Abstract

Continuous performance tasks are frequently associated with a vigilance decrement, particularly when target events are rare and after prolonged time on task. Here we characterized the time course of a performance decrement that happens more rapidly. Using the gradual-onset continuous performance task (the gradCPT), we presented participants with a long sequence of scenes that gradually faded in and out. Participants pressed a button as soon as they detected scenes in one category and ignored scenes in another category. We manipulated the novelty of stimuli, required response rate, and the prevalence rate of the target stimuli. Performance sensitivity declined moderately within and across three 8-minute-long blocks. Contrary to mindlessness accounts of vigilance decrement, the decline was not restricted to situations when target events were rare and the stimuli were repetitive. High motor response rates substantially impaired overall sensitivity and moderately increased performance decrement. Performance in the gradCPT did not correlate with individual differences in mindfulness, attentional lapses, media multitasking, or complex working memory span. The rapid and pervasively-observed decline in performance is consistent with attentional resource theories of vigilance decrement.

Keywords: Sustained attention, Vigilance decrement, Continuous performance task

Public significance statement

Small failures of visual attention can have catastrophic consequences in such attention-demanding activities as driving, security screening, or radar monitoring. This study shows that vigilance declines can occur very rapidly (within minutes), are unlikely due to boredom or mind-wandering, and appear to reflect limited neurocognitive attentional resource capacities. Future research should examine if providing temporary breaks, or designing a visual stream to contrast visual and conceptual information, can counteract swift declines in sustained visual attention.
Introduction

Human factors and cognitive research suggest that people have difficulty maintaining uniformly high levels of performance over an extended period of time. For example, a classic study showed that radar screeners tended to perform worse at the end of their shift than at the beginning (Mackworth, 1948). Such declines have been attributed to a vigilance decrement, which refers to a decreasing ability to maintain attention on the target task (See, Howe, Warm, & Dember, 1995). Fluctuations of attention have also been observed at shorter time scales than an entire shift. Such fluctuations often manifest as “mind wandering” in daily activities (Smallwood & Schooler, 2015). Most studies on attention, however, employ discrete tasks. The short breaks between discrete trials may allow the cognitive system to reset. Consequently, discrete tasks may underestimate the prevalence of vigilance decrement and do not represent real-world tasks for which performance decrements can be detrimental. The present study investigates the characteristics of sustained attention in continuous performance tasks. We focus on the time course of the performance decrement and the relationship between sustained attention and other cognitive functions.

Vigilance decrement

Traditional vigilance tasks are typically performed for 30 min or longer, and require participants to respond to infrequent targets that occur unpredictably among frequent nontargets (Ballard, 1996). Mackworth’s “clock” task, for example, presents participants with a clock hand that rotates around; occasionally the hand jumps twice as far as its normal pace. Response to these rare target events declines over the 2-hour watch time. Vigilance decrements are also observed in tasks with high target presentation rates. The time course of the decrement is more rapid with faster stimulus presentation rates (Parasuraman, 1979), especially when the signal salience is low (Helton & Warm, 2008). For example, Temple and colleagues (2000) presented participants with a stream of letter Ds, backward Ds, and Os against a noisy background. Participants’ sensitivity in detecting the infrequent Os declined over the 12-min watch time.

Several mechanisms have been proposed to underlie the change in performance over time. One of the oldest accounts, probability matching, posits that people adjust their response rate to equal the target prevalence rate, so that the response-to-target ratio approaches 1. When targets are rare, the initial response rate is higher than the target rate because untrained observers have not had enough experience to accurately gauge the target rate (Craig, 1978). This account attributes the decline in response rates to a response strategy, not changes in cognitive state. Such a strategy change should affect response criterion, but not detection sensitivity. Therefore, a probability matching account has difficulty explaining decrements in detection sensitivity with continued time on task.

A second account, response inhibition, was proposed to explain an increase in false alarms as response rates increase (Wilson, Finkbeiner, de Joux, Russell, & Helton, 2016), usually in tasks that require people to respond to frequent events and withhold responses to infrequent events (Esterman, Noonan, Rosenberg, & DeGutis, 2013). For example, in a simulated combat task, participants were asked to fire a laser shot at “foes” and to withhold fire at “friendlies” (Wilson, Head, de Joux, Finkbeiner, & Helton, 2015). As more foes were presented, the rate of friendly fire also increased. Stevenson, Russell, and Helton (2011) proposed that the ability to inhibit frequent responses demands attentional resources. The inability to sustain attention over time weakens the ability to inhibit responses to infrequent events.
The response inhibition account fits with a broader theoretical framework on vigilance decrement— the attention resource account (Grier et al., 2003). According to this account, the demand to continuously monitor incoming stimuli and make timely responses to targets drains attentional resources. Due to the high event rate and the lack of breaks, attentional resources are not replenished, resulting in a decline in performance (Helton & Warm, 2008). The precise nature of these attentional resources is unclear. However, based on signal detection findings and changes in blood flow (assessed by transcranial Doppler sonography) during a 40-min simulated air traffic control task, Hitchcock et al. (2003) suggest that it could correspond to the rate of cerebral blood flow to the prefrontal cortex. Support for this account comes from studies showing that vigilance decrement is greater when the monitoring task is more difficult, such as when the event rate is higher, when the signal salience is lower (Helton & Warm, 2008), or when the task is undertaken with a concurrent working memory load (Helton & Russell, 2011).

Other proposed accounts attribute the vigilance decrement to an inability to maintain cognitive control over time, either as a result of goal habituation (Ariga & Lleras, 2011) or from the onset of boredom and mindlessness (Manly, Robertson, Galloway, & Hawkins, 1999; Robertson, Manly, Andrade, Baddeley, & Yiend, 1997). These accounts do not attribute vigilance decrements to depletion of the postulated attentional processing resources, but to competition from other task goals. The longer one performs a single task and the more automated the task becomes, the less cognitive control is exerted, and the more likely one is to lapse into a state of mindlessness. This is especially true when the primary task is simple or targets are infrequent. Evidence for the mindlessness theory, however, has been mixed. Brief interruption of the task by a different task rescues vigilance decrement in some studies (Ariga & Lleras, 2011; Zijlstra, Roe, Leonora, & Krediet, 1999) but not in others (Altmann, Trafton, & Hambrick, 2014; Helton & Russell, 2012). A reduction in target frequency increases boredom and the vigilance decrement (Wilson et al., 2016), but challenging tasks that should have reduced boredom are often associated with greater, not less, vigilance decrement (Warm, Parasuraman, & Matthews, 2008). More recent theoretical accounts have combined elements of several theories to explain mind-wandering (e.g., Thomson, Besner, & Smilek, 2015; Kurzban, Duckworth, Kable, & Myers, 2013; see also Fortenbaugh, DeGutis, & Esterman, 2017). For example, the resource control theory (Thomson et al., 2015) proposes that although the amount of attentional resources available to an individual remains essentially constant (rather than becoming depleted) across time, executive control is needed to sustain active goal maintenance; such executive control is reduced with increasing time-on-task, allowing a default bias toward self-generated thought (mind wandering) to emerge, and leading to performance costs if the primary task is sufficiently attention-demanding.

