

1 Pupillometry and the vigilance decrement: Task-evoked but not baseline pupil measures reflect declining
2 performance in visual vigilance tasks

3 Joel T. Martin^{1,2}, Annalise H. Whittaker³ and Stephen J. Johnston¹

4

5 ¹School of Human and Health Sciences, Department of Psychology, University of Swansea, Swansea,
6 Wales, United Kingdom

7 ² Department of Psychology, University of York, United Kingdom, YO10 5DD

8 ³Dstl, Porton Down, Salisbury SP4 0JQ

9

10

11

12 Author Note

13 Joel T. Martin, Department of Psychology, University of York, United Kingdom, YO10 5DD

14 Email: joel.martin@york.ac.uk

15 Phone: +44 (0) 1904 323190

16 The research was funded by a grant from the Dstl

PUPILLOMETRIC VIGILANCE DECREMENT

17

Abstract

18 Baseline and task-evoked pupil measures are known to reflect the activity of the nervous system's central
19 arousal mechanisms. With the increasing availability, affordability and flexibility of video-based eye
20 tracking hardware, these measures may one day find practical application in real-time biobehavioral
21 monitoring systems to assess performance or fitness for duty in tasks requiring vigilant attention. But
22 real-world vigilance tasks are predominantly visual in their nature and most research in this area has taken
23 place in the auditory domain. Here we explore the relationship between pupil size—both baseline and
24 task-evoked—and behavioral performance measures in two novel vigilance tasks requiring visual target
25 detection: 1) a traditional vigilance task involving prolonged, continuous, and uninterrupted performance
26 ($n = 28$), and 2) a psychomotor vigilance task ($n = 25$). In both tasks, behavioral performance and task-
27 evoked pupil responses declined as time spent on task increased, corroborating previous reports in the
28 literature of a vigilance decrement with a corresponding reduction in task-evoked pupil measures. Also in
29 line with previous findings, baseline pupil size did not show a consistent relationship with performance
30 measures. We discuss our findings considering the adaptive gain theory of locus coeruleus function and
31 question the validity of the assumption that baseline (prestimulus) pupil size and task-evoked
32 (poststimulus) pupil measures correspond to the tonic and phasic firing modes of the LC.

33 *Keywords:* vigilance, psychomotor vigilance, sustained attention, pupillometry, locus coeruleus
34

PUPILLOMETRIC VIGILANCE DECREMENT

35 Pupillometry and the vigilance decrement: Task-evoked but not baseline pupil measures reflect declining
36 performance in visual vigilance tasks

1 – Introduction

38 The term *vigilance* has received varied usage in scientific research but broadly speaking it refers
39 to an organism's ability to sustain its attention over prolonged periods of time (Kahneman & Treisman,
40 1984; Raja Parasuraman et al., 1998; Raja Parasuraman & Davies, 1982; Joel S. Warm et al., 2008; Joel
41 S. Warm & Jerison, 1984). Although there is a long history of research into performance and continuous
42 work tasks (see Bills, 1943; Hockey, 2013 for reviews), Mackworth (1948, 1950) is frequently credited
43 for the first systematic studies of vigilance and the discovery that human detection performance on a
44 monotonous watch keeping task, under conditions similar to those experienced by radar and sonar
45 operators, declines as time spent on-task increases. This so-called *vigilance decrement* became the target
46 of numerous research efforts in human factors and experimental psychology which sought to understand
47 how factors specific to the task, the individual performing it, and the environment in which it is
48 performed, all contribute to failures of vigilant attention (Frankmann & Adams, 1962; Mackie, 1987;
49 Wiener, 1987). Signal detection theory (Green & Swets, 1974) has played a central role in the
50 psychophysical analysis of vigilance studies, with detection performance being characterized frequently
51 on the basis of the number of hits, misses, false alarms, correct rejections, and the derived measures of
52 sensitivity and criterion (e.g., Mackworth, 1970; Parasuraman & Davies, 1976). Detection latency also
53 features in analyses of vigilance task performance (e.g., Basner & Dinges, 2011; Broadbent, 1958; Buck,
54 1966) and biometric technologies such as electroencephalography (EEG) and functional magnetic
55 resonance imaging (fMRI) continue to shape our understanding of the neurophysiological mechanisms of
56 vigilant attention (for review, see: Fortenbaugh, DeGutis, & Esterman, 2017; Langner & Eickhoff, 2013;
57 Oken, Salinsky, & Elsas, 2006).

58 Another biometric technique that has been successfully applied to the study of vigilant attention is
59 cognitive pupillometry, the measurement of the size and reactivity of the eyes' pupils following exposure
60 to psychologically relevant stimuli. The pupils respond primarily to light, but when light levels are held

PUPILLOMETRIC VIGILANCE DECREMENT

61 constant, fluctuations in pupil size offer a window of insight into the brain's central arousal systems
62 (Joshi & Gold, 2020; Kahneman, 1973; Laeng et al., 2012). Specifically, nonluminance-mediated pupil
63 size changes are known to reflect the moment-to-moment activity of the locus coeruleus noradrenalin
64 system (LC-NA: Joshi et al., 2016; Rajkowski, Kubiak, & Aston-Jones, 1993), which has a central role in
65 the modulation of arousal and alertness (Berridge, 2008; Berridge et al., 2012; Berridge & Waterhouse,
66 2003) and in maintaining optimal levels of vigilance and performance (Petersen & Posner, 2012; Posner
67 & Petersen, 1990). Extensive single-cell recording studies in behaving rodents and monkeys show that the
68 noradrenergic neurons of the LC exhibit *phasic* and *tonic* modes of activation and that these distinct
69 modes correspond to different behavioral states (Aston-Jones & Cohen, 2005; Aston-Jones, Rajkowski, &
70 Cohen, 1999). The phasic mode is characterized by short bursts of activation in response to task-relevant
71 stimuli and supports task engagement and exploitation of the environment, whereas the tonic mode is
72 characterized by a sustained increase in baseline activation in response to diminishing task utility and
73 supports disengagement from the current task and exploration of the environment (Aston-Jones & Cohen,
74 2005).

75 The functional role of the LC-NA system and its association with the pupil has led many to
76 assume that baseline (i.e., prestimulus) and task-evoked measurements of pupil size may correspond to the
77 tonic and phasic firing modes of the LC, and that these measures may reflect changes in vigilant attention
78 over time. Indeed, changes in pupillometric response associated with changes in vigilant attention have
79 been noted previously. Beatty (1982) asked participants to monitor a string of tones presented at 3.2 s
80 intervals continuously for 48 min for the occurrence of target tones, which were slightly attenuated in
81 volume. Approximately 12 targets were presented at random intervals in every 5 min period, with 108
82 targets being presented across the whole task. As is common in vigilance tasks, while the task itself was
83 conducted as one continuous 48-minute procedure, the data were sub-divided into several periods of
84 watch and compared to determine time-related differences in performance. Detection accuracy decreased
85 in accordance with time spent on task, replicating the classic finding of a vigilance decrement described
86 by Mackworth (1948, 1950). Task-evoked pupillary responses to target stimuli mirrored these results,

PUPILLOMETRIC VIGILANCE DECREMENT

decreasing in amplitude across each third of the test; but baseline (prestimulus) measurements of pupillary activity (obtained prior to each target stimulus) showed little change. In more recent work we even see hints that pupil measures may serve to predict performance on a moment-to-moment basis. For example, Kristjansson et al. (2009) reported significant differences in pupil size and dilation rate for the fastest and slowest detection responses in a psychomotor vigilance task, suggesting that the pupil measures may provide sufficient reliable information to index alertness in real-time. But, despite the promising narratives of Kristjansson et al. and a handful of other studies (e.g., Unsworth & Robison, 2016; van den Brink et al., 2016), the nature of the relationship between pupil and performance measures remains unclear.

Since Beatty (1982), many studies have found performance decrements in long and demanding tasks that coincided with reduced task-evoked pupil responses (e.g., Hopstaken, van der Linden, et al., 2015; Hopstaken, Van der Linden, et al., 2015; Murphy, van Moort, & Nieuwenhuis, 2016; Unsworth & Robison, 2016) but, as noted by van den Brink et al. (2016), the literature is conflicted on the relationship between task performance and baseline pupil size. In some experiments, moments of off-task thought or poor task performance were associated with larger pupils at baseline (Franklin et al., 2013; Gilzenrat et al., 2010; Smallwood et al., 2011, 2012; Unsworth & Robison, 2016), whereas in other experiments poor task performance was associated with smaller pupils at baseline (Grandchamp, Braboszcz, & Delorme, 2014; Hopstaken, Van der Linden, et al., 2015; Kristjansson et al., 2009; Mittner et al., 2014; Van Orden, Jung, & Makeig 2000), or occurred following a gradual decrease in baseline pupil size (Grandchamp et al., 2014; Massar et al., 2016; McIntire, McKinley, & Goodyear, 2014; Murphy et al., 2011). Poor task performance was also found to occur with both relatively large and small baseline pupil size within experiments (van den Brink et al., 2016; Murphy et al., 2011; Smallwood et al., 2012; Unsworth & Robison, 2016), with one experiment reporting an increase in baseline pupil diameter as a function of time-on-task in a 37-min auditory vigilance task without breaks (Murphy et al., 2011).

Such discrepant findings on the relationship between baseline pupil size and task performance likely reflect the interplay of various methodological factors. Among the studies cited in the previous

PUPILLOMETRIC VIGILANCE DECREMENT

113 paragraph there is considerable variability in how performance was measured, with some focusing
114 primarily on RT measures, such as mean RTs (e.g., Smallwood et al., 2011, 2012), fraction of the slowest
115 or fastest RTs (e.g., Unsworth & Robison, 2016; van den Brink et al., 2016), or RT variability (e.g.,
116 Murphy et al., 2011); and others focusing more on perceptual sensitivity (i.e., d' : Beatty, 1982;
117 Hopstaken, van der Linden, et al., 2015; Hopstaken, Van der Linden, et al., 2015), or self-reported
118 measures of task engagement (e.g., Franklin et al., 2013; Grandchamp et al., 2014; Mittner et al., 2014).
119 Task demands also vary considerably across experiments, with some requiring only simple target
120 detection (e.g., Massar et al., 2016) and others requiring *simultaneous* (Beatty, 1982; van den Brink et al.,
121 2016; Gilzenrat et al., 2010; Murphy et al., 2011) or *successive* (Hopstaken et al., 2016; Hopstaken, van
122 der Linden, et al., 2015; Hopstaken, Van der Linden, et al., 2015; Smallwood et al., 2011, 2012)
123 discrimination¹. Further, some tasks called for prolonged continuous monitoring (Beatty, 1982; Murphy
124 et al., 2011), whereas others entailed intermittent breaks from the primary task (e.g., Hopstaken, Van der
125 Linden, et al., 2015; Smallwood et al., 2004; Unsworth & Robison, 2016), which even when very short
126 have the potential to improve performance by temporarily boosting motivation (Ariga & Lleras, 2011;
127 Ralph et al., 2016; Ross, Russell, & Helton, 2014). Finally, the stimuli varied substantially, and some may
128 have had undesirable behavioral or pupillometric consequences. For example, the ‘running counter’
129 stimulus used in Massar et al.’s (2016) psychomotor vigilance task provides feedback which could enable
130 participants to detect declines in their performance and adopt compensatory strategies (Thorne et al.,
131 2005); and for studies using visual stimuli (e.g., van den Brink et al., 2016; Smallwood et al., 2011,

¹ The distinction between successive and simultaneous discrimination tasks was first made by Parasuraman (1979). Successive tasks are absolute judgement tasks where observers must compare the current sensory input with a template in working memory in order to determine whether a particular stimulus is, or is not, a critical signal. Simultaneous tasks on the other hand are comparative judgement tasks, where each stimulus contains all of the information required to determine whether it is (or is not) a signal. Due to the involvement of working memory, successive tasks are thought to be more resource demanding than simultaneous tasks.

PUPILLOMETRIC VIGILANCE DECREMENT

132 2012), differences in visual attributes such as luminance, color and contrast may have contributed to
133 pupillometric and behavioral variance (Barbur, Harlow, & Sahraie, 1992; Goldwater, 1972).

134 The increasing availability, affordability, and flexibility of video-based eye tracking hardware
135 means that pupils' predictive power for vigilant attention may one day find practical application in
136 passive, real-time biobehavioral monitoring systems to assess performance or fitness for duty. Such
137 systems would be of particular use in scenarios where traditional reaction time assessments cannot easily
138 be administered. Presently, however, it remains unclear which pupil measures, if any, would be suitable
139 for an application of this kind. Focusing on the issues raised above, here we examine the relationship
140 between pupil measures—both baseline and task-evoked—and task performance in a novel
141 implementation of two well-established vigilance task paradigms: 1) continuous, uninterrupted vigilance,
142 and 2) psychomotor vigilance.

2 – Experiment 1

143 Since Mackworth (1948, 1950), experimental vigilance tasks have generally aimed to simulate
144 the conditions of real-world scenarios where monotonous repetitive tasks have become commonplace due
145 to automation and industrial mechanisation. Though vigilance tasks can vary in many ways, the defining
146 characteristic is that observers must remain alert and respond to critical signals presented against a
147 background of noncritical signals over prolonged, unbroken stretches of time—usually at least 30 mins
148 (Frankmann & Adams, 1962; Parasuraman & Davies, 1976). Key differences between tasks known to
149 influence performance are the sensory modality of stimulus presentation (e.g., auditory, visual), the
150 psychophysical dimensions used to define critical signals (e.g., brightness, loudness), and whether the
151 detection of targets requires successive or simultaneous discrimination (Parasuraman, 1979; Warm et al.,
152 2008); but performance ultimately depends on complex interactions between factors relating to the task,
153 the environment, and the individual (Ballard, 1996). To date, most vigilance tasks conducted with
154 pupillometry have presented stimuli in the auditory modality, probably to avoid methodological
155 confounds associated with the effects of visual stimulation and optical distortion of raw pupil
156 measurements inherent to video-based systems. But vigilance tasks in the real world are predominantly
157

PUPILLOMETRIC VIGILANCE DECREMENT

158 visual, and if pupil measures are to serve a useful purpose for tracking vigilant attention in real world
159 settings, they must be robust to the effects of visual stimuli.

160 In Experiment 1 we explored the relationship between pupil and performance measures (RT,
161 accuracy, d' and c) in a canonical vigilance task with stimuli presented in the visual modality. To our
162 knowledge, McIntire et al. (2014) is the only previous example of such a study, though the analysis was
163 correlational and simply explored how average pupil size and percentage of hits showed a similar decline
164 across four successive 10-min periods of continuous task performance. This is in contrast to auditory
165 vigilance experiments (e.g., Beatty, 1982; Gilzenrat et al., 2010; Murphy et al., 2011), where there has
166 been considerable focus on event-related pupil measurements. The present experiment therefore aimed to
167 examine event-related pupil responses in a novel vigilance task with visual stimuli, whilst controlling
168 appropriately for the effects of eye movements and luminance confounds.

