G., & Cohen, J. D. (2005). AN INTEGRATIVE THEORY OF LOCUS COERULEUS-NOREPINEPHRINE FUNCTION: Adaptive Gain and Optimal Performance. *Annual Review of Neuroscience*, *28*(1), 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

Baird, B., Smallwood, J., Lutz, A., & Schooler, J. W. (2014). The Decoupled Mind: Mind-wandering Disrupts Cortical Phase-locking to Perceptual Events. *Journal of Cognitive Neuroscience*, *26*(11), 2596–2607. https://doi.org/http://dx.doi.org/10.1162/jocn\_a\_00656

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.1177/009286150103500418

Bernhardt, B. C., Smallwood, J., Tusche, A., Ruby, F. J. M., Engen, H. G., Steinbeis, N., & Singer, T. (2014). Medial prefrontal and anterior cingulate cortical thickness predicts shared individual differences in self-generated thought and temporal discounting. *NeuroImage*, *90*, 290–297. https://doi.org/10.1016/j.neuroimage.2013.12.040

Bradshaw, J. L. (1968). Pupillary changes and reaction time with varied stimulus uncertainty. *Psychon. Sci.*, *13*(2), 69–70.

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 of the United States of America*, *106*(21), 8719–8724. https://doi.org/10.1073/pnas.0900234106

Engert, V., Smallwood, J., & Singer, T. (2014). Mind your thoughts: Associations between self-generated thoughts and stress-induced and baseline levels of cortisol and alpha-amylase. *Biological Psychology*, *103*, 283–291. https://doi.org/10.1016/j.biopsycho.2014.10.004

Fox, J. (2003). Eﬀect Displays in R for Generalised Linear Models. *Journal of Statistical Software*, *8*(15), 1–27. https://doi.org/10.2307/271037

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. *Quarterly Journal of Experimental Psychology (2006)*, *66*(May 2015), 2289–94. https://doi.org/10.1080/17470218.2013.858170

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective & Behavioral Neuroscience*, *10*(2), 252–69. https://doi.org/10.3758/CABN.10.2.252

Golchert, J., Smallwood, J., Jefferies, E., Seli, P., Huntenburg, J. M., Liem, F., … others. (2016). Individual variation in intentionality in the mind-wandering state is reflected in the integration of the default-mode, fronto-parietal, and limbic networks. *NeuroImage*.

Grandchamp, R., Braboszcz, C., & Delorme, A. (2014). Oculometric variations during mind wandering. *Frontiers in Psychology*, *5*(FEB). https://doi.org/10.3389/fpsyg.2014.00031

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. *Science*, *306*(5702), 1776–1780. https://doi.org/10.1126/science.1103572

Kane, M. J., Brown, L. H., McVay, J. C., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007). For whom the mind wanders, and when: an experience-sampling study of working memory and executive control in daily life. *Psychological Science : A Journal of the American Psychological Society / APS*, *18*(7), 614–21. https://doi.org/10.1111/j.1467-9280.2007.01948.x

Karapanagiotidis, T., Bernhardt, B. C., Jefferies, E., & Smallwood, J. (2016). Tracking thoughts: Exploring the neural architecture of mental time travel during mind-wandering. *NeuroImage*.

Konishi, M., McLaren, D. G., Engen, H., & Smallwood, J. (2015). Shaped by the Past: The Default Mode Network Supports Cognition that Is Independent of Immediate Perceptual Input. *PloS One*, *10*(6), e0132209. https://doi.org/10.1371/journal.pone.0132209

Konishi, M., & Smallwood, J. (2016). Shadowing the wandering mind: how understanding the mind-wandering state can inform our appreciation of conscious experience. *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*(4), 233–246.

Kuchinsky, S. E., Pandža, N. B., & Haarmann, H. J. (2016). Linking Indices of Tonic Alertness: Resting-State Pupil Dilation and Cingulo-Opercular Neural Activity. In *International Conference on Augmented Cognition* (pp. 218–230).

Levinson, D. B., Smallwood, J., & Davidson, R. J. (2012). The persistence of thought: evidence for a role of working memory in the maintenance of task-unrelated thinking. *Psychological Science*, *23*(4), 375–80. https://doi.org/10.1177/0956797611431465

Mason, M. F., Norton, M. I., Horn, J. D. Van, Wegner, D. M., Grafton, S. T., Macrae, C. N., … Macrae, C. N. (2007). Wandering Minds: Stimulus-Independent Thought. *Science*, *315*(January), 393–395. https://doi.org/10.1126/science.1131295

Matthew A. Killingsworth and Daniel T. Gilbert. (2010). A Wandering Mind Is an Unhappy Mind. https://doi.org/10.1126/science.1192439

Mcvay, J. C., & Kane, M. J. (2009). Conducting the train of thought: Working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(1), 196–204. https://doi.org/10.1037/a0014104

Medea, B., Karapanagiotidis, T., Konishi, M., Ottaviani, C., Margulies, D., Bernasconi, A., … Smallwood, J. (2016). How do we decide what to do? Resting-state connectivity patterns and components of self-generated thought linked to the development of more concrete personal goals. *Experimental Brain Research*, pp. 1–13. https://doi.org/10.1007/s00221-016-4729-y