One problem in understanding the mechanisms underlying vigilance decrement is the diversity of the tasks used in its study and the limited integration of those tasks into the cognitive literature. A novel continuous performance task used in recent cognitive and brain imaging studies promises to build broader connections to cognitive and brain functions. As reviewed next, this task produces notably high test-retest reliability in the assessment of attention function, and has been used to uncover a brain network index of sustained attention ability. It is cross-validated across people with normal attention function and those with attention deficit hyperactivity disorder (ADHD). The task also provides a new avenue for understanding the various mechanisms that contribute to vigilance decrement.

Gradual onset continuous performance task (gradCPT)
Intriguing new findings on sustained attention have emerged using a new experimental paradigm: the gradual onset continuous performance task (gradCPT; Esterman et al., 2013). Like traditional continuous performance tasks (CPT), the gradCPT presents participants with a long sequence of stimuli comprised of targets and nontargets. Participants are instructed to press a button for 90% of the stimuli (e.g., city images) and withhold responses for the other stimuli that rarely occur (e.g., mountain images). To increase the task demands relative to that found in standard CPT tasks, Esterman et al. (2013) created blends for temporally adjacent images, such that one image gradually emerges into clarity as the preceding image fades out and the next image fades in. This procedure creates uncertainty as to when an image is presented (i.e., its onset is gradual); it also reduces the duration during which the image is clearly in view, and increases interference from adjacent images. Healthy adults typically miss between 10% to 30% of the infrequent stimuli. Erroneous omission of responses to frequent stimuli is in the range of 1–3% (Esterman et al., 2013; Rosenberg et al., 2013). When these two types of errors are combined in a sensitivity measure, $d'$, substantial and consistent individual differences are observed in healthy adults: split-half reliability exceeded .90 in Rosenberg et al. (2016a). These individual differences in behavior have corresponding effects in the brain. An index extracted from the connectivity patterns among 268 nodes in the brain during an fMRI scan reliably predicts an individual’s $d'$ in the gradCPT (Rosenberg et al., 2016a; see also Fortenbaugh, Rothlein, McGlinchey, DeGutis, & Esterman, 2018). This index, considered a global measure of one’s ability to sustain attention, has been validated in both functional scans when people engage in the gradCPT, and resting scans when people rest in the scanner. The role of the gradCPT in uncovering a global measure of sustained attention makes this paradigm an especially promising approach in studies on attention.

The novelty of the gradCPT paradigm, however, also means that few characteristics about sustained attention are established in this task. Chief among them are the time course and the mechanisms underlying performance decrement, and the relationship between sustained attention and other cognitive functions. We next discuss these two issues and our consequent experiment goals.

A striking finding from the gradCPT is the rapidity of performance decrement. In Esterman et al. (2013), for example, participants engaged in the gradCPT inside an fMRI scanner. There were three gradCPT runs, each lasting 8 minutes. Esterman et al. observed significant decline in $d'$ within the 8-min runs. Some aspects of the fMRI environment may have contributed to the rapidity of the performance decrement. These include lying supine in a position that may induce sleepiness, and exposure to repetitive and loud scanner noises. However, similar findings were observed in studies that tested participants outside of the scanner, in typical upright positions and in quiet lab environments (Esterman, Reagan, Liu, Turner, & DeGutis, 2014; Esterman et al., 2016; Rosenberg et al., 2013; van den Brink et al., 2016). However, not all behavioral studies reveal clear evidence of a reduction in $d'$. In the study of Rosenberg et al. (2013), for example, response time variability and commission errors increased, but omission errors and mean response time did not increase, raising questions about whether the rapid change in performance reflected a true decrement in vigilance.

The first goal of the present study is to provide an independent replication of the vigilance-related performance decrement in the gradCPT task. The experiments reported here use temporal parameters comparable to those used by Esterman et al. (2013), where participants engaged in 3 blocks of the gradCPT task, each lasting 8 minutes. We aim to characterize performance changes across and within the blocks. Additional experiments allow us to evaluate the contribution of several factors to gradCPT performance. These factors are related to the
various accounts previously postulated to contribute to vigilance performance: stimulus novelty, response rate, and the prevalence rate of targets. The standard gradCPT employs a small set of visual stimuli that appear repetitively throughout the experiment. The lack of novelty in the stimuli may have led to visual adaptation, boredom, and an increase in mindlessness. In addition, stimuli requiring a response occur at a high rate, yielding frequent and repetitive motor responses that may be difficult to inhibit. Finally, targets that induce a change of response occur infrequently (e.g., 10%), a condition that contributes to boredom and mismatches untrained observers’ default response rates. These three factors, individually or jointly, may contribute to performance decrement owing to mindlessness, impulsivity in responses, and probability (mis)match. By manipulating stimulus novelty, frequency of responses, and the occurrence rate of target stimuli, the experiments reported here shed light on the mechanisms underlying performance decrement.

Beside performance decrement, a second important question regarding sustained attention is its relationship to other cognitive functions. The gradCPT has been instrumental in indexing the sustained attention network (Rosenberg et al., 2016a). In fact, the gradCPT has been linked to a wide range of other cognitive tasks, including the Attention Network Task (Rosenberg et al., 2018), the stop signal task (Rosenberg et al., 2016b), and reading comprehension (Jangraw et al., 2018). Yet performing the gradCPT calls on a unique combination of visual, motor, and attentional components. Brain regions that carry heavy weights in the sustained attention network index are in the cerebellum, the temporal lobe and the occipital cortex, rather than the dorsal frontoparietal regions more commonly associated with spatial attention and cognitive control. This discrepancy raises questions about how sustained attention and performance on the gradCPT relate to other cognitive functions, including ones closely related to cognitive control. Rosenberg et al. (2013) explored this question by correlating individual differences on the gradCPT task with self-reported mindfulness and daily experience of attention lapses. They found significant correlations between commission errors in the gradCPT and self-reported attention-related cognitive errors. These findings, however, have not been independently corroborated. Here we attempt such a test.