169 Our task required participants to continuously monitor four centrally presented equiluminant
170 visual stimuli for 30 min in order to detect and respond to brief targets occurring with temporal and
171 spatial uncertainty against a high background event rate. A relatively high number of targets—6 per
172 min—was used to ensure that a suitable amount of event-related data would be generated for the analysis
173 (e.g., Mackie, 1987). We predicted that performance measures (e.g., RT, accuracy, d') across successive
174 10 min task blocks would betray a classic vigilance decrement as has been reported widely in the
175 literature (Frankmann & Adams, 1962; Mackie, 1987; Mackworth 1948, 1950; Wiener, 1987). Second,
176 based on the most consistent findings from pupillometric studies of vigilance (e.g., Hopstaken, van der
177 Linden, et al., 2015; Hopstaken, Van der Linden, et al., 2015; Murphy, van Moort, & Nieuwenhuis, 2016;
178 Unsworth & Robison, 2016), we predicted that the magnitude of task-evoked pupil size changes would
179 decrease, in line with performance measures, across the duration of the task. Considering the discrepant
180 findings in the literature, we did not make specific predictions about baseline or ‘tonic’ pupil size for this
181 experiment.

182 **2.1 – Materials and methods.**

PUPILLOMETRIC VIGILANCE DECREMENT

183 **2.1.1 – Participants.** Twenty-eight participants (23 females; age range 18-32 years, $M = 20.07$,
184 $SD = 2.8$) completed the experiment voluntarily or in exchange for course credit. All participants were
185 students at Swansea University reporting normal or corrected-to-normal acuity and color vision. The
186 experimental protocol was approved by the Ministry of Defence Research Ethics Committee and the
187 Department of Psychology Ethics Committee at Swansea University. Written informed consent was
188 obtained from each participant.

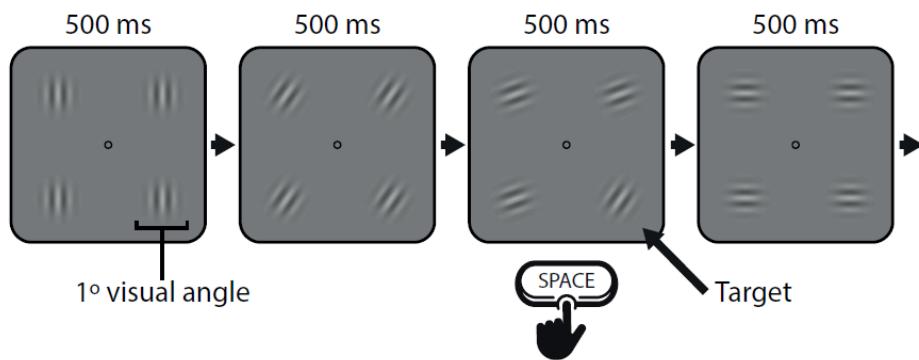
189 **2.1.2 – Design.** Task design reflected the core principles of classic experimental vigilance
190 paradigms (Baddeley & Colquhoun, 1969; Mackworth, 1948, 1950; Parasuraman & Davies, 1976).
191 Participants were asked to monitor four low-contrast gratings arranged squarely around a central fixation
192 circle. The gratings rotated synchronously in a clockwise ticking motion at a rate of 120 ticks per minute
193 (30° rotation per tick) and targets were defined as instances where one became briefly out of sync with
194 the others (i.e., it missed a tick: see Figure 1). The task lasted for 30 min, during which time continuous
195 monitoring was required. Six targets were presented every minute (180 overall) at pseudorandom
196 intervals, subject to the following constraints: 1) the time between targets was at least 6 s and at most 30
197 s, 2) targets did not occur within 2 s of the beginning or end of the task. Targets occurred equally often at
198 all of the four locations, although this was randomized across the whole experiment so that spatial
199 uncertainty as to the location of the target would contribute to task difficulty (Broadbent, 1958; Mackie,
200 1987; Warm, Parasuraman, & Matthews, 2008). All participants completed one trial of this experiment in
201 a single testing session lasting approximately 40 min.

202 **2.1.3 – Stimuli and apparatus.** Four gratings enveloped within a cosine window (spatial
203 frequency = 0.1, $SD = 12$, 39% contrast), each spanning $1^\circ \times 1^\circ$ of visual angle, were arranged in a square
204 pattern around a central fixation circle on a grey background (Figure 1). The gratings were generated
205 using an online tool² and saved in JPEG format. The task was administered on a 24-in. Ilyama monitor

² <https://www.cogsci.nl/gabor-generator>

PUPILLOMETRIC VIGILANCE DECREMENT

206 running at a resolution of 1024×768 (1:1 aspect ratio) with a refresh rate of 144 Hz, and button
207 responses were collected on a standard computer keyboard. The monitor and eye tracker were enclosed
208 such that the only direct illumination came from the display screen and the participant could not see
209 anything in their periphery. A viewing distance of 40 cm was maintained by a chin rest and forehead bar.
210 Using a colorimeter (ColorCAL MKII, Cambridge Research Systems), the surface luminance of the grey
211 background was recorded as 73.54 cd/m^2 , and the surrounding dark light of the unused portion of screen
212 as 0.53 cd/m^2 . Pupil size and gaze data were recorded monocularly (left eye) with an EyeLink 1000 (SR
213 Research, Mississauga, Ontario, Canada) system in tower mount configuration (32 mm lens) sampling at
214 250 Hz. According to the user manual, the system resolves pupil diameter to within 0.2 mm (SR Research
215 Ltd., 2010). Eye level and camera position remained constant throughout the recording session for each
216 participant. Stimulus presentation was managed with Experiment Builder (SR Research, Mississauga,
217 Ontario, Canada).



218
219 *Figure 1.* Stimuli and trial sequence for Experiment 1. Participants were asked to respond by pressing
220 space every time one of the gratings did not follow the standard sequence (6 times per min).

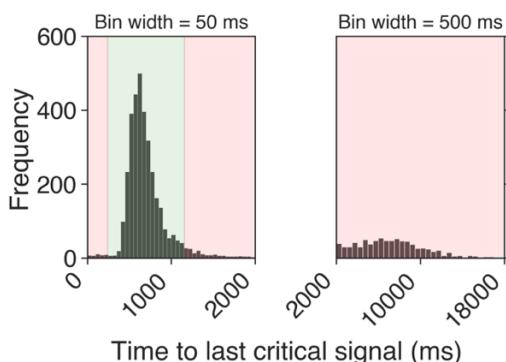
221 **2.1.4 – Procedure.** On arrival at the lab, participants were told that for the next 30 min they
222 would be required to complete a vigilance task that involved monitoring four circular patches rotating
223 with a ticking motion at the center of the screen. It was explained that, from time to time, one of the
224 patches would briefly become out of phase with the others, and that this was a target to which they had to
225 respond. Participants were not given any further information about the frequency or temporal and spatial

PUPILLOMETRIC VIGILANCE DECREMENT

226 uncertainty of the targets. Once comfortable with the definition of a target, they were instructed that their
227 task was to press the space bar every time they noticed such an event. Participants were forewarned that
228 the task was monotonous, but were asked to try and respond as quickly and accurately as possible. They
229 were also instructed to maintain central fixation on the screen. A 5-point calibration and validation routine
230 was performed prior to starting the experiment.

231 **2.1.5 – Task performance and pupillometry.** Task performance was assessed with RT, accuracy
232 (i.e., percent hits and false alarms), and the signal detection theory measures sensitivity (d') and response
233 bias (c). Hits, correct rejections, misses and false alarms were determined by an iterative algorithm which
234 assigned button responses to stimulus events. For each button response, the RT in milliseconds to the last
235 target was calculated. If the RT was greater than the minimum time between targets (6 s), or if no target
236 had yet been presented, the button response was allocated to the nearest elapsed neutral event and labelled
237 as a false alarm. All remaining button responses were then grouped together and a permissible range for
238 hits was determined as ± 2 median absolute deviations (Leys et al., 2013) from the group-level median
239 RT. Accordingly, all button responses that occurred within 225 to 1156 ms of targets were counted as
240 hits, and those with RTs outside this range, as previous, were allocated to the nearest neutral event and
241 counted as false alarms. The resulting distribution of RTs and the permissible hit range is illustrated in
242 Figure 2. Finally, targets without a valid button response were counted as misses, and all remaining
243 neutral events as correct rejections. This process of dealing with behavioral responses in sustained-
244 attention tasks with high event rates is similar to that used by Esterman et al. (2016) and van den Brink et
245 al. (2016). For the purposes of analysis, the complete experiment was decomposed into three 10 min
246 periods of watch.

PUPILLOMETRIC VIGILANCE DECREMENT



247
248 *Figure 2.* Distribution of RTs for Experiment 1 across all participants in 50 ms bins (left) and 500 ms bins (right). The green shaded area demarcates the permissible hit range (225 to 1156 ms after the target), and
249
250 the red shaded areas show the range where button responses were considered false alarms.

251 Individual pupil traces were extracted for all signal detection theory outcomes. For misses and
252 correct rejections the pupil data were time locked to the stimulus event and baseline pupil size was
253 defined as the average pupil size in a 500 ms window prior to the event. For hits and false alarms, the data
254 were time locked to the corresponding button event and the baseline period was offset by a further 500 ms
255 (i.e., from -1000 to -500 ms prior to the button response) to minimize contamination from fluctuations in
256 pupil size associated with movement preparation and execution (Einhäuser, Koch, & Carter, 2010; Hupé,
257 Lamirel, & Lorenceau, 2009; Martin, Whittaker, & Johnston, 2020; Richer & Beatty, 1985).

258 Baseline and task-evoked pupil measures were also derived from the pupil traces to probe the
259 effects of time-on-task. Baseline measures were defined as the average of the z -transform of pupil size
260 (across the whole experiment) in the baseline periods, and task-evoked measures were defined as the
261 average percent modulation in the peristimulus intervals (i.e., the portion of the pupil trace not including
262 the baseline). We also averaged the z -transform of pupil size into 1-min bins for an overarching look at
263 how pupil size varied across the whole task.

264 **2.1.6 – Data processing and statistical analysis.** Pupil data were processed and analyzed using
265 custom python scripts. Eye-blinks were detected using the standard EyeLink parsing algorithm and
266 reconstructed with linear interpolation prior to smoothing with a 3rd-order Butterworth filter (4 Hz cut-

PUPILLOMETRIC VIGILANCE DECREMENT

267 off). Frequently, pre- and post-blink samples were noticeably part of the blink artifact, so we extended the
268 blink endpoints by 100 ms in each direction. The average amount of data replaced by blink interpolation
269 across all participants that were included in the analysis was 6.5%. Pupil data were then down-sampled to
270 50 Hz, baseline corrected at the trial level with the subtractive procedure (Mathôt et al., 2018), and
271 converted to units of percent signal change.

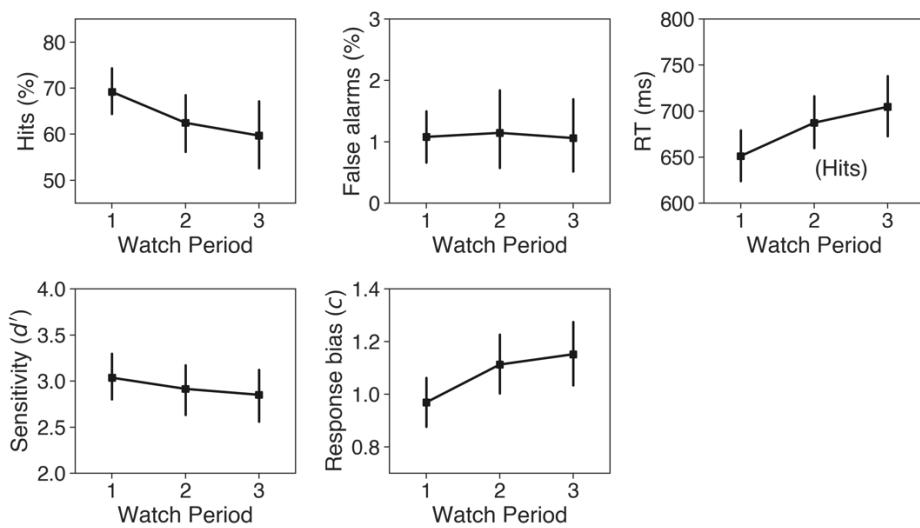
272 To examine the general pattern of pupil measures for each of the signal detection theory
273 outcomes, stimulus- (misses and correct rejections) and button-locked (hits and false alarms) pupil traces
274 from across the whole experiment were compared using two-tailed nonparametric permutation tests with
275 cluster-based correction for the multiple comparisons problem (Maris & Oostenveld, 2007). This
276 approach does not depend on theoretical assumptions about the data and reduces experimenter bias
277 associated with choosing a time-period over which to compute summary statistics. For the permutation
278 tests, *t*-tests were used to compare two conditions (significance thresholds for test statistics determined
279 theoretically from the appropriate degrees of freedom at $\alpha = .05$) and we follow the guidance of
280 Sassenhagen and Draschkow (2019) for reporting and interpretation. To examine time-on-task effects,
281 performance (RT, accuracy, d' , c) and scalar pupil measures were averaged within 10-min watch periods
282 and analyzed with repeated measures ANOVAs. Where Mauchly's *W* indicated that the assumption of
283 sphericity was violated, *p*-values were adjusted using the Greenhouse-Geisser correction. An alpha level
284 of .05 was used for all statistical tests. The mean and standard deviation of horizontal ($M = 527$, $SD = 20$)
285 and vertical ($M = 379$, $SD = 26$) gaze position for all samples included in the analysis indicate that
286 participants maintained steady fixation at the center of the screen throughout the task.

287 **2.1.7 – Exclusions.** Pupil data associated with stimulus and button events were discarded if the
288 participant blinked during the baseline period or if more than 25% of the data across the epoch of interest
289 were interpolated. Overall, this led to the discarding of pupil data for 17.89% of stimulus-locked (i.e.,
290 misses and correct rejections) epochs and 14.32% of button-locked (i.e., hits and false alarms) epochs. No
291 participants were excluded from the analysis.