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. *Journal of Neuroscience*, *34*(July), 16286–16295. https://doi.org/http://dx.doi.org/10.1523/JNEUROSCI.2062-14.2014

Mittner, M., Hawkins, G. E., Boekel, W., & Forstmann, B. U. (2016). A Neural Model of Mind Wandering. *Trends in Cognitive Sciences*, *20*(8), 570–578. https://doi.org/10.1016/j.tics.2016.06.004

Morad, Y., Lemberg, H., Yofe, N., & Dagan, Y. (2000). Pupillography as an objective indicator of fatigue. *Current Eye Research*, *21*(1), 535–542. https://doi.org/citeulike-article-id:10485869\rdoi: 10.1076/0271-3683(200007)2111-ZFT535

Murphy, P. R., O’Connell, R. G., O’Sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human Brain Mapping*, *35*(8), 4140–4154. https://doi.org/10.1002/hbm.22466

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. https://doi.org/10.1111/j.1469-8986.2011.01226.x

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Peirce, J. W. (2008). Generating Stimuli for Neuroscience Using PsychoPy. *Frontiers in Neuroinformatics*, *2*(January), 10. https://doi.org/10.3389/neuro.11.010.2008

Poerio, G. L., Totterdell, P., & Miles, E. (2013). Mind-wandering and negative mood: Does one thing really lead to another? *Consciousness and Cognition*, *22*(4), 1412–1421. https://doi.org/10.1016/j.concog.2013.09.012

Ruby, F. J. M., Smallwood, J., Engen, H., & Singer, T. (2013). How Self-Generated Thought Shapes Mood-The Relation between Mind-Wandering and Mood Depends on the Socio-Temporal Content of Thoughts. *PLoS ONE*, *8*(10). https://doi.org/10.1371/journal.pone.0077554

Ruby, F. J. M., Smallwood, J., Sackur, J., & Singer, T. (2013). Is self-generated thought a means of social problem solving? *Frontiers in Psychology*, *4*(DEC). https://doi.org/10.3389/fpsyg.2013.00962

Rummel, J., & Boywitt, C. D. (2014). Controlling the stream of thought: Working memory capacity predicts adjustment of mind-wandering to situational demands. *Psychonomic Bulletin & Review*, *21*(5), 1309–1315. https://doi.org/10.3758/s13423-013-0580-3

Seli, P., Jonker, T. R., Cheyne, J. A., Cortes, K., Smilek, D., Seli, P., … 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*(3), 703–709. https://doi.org/10.1037/xhp0000029

Seli, P., Risko, E. F., & Smilek, D. (2016). On the Necessity of Distinguishing Between Unintentional and Intentional Mind Wandering. *Psychological Science*, 0956797616634068-. https://doi.org/10.1177/0956797616634068

Seli, P., Risko, E. F., Smilek, D., & Schacter, D. L. (2016). Mind-Wandering With and Without Intention. *Trends in Cognitive Sciences*, *20*(8), 605–617. https://doi.org/10.1016/j.tics.2016.05.010

Smallwood, J. (2013). Distinguishing how from why the mind wanders: a process-occurrence framework for self-generated mental activity. *Psychological Bulletin*, *139*(3), 519–35. https://doi.org/10.1037/a0030010

Smallwood, J., & Andrews-Hanna, J. (2013). Not all minds that wander are lost: The importance of a balanced perspective on the mind-wandering state. *Frontiers in Psychology*, *4*(AUG), 1–6. https://doi.org/10.3389/fpsyg.2013.00441

Smallwood, J., Brown, K. S., Baird, B., Mrazek, M. D., Franklin, M. S., & Schooler, J. W. (2012). Insulation for daydreams: A role for tonic norepinephrine in the facilitation of internally guided thought. *PLoS ONE*, *7*(4), 1–5. https://doi.org/10.1371/journal.pone.0033706

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). https://doi.org/10.1371/journal.pone.0018298

Smallwood, J., Karapanagiotidis, T., Ruby, F., Medea, B., Caso, I. De, Konishi, M., … Jefferies, E. (2016a). Representing representation: Integration between the temporal lobe and the posterior cingulate influences the content and form of spontaneous thought. *PLoS ONE*, *11*(4).

Smallwood, J., Karapanagiotidis, T., Ruby, F., Medea, B., Caso, I. De, Konishi, M., … Jefferies, E. (2016b). Representing representation: Integration between the temporal lobe and the posterior cingulate influences the content and form of spontaneous thought. *PLoS ONE*, *11*(4). https://doi.org/10.1371/journal.pone.0152272

Smallwood, J., McSpadden, M., & Schooler, J. W. (2008). When attention matters: the curious incident of the wandering mind. *Memory & Cognition*, *36*(6), 1144–50. https://doi.org/10.3758/MC.36.6.1144

Smallwood, J., Nind, L., & O’Connor, R. C. (2009). When is your head at? An exploration of the factors associated with the temporal focus of the wandering mind. *Consciousness and Cognition*, *18*(1), 118–125. https://doi.org/10.1016/j.concog.2008.11.004