In this study, we re-examined the correlation between gradCPT performance and other cognitive measures. We conducted a confirmatory analysis on the correlation between gradCPT performance, self-reported attention lapses in daily life, and a self-reported mindfulness trait. In addition, we included a standardized assessment of complex working memory span – the operation span task – as an index of working memory and cognitive control. If individual differences in the gradCPT share common mechanisms with individual differences in working memory and cognitive control, then performance on the gradCPT should correlate with complex working memory span. Individual differences in everyday media multitasking propensities were also measured as an additional trait index of cognitive control (Ophir et al., 2009). Together, these systematic empirical probes and theoretical development yield insights about key characteristics of sustained attention in a demanding continuous performance task.

Experiment 1

Experiment 1 aims to replicate the rapidity of performance decrement in the gradCPT. Participants performed three 8-minute blocks of the gradCPT while seated at a desktop computer. The blocks were separated by a short break of 30 seconds. Similar to Esterman et al. (2013, 2016), the visual stimuli were natural scenes from two categories: mountains and cities. Participants were asked to press a button as quickly as possible when they saw a city image,
which occurred on 90% of the trials, and to withhold responses for mountain images, which occurred on 10% of the trials. We examined how performance changed within and across blocks. Because a small set of stimuli were used repetitively, the visual properties of the task may induce habituation and boredom. In addition, owing to the high required response rate, the procedure may induce habitual motor responses that are difficult to inhibit. Finally, the rarity of “no-go” trials is likely to produce the rare target effect, manifested as a high likelihood of missing the rare occurrence of mountains (Wolfe, Horowitz, & Kenner, 2005). Together with the possibility of attentional resource depletion, these conditions are favorable for producing performance decrement over time.

In addition to the gradCPT, participants completed a computerized complex working memory span task, and paper-and-pencil questionnaires assessing mindfulness and media multitasking traits. The individual differences results will be reported separately from the gradCPT results, after all experiments have been described.

Method

Participants. We planned to test 32 participants in each experiment. For power analyses we used previously reported effect sizes of $\eta^2_p = .54$ for vigilance decrement (Esterman et al., 2016), and $r = .473$ for the correlation between error rates in the gradCPT and self-reported attention lapse (Rosenberg et al., 2013). G*Power analysis (Faul, Erdfelder, Lang, & Buchner, 2007) showed that if those effect sizes apply, our sample size corresponds to a power greater than .95 for detecting a vigilance decrement, and a power greater than .80 for detecting a correlation between gradCPT and attention lapse.

Thirty-two participants, 20 females and 12 males, completed Experiment 1. All participants had normal or corrected-to-normal visual acuity and received extra course credit or $10/hour for their time. Their ages ranged from 18-35 years, with a mean age of 20.3 years. The study was approved by the University of Minnesota IRB and all participants provided written informed consent.

Materials and apparatus. Participants were tested individually in a room with normal interior lighting. The experiments were programmed using Psychtoolbox (Brainard, 1997; Pelli, 1997) implemented in MATLAB (http://www.mathworks.com). Stimuli were presented on a 17’’ CRT monitor with a spatial resolution of 1024 × 768 pixels and a refresh rate of 75 Hz. Viewing distance was unconstrained but was approximately 40 cm.

The entire stimulus set comprised 10 mountain images and 10 city scenes in grayscale. The scenes were cropped to occupy a circle with a radius of 7.2°. Each scene was displayed at the center of the monitor with a white background.

Design and procedure: GradCPT. Following Esterman et al. (2013), participants were tested in three, 8-minute runs of the gradCPT task. Each run contained 600 images presented at a pace of 800ms/image, for a total of 8 minutes. Ninety-percent of the images were chosen randomly from the 10 city scenes, and the remaining 10% of the images were chosen randomly from the 10 mountain images, with the constraint that an image would not repeat consecutively. To create gradual onsets, scenes in adjacent temporal positions were linearly interpolated pixel-by-pixel, from 100% of the values used in scene 1 and 0% of the values representing scene 2 at time zero, to 100% of scene 2 and 0% of scene 1 at time 800ms, with a gradual change in the percentage every 50ms. Participants were asked to press “c” whenever they saw a city scene, and to withhold responses to mountains (Figure 1).
Prior to the main experiment, participants practiced the task for 1 minute, followed by three 8-min blocks of the gradCPT. A timer counted down from 30s to 1s between blocks, for a break time of 30s.

**Figure 1.** An example trial sequence used in the gradCPT task. Images from adjacent time points were blended to produce gradual onset. A slightly different percentage of blends was presented every 50 ms, using linear interpolation between adjacent images. Sample trials can be found at http://jianglab.psych.umn.edu/gradCPT/gradCPTExp1Demo.mov.

**Response classification.** Owing to the rapid rate of presentation and the gradual transition of images, responses were almost always made as one image faded in and another faded out. We adopted the procedure used by Esterman et al. (2013) to attribute each response to a specific image. First, unambiguously correct responses to the \( n \)th image were those that occurred after the image in the \( n \)th trial was more than 70% coherent, but before the image on \( n+1 \) trial reached 40% coherence. Second, the remaining responses were assigned to the most adjacent trial that received no response. If multiple responses were made in a trial, the fastest one was assigned to that trial.

**Additional measures.** Following the gradCPT, three paper-and-pencil questionnaires were administered: a 15-item mindfulness questionnaire using the Mindful Attention Awareness Scale (MAAS; Brown & Ryan, 2003), a 12-item daily attention lapses questionnaire using the Attention-Related Cognitive Errors Scale (ARCES; Cheyne, Carriere, & Smilek, 2006), and a questionnaire assessing participants’ habits of using multiple media concurrently, measured with the Media Multi-tasking Index (MMI; Ophir, Nass, & Wagner, 2009; Pea et al., 2012). Finally, participants completed a computerized complex working memory span task adapted from Lewandowsky, Oberauer, Yang, and Ecker (2010). In this task, participants evaluated whether a math equation was true or false while remembering a letter. After a variable number of interleaved equations and letters were presented, participants reported the sequence of letters they saw earlier. The sequence varied in length from 4 to 8. Participants were asked to remember the letters in the order they were presented and to not skip any letters during recall. Operation span was defined as the mean number of letters that participants recalled in the correct position.