292 **2.2 – Results.**

PUPILLOMETRIC VIGILANCE DECREMENT

293 **2.2.1 – Task performance.** The average number of button responses made during the task was
294 151 ($SD = 51$). With respect to target events, the average percentage of hits and false alarms was 63.77%
295 ($SD = 14.66\%$) and 1.09% ($SD = 1.43\%$), respectively. These data, together with the remaining behavioral
296 data, are summarized in Figures 2 and 3. Figure 2 shows the RT distribution for the whole experiment
297 including the permissible hit range (225 – 1156 ms post-target), and Figure 3 shows accuracy (percentage
298 of hits and false alarms), RT for hits, sensitivity (d') and response bias (c) as a function of Watch Period.
299 The average RT for all hits was 666 ms ($SD = 156$ ms). Average sensitivity and response bias across the
300 whole experiment were 2.93 ($SD = 0.67$) and 1.07 ($SD = 0.26$), respectively, indicating that perceptual
301 sensitivity to targets was good, but also that participants were generally biased to withhold responses to
302 targets.



303
304 Figure 3. Task performance as a function of Watch Period in Experiment 1: percentage of hits (top-left)
305 and false alarms (top-middle), RTs for hits (top-right), sensitivity (bottom-left) and response bias
306 (bottom-middle). Error bars reflect 95% confidence intervals (bootstrapped, 1000 iterations).

307 To examine the effects of time-on-task, one-factor (Watch Period) repeated measures ANOVAs
308 were conducted for all performance measures. First, the percentage of hits and false alarms were
309 analyzed. There was a significant main effect of Watch Period on the percentage of hits, $F(2, 54) = 6.14$,

PUPILLOMETRIC VIGILANCE DECREMENT

310 $p = .004$, $\eta_p^2 = 0.19$, with Bonferroni-corrected t -tests showing that participants attained a significantly
311 higher percentage of hits in Watch Period 1 ($M = 69.2\%$, $SD = 14.0\%$) compared to Watch Period 3 ($M =$
312 59.6% , $SD = 19.7\%$), $t(28) = 2.96$, $SEM = 3.21$, $p = .019$. The difference between Watch Period 1 and
313 Watch Period 2 ($M = 62.46\%$, $SD = 16.7\%$) was marginally significant, $t(28) = 2.37$, $SEM = 2.83$, p
314 = .076, and the remaining comparison (2 vs. 3) was not significant ($p = .662$). The effect of Watch Period
315 on the percentage of false alarms was not significant ($p = .894$).

316 ANOVA on the RT data for hits revealed a significant main effect of Watch Period, $F(2, 54) =$
317 16.51 , $p < .001$, $\eta_p^2 = 0.38$. Post hoc analysis with Bonferroni adjustment confirmed that RT was
318 significantly faster in the first Watch Period ($M = 651$ ms, $SD = 79$ ms) compared to the second ($M = 687$
319 ms, $SD = 78$ ms), $t(28) = 4.95$, $SEM = 7.31$, $p < .001$, and the third ($M = 704$ ms, $SD = 89$ ms), $t(28) =$
320 5.47 , $SEM = 9.79$, $p < .001$, but that there was no significant difference in RT between Watch Period 2
321 and Watch Period 3 ($p = .384$).

322 The same analysis was repeated for the signal detection measures sensitivity (d') and response
323 bias (c). There was no significant main effect of Watch Period on sensitivity ($p = .193$), indicating that
324 participants' ability to discriminate targets did not change throughout the task, but there was a significant
325 main effect on response bias, $F(1.64, 44.31) = 7.38$, $p = .003$, $\eta_p^2 = 0.22$. Bonferroni-corrected post hoc t -
326 tests showed that response bias in Watch Period 1 ($M = 0.96$, $SD = 0.26$) was significantly lower than it
327 was in Watch Period 2 ($M = 1.11$, $SD = 0.31$), $t(28) = 2.8$, $SEM = 0.05$, $p = .028$, and Watch Period 3 (M
328 = 1.15 , $SD = 0.33$), $t(28) = 3.11$, $SEM = 0.06$, $p = .013$, but that there was no significant difference
329 between Watch Period 2 and Watch Period 3 ($p = .937$). This suggests that participants became more
330 conservative as the task progressed and were therefore more reluctant to report that a target was present.
331 As predicted, these performance data are consistent with the classic vigilance decrement.

332 **2.2.2 – Pupil data.** The z -transform of pupil data declined sharply in the first few minutes of the
333 task and then increased steadily until the end. A cluster in the observed data extending from the 4th to the
334 12th minute differed significantly from the population mean (top panel, Figure 4). This time-on-task effect
335 is well-noted in the literature for many different types of experiment (Fried et al., 2014; Hopstaken et al.,

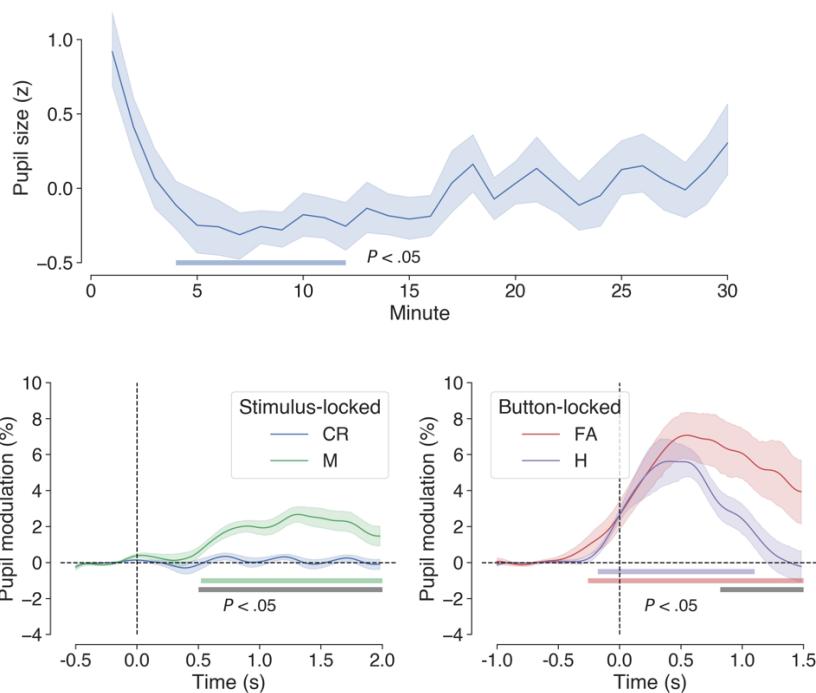
PUPILLOMETRIC VIGILANCE DECREMENT

336 2016; Hopstaken, Van der Linden, et al., 2015; Massar et al., 2016; McIntire et al., 2014; Unsworth &
337 Robison, 2016; van den Brink et al., 2016) and may reflect changes in overall arousal state or, more
338 specifically, the transition from phasic to tonic modes of LC output (Joshi & Gold, 2020).

339 Event-related pupil data were time-locked to button events for hits and false alarms and to
340 stimulus events for misses and correct rejections. This was to ensure the comparability of pupil data that
341 were consistently affected by motor acts. The grand-average pupil traces for each of these behavioral
342 outcomes are shown in Figure 4. Both button-locked outcomes (hits and false alarms) showed the usual
343 pattern of pupil modulation associated with the preparation and execution of motor responses (e.g., see
344 Einhäuser et al., 2010; Hupé et al., 2009; Martin et al., 2020; Richer & Beatty, 1985), with dilation
345 beginning up to 500 ms before the motor act and peaking shortly afterwards. Permutation tests revealed
346 significant modulation from baseline for hits and false alarms, as well as a significant difference between
347 these two outcomes (lower-right panel of Figure 4). The differences between the two traces can be
348 summarized as follows. For hits, there was an average pupil modulation of 2.04% and a peak modulation
349 of 5.62% with a latency of 400 ms from the button press, whereas for false alarms these values were
350 3.55%, 7.07%, and 540 ms, respectively. For the stimulus-locked pupil measures (misses and correct
351 rejections), only misses resulted in significant modulation from baseline in the poststimulus period, and
352 there was also a significant difference in pupil modulation between misses and correct rejections (bottom-
353 left panel of Figure 4). For misses, there was an average modulation of 1.25% and a peak modulation of
354 2.68% with a latency 1320 ms. The time-course for correct rejections resembled a flattened sine wave
355 (average modulation of -0.05%, peak modulation of 0.34%) in phase with the onset of stimulus events.
356 This periodic pattern is redolent of van den Brink et al.'s (2016) pupillometry data, which were observed
357 in a task with a similar event-related design. We attribute this to task-correlated blinking and the blink-
358 induced pupillary response (Knapen et al., 2016).

359

PUPILLOMETRIC VIGILANCE DECREMENT



360

361

362 *Figure 4.* Grand average pupil data in Experiment 1. The top panel shows the average z -transformed pupil
363 data between subjects in 1-min bins across the whole experiment. A cluster extending from the 4th to the
364 12th minute differed significantly from the population norm (blue colored bar). The bottom panels show
365 stimulus- (left: correct rejections and misses) and button-locked (right: hits and false alarms) averages
366 expressed as %-change from baseline, with colored horizontal bars indicating clusters of significant
367 modulation from baseline and gray bars showing significant differences between traces (1024
368 permutations, $p < .05$, cluster-corrected for multiple comparisons). Shaded areas surrounding the pupil
369 traces denote the standard error of the mean (SEM: bootstrapped, 5000 iterations).

370 After examining the stimulus- and button-locked pupil traces across the whole experiment we
371 probed the effects of time-on-task by analyzing the scalar representations of pupil data with two-way
372 (Outcome \times Watch Period) repeated measures ANOVA. These data are displayed in Figure 5. For the
373 stimulus-locked (i.e., misses and correct rejections) baseline pupil data, there were no significant effects
374 of Watch Period ($F[2, 54] = 1.99, p = 0.146$) or Outcome ($F[1, 27] = 0.12, p = .737$), and the Outcome \times

PUPILLOMETRIC VIGILANCE DECREMENT

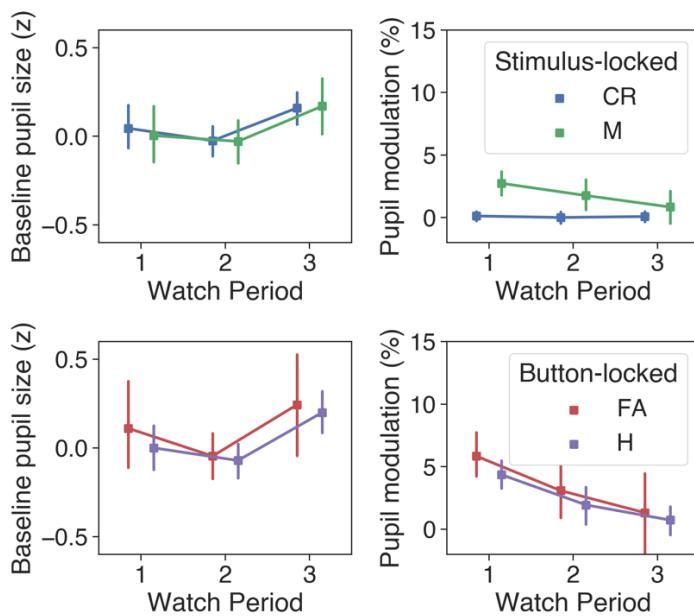
375 Watch Period interaction was not significant ($F[1.57, 42.39] = 0.37, p = .644$). For the stimulus-locked,
376 task-evoked pupil modulations there was a significant main effect of Outcome, $F(1, 27) = 37.68, p < .001$,
377 $\eta_p^2 = 0.58$, but the main effect of Watch Period ($F[2, 54] = 2.10, p = .132$) and the Outcome \times Watch
378 Period interaction ($F[2, 54] = 2.01, p = .144$) were not significant. Simple main effects analysis showed
379 that misses resulted in greater pupil modulation than correct rejections during Watch Period 1 ($F = 30.69$,
380 $p < .001$) and Watch Period 2 ($F = 8.02, p = .009$), but not during Watch Period 3 ($F = 1.20, p = 0.283$).

381 For button-locked (i.e., hits and false alarms) baseline pupil data, the main effect of Watch Period
382 was not significant, $F(1.58, 33.25) = 2.32, p = 0.124$, and the effect of Outcome was marginally
383 significant, $F(1, 21) = 4.03, p = .058$, with baseline pupil size being greater on average for false alarms
384 compared to hits. The Outcome \times Watch Period interaction did not significantly effect baseline pupil size,
385 $F(2, 42) = 0.141, p = .869$. For the button-locked pupil modulations, the main effect of Outcome was not
386 significant, $F(1, 21) = 0.004, p = .953$, but there was a significant main effect of Watch Period $F(2, 42) =$
387 $8.86, p < .001, \eta_p^2 = 0.3$. Post hoc *t*-tests revealed that the average %-modulation for Watch Period 1 was
388 greater than it was for Watch Period 2 ($MD = 2.32, t[22] = 2.95, p = .023$) or Watch Period 3 ($MD = 4.04$,
389 $t[22] = 3.74, p = .004$), but that Watch Period 2 and Watch Period 3 did not differ significantly ($p = .297$).
390 The Outcome \times Watch Period interaction for button-locked pupil modulations was not significant
391 ($F[1.52, 31.87] = 0.09, p = .861$).

392 Overall, these patterns in the pupil data are consistent with the prediction that the magnitude of
393 task-evoked responses will mirror behavioral performance and decline as time-on-task increased.

394 **2.2.3 – Correlational analyses.** Across all button responses included in the analysis, RT did not
395 significantly correlate with prestimulus baseline pupil size, $r(4102) = -.009, p = .559$, or task evoked pupil
396 size, $r(4102) = -.021, p = .073$. In line with previous literature (e.g., see de Gee et al., 2014), there was a
397 significant negative correlation between baseline and task evoked pupil size $r(4102) = -.273, p < .001$.
398

PUPILLOMETRIC VIGILANCE DECREMENT



399

400 *Figure 5.* Stimulus- (M: misses; CR: correct rejections) and button-locked (H: hits; FA: false alarms)
401 pupil averages across each Watch Period in Experiment 1. Error bars reflect 95% confidence intervals
402 (bootstrapped, 1000 iterations).

403 **2.3 – Discussion.** This experiment examined the relationship between task performance and pupil
404 size in a prolonged, uninterrupted vigilance task with visually presented stimuli. Participants monitored
405 four centrally located equiluminant gratings continuously for 30 min under the instruction to respond by
406 pressing the space bar every time they detected a target. The task required successive discrimination
407 (Parasuraman, 1979), but the high background event rate, the brief target duration, and the temporal and
408 spatial uncertainty of the targets added elements of difficulty (Broadbent, 1958; Mackie, 1987; Warm et
409 al., 2008). In line with robust trends in the literature, we predicted that task performance and the
410 magnitude of task-evoked pupillary responses would decrease as time-on-task increased.