Smallwood, J., & O’Connor, R. C. (2011). Imprisoned by the past: Unhappy moods lead to a retrospective bias to mind wandering. *Cognition & Emotion*, *9931*(930884466), 1–10. https://doi.org/10.1080/02699931.2010.545263

Smallwood, J., Ruby, F. J. M., & Singer, T. (2013). Letting go of the present: Mind-wandering is associated with reduced delay discounting. *Consciousness and Cognition*, *22*(1), 1–7. https://doi.org/10.1016/j.concog.2012.10.007

Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin*, *132*(6), 946–958. https://doi.org/10.1037/0033-2909.132.6.946

Smallwood, J., & Schooler, J. W. (2015). The Science of Mind Wandering: Empirically Navigating the Stream of Consciousness. *Annual Review of Psychology*, *66*(1), 487–518. https://doi.org/10.1146/annurev-psych-010814-015331

Stanners, R. F., Coulter, M., Sweet, A. W., & Murphy, P. (1979). The pupillary response as an indicator of arousal and cognition. *Motivation and Emotion*, *3*(4), 319–340. https://doi.org/10.1007/BF00994048

Stawarczyk, D., Majerus, S., Maquet, P., & D’Argembeau, A. (2011). Neural correlates of ongoing conscious experience: Both task-unrelatedness and stimulus-independence are related to default network activity. *PLoS ONE*, *6*(2). https://doi.org/10.1371/journal.pone.0016997

Teasdale, J. D., Proctor, L., Lloyd, C. A., & Baddeley, A. D. (1993). Working-Memory and Stimulus-Independent Thought - Effects of Memory Load and Presentation Rate. *European Journal of Cognitive Psychology*, *5*(4), 417–433. https://doi.org/10.1080/09541449308520128

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*(1), 14–26. https://doi.org/10.1016/j.concog.2014.04.001

Unsworth, N., & Robison, M. K. (2016). Pupillary correlates of lapses of sustained attention. *Cognitive, Affective, & Behavioral Neuroscience*, (April), 601–615. https://doi.org/10.3758/s13415-016-0417-4

Van Orden, K. F., Jung, T. P., & Makeig, S. (2000). Combined eye activity measures accurately estimate changes in sustained visual task performance. *Biological Psychology*, *52*(3), 221–240. https://doi.org/10.1016/S0301-0511(99)00043-5

Wickham, H. (2009). Elegant Graphics for Data Analysis. *Media*, *35*(July), 211. https://doi.org/10.1007/978-0-387-98141-3

Wilhelm, B., Wilhelm, H., Lüdtke, H., Streicher, P., & Adler, M. (1998). Pupillographic assessment of sleepiness in sleep-deprived healthy subjects. *Sleep*, *21*(3), 258–65. https://doi.org/citeulike-article-id:10484957

Yoss, R. E., Moyer, N. J., & Hollenhorst, R. W. (1970). Pupil size and spontaneous pupillary waves associated with alertness, drowsiness, and sleep. *Neurology*, *20*(6), 545.

**Figure Legends**

**Figure 1 – Task paradigm and behavioural results**

Top panel illustrates our paradigm: in both conditions, after a certain number of Non-Targets (NTs) participants were faced with a target decision (in Experiment 1 & 2), or a thought probe (only in Experiment 2). In the 0-back condition, the decision is based on the presently perceived stimulus (*is* the square on the left or the right?); the NTs are thus irrelevant to the task, unconstraining the participant’s external attention for long periods of the 0-back condition. Conversely, in the 1-back condition the target decision is based on the previously attended NT (*was* the square on the left or the right?). Not knowing when a target will be presented, participants must maintain external attention on the task in order to perform accurately. We thus selected the time window (~3.5 seconds) of the NT and fixation cross immediately preceding a target or a thought probe to analyse the effects of average pupil size on behaviour and internal reports.

Bottom grey panel shows how our conditions modulate behaviour and on-task reports. In both studies participants are slower and less accurate in the 1-back condition; additionally, their pupils are larger and they report being more on-task.

**Figure 2 – Relationship between self-reports of mind-wandering and baseline pupil size**

Grey panel (right) illustrates the PCA decomposition of the thought probes in the 6 components of thought, divided by content and form. In the white panel are illustrated the main effects of task condition and baseline pupil size on the 6 components of thought. Task condition predicts reports for 4 components: in the 0-back condition, participants report being more off task, having more thoughts about the past (P), the future (F), more intrusive thoughts (IN), and thinking more in images than in words (I/W). Additionally, smaller pupils predicted more reports of off task and past thoughts (P), and of intrusive thoughts (IN).

**Figure 3 – Relationship between reaction times and baseline pupil size**

Illustrated on white background are the main effects of task condition and baseline pupil size on reaction times (log-transformed) and on-task reports. Larger pupils, and task condition, predict slower reaction times in both of our studies (top two panels). Additionally, larger pupils predict reports of being on-task in the 0-back condition (bottom right panel). Grey panel shows task learning effects: over successive days, participants’ RTs become faster for both conditions, in both studies.