For each participant, we extracted a single index from each of the four measures: complex working memory span, mindfulness, attention lapses, and multimedia use. Correlations of these indices with the gradCPT performance will be reported after all experiments have been described.

**Results**
To examine changes in performance within and across blocks, we divided each 8-minute block into four 2-minute time bins. Participants were instructed to respond to cities (90% of the trials) and to withhold responses to mountains (10% of the trials). Errors were of two types: an erroneous response to mountains, or P(“go response” | “no-go stimulus”), or an omission of response to cities, or P(“no response” | “go stimulus”). These errors are commonly combined to produce a single sensitivity index, $d'$ or its nonparametric equivalent $A'$. The parametric assumptions underlying $d'$ – similar variance for noise and for noise plus signal distributions – are often violated in experiments involving infrequent targets (Wolfe et al., 2007). Here we chose to report results from the nonparametric index $A'$ (Grier, 1971; see also Helton & Russell, 2011). Qualitatively similar results were obtained for $A'$ and $d'$. Figure 2 (left) presents the mean $A'$ across blocks. The appendix lists $d'$, response bias $c$, false alarms, and misses in this and subsequent experiments.

We conducted a repeated-measures ANOVA on $A'$, including block (1-3) and time bin (1-4) as within-subject factors. We report the multivariate test results because it does not assume sphericity in the ANOVA (Lane, 2016). Results revealed significant main effects of block, $F(2, 30) = 11.01, p = 2.59 \times 10^{-4}, n_p^2 = .42$, with a significant linear trend, $F(1, 31) = 13.24, p = 9.86 \times 10^{-4}, n_p^2 = .30$, and time bin, $F(3, 29) = 4.83, p = 7.60 \times 10^{-3}, n_p^2 = .33$, with a significant linear trend, $F(1, 31) = 11.83, p = 1.68 \times 10^{-3}, n_p^2 = .28$, and no interaction, $F = .65, p = .69$. $A'$ declined both across the three blocks, and across the four time-bins within single blocks. These data showed that performance declined over time, and the decline occurred rapidly, detectable within single 8-minute blocks.

![Figure 2. Results from Experiment 1. Participants were asked to press a button to cities (90% of the trials) and to withhold responses to mountains (10% of the trials). (a). Data from all trials combined. (b). Separate plots of unambiguous and ambiguous responses. Error bars show ± 1 S.E. of the mean across participants.](image)

A unique methodological challenge of the gradCPT task was the attribution of responses to trials. Images from one trial gradually transitioned to images from the next trial. Any response participants made could correspond to either image. As indicated in the method section, we adopted the approach previously used by Esterman et al. (2013) to classify responses. Most responses (82%) were unambiguous and could be clearly attributed to one image or the other. To verify that the pattern of results presented so far was not predominantly carried by ambiguous trials, we separately analyzed unambiguous and ambiguous trials. As shown in Figure 2 (right), the results held for both types of trials. Response classification did not
interact with block, $F(2, 30) = .13, p = .88, n_p^2 = .01$, or with time bin, $F(3, 29) = 2.00, p = .14, n_p^2 = .17$. This was also true of subsequent experiments, so data from all trials were combined in all subsequent experiments.

**Discussion**

Using a well-powered sample, and an independent replication, Experiment 1 provided clear evidence of performance decrement in the gradCPT task. Detection sensitivity $A'$ declined significantly both across the three blocks (each 8-minute long), and within the four 2-minute time-bins of each block. In the next experiment, we examine whether high response rates (that is, the frequency of targets requiring a response) contributed to performance on the gradCPT.

**Experiment 2**

Experiment 2 manipulated the frequency of motor responses while keeping other aspects of the task comparable to those of Experiment 1. The required response rate in Experiment 1 was 90% - participants responded to all of the city images, which occurred 90% of the time. In Experiment 2, the required response rate was 10%. Specifically, we asked participants to respond to the 10% of trials with mountain images and withhold responses to the 90% of trials depicting city images. Because the frequency of required responses was much lower, repetitive motor habit should be slower to emerge. We examined how a change in the required response rate affected (i) overall $A'$, and (ii) the degree of performance decline over time. If response inhibition contributes to overall performance and to the performance decrement, then overall $A'$ should be higher, and its decrement over time should be less obvious, in Experiment 2 than it was in Experiment 1. On the other hand, a reduction in response rate lessens task load and may increase boredom and the emergence of mindlessness. If mindlessness is a key contributor to gradCPT performance, then $A'$ may be lower and the decrement in $A'$ over time may be more severe in Experiment 2 than in Experiment 1.

**Method**

**Participants.** Thirty-two new participants completed Experiment 2. There were 23 females and 9 males with a mean age of 20.7 years.

**Design and Procedure.** This experiment was the same as Experiment 1, except for the altered task-response requirement in the gradCPT. Instead of pressing a button for city images, participants were asked to press a button for mountain images and to withhold responses to city images. The ratio of city images to mountain images remained 9:1.

**Results**

Figure 3 shows results from Experiment 2. An ANOVA including block (1-3) and time bin (1-4) as within-subject factors revealed no main effect of block on detection sensitivity $A'$, $F(2, 30) = 2.18, p = .13, n_p^2 = .13$. The main effect of time bin failed to reach significance, $F(3, 29) = 1.28, p = .30, n_p^2 = .12$. Block and time bin did not interact, $F(6, 26) = 1.08, p = .40, n_p^2 = .20$. Trend-analysis showed a marginally significant linear trend of block, $F(1, 31) = 4.48, p = .04, n_p^2 = .13$, and of time bin, $F(1, 31) = 3.05, p = .09, n_p^2 = .09$. 
Figure 3. Results from Experiment 2. Participants were asked to press a button whenever they saw mountain images (10% of the trials) and to make no responses to city images (90% of the trials). Error bars show ± 1 S.E. of the mean across participants. Data from Experiment 1 are re-plotted for comparison.