411 Key behavioral measures were indicative of a classic vigilance decrement. Both the percentage of
412 hits and the RT for hits changed across each successive 10 min watch period in a manner reflecting
413 declining vigilance. These findings are consistent with well-established findings in the literature regarding
414 the effects of time-on-task on detection performance under conditions of prolonged monitoring (e.g.,

PUPILLOMETRIC VIGILANCE DECREMENT

415 Broadbent, 1953; Broadbent & Gregory, 1965; Buck, 1966; Mackworth, 1948, 1950, Parasuraman &
416 Davies, 1976, 1982; Warm et al., 2008). The signal detection measures, sensitivity (d') and response bias
417 (c), were calculated to gain further insight into the cause of the declining percentage of hits. Given that
418 the nature of the task was in making trivially easy judgements about suprathreshold stimuli, it is not
419 surprising that sensitivity remained at ceiling throughout. However, there was a conservative shift in
420 response bias, suggesting that the decline in accuracy was linked to the participants becoming less willing
421 to report a detection, rather than a diminishing ability to discriminate targets from nontargets (Green &
422 Swets, 1974). This is consistent with previous reports that the vigilance decrement in tasks with high
423 event rates is more closely related to changes in the strictness of the decision criterion over time, rather
424 than perceptual sensitivity (e.g., Baddeley & Colquhoun, 1969; Broadbent, 1971; Colquhoun, 1961;
425 Parasuraman & Davies, 1976).

426 Event-related pupil data were extracted for all signal detection outcomes in order to gain insight
427 into the cognitive processing associated with these events. For misses and correct rejections, data were
428 time-locked to the onset of the relevant stimulus event (bottom-left panel of Figure 4). Notably, misses
429 resulted in reliably greater pupil dilation than correct rejections, a similar observation to that made by
430 Beatty (1982) in his auditory vigilance experiment. As Beatty (1982) suggested, from a signal detection
431 perspective, an enhanced pupillometric response to missed targets may reflect increased processing of
432 sensory information for stimuli that fall close to the decision criterion. In this vein, the pupil dilation
433 following missed targets may in part reflect subconscious processing of the target stimuli (Laeng et al.,
434 2012). However, in the context of this experiment, such an interpretation must be tempered against the
435 possibility that pupil modulation for misses was linked to the detection, decision and motor effects of
436 neighboring button-presses falling just outside of the permissible hit range. Previous research indicates
437 that the pupillometric effects associated with motor acts can begin to emerge up to 1000 ms prior to the
438 act itself (e.g., Einhäuser et al., 2010; Hupé et al., 2009; Martin et al., 2020; Richer & Beatty, 1985),
439 which means that false alarms whose RT fell just outside the permissible hit range (upper bound of 1156
440 ms) may have contributed to the pupil dilation for missed targets. For hits and false alarms, pupil data

PUPILLOMETRIC VIGILANCE DECREMENT

441 were time-locked to the button response and showed typical patterns of modulation associated with motor
442 preparation and execution. Pupil dilation was also significantly greater for false alarms compared to hits
443 (lower-right panel of Figure 4). As suggested by Murphy et al. (2011), who had similar findings, this
444 difference may reflect the cognitive effects of a self-regulatory performance monitoring process.
445 However, the larger pupil dilation for false alarms may also be associated with the higher degree of
446 uncertainty that accompanies these events compared to correct detections (Yu & Dayan, 2005).

447 To examine the effects of time-on-task on pupil dynamics, scalar values of baseline and task-
448 evoked pupil size were calculated for all stimulus and button events in each third of the task (Figure 5).
449 As with Beatty (1982), baseline pupil size for all outcomes was relatively unchanged across the duration
450 of the task, suggesting that the mode of organismic activation linked to fluctuations in tonic pupil
451 diameter was not related to the central processes underpinning the vigilance decrement. In contrast to this,
452 however, task-evoked pupil modulations for misses, hits and false alarms exhibited a marked decline in
453 magnitude across each successive Watch Period. This pattern of change parallels the decline in vigilance
454 indexed by the percentage of hits and RT for hits, and is therefore consistent with the findings from the
455 vigilance experiments of Beatty (1982) and Murphy et al. (2011), as well as various other pupillometric
456 studies of tasks requiring sustained attention (e.g., Hopstaken, van der Linden, et al., 2015; McIntire et al.,
457 2014; Unsworth & Robison, 2016).

458 Task-evoked pupillary responses have been linked to phasic activation of the LC-NA system by
459 neurophysiological and behavioral studies in both human and non-human primates (e.g., Alnaes et al.,
460 2014; Aston-Jones & Cohen, 2005; Beatty, 1982; de Gee et al., 2017; Einhäuser, Stout, Koch, & Carter,
461 2008; Gilzenrat et al., 2010; Jepma & Nieuwenhuis, 2011; Joshi, Li, Kalwani, & Gold, 2016; Murphy,
462 O'Connell, O'Sullivan, Robertson, & Balsters, 2014; Murphy et al., 2011; Phillips, Szabadi, & Bradshaw,
463 2000; Rajkowski, Kubiak, & Aston-Jones, 1993; Thorne et al., 2005; Urai, Braun, & Donner, 2017;
464 Varazzani, San-Galli, Gilardeau, & Bouret, 2015). Further, single-unit recording studies in animals have
465 found that phasic activation of the LC-NA system occurs typically in response to task-related events
466 during periods of high performance (e.g., Aston-Jones, Chiang, & Alexinsky, 1991; Aston-Jones,

PUPILLOMETRIC VIGILANCE DECREMENT

467 Rajkowski, Kubiak, & Alexinsky, 1994; Rajkowski, Kubiak, & Aston-Jones, 1994). For example, the
468 study by Aston-Jones et al. (1994) revealed that noradrenergic neurons in monkey LC are phasically
469 activated by infrequent target cues during a vigilance task, and also that the amplitude of these phasic
470 responses diminishes over time. With respect to the well-known functional association between task-
471 evoked pupil responses and phasic LC activation (Laeng et al., 2012), the findings from these animal
472 studies fit well with those from the present study.

473 From a theoretical standpoint, it is difficult to ascertain whether the findings of this study fit best
474 with a resource depletion, mind wandering, or resource control-failure account of the vigilance decrement
475 (see Caggiano & Parasuraman, 2004; Smallwood & Schooler, 2006; Thomson, Besner, & Smilek, 2015).
476 Key indicators would be whether participants experienced the task as being effortful and the extent to
477 which they engaged in task-unrelated thought, but we did not obtain these data as it would have required
478 the use of intermittent thought probes (e.g., Hopstaken et al., 2015; Smallwood et al., 2004; Unsworth &
479 Robison, 2016), which involve temporary disengagement and therefore undermine a key aspect of
480 classical vigilance task design—the requirement for continuous monitoring (Parasuraman & Davies,
481 1976). Thought probes may also serve as ‘mini breaks’, which can disrupt task monotony and alleviate
482 the vigilance decrement (Ariga & Lleras 2011; Ralph et al., 2016; Ross et al., 2014). In the absence of
483 subjective reports, only the pupil data and the nature of the task can serve as a basis for inferring the cause
484 of the vigilance decrement. First, the task itself was prolonged and monotonous, and participants were
485 without a strong incentive to maintain high levels of vigilance—conditions that provide fertile grounds
486 for mind wandering (Smallwood & Schooler, 2006). Second, the magnitude of task-evoked dilations
487 reduced across the course of the task, an effect that has previously been linked to disengagement
488 (Hopstaken, van der Linden, et al., 2015) and mind-wandering (Smallwood et al., 2011). Taken together,
489 this suggests that the vigilance decrement in the current experiment may have been linked primarily to
490 mind-wandering, or a resource control-failure leading to thought intrusion, but further data would be
491 required to confirm this.

PUPILLOMETRIC VIGILANCE DECREMENT

492 In sum, the present study replicated the well-known vigilance decrement—the reduction in
493 detection performance that takes place during conditions of prolonged and continuous monitoring.
494 Mirroring this behavioral effect, task evoked pupil responses declined across the duration of the task, but
495 baseline pupil size was mostly unchanged, suggesting that the vigilance decrement may have been linked
496 to gradual disengagement of attention as the task progressed, rather than a change in organismic arousal
497 state.

3 – Experiment 2

499 Traditional experimental vigilance tasks aim to emulate the conditions of real-world operator
500 settings, but another form of vigilance task—the Psychomotor vigilance task (PVT: Wilkinson &
501 Houghton, 1982)—aims to quickly assess declines in vigilant attention associated with sleep loss,
502 circadian factors and other environmental stressors (Basner & Dinges, 2011; Basner, Mollicone, &
503 Dinges, 2011; Blatter et al., 2006; Caldwell, Prazinko, & Caldwell, 2003; Dinges et al., 1997; Van
504 Dongen & Dinges, 2005; Graw et al., 2004). In a PVT, instead of responding to infrequent signals over a
505 prolonged period, subjects must make speeded responses to more regular signals occurring at
506 pseudorandom intervals over a short period of time, usually 10 min or less. Wilkinson and Houghton's
507 (1982) original version of this task was administered on a small hand-held battery-powered device
508 displaying a millisecond counter set to '000'. The subject held the device and quickly pressed a button
509 every time the counter began to increment, which happened at intervals ranging between 1-10 s. Upon
510 detection of a response, the timer froze for 1.5 s, and the RT was saved before the timer reset to '000'. A
511 variety of performance metrics can be derived from the data produced by this task, but analysis
512 commonly focuses on mean and median RT, the fastest and slowest 10% of trials, and the proportion of
513 'lapses', which are usually defined as RTs greater than 500 ms (Basner & Dinges, 2011).

514 In Experiment 2 we sought to examine how pupil measures relate to PVT task performance, but
515 with a novel stimulus approach optimized for pupillometry. Most PVTs utilize the prototypical stimulus
516 of a running millisecond timer that counts up from zero, but as noted by Thorne et al. (2005), this may
517 have undesirable consequences. From a behavioral and pupillometric perspective, the two most relevant

PUPILLOMETRIC VIGILANCE DECREMENT

518 points made by Thorne et al. are as follows. First, the intensity of the stimulus changes in a nonlinear
519 fashion as the running counter increases, which could be a source of increasing variance in both
520 behavioral and pupil measures. Second, the counter provides feedback to participants whether feedback is
521 desired or not, which may enable them to monitor their own performance during the session and increase
522 attention or effort to compensate for a noticed decline—a boon that would typically not be available in
523 real-world settings. To avoid these issues, Thorne et al. (2005) devised a version of the PVT with a
524 luminance-based graphic stimulus comprising two alternating black and white circular annuli resembling
525 a target or “bull's eye”. This PVT was administered on a dedicated hand-held device and produced highly
526 comparable results despite its different stimulus characteristics.

527 Since a luminance-based stimulus such as that used by Thorne et al. (2005) would trigger the
528 pupillary light reflex, the RT-initiating stimulus opted for in the present PVT experiment was a change in
529 the orientation of a low-contrast grating. Participants were instructed to monitor a grating at the center of
530 a screen and respond as quickly as possible by pressing space whenever it flipped on its side (i.e., when it
531 rotated 90°). As in other PVTs, the vigilance element of the task was instantiated with time-on-task (~13
532 min) and ISI (4-12 s) parameters. Due to the use of a low intensity stimulus, we expected that RTs in the
533 current PVT would be slower on average than for PVTs using a running counter stimulus. However,
534 because our approach avoids the confounds of variable stimulus intensity and feedback, behavioral and
535 pupil measures should more faithfully reflect changes in vigilant attention. Based on the general findings
536 outlined in the introduction we predicted that time-on-task, both within and between successive blocks of
537 the PVT, would lead to declining performance and a decrease in pupil size. Also, following the findings
538 of Kristjansson, Stern, Brown, and Rohrbaugh (2009) and Unsworth and Robison (2016), who employed
539 similar tasks, we expected that poorest performance would be associated with smaller pupils at baseline.

540 **3.1 – Materials and methods.**

541 **3.1.1 – Participants.** Twenty-five participants (18 females; age range 18-36 years, $M = 22.96$, SD
542 = 4.65) completed the experiment voluntarily or in exchange for course credit. All participants were
543 students at Swansea University reporting normal or corrected-to-normal acuity and color vision. The

PUPILLOMETRIC VIGILANCE DECREMENT

544 experimental protocol was approved by the Department of Psychology Ethics Committee at Swansea
545 University and the Ministry of Defence Research Ethics Committee. Written informed consent was
546 obtained from each participant.

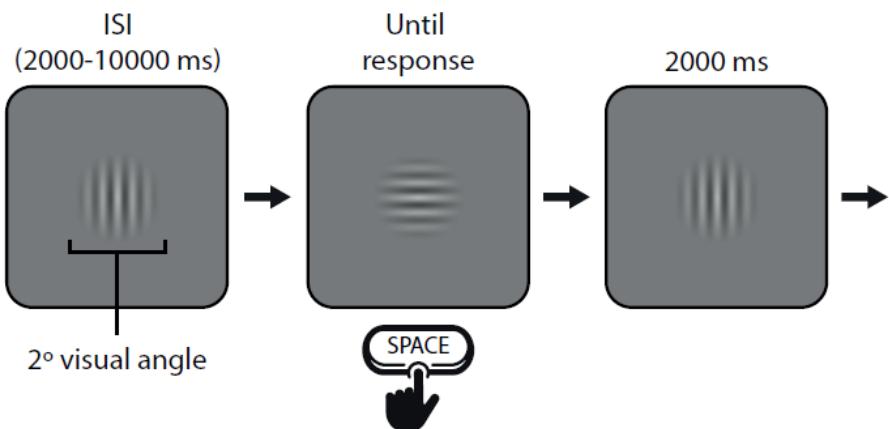
547 **3.1.2 – Design.** Performance and pupil measures were analyzed in a repeated measures design as
548 a function of Trial Group (1, 2, 3, 4, 5: sets of 18 contiguous trials within a PVT block) and Block (1, 2,
549 3: successive blocks of 90 trials)—two factors whose purpose was to capture within- and between-block
550 effects of time-on-task. The ISI, defined as the period between the last button response and the next flip of
551 the grating, varied between 4-12 s. This period included a fixed component of 2000 ms and a random
552 component varying between 2-10 s. The random component was constrained such that a third of the
553 intervals would be short (2000-4666 ms), a third medium (4666-7333 ms), and a third long (7333-10000
554 ms). The grating was present at the center of the screen throughout the task.

555 **3.1.3 – Stimuli and apparatus.** The stimulus was a grating enveloped within a cosine window
556 (spatial frequency = 0.1, $SD = 12$, 39% contrast) at the center of the screen, spanning $2^\circ \times 2^\circ$ of visual
557 angle. It was generated using the same process as specified for Experiment 1. All other details relating to
558 hardware and software were the same as for Experiment 1.

559 **3.1.4 – Procedure.** Participants completed 3 consecutive blocks of the PVT in a single testing
560 session, taking a forced break of only 1 min between blocks. In each block, participants were instructed to
561 monitor the ‘circular stimulus’ at the center of the screen and to respond as quickly as possible by
562 pressing the space bar whenever it flipped on its side, thereby to reset it to its original position (see Figure
563 6). Each block lasted approximately 13 min ($M = 12.8$, $SD = 0.63$), with some small variability arising
564 from differences in RT and the random element of the ISI. Continuous recordings of gaze position and
565 pupil data were obtained for each trial, and RT was defined as the time in milliseconds between the flip of
566 the grating and the subsequent button response. A 5-point calibration and validation routine was
567 performed at the start of each block. The task was performed in a dimly lit room.

568

PUPILLOMETRIC VIGILANCE DECREMENT



569

570 *Figure 6.* Stimuli and trial sequence for Experiment 2. Participants monitored a grating and responded by
571 pressing space every time it flipped on its side (every 4-12 s).

572 **3.1.5 – Data processing and statistical analysis.** Pupil data were processed using the same
573 general approach as described for Experiment 1. The average amount of data replaced by blink
574 interpolation across all participants included in the analysis was 6.82%. Segments of pupil data 2500 ms
575 in length were extracted for each trial, time-locked to the button response (-1000 to 1500 ms). These data
576 were expressed as %-modulation from a baseline defined as the average pupil size in a 500 ms period
577 prior to the RT-initiating stimulus event.

578 To assess task performance we focused on 1/RT and lapse frequency, which are among the most
579 sensitive measures of alertness in PVTs (Basner & Dinges, 2011). Lapses in PVTs are traditionally
580 defined as RT greater than 500 ms but due to our novel take on the task we defined lapses as RT greater
581 than two median absolute deviations (Leys et al., 2013) from each participant median, which resulted in
582 an average lapse threshold of 585 ms ($SD = 109$) across participants. Our pupil measures of interest were
583 baseline pupil size and the task-evoked pupil response. Baseline pupil size was defined as the average z -
584 transform of pupil size in the baseline period, whereas the task-evoked pupil response was defined as the
585 average percentage of pupil modulation around the time of the button response (-500 to 1500 ms). The
586 data for each of these four variables were analyzed separately using two-factor (Trial Group \times Block)
587 repeated measures ANOVA. Where Mauchly's W indicated that the assumption of sphericity was

PUPILLOMETRIC VIGILANCE DECREMENT

588 violated, p -values were adjusted using the Greenhouse-Geisser correction. We conducted further analyses
589 on pupil measures for the fastest and slowest 20% RTs to determine whether pupil size at baseline was
590 indicative of faster or slower detection responses (Kristjansson et al., 2009; Unsworth & Robison, 2016),
591 and more generally how the extremes of performance are reflected in the pupil data. The mean and
592 standard deviation of horizontal ($M = 518$, $SD = 10$) and vertical ($M = 388$, $SD = 14$) gaze position for all
593 samples included in the analysis indicate that participants maintained steady fixation at the center of the
594 screen throughout the task.

595 **3.1.6 – Exclusions.** Two participants were excluded from the analysis for yielding poor quality
596 pupil data (both had over 50% interpolated data for baselines and over 60% interpolated data for task-
597 evoked responses). The general pattern of results was the same both with and without the exclusion of
598 these participants. For the pupil analyses, trials were excluded if there was a blink in the baseline or if
599 more than 25% of data were interpolated across the whole epoch (28.06% of trials).

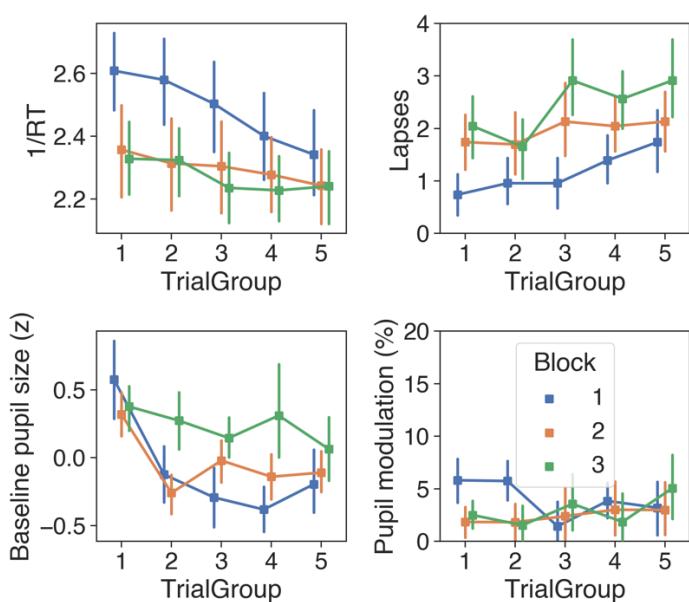
600 3.2 – Results.

601 **3.2.1 – Task performance.** Average RT across all participants was 420 ms ($SD = 79$) for non-
602 lapse trials and 960 ms ($SD = 1463$) for lapse trials. ANOVA on mean 1/RT (i.e., the reciprocal transform
603 of RT) revealed a significant Trial Group \times Block interaction, $F(8, 176) = 3.16, p = .002, \eta_p^2 = 0.13$, the
604 nature of which is illustrated in the top-left panel of Figure 7. Simple effects showed that 1/RT decreased
605 significantly across Trial Group in each Block (Block 1: $F = 16.65, p < .001$; Block 2: $F = 2.89, p = .027$;
606 Block 3: $F = 3.40, p < .012$), which is consistent with the prediction that performance would decline as
607 time-on-task increased. Post hoc analysis with Bonferroni adjustment showed that, in Trial Groups 1-3,
608 1/RT was significantly greater in Block 1 compared to Blocks 2 and 3 (all $p < .05$) and that 1/RT in Trial
609 Group 4 was significantly greater for Block 1 compared to Block 3 ($p < .05$). No other comparisons were
610 significant ($p > .05$). Therefore, as indexed by 1/RT, performance was best overall in Block 1 compared
611 to Block 2 and Block 3, but the magnitude of this effect decreased across Trial Groups.

612 Average lapse frequency across participants was 27.6 ($SD = 7.8$). ANOVA showed that the main
613 effect of lapse frequency was significant for Trial Group, $F(4, 88) = 5.43, p < .001, \eta_p^2 = 0.20$, and for

PUPILLOMETRIC VIGILANCE DECREMENT

614 Block, $F(2, 44) = 16.74, p < .001, \eta_p^2 = 0.43$ (top-right panel of Figure 7), but that the Trial Group \times
615 Block interaction was not significant ($F[8, 176], p = .541$). Simple effects for Trial Group showed that the
616 number of lapses increased significantly throughout Block 1 ($F = 3.16, p = .018$) and Block 3 ($F = 3.01, p$
617 = .022), but not Block 2 ($F = 0.60, p < .662$). Lapse frequency therefore followed the same general pattern
618 as 1/RT, and together these data are consistent with the prediction that performance would decline as
619 time-on-task increased.

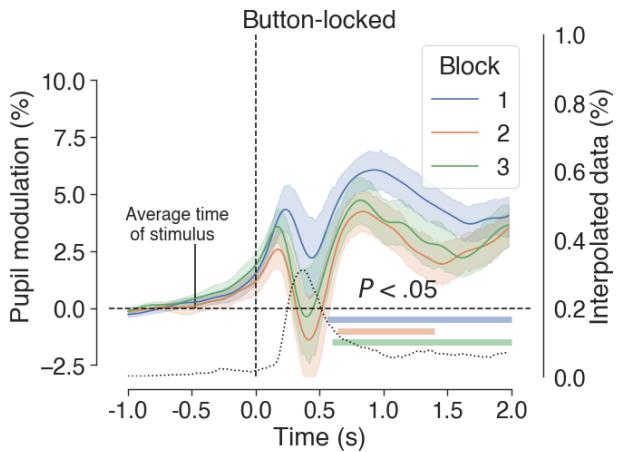


620
621 *Figure 7.* Performance (top row) and pupil (bottom row) measures across Trial Group and Block in
622 Experiment 2, with error bars showing 95% confidence intervals (bootstrapped, 1000 iterations).

623 **3.2.2 – Pupil data.** Grand-average button-locked pupil traces for each Block are shown in Figure
624 8. The pupil began to dilate slowly following the stimulus event and then rapidly after the button-press. In
625 the 1500 ms following the button-press there was an average modulation of 5.22% and a peak latency of
626 880 ms. A conspicuous trough in the pupil traces after the button-press coincides with a transient but
627 marked increase in the percentage of interpolated data. This artifact resembles the blink-induced pupillary
628 response (e.g., Knapen et al., 2016) and is therefore indicative of task-correlated blinking (i.e.,

PUPILLOMETRIC VIGILANCE DECREMENT

629 participants tended to blink after button presses). We did not correct this artifact with linear interpolation
630 as it would involve altering too much data and excluding more trials.



631
632 *Figure 8.* Average button-locked pupil traces for each Block in Experiment 2. The black dotted trace
633 shows the percentage of interpolated data, which indicates task-correlated blinking. Shaded areas
634 surrounding the colored traces show the SEM (bootstrapped, 5000 iterations) and colored horizontal bars
635 denote clusters of significant modulation from baseline, as revealed by nonparametric permutation tests
636 (1024 permutations, $p < .05$, cluster-corrected for multiple comparisons).

637 ANOVA on the baseline pupil measures revealed a significant Trial Group \times Block interaction,
638 $F(3.81, 76.10) = 3.47, p = .013, \eta_p^2 = 0.15$, which is displayed in the bottom-left panel of Figure 7. Simple
639 effects analysis for Trial Group showed that baseline pupil size decreased significantly across Block 1 (F
640 $= 13.57, p < .001$) and Block 2 ($F = 7.02, p < .001$), but not Block 3 ($F = 0.89, p = .470$). Bonferroni-
641 corrected post hoc *t*-tests revealed that baseline pupil size was significantly greater in Trial Group 4 for
642 Block 3 compared to Block 1 ($p = .007$), but no other between Block comparisons were significant (p
643 $> .05$).

644 For measures of pupil modulation, there was a significant main effect of Block, $F(2, 42) = 5.93, p$
645 $= .010, \eta_p^2 = 0.23$, but the effect of Trial Group ($F[4, 80] = 1.86, p = .125$) and the Trial Group \times Block
646 interaction ($F[4.83, 96.63] = 1.87, p = .109$) were not significant. Simple main effects showed that pupil

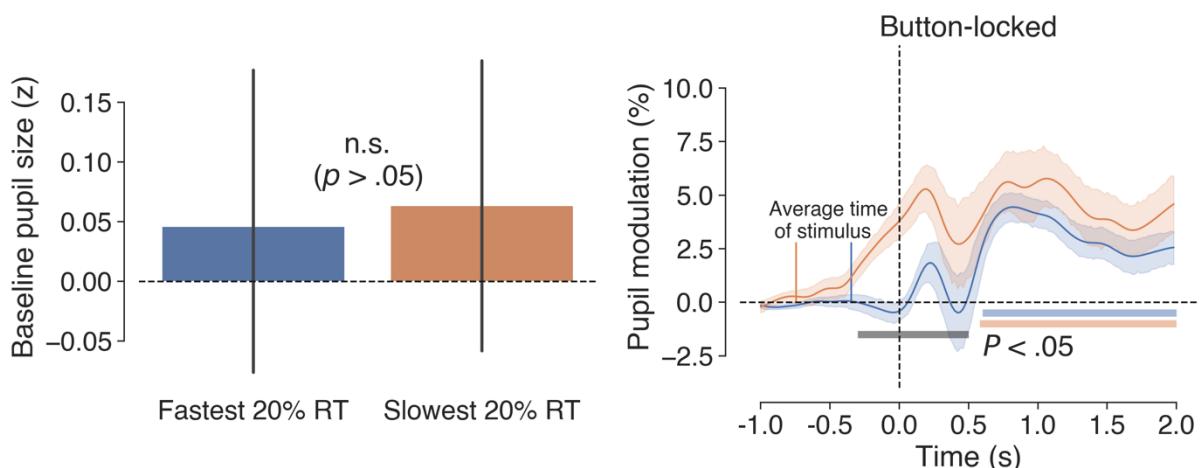
PUPILLOMETRIC VIGILANCE DECREMENT

647 modulation was greater in Block 1 for Trial Group 1 ($F = 8.97, p < .001$) and 2 ($F = 9.07, p < .001$), but
648 not for Trial Groups 3 to 5 ($p > .05$).

649 **3.2.3 – Correlational analyses.** Across all trials included in the analysis, RT correlated
650 significantly with task evoked pupil size, $r(4466) = -.159, p < .001$, but not with baseline pupil size,
651 $r(4466) = -.011, p = .459$. The classic negative correlation between baseline and task evoked pupil size
652 (e.g., see de Gee et al., 2014) was also present, $r(4466) = -.366, p < .001$.

653 **3.2.4 – Fastest vs. slowest RTs.** To explore how the extremes of performance are reflected in the
654 pupil data we conducted further analysis on trials with the fastest ($M = 347$ ms, $SD = 43$ ms) and slowest
655 ($M = 746$ ms, $SD = 43$ ms) 20% RTs. Figure 9 displays the pupillometry results for these extreme
656 quintiles. Baseline pupil size did not differ significantly ($p > .05$, left panel of Figure 9), but there was a
657 significant difference in the button-locked pupil traces ($p < .05$, cluster-corrected permutation test, right
658 panel of Figure 9) marked by a cluster spanning the button event. This difference clearly pertained to the
659 timing and magnitude of pupil dilation. For the slowest RTs, dilation began prior to the button response,
660 whereas for the faster RTs dilation did not begin until afterwards. Further, the average modulation in the
661 2000 ms post-button period was greater on average for the slowest ($M = 4.39\%, SD = 5.46\%$) than for the
662 fastest RTs ($M = 2.45\%, SD = 3.23\%$). These data do not corroborate previous reports of pretrial baseline
663 predicting performance (e.g., Kristjansson et al., 2009; Unsworth & Robison, 2016), but rather they
664 suggest that, at least within the context of our experiment, the pattern of pupil dilation prior to a detection
665 response may be the more relevant predictor. In this respect, our data are in line with recent PVT studies
666 where the fastest RTs were associated with larger pupil dilations in the ISI (Unsworth et al., 2020;
667 Unsworth & Robison, 2018).