To directly examine effects of required response rates on the gradCPT, we conducted an ANOVA using required response rate (Experiments 1 vs. 2) as a between-subject factor, and block and time bin as within-subject factors. This comparison showed a significant main effect of response rate, as $A'$ was significantly higher when the required response rate was lower, $F(1, 62) = 36.40, p = 9.77 \times 10^{-8}, n_{p^2} = .37$. There were significant main effects of block, $F(2, 61) = 12.26, p = 3.3 \times 10^{-3}, n_{p^2} = .29$, and time bin, $F(3, 60) = 5.72, p = 1.65 \times 10^{-3}, n_{p^2} = .22$. Response rate significantly interacted with block, as the decrement across blocks was less salient in Experiment 2 than in Experiment 1, $F(2, 61) = 3.74, p = 2.94 \times 10^{-2}, n_{p^2} = .11$. Neither the interaction between time bin and response rate, $F(3, 60) = 1.51, p = .22, n_{p^2} = .07$, nor the interaction between block and time bin, $F(6, 57) = .52, p = .79, n_{p^2} = .05$, nor the three-way interaction between block, time bin, and response rate, $F(6, 57) = 1.10, p = .37, n_{p^2} = .10$, reached significance.

Discussion

Instead of responding to city images that occurred 90% of the time, as was done in Experiment 1, participants in Experiment 2 responded to mountain images that occurred 10% of the time. We found two effects of required response rates on the gradCPT. First, overall performance was higher in Experiment 2 than in Experiment 1, suggesting that high rates of motor responses were detrimental to performance. The effect size, calculated as Cohen’s $f^2$, was .587, corresponding to a large effect size. This finding is consistent with previous studies that varied response rates in continuous performance tasks (Wilson et al., 2016). It suggests that when response rates are high, response inhibition is a salient requirement of the task. Second, performance decrement failed to reach statistical significance following the reversal of the response requirement. This is the smaller of the two effects - Cohen’s $f^2$ was .124, corresponding to a small to medium effect size. The small effect is perhaps not surprising – given the brevity of the task, performance decrement within and across blocks was moderate even in Experiment 1. This leaves little room for a further reduction in effect size. Note, however, that some theories of
vigilance decrement predict that vigilance decrement should increase in Experiment 2 owing to the likely onset of boredom and comparatively lower task load (Manly et al., 1999; Robertson et al., 1997). Our results do not support this prediction. Instead, when the required response rate decreased, performance decrement over time became less notable in the relatively short task period used in our study.

Our study suggests that a change in response rate did not simply influence response criterion. If participants had simply developed a strong tendency to make a response, this should have increased hit rates as well as false alarms, leaving detection sensitivity unaltered. Instead, as motor response rates increased, greater cognitive control is needed to inhibit that response. This increased the overall task demand, impairing overall performance and exacerbating the decrement function over time. Response inhibition therefore constitutes a significant contributing factor in the gradCPT.

Experiment 3

Experiment 3 examined the interaction between the gradCPT performance decrement and the novelty of the stimulus images presented during the task. Mind wandering, the tendency for one to have task-unrelated thoughts, is greater for simple and repetitive tasks than for complex ones (Antrobus, Singer, Goldstein, & Fortgang, 1970). One may therefore expect performance decrement to be more pronounced for repeated stimulus images than for novel ones. To this end, we alternated 2-minute bins of repetitive periods with novel periods. In the repetitive periods, the same set of 12 city images were repeatedly sampled and used. In the novel periods, completely novel city images were used without any repetitions. We examined whether repetition of images modulated performance on the gradCPT. Other than the novelty of images, the required response rate reverted back to 90% as in Experiment 1.

Method

Participants. Thirty-two new participants completed Experiment 3. There were 27 females and 5 males with a mean age of 20.6 years.

Materials and stimuli. The stimulus set was comprised of 1,096 city images and 24 mountain images. Out of the total set, 1,072 city and 12 mountain images were randomly assigned to the stimuli for the gradCPT, and the rest were used for a surprise memory test that was administered after participants completed the gradCPT.

Design and Procedure. Similar to Experiment 1, in each block the city images appeared on 90% of the trials, to which participants were instructed to make a button press, and the mountain images appeared on the remaining 10% of the trials, to which no response should be given. Participants completed 4 blocks of the task, each of 8 minutes, with a 30s break between blocks. Each block of the 8 minutes was divided into four 2-minute-long time bins. These four bins alternated between novel and repetitive bins, and the order of alteration was counterbalanced across the four blocks. In the repetitive bin, city images were drawn from a fixed set of 12 images chosen at random from the total set (a different set of 12 images was used for different participants). In the novel bin, city images were chosen from the larger pool such that no image would repeat within a bin, and none of the images had appeared in the repetitive bins. Half of the mountain images were used during the repetitive bins and the other half were used in the novel bins.

At the completion of the gradCPT task, participants were given a surprise memory test. They were presented with 36 city images and 24 mountain images, one at a time, and asked to...
indicate whether each image was old (i.e., had been shown during the gradCPT task) or new (i.e., was never presented during the gradCPT task). Out of the 36 city images, 12 were randomly chosen from the novel bins, 12 from the repetitive bin, and the remaining 12 were new city images. Out of the 24 mountain images, 12 were from the gradCPT task and the rest were new.

Results

Because the novel (N) and repetitive (R) bins were presented in alternating orders, as NNRR, NRRN, NRNR or RNRN, the four time bins within a block did not comprise the same stimuli. That is, owing to the interleaved presentation of different stimulus types and the counterbalancing of orders, the novel and repetitive time bins did not neatly fall into a specific time bin across blocks. We therefore first report an analysis where stimuli within a block were averaged. In this analysis, we averaged data from the two novel bins of each block, and averaged data from the two repetitive bins of each block, and evaluated changes in performance across blocks for the two types of trials. A subsequent analysis tested time bin by separating the first and second occurrence of a condition (i.e., novel or repetitive) within each block.

1. Sensitivity A’ in the gradCPT

Performance on the gradCPT was insensitive to the novelty of the stimuli (Figure 4a). For A’, an ANOVA including stimulus type (novel vs. repetitive) and block (1-4) revealed no main effect of stimulus type, $F(1, 31) = .06, p = .81, n_p^2 = .002$, and no interaction between stimulus type and block, $F(3, 29) = .14, p = .94, n_p^2 = .01$. Performance declined across the four blocks, producing a significant main effect of block, $F(3, 29) = 9.68, p = 1.38\times10^{-4}, n_p^2 = .50$, accompanied by a significant linear trend, $F(1, 31) = 17.31, p = 2.33\times10^{-4}, n_p^2 = .36$.

![Figure 4](image_url)

Figure 4. Results from Experiment 3. City images used within a 2-min time bin were either sampled from a fixed set of 12 (“repetitive”) or a large set of 1,060 (“novel”) images. (a) Data from the novel and the repetitive bins were combined within a block. (b) Data shown separately for the first and second occurrence of each stimulus condition within a block. Error bars show ± 1 S.E. of the mean across participants.

In an additional analysis, we examined the first and second occurrence of a specific stimulus type (e.g., novel or repetitive) within a block. We ran an ANOVA including time bin (the first or second bin), stimulus type (novel vs. repetitive), and block (1-4). Consistent with the main analysis, there was no main effect of stimulus type, $F(1, 31) = .05, p = .82, n_p^2 = .002$, and no
interaction between stimulus type and block, $F(3, 29) = .40, p = .76, n_p^2 = .04$, or between stimulus type and time bin, $F(1, 31) = 2.22, p = .15, n_p^2 = .07$. Performance declined within a block, i.e., worse performance in the second than the first time bin, $F(1, 31) = 9.77, p = 3.84 \times 10^{-3}, n_p^2 = .236$, as well as across blocks, $F(3, 29) = 9.04, p = 2.20 \times 10^{-4}, n_p^2 = .48$, and there was no interaction between time bin and block, $F(3, 93) = .67, p = .58, n_p^2 = .07$.

2. **Old and new recognition memory**

Given the lack of an effect of stimulus novelty on gradCPT sensitivity, one might wonder whether participants processed each image at the level of exemplars. In the recognition test we examined their memory for the images. For mountain stimuli, participants were always exposed to the same set of 12 mountain images in both the novel and repetitive runs. They recognized 85% of the old images as “old”, which was significantly higher than the false alarm rate of 31% for new mountain images, $t(31) = 14.06, p = 5.37 \times 10^{-15},$ Cohen’s $d = 2.49$.

For city stimuli, participants were exposed to 1,060 novel images in the novel periods and 12 images in the repetitive periods. They recognized 84% of the repeated set, which was significantly higher than the false alarm rate of 39% for foil city images, $t(31) = 12.61, p = 9.69 \times 10^{-14},$ Cohen’s $d = 2.23$. In contrast, recognition rate for the stimuli used in the novel periods was only 46%, a level not significantly higher than the false alarm rate for foils, $t(31) = 1.70, p = .10$, Cohen’s $d = 0.31$. Memory for the repetitive city images was significantly higher than for the non-repeated ones, $t(31) = 10.22, p = 1.92 \times 10^{-11},$ Cohen’s $d = 1.81$.

**Discussion**

Even though the use of repetitive stimuli may often induce habituation and boredom, Experiment 3 showed that this factor did not substantially contribute to performance in the gradCPT. For both periods with novel images and those with repetitive images, we observed significant decline in performance across blocks. Memory performance in the recognition test suggests that the images had been encoded at the exemplar level. This finding runs counter to the idea that an increase in boredom, owing to stimulus repetition, contributes to vigilance decrement. Stimulus repetitiveness is not a major contributor to performance decrement in the gradCPT.

**Experiment 4**

In the first three experiments targets – mountains – rarely occurred among a sequence of nontargets – cities. When errors for the two types of trials were separately examined, error rate was much higher for mountains than for cities, and this finding held regardless of whether mountains entailed a no-go response (Experiments 1 and 3) or a go response (Experiment 2), all $ps < .01$. This finding is reminiscent of the “rare target effect” seen in visual search, where people frequently miss target objects if the target prevalence is low (Wolfe et al., 2005). In some experiments, the rare target effect increases with time on task, becoming more pronounced in the second half of the experiment than in the first (Wolfe et al., 2007, Experiment 1). If performance decrement in the gradCPT is partly driven by rapid onset of the rare target effect, then such decrement should be less obvious in experiments where the target to nontarget ratio is 1:1. Experiment 4 tested this prediction by increasing the target prevalence rate to 50%.

As the target prevalence rate changed, the required motor response rate also changed. The 50% response rate was intermediate between that used in Experiment 1 (90% response rate) and Experiment 2 (10% response rate). If the task-required response rate is a major factor in
determining performance on the gradCPT, then the overall $A'$ in Experiment 4 should fall somewhere between that observed in the first two experiments. The degree of performance decrement over time may also be intermediate, though we may be unable to detect such a small effect size.

**Methods**

*Participants.* Thirty-two new participants, including 22 females and 10 males, completed Experiment 4. Their mean age was 20 years.

*Design and Procedure.* This experiment was identical to Experiment 1 except for the following changes. The stimulus stream was composed of an equal number of mountain and city images. Half of the participants were asked to press “c” whenever they saw city images and to withhold responses to mountains. The other half were told to press “m” whenever they saw mountain images and to withhold responses to cities.

**Results**

Figure 5 presents the primary sensitivity results from Experiment 4.

![Figure 5](image)

*Figure 5.* Results from Experiment 4. There were an equal number of city and mountain trials. Half of the participants were instructed to respond to city images (while withholding responses to mountains) and the other half were instructed to respond to mountain images (withholding responses to cities). Error bars show ± 1 S.E. of the mean across participants.

As can be seen from Figure 5, sensitivity $A'$ in the gradCPT significantly declined across blocks. An ANOVA including block and time bin revealed a significant main effect of block, $F(2, 30) = 6.64, p = 4.10 \times 10^{-3}, n_p^2 = .31$, but no main effect of time bin, $F(3, 29) = .68, p = .57, n_p^2 = .07$. The interaction between the two factors failed to reach significance, $F(6, 26) = 1.87, p = .13, n_p^2 = .30$. Trend analysis showed a significant linear decline in $A'$ across blocks, $F(1, 31) = 11.01, p = 2.32 \times 10^{-3}, n_p^2 = .26$, but not a linear effect of time bin, $F(1, 31) = 1.59, p = .22, n_p^2 = .05$.

A direct comparison between Experiment 1 (90% go trials) and Experiment 4 (50% go trials) showed only a significant main effect of experiment: Participants in Experiment 4 had
higher $A'$, $F(1, 62) = 16.67, p = 1.29\times10^{-4}, \eta^2_p = .21$. The interaction between experiment and block was marginally significant, $F(2, 61) = 3.00, p = .06, \eta^2_p = .09$, with a significant linear trend in the interaction term, $F(1, 62) = 4.43, p = 3.94\times10^{-2}, \eta^2_p = .07$, suggesting that performance decrement was more pronounced when response rate was higher. Comparing Experiment 2 (10% go trials) and Experiment 4 (50% go trials) also showed a significant main effect of experiment in that participants in Experiment 4 had lower $A'$, $F(1, 62) = 12.65, p = 7.25\times10^{-4}, \eta^2_p = .17$. None of the interaction effects reached significance.