PUPILLOMETRIC VIGILANCE DECREMENT



668

669 *Figure 9.* Baseline and button-locked pupil measures for the fastest ($M = 347$ ms, $SD = 43$ ms) and
670 slowest ($M = 746$ ms, $SD = 43$ ms) 20% RTs in Experiment 2. The left panel shows mean prestimulus
671 baseline pupil size with 95% confidence intervals (bootstrapped, 1000 iterations) and the right panel
672 shows pupil dilations time locked to button responses, with shaded areas surrounding the pupil traces
673 denoting the SEM (bootstrapped, 5000 iterations). Colored horizontal bars in the right-hand panel denote
674 clusters of significant modulation from baseline for the respective traces (grey bar represents the
675 difference between the traces), as revealed by nonparametric permutation tests (1024 permutations, p
676 $< .05$, cluster-corrected for multiple comparisons).

677 **3.3 – Discussion.** This experiment sought insight into the relationship between pupil size and
678 performance measures in a novel PVT. Participants monitored a low contrast grating for a sudden 90°
679 rotation and responded with a button press as quickly as possible after the event. We adopted an atypical
680 stimulus approach to avoid confounds associated with the canonical running counter stimulus—namely
681 its variable intensity and the performance feedback that it provides—which could potentially contribute to
682 variance in behavioral and pupillometric measures (Thorne et al., 2005). Participants completed three
683 successive blocks of the task taking only a 1-min break in between, and changes in performance and pupil
684 measures were explored both within and between blocks. We predicted that performance and pupil size

PUPILLOMETRIC VIGILANCE DECREMENT

685 would decrease as time-on-task increased, and that worse performance would be associated with smaller
686 pupils at baseline.

687 The initial point to note is that our novel stimulus approach led to longer RTs than are typically
688 observed in PVTs that use the canonical running counter stimulus. In these tasks, average RT for
689 subjectively alert participants is generally in the range of 200 to 300 ms (e.g., Basner, Mollicone, &
690 Dinges, 2011; Blatter et al., 2006; Dorrian, Roach, Fletcher, & Dawson, 2007; Loh, Lamond, Dorrian,
691 Roach, & Dawson, 2004; Matsangas, Shattuck, & Brown, 2016; McClelland, Pilcher, & Moore, 2010;
692 Wilkinson & Houghton, 1982), whereas in the current PVT, also with subjectively alert participants,
693 average RT was 420 ms. We attribute this to differences in stimulus intensity. The running counter
694 stimulus is dynamic and constantly changing, providing a constantly refreshed cue for the participant to
695 respond, whereas a change in the orientation of a low contrast grating is more subtle and discrete, and
696 issues no refreshing cue to respond.

697 As predicted, the main performance measures exhibited typical time-on-task effects, with 1/RT
698 decreasing and the number of lapses increasing as time-on-task increased. This general pattern was
699 observed within and between each block of the PVT for both performance measures. For 1/RT, the
700 biggest change was between the first block and the two subsequent blocks, with the difference being
701 largest across the first three groups of trials. Lapse frequency increased gradually within each block and
702 between successive blocks. These patterns in the performance data were statistically robust even without
703 the state manipulations (e.g., time of day, sleep deprivation) and large number of repeated tests that are
704 often integral to the design of mainstream PVT research (e.g., Basner et al., 2011; Blatter et al., 2006;
705 Dorrian et al., 2003; Graw, Kräuchi, Knoblauch, Wirz-Justice, & Cajochen, 2004; Loh et al., 2004;
706 Manousakis, Maccora, Ftouni, & Anderson, 2017).

707 As regards the pupil data, the pattern of within-block declining baseline pupil size broadly
708 reflected the decline in task performance, corroborating findings from previous PVT studies (Massar et
709 al., 2016; Unsworth & Robison 2016) as well as various other studies which examined pupil and
710 performance measures in vigilance or sustained attention (van den Brink et al., 2016; Grandchamp et al.,

PUPILLOMETRIC VIGILANCE DECREMENT

711 2014; Hopstaken, van der Linden, et al., 2015; McIntire et al., 2014; Van Orden et al., 2000). However,
712 the relationship between baseline pupil size and task performance was not clear cut. Participants
713 performed best and had the largest baseline pupil size at the beginning of each block, but the sharp drop in
714 baseline pupil size which occurred between the first and second Trial Group did not have a commensurate
715 drop in performance. Further, baseline pupil size was largest overall and showed the least variability
716 across Trial Groups in Block 3, where performance was at its worst. In a similar fashion, the task-evoked
717 pupil responses were largest at the beginning of Block 1, where performance was best, but were less
718 consistent with respect to the performance data at other times. These patterns in the pupil data are in line
719 with the general prediction that pupil size would decrease as time-on-task increased, but they run contrary
720 to the prediction that worse performance would be reflected in smaller pupils at baseline.

721 Previous experiments offer conflicting evidence as to whether optimal task performance is
722 associated with larger or smaller pupils at baseline (e.g., Kristjansson et al., 2009; Unsworth & Robison,
723 2016). To address this issue, we compared baseline and task-evoked pupil responses for the trials with the
724 fastest and slowest 20% RTs. Whilst there was no significant difference in baseline pupil size between
725 these two groups of trials, there was a clear difference between the observed pupil traces. For the faster
726 RTs, dilation did not begin until after the button response was made, whereas for the slower RTs, dilation
727 was apparent around 500 ms before the button response and increased gradually until it peaked shortly
728 afterwards. The finding of gradual dilation prior to an overt detection response, which is well documented
729 in the literature, has been linked to cognitive factors associated with target recognition and decision
730 making (e.g., see Einhäuser et al., 2010; Martin et al., 2020; Richer & Beatty, 1985). The reason we see
731 this only for the slowest and not the fastest trials probably reflects the difference in RT and the fact that
732 genuine cognitive effects on pupil size tend not to develop until at least 220 ms from the causal event
733 (Mathôt et al., 2015, 2018). For the fastest trials, pupil modulation effects relating to target recognition
734 and decision making were likely mixed in with the motor component.

735 The findings from the current experiment are generally consistent with previous studies showing
736 time-on-task effects on performance and pupil size, but they do not align perfectly with a specific theory

PUPILLOMETRIC VIGILANCE DECREMENT

737 of vigilance. The understimulating and unrewarding nature of the task does however provide ripe
738 conditions for mind-wandering, suggesting that this may have been partly responsible for the decline in
739 performance. Previous studies have also reported larger pupils at baseline during periods of mind-
740 wandering and poor task performance (e.g., Franklin, Broadway, Mrazek, Smallwood, & Schooler, 2013;
741 Smallwood et al., 2011, 2012; Unsworth & Robison, 2016), which is the pattern that was observed in
742 Block 3 of the current experiment. Research also suggests that very short breaks can reduce mind-
743 wandering and lead to performance improvements by temporarily boosting motivation (e.g., Ariga &
744 Lleras, 2011; Ralph, Onderwater, Thomson, & Smilek, 2016; Ross, Russell, & Helton, 2014), which fits
745 with the pattern of data in the current experiment, where participants' performance was restored to more-
746 optimal levels after taking a 1-min break in between each block.

747 We recognize that various factors relating to the individual state of the participants could have
748 influenced the results of the present experiment. For example, performance in PVTs is affected by sleep
749 pressure (Blatter et al., 2006), time-of-day and its interaction with circadian rhythms (Van Dongen &
750 Dinges, 2005; Graw et al., 2004), the consumption of stimulants such as caffeine (Van Dongen et al.,
751 2001), and individual differences in intrinsic alertness (Unsworth et al., 2020). The current experiment did
752 not control for any of such factors, but this could easily be achieved in a subsequent study. For instance,
753 circadian effects could be controlled for by excluding strong “morning and evening types” (Horne, Brass,
754 & Pettitt, 1980) and by testing participants at the same times during the day, after they have reported
755 having similar amounts of sleep. Alternatively, one could examine how performance and pupillometry
756 vary with respect to individual differences in a broad range of cognitive and self-reported personality
757 factors (e.g., Unsworth et al., 2019, 2020).

758 Finally, we note that our novel take on the PVT limits the extent to which it can be directly
759 compared to a more traditional PVT. The use of an alternative stimulus was desirable to avoid certain
760 confounds, but the experiment also differed in terms of block length and ISI. In their general
761 recommendations for the standardized design and analysis of PVTs, Basner and Dinges (2011) suggest
762 using an ISI of 2-10 s and having a fixed block length of 10 min. Due to the way the current experiment

PUPILLOMETRIC VIGILANCE DECREMENT

763 was implemented, ISI was 4-12 seconds and block length was variable ($M = 12.8$ min, $SD = 0.63$ min).
764 Future experiments may wish to bring our approach closer to the task specifications set out by Basner and
765 Dinges (2011), which would broaden the basis for comparison of experimental findings in the wider
766 literature.

4 – General discussion

768 Recent pupillometric studies of vigilance and sustained attention suggest that measurements of
769 pupil size could potentially be used in operational settings to monitor performance, and perhaps even to
770 predict and prevent errors associated with lapses of attention before they occur. But the literature in this
771 area—especially regarding visual tasks—is sparse, and differences in methodology and task requirements
772 have led to conflicting findings. The purpose of the current study was to further explore the relationship
773 between pupil size and performance measures in the context of well-established task frameworks from the
774 vigilance literature.

775 The most consistent finding across both experiments regarding the relationship between pupil size
776 and monitoring performance was that, in line with previous experimental findings (e.g., Beatty, 1982;
777 Hopstaken et al., 2015; Unsworth & Robison, 2016) and the predictions of established theory (Aston-
778 Jones & Cohen, 2005), task-evoked pupil responses were generally more pronounced when performance
779 was best. This trend was most consistent in Experiment 1, where the decline in detection performance was
780 mirrored by a decline in the magnitude of task-evoked responses associated with hits, misses, and false
781 alarms. In Experiment 2, the relationship between task-evoked responses and performance measures was
782 less consistent, although the largest responses did occur when performance was best (i.e., at the beginning
783 of Block 1). In general, these findings suggest that changes in task-evoked pupil responses may serve as
784 an accurate indication of general task engagement, with a decline in their magnitude over time reflecting
785 cognitive disengagement from the task and an increased likelihood of suboptimal performance.

786 Our baseline pupil measures did not show a consistent relationship with performance. In
787 Experiment 1, baseline pupil size was mostly unchanged across three successive periods of watch, despite
788 a marked decrement in performance. In Experiment 2, baseline pupil size showed an overall decline

PUPILLOMETRIC VIGILANCE DECREMENT

789 within each Block, although the slope became less pronounced with each successive Block. Interestingly,
790 baseline pupil size was biggest overall at the beginning of each Block, where task performance was best,
791 suggesting that it reflects heightened arousal, alertness, and focused attention. But, by this account, our
792 baseline measures in the PVT reflect combinations of autonomic tone as well as task-related factors,
793 which means that they are not serving uniquely as a window of insight into the “tonic” mode of LC
794 activation, as is often explicitly or implicitly assumed (see below). The lack of consistency in our baseline
795 measures and their relationship with performance metrics is not unprecedented in light of the literature
796 reviewed in the introduction, which indicates that the relationship is complex and in need of further
797 characterization. One possibility raised by van den Brink et al. (2016) is that the effects of time-on-task
798 on baseline pupil size obscure a more nuanced relationship with performance. In their gradual-onset
799 performance task, after regressing out the effects of time-on-task from the baseline pupil data, the authors
800 observed a quadratic relationship with performance, such that performance was optimal when baseline
801 pupil size was at intermediate levels. This idea dovetails with the Yerkes-Dodson law (Yerkes & Dodson,
802 1908) of optimum arousal, whereby the relationship between task performance and arousal is described
803 by an inverted-U function, such that poor performance is associated with both under- and over-arousal,
804 and optimum performance occurs at a “sweet spot” on the arousal curve.

805 We refrained from using the words “tonic” and “phasic” to describe our pupil measures because
806 we are aware of numerous caveats to the assumption that baseline and task-evoked measures map neatly
807 onto the different modes of LC output. Joshi and Gold (2020) discuss this issue in detail and emphasize
808 that, in the context of LC activation, the terms “tonic” and “phasic” differentiate between distinct modes
809 of activation, and not simply between baseline and transient activity (Aston-Jones & Cohen, 2005).
810 Further, the operational definition of “tonic” and “phasic” pupil measures varies substantially between
811 publications. Also, the precise neural mechanisms of the relationship between pupil measures and LC
812 activation are presently unclear and it is possible that a third variable, as of yet not understood, may
813 account for the observed pupil-LC link (Costa & Rudebeck, 2016).

PUPILLOMETRIC VIGILANCE DECREMENT

814 In conclusion, the results of our two vigilance experiments support the general notion that
815 changes in task-evoked pupil measures can be used to gain insight into monitoring performance in long
816 and demanding tasks where the emphasis is on additive effects over a series of trials. But there is clearly a
817 need for further research to determine the practical feasibility of utilizing pupil size as a
818 psychophysiological marker of attentional lapses in real-time monitoring systems. Characterizing the
819 precise relationship between different measures of behavioral performance, task-related factors and
820 patterns of pupil behavior will be a crucial next step in this regard.

821 **Acknowledgements**

822 The research was funded by a grant from the Defence Science and Technology Laboratory
823 (DSTLX1000083208)—an executive government agency that ensures innovative science and technology
824 contribute to the defence and security of the UK. We thank three anonymous reviewers for their
825 comments and suggestions.

826

PUPILLOMETRIC VIGILANCE DECREMENT

References

828 Alnaes, D., Sneve, M. H., Espeseth, T., Endestad, T., van de Pavert, S. H. P., & Laeng, B. (2014). Pupil
829 size signals mental effort deployed during multiple object tracking and predicts brain activity in the
830 dorsal attention network and the locus coeruleus. *Journal of Vision*, 14(2014), 1–20.
831 <https://doi.org/10.1167/14.4.1>

832 Ariga, A., & Lleras, A. (2011). Brief and rare mental “breaks” keep you focused: Deactivation and
833 reactivation of task goals preempt vigilance decrements. *Cognition*, 118(3), 439–443.
834 <https://doi.org/10.1016/j.cognition.2010.12.007>

835 Aston-Jones, G., Chiang, C., & Alexinsky, T. (1991). Discharge of noradrenergic locus coeruleus neurons
836 in behaving rats and monkeys suggests a role in vigilance. *Progress in Brain Research*, 88(C), 501–
837 520. [https://doi.org/10.1016/S0079-6123\(08\)63830-3](https://doi.org/10.1016/S0079-6123(08)63830-3)

838 Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine
839 function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, 28, 403–450.
840 <https://doi.org/10.1146/annurev.neuro.28.061604.135709>

841 Aston-Jones, G., Rajkowski, J., & Cohen, J. D. (1999). Role of locus coeruleus in attention and
842 behavioral flexibility. *Biological Psychiatry*, 46(9), 1309–1320. [https://doi.org/10.1016/S0006-3223\(99\)00140-7](https://doi.org/10.1016/S0006-3223(99)00140-7)

843 Aston-Jones, G., Rajkowski, J., Kubiak, P., & Alexinsky, T. (1994). Locus coeruleus neurons in monkey
844 are selectively activated by attended cues in a vigilance task. *The Journal of Neuroscience*, 14(July),
845 4467–4480.