An ANOVA including all three experiments (Experiments 1, 2, and 4) as a between-subjects factor showed that experiment significantly interacted with blocks, $F(4, 186) = 2.46, p = 4.68\times10^{-2}, \eta^2_p = .05$, suggesting that performance decrement was not identical across different response rates (90%, 50%, and 10% for Experiments 1, 4, and 2 respectively). This interaction was accompanied by a marginally significant linear trend in the interaction term, $F(2, 93) = 2.36, p = 9.96\times10^{-2}, \eta^2_p = .05$. The interaction between experiment and time bin was not significant, $F(6, 184) = .88, p = .51, \eta^2_p = .03$, without any significant trend.

Discussion

Experiment 4 examined the role of infrequent targets on gradCPT performance. Using a target-to-nontarget ratio of 1:1, the experimental condition was no longer subjected to the rare-target effect, the phenomenon in which infrequent targets are often missed (Wolfe et al., 2005). Despite the use of high target prevalence, Experiment 4 observed significant decline in $A'$ across blocks. This finding suggests that performance decline is a pervasive finding and not specific to situations with rare targets. The overall level of performance, however, varied across experiments. The order of performance level, highest in Experiment 2, intermediate in Experiment 4, and lowest in Experiment 1, was inversely related to the required motor response rate. Performance decrement was less obvious when the required response rates were lower. There was also suggestive evidence that response inhibition contributed to performance decrement, as revealed by the interaction between Experiment (1, 4, and 2) and block. These findings suggest that response inhibition contributes to performance on the gradCPT.

Additional evidence that response inhibition affected the gradCPT is provided by an analysis of the two types of errors. Appendix Table A lists the two types of errors across time: (i) Commission error rate is the proportion of trials on which participants erroneously pressed a key to an image that should not have received a response, and (ii) omission error rate is the proportion of trials on which participants failed to respond to an image that should have received a response. If response inhibition became increasingly demanding later in the experiment, then commission error rates should increase over time, but omission error rates may be less likely to increase over time. This was the case in Experiment 1 – an increase in commission error, $F(2, 30) = 7.63, p = 2.10\times10^{-3}, \eta^2_p = .34$, without an increase in omission error, $F(2, 30) = 1.76, p = .19, \eta^2_p = .11$, across blocks. As the required response rates declined to 10% in Experiment 2, we no longer observed an increase in either commission, $F(2, 30) = .49, p = .62, \eta^2_p = .03$, or omission errors, $F(2, 30) = 2.39, p = .11, \eta^2_p = .14$, across blocks.

Rapidity of performance decrement in the first 8-min of the vigil

To test the extremely fast onset of performance decrement, we examined the effect of time bins in Block 1 (see Figure 6). Note that time bin on its own produced inconsistent statistical results across individual experiments. Here we evaluated the effect of time bin in the first block when combining across the full dataset. Because stimulus novelty had no effect on performance in Experiment 3, trials were analyzed regardless of whether they came from...
repetitive or novel periods. An ANOVA including time bin (1-4) and experiment (1-4) revealed a main effect of time bin, $F(3, 122) = 7.75, p = 8.8 \times 10^{-5}, n_p^2 = .16$, accompanied by a significant linear trend, $F(1, 124) = 22.04, p = 7.0 \times 10^{-6}, n_p^2 = .15$. The main effect of experiment was significant, $F(3, 124) = 19.18, p = 2.80 \times 10^{-10}, n_p^2 = .32$. Performance related inversely to required response rate. Time bin did not interact with experiment, $F(9, 372) = .48, p = .89, n_p^2 = .01$. Planned comparisons showed that, considered individually and on their own, none of the experiments observed a significant main effect of time bin in the first block, $F(3, 29) = 2.12, p = .12, n_p^2 = .18$ in Experiment 1, $F(3, 29) = 2.26, p = .10, n_p^2 = .19$ in Experiment 2, $F(3, 29) = 2.24, p = .11, n_p^2 = .19$ in Experiment 3, $F(3, 29) = 1.65, p = .20, n_p^2 = .15$ in Experiment 4. The effect, however, was detectable when the full sample was considered. Thus, we observed a small, but consistent, decline in $A'$ in the first 8-min of the vigil. The magnitude of decline is comparable to previous work on vigilance decrement using other paradigms (Ariga & Lleras, 2011; Helton & Russell, 2011, 2012; Temple et al., 2000). Although the moderate size of the decline may not be practically significant in some activities where there is ample opportunity for error correction and little is at stake, the rapidity of the decrement may have important implications for other activities for which small errors can lead to catastrophe, such as in air traffic control, or many types of security screening.

![Figure 6](image_url)  
**Figure 6.** Performance change in the first 8-min of the vigil. Data from the first block were separated into 2-min time bins. Error bars show ± 1 S.E. of the mean across participants.

**Relationship to other cognitive functions**  
In an exploratory analysis, Rosenberg et al. (2013) observed significant correlations between performance on the gradCPT and self-reported ratings relating to sustained attention. Specifically, individuals who indicated they had higher cognitive errors in daily life (assessed by the ARCES) demonstrated higher commission error rates in the gradCPT, with a reported $r$ of 0.47. In addition, individuals who indicated more frequent experiences of mind wandering (measured by the MAAS) missed targets more frequently, with a reported $r$ of -0.72, when there were task-irrelevant distractors in the background. Our study was not a strict replication of Rosenberg et al. (2013) given differences in stimuli: we used scenes as the main task and no background noise, Rosenberg used faces as the main task and included background noise for
some participants. Nonetheless, it is worth examining whether correlations reported between gradCPT and mental lapses in daily life also holds in our study.

Therefore, we performed a planned analysis on the correlation between the commission error rates in the gradCPT and the ARCES, and the omission error rates and the MAAS scores from Experiment 1 — that is, the experiment that had the same ratio of “go” and “no-go” trials (9:1) as in Rosenberg et al. (2013). As indicated in the method section, the sample size was adequate to detect an effect size comparable to that reported in the Rosenberg et al. study. In addition, we also conducted four planned correlation analyses, between the gradCPT A’ and (i) self-reported mindfulness, as measured by MAAS, (ii) self-reported attention-related cognitive errors (ARCES), (iii) concurrent multi-media usage in daily life, and (iv) complex working memory span. These further correlational analyses were performed on data from Experiments 1, 2, and 4. Experiment 3 did not assess the self-reported questionnaires and the working memory span. Because no previous studies performed the same exploratory analyses as we did, a priori effect size was unknown. Examining the correlational analyses across three separate experiments provided some evidence for statistical consistency.