846 Baddeley, A. D., & Colquhoun, W. P. (1969). Signal probability and vigilance: A reappraisal of the
847 “signal-rate” effect. *British Journal of Psychology*, 60(2), 169–178. <https://doi.org/10.1111/j.2044-8295.1969.tb01189.x>

848 Ballard, J. C. (1996). Computerized assessment of sustained attention: A review of factors affecting
849 vigilance performance. *Journal of Clinical and Experimental Neuropsychology*, 18(6), 843–863.
850 <https://doi.org/10.1080/01688639608408307>

PUPILLOMETRIC VIGILANCE DECREMENT

853 Barbur, J. L., Harlow, A. J., & Sahraie, A. (1992). Pupillary responses to stimulus structure, colour and
854 movement. *Ophthalmic & Physiological Optics*, 12(2), 137–141. <https://doi.org/10.1111/j.1475-1313.1992.tb00276.x>

855 Basner, M., & Dinges, D. F. (2011). Maximizing sensitivity of the psychomotor vigilance test (PVT) to
856 sleep loss. *Sleep*, 34(5), 581–591. <https://doi.org/10.1093/sleep/34.5.581>

857 Basner, M., Mollicone, D., & Dinges, D. F. (2011). Validity and sensitivity of a brief psychomotor
858 vigilance test (PVT-B) to total and partial sleep deprivation. *Acta Astronautica*, 69(11–12), 949–
859 959. <https://doi.org/10.1016/j.actaastro.2011.07.015>

860 Beatty, J. (1982). Phasic not tonic pupillary responses vary with auditory vigilance performance.
861 *Psychophysiology*, 19(2), 167–172. <https://doi.org/10.1111/j.1469-8986.1982.tb02540.x>

862 Berridge, C. W. (2008). Noradrenergic modulation of arousal. *Brain Research Reviews*, 58(1), 1–17.
863 <https://doi.org/10.1016/j.brainresrev.2007.10.013>

864 Berridge, C. W., Schmeichel, B. E., & España, R. A. (2012). Noradrenergic modulation of
865 wakefulness/arousal. *Sleep Medicine Reviews*, 16(2), 187–197.
866 <https://doi.org/10.1016/j.smrv.2011.12.003>

867 Berridge, C. W., & Waterhouse, B. D. (2003). The locus coeruleus-noradrenergic system: Modulation of
868 behavioral state and state-dependent cognitive processes. *Brain Research Reviews*, 42(1), 33–84.
869 [https://doi.org/10.1016/S0165-0173\(03\)00143-7](https://doi.org/10.1016/S0165-0173(03)00143-7)

870 Bills, A. G. (1943). *The psychology of efficiency: A discussion of the hygiene of mental work*. Harper &
871 Bros.

872 Blatter, K., Graw, P., Münch, M., Knoblauch, V., Wirz-Justice, A., & Cajochen, C. (2006). Gender and
873 age differences in psychomotor vigilance performance under differential sleep pressure conditions.
874 *Behavioural Brain Research*, 168(2), 312–317. <https://doi.org/10.1016/j.bbr.2005.11.018>

875 Broadbent, D. E. (1953). Noise, paced performance and vigilance tasks. *British Journal of Psychology*,
876 44(4), 295–303. <https://doi.org/10.1111/j.2044-8295.1953.tb01210.x>

877 Broadbent, D. E. (1958). *Perception and communication*. Oxford University Press.

PUPILLOMETRIC VIGILANCE DECREMENT

879 Broadbent, D. E. (1971). *Decision and stress*. Academic Press.

880 Broadbent, D. E., & Gregory, M. (1965). Effects of noise and of signal rate upon vigilance analysed by
881 means of decision theory. *Human Factors*, 7, 155–162.

882 <https://doi.org/10.1177/001872086500700207>

883 Buck, L. (1966). Reaction time as a measure of perceptual vigilance. *Psychological Bulletin*, 65(5), 291–
884 304.

885 Caggiano, D. M., & Parasuraman, R. (2004). The role of memory representation in the vigilance
886 decrement. *Psychonomic Bulletin & Review*, 11(5), 932–937.
887 <http://www.ncbi.nlm.nih.gov/pmc/articles/1351035/>&tool=pmcentrez&rendertype=abstract

888

889 Caldwell, J. A., Prazinko, B., & Caldwell, J. L. (2003). Body posture affects electroencephalographic
890 activity and psychomotor vigilance task performance in sleep-deprived subjects. *Clinical
891 Neurophysiology*, 114(1), 23–31. [https://doi.org/10.1016/S1388-2457\(02\)00283-3](https://doi.org/10.1016/S1388-2457(02)00283-3)

892 Colquhoun, W. V. (1961). the effect of ‘unwanted’ signals on performance in a vigilance task.
893 *Ergonomics*, 4(1), 41–51. <https://doi.org/10.1080/00140136108930506>

894 Costa, V. D., & Rudebeck, P. H. (2016). More than meets the eye: The relationship between pupil size
895 and locus coeruleus activity. *Neuron*, 89(1), 8–10. <https://doi.org/10.1016/j.neuron.2015.12.031>

896 de Gee, J. W., Colizoli, O., Kloosterman, N. A., Knapen, T., Nieuwenhuis, S., & Donner, T. H. (2017).
897 Dynamic modulation of decision biases by brainstem arousal systems. *eLife*, 6(Lc), 1–36.
898 <https://doi.org/10.7554/eLife.23232>

899 de Gee, J. W., Knapen, T., & Donner, T. H. (2014). Decision-related pupil dilation reflects upcoming
900 choice and individual bias. *Proceedings of the National Academy of Sciences, USA*, 111(5), E618–
901 25. <https://doi.org/10.1073/pnas.1317557111>

902 Dinges, D. F., Pack, F., Williams, K., Gillen, K. A., Powell, J. W., Ott, G. E., Aptowicz, C., & Pack, A. I.
903 (1997). Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance
904 decrements during a week of sleep restricted to 4-5 hours per night. *Sleep*, 20(4), 267–277.

PUPILLOMETRIC VIGILANCE DECREMENT

905 <http://www.ncbi.nlm.nih.gov/pubmed/9231952>

906 Dorrian, J., Lamond, N., Holmes, A. L., Burgess, H. J., Roach, G. D., Fletcher, A., & Dawson, D. (2003).

907 The ability to self-monitor performance during a week of simulated night shifts. *Sleep*, 26(7), 871–

908 877.

909 Dorrian, J., Roach, G. D., Fletcher, A., & Dawson, D. (2007). Simulated train driving: Fatigue, self-

910 awareness and cognitive disengagement. *Applied Ergonomics*, 38(2), 155–166.

911 <https://doi.org/10.1016/j.apergo.2006.03.006>

912 Einhäuser, W., Koch, C., & Carter, O. L. (2010). Pupil dilation betrays the timing of decisions. *Frontiers*

913 in *Human Neuroscience*, 4(February), 18. <https://doi.org/10.3389/fnhum.2010.00018>

914 Einhäuser, W., Stout, J., Koch, C., & Carter, O. L. (2008). Pupil dilation reflects perceptual selection and

915 predicts subsequent stability in perceptual rivalry. *Proceedings of the National Academy of Sciences*,

916 USA, 105(5), 1704–1709. <https://doi.org/10.1073/pnas.0707727105>

917 Esterman, M., Grossi, M., Liu, G., Mitko, A., Morris, R., & DeGutis, J. (2016). Anticipation of monetary

918 reward can attenuate the vigilance decrement. *PLoS ONE*, 11(7), e0159741.

919 <https://doi.org/10.1371/journal.pone.0159741>

920 Fortenbaugh, F. C., DeGutis, J., & Esterman, M. (2017). Recent theoretical, neural, and clinical advances

921 in sustained attention research. *Annals of the New York Academy of Sciences*, 1396, 70–91.

922 <https://doi.org/10.1111/nyas.13318>

923 Franklin, M. S., Broadway, J. M., Mrazek, M. D., Smallwood, J., & Schooler, J. W. (2013). Window to

924 the wandering mind: Pupilometry of spontaneous thought while reading. *The Quarterly Journal of*

925 *Experimental Psychology*, 66(12), 2289–2294. <https://doi.org/10.1080/17470218.2013.858170>

926 Frankmann, J. P., & Adams, J. A. (1962). Theories of vigilance. *Psychological Bulletin*, 59(4), 257–272.

927 <https://doi.org/10.1037/h0046142>

928 Fried, M., Tsitsiashvili, E., Bonneh, Y. S., Sterkin, A., Wygnanski-Jaffe, T., Epstein, T., & Polat, U.

929 (2014). ADHD subjects fail to suppress eye blinks and microsaccades while anticipating visual

930 stimuli but recover with medication. *Vision Research*, 101, 62–72.

PUPILLOMETRIC VIGILANCE DECREMENT

931 <https://doi.org/10.1016/j.visres.2014.05.004>

932 Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in
933 control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective*
934 & *Behavioral Neuroscience*, 10(2), 252–269. <https://doi.org/10.3758/CABN.10.2.252>

935 Goldwater, B. C. (1972). Psychological significance of pupillary movements. *Psychological Bulletin*,
936 77(5), 340–355. <http://psycnet.apa.org/journals/bul/77/5/340/>

937 Grandchamp, R., Braboszcz, C., & Delorme, A. (2014). Oculometric variations during mind wandering.
938 *Frontiers in Psychology*, 5(FEB), 1–10. <https://doi.org/10.3389/fpsyg.2014.00031>

939 Graw, P., Kräuchi, K., Knoblauch, V., Wirz-Justice, A., & Cajochen, C. (2004). Circadian and wake-
940 dependent modulation of fastest and slowest reaction times during the psychomotor vigilance task.
941 *Physiology and Behavior*, 80(5), 695–701. <https://doi.org/10.1016/j.physbeh.2003.12.004>

942 Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics* (Vol. 1). RF Krieger.

943 Hockey, R. J. (2013). *The psychology of fatigue: Work, effort and control*. Cambridge University Press.

944 Hopstaken, J. F., van der Linden, D., Bakker, A. B., & Kompier, M. A. J. (2015). The window of my
945 eyes: Task disengagement and mental fatigue covary with pupil dynamics. *Biological Psychology*,
946 110, 100–106. <https://doi.org/10.1016/j.biopsych.2015.06.013>

947 Hopstaken, J. F., Van der Linden, D., Bakker, A. B., & Kompier, M. A. J. (2015). A multifaceted
948 investigation of the link between mental fatigue and task disengagement. *Psychophysiology*, 52(3),
949 305–315. <https://doi.org/10.1111/psyp.12339>

950 Hopstaken, J. F., Van der Linden, D., Bakker, A. B., Kompier, M. A. J., & Leung, Y. K. (2016). Shifts in
951 attention during mental fatigue: Evidence from subjective, behavioral, physiological, and eye-
952 tracking data. *Journal of Experimental Psychology: Human Perception and Performance*, 42(6),
953 878–889. <https://doi.org/10.1037/xhp0000189>

954 Horne, J. A., Brass, C. G., & Pettitt, A. N. (1980). Circadian performance differences between morning
955 and evening “types.” *Ergonomics*, 23(January), 29–36. <https://doi.org/10.1080/00140138008924715>

956 Hupé, J.-M., Lamirel, C., & Lorenceau, J. (2009). Pupil dynamics during bistable motion perception.

PUPILLOMETRIC VIGILANCE DECREMENT

957 *Journal of Vision*, 9(7), 10. <https://doi.org/10.1167/9.7.10>

958 Jepma, M., & Nieuwenhuis, S. (2011). Pupil diameter predicts changes in the exploration-exploitation
959 trade-off: Evidence for the adaptive gain theory. *Journal of Cognitive Neuroscience*, 23(7), 1587–
960 1596. <https://doi.org/10.1162/jocn.2010.21548>

961 Joshi, S., & Gold, J. I. (2020). Pupil Size as a Window on Neural Substrates of Cognition. *Trends in*
962 *Cognitive Sciences*, 24(6), 466–480. <https://doi.org/10.1016/j.tics.2020.03.005>

963 Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between pupil diameter and neuronal
964 activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron*, 89(1), 221–234.
965 <https://doi.org/10.1016/j.neuron.2015.11.028>

966 Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.
967 <https://doi.org/10.2307/1421603>

968 Kahneman, D., & Treisman, A. M. (1984). Changing views of attention and automaticity. In R.
969 Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 29–61). Academic Press.

970 Knapen, T., de Gee, J. W., Brascamp, J., Nuitjen, S., Hoppenbrouwers, S., & Theeuwes, J. (2016).
971 Cognitive and ocular factors jointly determine pupil responses under equiluminance. *PLoS ONE*,
972 11(5), e0155574. <https://doi.org/10.1371/journal.pone.0155574>

973 Kristjansson, S. D., Stern, J. A., Brown, T. B., & Rohrbaugh, J. W. (2009). Detecting phasic lapses in
974 alertness using pupillometric measures. *Applied Ergonomics*, 40(6), 978–986.
975 <https://doi.org/10.1016/j.apergo.2009.04.007>

976 Laeng, B., Sirois, S., & Gredeback, G. (2012). Pupillometry: a window to the preconscious? *Perspectives*
977 *on Psychological Science*, 7(1), 18–27. <https://doi.org/10.1177/1745691611427305>

978 Langner, R., & Eickhoff, S. B. (2013). Sustaining attention to simple tasks: A meta-analytic review of the
979 neural mechanisms of vigilant attention. *Psychological Bulletin*, 139(4), 870–900.
980 <https://doi.org/10.1037/a0030694>

981 Leys, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation
982 around the mean, use absolute deviation around the median. *Journal of Experimental Social*

PUPILLOMETRIC VIGILANCE DECREMENT

983 *Psychology*, 49(4), 764–766. <https://doi.org/10.1016/j.jesp.2013.03.013>

984 Loh, S., Lamond, N., Dorrian, J., Roach, G. D., & Dawson, D. (2004). The validity of psychomotor
985 vigilance tasks of less than 10-minute duration. *Behavior Research Methods, Instruments, &*
986 *Computers*, 36(2), 339–346. <https://doi.org/10.3758/BF03195580>

987 Mackie, R. R. (1987). Vigilance research--are we ready for countermeasures? *Human Factors*, 29(6),
988 707–723. <https://doi.org/10.1177/001872088702900610>

989 Mackworth, J. F. (1970). *Vigilance and attention*. Penguin.

990 Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search. *The Quarterly*
991 *Journal of Experimental Psychology*, 1(1), 6–21. <https://doi.org/10.1080/17470214808416738>

992 Mackworth, N. H. (1950). *Researches on the measurement of human performance*. H. M. S. O.

993 Manousakis, J., Maccora, J., Ftouni, S., & Anderson, C. (2017). It's in the eyes: Pupillary response
994 provides a physiological marker of alertness and performance impairment. *Journal of Sleep*
995 *Research*, 26(Suppl. 1), 28. https://doi.org/10.1111/jsr.59_12618

996 Martin, J. T., Whittaker, A. H., & Johnston, S. J. (2020). Component processes in free-viewing visual
997 search: Insights from fixation-aligned pupillary response averaging. *Journal of Vision*, 20(7), 5.
998 <https://doi.org/10.1167/jov.20.7.5>

999 Massar, S. A. A., Lim, J., Sasmita, K., & Chee, M. W. L. (2016). Rewards boost sustained attention
1000 through higher effort: A value-based decision making approach. *Biological Psychology*, 120, 21–27.
1001 <https://doi.org/10.1016/j.biopsych.2016.07.019>

1002 Mathôt, S., Fabius, J. H., Van Heusden, E., & Van der Stigchel, S. (2018). Safe and sensible baseline
1003 correction of pupil-size data. *Behavior Research Methods*, April, 1–25.
1004 <https://doi.org/10.3758/s13428-017-1007-2>

1005 Mathôt, S., Siebold, A., Donk, M., Vitu, F., & Siebold, A. (2015). Large pupils predict goal-driven eye
1006 movements. *Journal of Experimental Psychology: General*.
1007 <https://doi.org/http://dx.doi.org/10.1037/a0039168>

1008 Matsangas, P., Shattuck, N. L., & Brown, S. (2016). Preliminary validation study of the 3-min wrist-worn

PUPILLOMETRIC VIGILANCE DECREMENT

1009 psychomotor vigilance test. *Behavior Research Methods*, 1–10. <https://doi.org/10.3758/s13428-016-0821-2>

1010

1011 McClelland, L. E., Pilcher, J. J., & Moore, D. D. (2010). Oculomotor measures as predictors of

1012 performance during sleep deprivation. *Aviation Space and Environmental Medicine*, 81(9), 833–

1013 842. <https://doi.org/10.3357/ASEM.2653.2010>

1014 McIntire, L. K., McKinley, R. A., & Goodyear, C. (2014). Detection of vigilance performance with

1015 pupillometry. *Applied Proceedings of the Symposium on Eye Tracking Research and Applications*,

1016 167–174. <https://doi.org/10.1145/2578153.2578177>

1017 Mittner, M., Boekel, W., Tucker, A. M., Turner, B. M., Heathcote, A., & Forstmann, B. U. (2014). When

1018 the brain takes a break: A model-based analysis of mind wandering. *The Journal of Neuroscience*,

1019 34(49), 16286–16295. <https://doi.org/10.1523/JNEUROSCI.2062-14.2014>

1020 Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil

1021 diameter covaries with BOLD activity in human locus coeruleus. *Human Brain Mapping*, 4154,

1022 4140–4154. <https://doi.org/10.1002/hbm.22466>

1023 Murphy, P. R., Robertson, I. H., Balsters, J. H., & O'Connell, R. G. (2011). Pupillometry and P3 index

1024 the locus coeruleus-noradrenergic arousal function in humans. *Psychophysiology*, 48(11), 1532–

1025 1543. <https://doi.org/10.1111/j.1469-8986.2011.01226.x>

1026 Murphy, P. R., van Moort, M. L., & Nieuwenhuis, S. (2016). The pupillary orienting response predicts

1027 adaptive behavioral adjustment after errors. *PLoS ONE*, 11(3), 1–11.

1028 <https://doi.org/10.1371/journal.pone.0151763>

1029 Oken, B. S., Salinsky, M. C., & Elsas, S. M. (2006). Vigilance, alertness, or sustained attention:

1030 physiological basis and measurement. *Clinical Neurophysiology*, 117(9), 1885–1901.

1031 <https://doi.org/10.1016/j.clinph.2006.01.017>

1032 Parasuraman, Raja. (1979). Memory load and event rate control sensitivity decrements in sustained

1033 attention. *Science*, 205(4409), 924–927. <https://doi.org/10.1126/science.472714>

1034 Parasuraman, Raja, & Davies, D. R. (1976). Decision theory analysis of response latencies in vigilance.

PUPILLOMETRIC VIGILANCE DECREMENT

1035 *Journal of Experimental Psychology: Human Perception and Performance*, 2(4), 578–590.

1036 <http://www.ncbi.nlm.nih.gov/pubmed/1011007>

1037 Parasuraman, Raja, & Davies, D. R. (1982). *The psychology of vigilance*. Academic Press.

1038 Parasuraman, Raja, Warm, J. S., & See, J. E. (1998). Brain systems of vigilance. In R. Parasuraman (Ed.),
1039 *The attentive brain* (pp. 221–256).

1040 <http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1998-07668-010&site=ehost-live>

1041 Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual
1043 Review of Neuroscience*, 35(1), 73–89. <https://doi.org/10.1146/annurev-neuro-062111-150525>

1044 Phillips, M. A., Szabadi, E. U., & Bradshaw, C. M. (2000). Comparison of the effects of clonidine and
1045 yohimbine on spontaneous pupillary fluctuations in healthy human volunteers.

1046 *Psychopharmacology*, 150(1), 85–89. <https://doi.org/10.1007/s002130000398>

1047 Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of
1048 Neuroscience*, 13, 25–42. <https://doi.org/10.1146/annurev.ne.13.030190.000325>

1049 Rajkowski, J., Kubiak, P., & Aston-Jones, G. (1993). Correlations between locus coeruleus (LC) neural
1050 activity, pupil diameter and behavior in monkey support a role of LC in attention. *Society for
1051 Neuroscience Abstracts*, 19, 47.

1052 Rajkowski, J., Kubiak, P., & Aston-Jones, G. (1994). Locus coeruleus activity in monkey: Phasic and
1053 tonic changes are associated with altered vigilance. *Brain Research Bulletin*, 35(5–6), 607–616.
1054 [https://doi.org/10.1016/0361-9230\(94\)90175-9](https://doi.org/10.1016/0361-9230(94)90175-9)

1055 Ralph, B. C. W., Onderwater, K., Thomson, D. R., & Smilek, D. (2016). Disrupting monotony while
1056 increasing demand: Benefits of rest and intervening tasks on vigilance. *Psychological Research*,
1057 81(1948), 432–444. <https://doi.org/10.1007/s00426-016-0752-7>

1058 Richer, F., & Beatty, J. (1985). Pupillary dilations in movement preparation and execution.
1059 *Psychophysiology*, 22(2), 204–207. <https://doi.org/10.1111/j.1469-8986.1985.tb01587.x>

1060 Ross, H. A., Russell, P. N., & Helton, W. S. (2014). Effects of breaks and goal switches on the vigilance

PUPILLOMETRIC VIGILANCE DECREMENT

1061 decrement. *Experimental Brain Research*, 232(6), 1729–1737. <https://doi.org/10.1007/s00221-014-3865-5>

1062

1063 Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not

1064 establish significance of effect latency or location. *Psychophysiology*, 56(6), 1–8.

1065 <https://doi.org/10.1111/psyp.13335>

1066 Smallwood, J., Brown, K. S., Baird, B., Mrazek, M. D., Franklin, M. S., & Schooler, J. W. (2012).

1067 Insulation for daydreams: A role for tonic norepinephrine in the facilitation of internally guided

1068 thought. *PLoS ONE*, 7(4), e33706. <https://doi.org/10.1371/journal.pone.0033706>

1069 Smallwood, J., Brown, K. S., Tipper, C., Giesbrecht, B., Franklin, M. S., Mrazek, M. D., Carlson, J. M.,

1070 & Schooler, J. W. (2011). Pupillometric evidence for the decoupling of attention from perceptual

1071 input during offline thought. *PLoS ONE*, 6(3), e18298.

1072 <https://doi.org/10.1371/journal.pone.0018298>

1073 Smallwood, J., Davies, J. B., Heim, D., Finnigan, F., Sudberry, M., O'Connor, R., & Obonsawin, M.

1074 (2004). Subjective experience and the attentional lapse: Task engagement and disengagement during

1075 sustained attention. *Consciousness and Cognition*, 13(4), 657–690.

1076 <https://doi.org/10.1016/j.concog.2004.06.003>

1077 Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychology of Consciousness: Theory,*

1078 *Research, and Practice*, 132(6), 946–958. <https://doi.org/10.1037/0033-2909.132.6.946>

1079 SR Research Ltd. (2010). *EyeLink user manual. Version 1.5.2*.

1080 Thomson, D. R., Besner, D., & Smilek, D. (2015). A resource-control account of sustained attention:

1081 Evidence from mind wandering and vigilance paradigms. *Perspectives on Psychological Science*,

1082 10(1), 82–96. <https://doi.org/10.1177/1745691614556681>

1083 Thorne, D. R., Johnson, D. E., Redmond, D. P., Sing, H. C., Belenky, G., & Shapiro, J. M. (2005). The

1084 Walter Reed palm-held psychomotor vigilance test. *Behavior Research Methods*, 37(1), 111–118.

1085 <https://doi.org/10.3758/BF03206404>

1086 Unsworth, N., Miller, A. L., & Robison, M. K. (2020). Individual differences in lapses of sustained

PUPILLOMETRIC VIGILANCE DECREMENT

1087 attention: Oculometric indicators of intrinsic alertness. *Journal of Experimental Psychology: Human*
1088 *Perception and Performance*, 46(6), 569–592. <https://doi.org/10.1037/xhp0000734>

1089 Unsworth, N., & Robison, M. K. (2016). Pupillary correlates of lapses of sustained attention. *Cognitive,*
1090 *Affective & Behavioral Neuroscience*, 16(April), 601–615. <https://doi.org/10.3758/s13415-016-0417-4>

1091 Unsworth, N., & Robison, M. K. (2018). Tracking arousal state and mind wandering with pupillometry.
1092 *Cognitive, Affective and Behavioral Neuroscience*, 18(4), 638–664. <https://doi.org/10.3758/s13415-018-0594-4>

1093 Unsworth, N., Robison, M. K., & Miller, A. L. (2019). Individual differences in baseline oculometrics:
1094 Examining variation in baseline pupil diameter, spontaneous eye blink rate, and fixation stability.
1095 *Cognitive, Affective and Behavioral Neuroscience*, 19(4), 1074–1093.
1096 <https://doi.org/10.3758/s13415-019-00709-z>

1097 Urai, A. E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and
1098 alters serial choice bias. *Nature Communications*, 8, 14637. <https://doi.org/10.1038/ncomms14637>

1099 van den Brink, R. L., Murphy, P. R., & Nieuwenhuis, S. (2016). Pupil diameter tracks lapses of attention.
1100 *PLoS ONE*, 11(10), 1–16. <https://doi.org/10.1371/journal.pone.0165274>

1101 Van Dongen, H. P. A., & Dinges, D. F. (2005). Sleep, circadian rhythms, and psychomotor vigilance.
1102 *Clinics in Sports Medicine*, 24(2), 237–249. <https://doi.org/10.1016/j.csm.2004.12.007>

1103 Van Dongen, H. P. A., Price, N. J., Mullington, J. M., Szuba, M. P., Kapoor, S. C., & Dinges, D. F.
1104 (2001). Caffeine eliminates psychomotor vigilance deficits from sleep inertia. *Sleep*, 24(7), 813–
1105 819. <https://doi.org/10.1093/sleep/24.7.813>

1106 Van Orden, K. F., Jung, T. P., & Makeig, S. (2000). Combined eye activity measures accurately estimate
1107 changes in sustained visual task performance. *Biological Psychology*, 52(3), 221–240.
1108 [https://doi.org/10.1016/S0301-0511\(99\)00043-5](https://doi.org/10.1016/S0301-0511(99)00043-5)

1109 Varazzani, C., San-Galli, A., Gilardeau, S., & Bouret, S. (2015). Noradrenaline and dopamine neurons in
1110 the reward/effort trade-off: A direct electrophysiological comparison in behaving monkeys. *The*

PUPILLOMETRIC VIGILANCE DECREMENT

1113 *Journal of Neuroscience*, 35(7866), 7877. <https://doi.org/10.1523/JNEUROSCI.0454-15.2015>

1114 Warm, Joel S., & Jerison, H. J. (1984). The psychophysics of vigilance. In J. S. Warm (Ed.), *Sustained*
1115 *attention in human performance* (pp. 15–59). John Wiley & Sons.

1116 Warm, Joel S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is
1117 stressful. *Human Factors*, 50(3), 433–441. <https://doi.org/10.1518/001872008X312152>

1118 Wiener, E. L. (1987). Application of vigilance research: Rare, medium, or well done? *Human Factors*,
1119 29(6), 725–736. <https://doi.org/10.1177/001872088702900611>

1120 Wilkinson, R. T., & Houghton, D. (1982). Field test of arousal: a portable reaction timer with data
1121 storage. *Human Factors*, 24(4), 487–493. <https://doi.org/10.1177/001872088202400409>

1122 Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation.
1123 *Journal of Comparative Neurology and Psychology*, 18(5), 459–482.

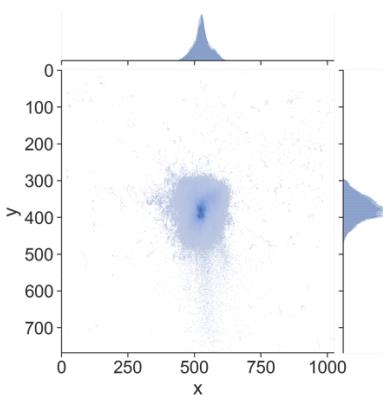
1124 <https://doi.org/10.1002/cne.920180503>

1125 Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, 46(4), 681–692.
1126 <https://doi.org/10.1016/j.neuron.2005.04.026>

1127

1128

1129



1130

1131 *Figure X*. Joint histogram for x and y of gaze position in screen pixel coordinates.