(1) Confirmatory analysis: correlation between error rates in the gradCPT and self-reported questionnaires

Participants who showed higher commission error rates in one block also tended to have higher commission errors in another block. The average correlation across blocks (Pearson’s r) was 0.82 (all ps < 1.62×10^-7 in Experiment 1), suggesting that commission errors were stable within an individual. The correlation between commission errors and ARCES was not replicated, Pearson’s r < .10, p > .57 (Figure 7). This analysis fails to support the idea that commission errors in the gradCPT are related to one’s self-reported attention-related cognitive errors. Test-retest reliability for the (less frequent) omission errors across blocks was likewise high in Experiment 1. Participants who had missed targets in one block were more likely to miss targets in another block, with an average correlation across blocks of .64 (p < .001). The correlation between omission errors and MAAS was not replicated, Pearson’s r < .06, p > .75 (Figure 7). Our data fail to support the idea that omission errors in the gradCPT are related to self-reported mindfulness. These results were replicated after we removed one outlier in Experiment 1 whose sensitivity (A’) was at chance level.
(2) **Exploratory analysis: gradCPT and other measures**

Performance on the gradCPT was instrumental in uncovering a global index of sustained attention using fMRI (Rosenberg et al., 2016a). How does the gradCPT performance relate to other cognitive functions? Four other individual differences measures were obtained in Experiments 1, 2, and 4: mindfulness (MAAS score), self-reported cognitive errors (ARCES score), multimedia usage (MMI score), and complex working memory span (Ospan). Figure 8 shows separate scatterplots, and the obtained Pearson’s correlations, between sensitivity ($A'$) on the gradCPT and each of the four indices. Similar results were obtained when $d'$ was used. Bonferroni corrected critical alpha level is 0.0125 owing to multiple comparisons.

These analyses showed no consistent correlations between gradCPT performance and mindfulness, multimedia use, working memory span, and self-reported attention-related cognitive errors (Figure 8). The lack of correlation was not due to the unreliability of the individual measures. For example, test-retest reliability of gradCPT $A'$, as measured by the correlation across different blocks of the task, was .84 in Experiment 1, .71 in Experiment 2, .71 in Experiment 3, and .78 in Experiment 4, all highly significant, $p$ values ranged from $4.68 \times 10^{-13}$ to $3.61 \times 10^{-3}$. In addition, two questionnaires assessing attention lapses versus the opposing mental state of mindful awareness in daily life yielded, as expected, significantly negatively correlated results, $r = -.58$, $p = 5.34 \times 10^{-4}$ in Experiment 1, $r = -.50$, $p = 3.75 \times 10^{-3}$ in Experiment 2, and $r = -.57$, $p = 6.34 \times 10^{-4}$ in Experiment 4. Finally, both working memory span (Lewandowsky et al., 2010) and multimedia usage (Ophir et al., 2009; Pea et al., 2012) have established validity in previous studies. These findings suggest that cognitive mechanisms underlying sustained attention in the gradCPT are not strongly associated with those underlying mindfulness, attention lapses in daily life, or importantly, complex working memory span. We conducted the same analyses after removing one outlier in Experiment 1 whose sensitivity ($A'$) was at chance level and the results were replicated.
Figure 8. Scatter plots illustrating the relation between sensitivity ($A'$) on the gradCPT and the mean score in the Mindful Attention Awareness Scale (MAAS; $p = .74$, $.17$, $.48$ in Experiments 1, 2, and 4, respectively), the Attention-Related Cognitive Errors Scale (ARCES; $p = .46$, $.38$, and $.02$ in Experiments 1, 2, and 4, respectively), the Media Multi-tasking Index (MMI; $p = .38$, $.18$, and $.42$ in Experiments 1, 2, and 4, respectively), and the complex working memory span (Ospan; $p = .57$, $.27$, and $.10$ in Experiments 1, 2, and 4, respectively).

RT variability
Performance decrement in continuous performance tasks has sometimes been reflected as increased variability in response time. In fact, Esterman et al. (2013) divided the gradCPT runs into moments when participants were “in the zone” or “out of the zone” based on a median split of RT variability. Phases of the continuous performance task that, for a given
participant, are associated with low RT variability are considered as on-task, “in the zone” moments. In contrast, phases of task performance associated with high RT variability are considered “out of the zone” moments. We conducted similar analyses on RT variability and found similar results as in Esterman et al. (2013): periods associated with high RT variability also corresponded to greater commission errors. Details of these analyses can be obtained from the authors.

General Discussion

In four experiments using the gradual onset continuous performance task (gradCPT) we observed performance decline over time. Consistent with Esterman et al. (2013), the decline occurred rapidly, including in the first 8-minute block. Our study further tested three factors that may have contributed to the decline in performance: required response rate, the repetitive nature of the task stimuli, and the low target prevalence. Contrary to accounts of vigilance decrement that emphasize the contributions of mindlessness (Manly et al., 1999; Robertson et al., 1997), performance decrement was not greater for repetitive than for novel task stimuli or when the target prevalence rate was lower. These findings, taken together with the outcomes of other studies manipulating task difficulty and secondary working memory load, suggest that vigilance decrement is unlikely accounted for by boredom and mindlessness (Helton & Warm, 2008).

Probability matching

Our findings also do not support the idea that changes in response over time were primarily due to probability matching (Craig, 1978). Untrained observers may overestimate target rates initially, especially when the target rate is low. However, the decrement in performance in our study was found even when the target rate was as high as 50%. To directly assess how participants adjusted their rate of responding, we calculated the proportion of trials on which participants made a response, regardless of whether it was correct. These proportions combined hits (correct responses to go stimuli) and false alarms (responses to no-go stimuli). Table 1 presents these combined (hit plus false alarm) response rates for the first three blocks in each of the four experiments. Perfect probability matching should yield 90% response in Experiment 1 and 85% in Experiment 3. The actual response rate, about 88% in Experiment 1 and 85% in Experiment 3, underestimated the proportion of trials that a response should be made, t(31) = -3.17, p = 3.38×10⁻³, Cohen’s d = -.56 in Experiment 1, t(31) = -5.27, p = 1.00×10⁻³, Cohen’s d = -.93 in Experiment 3. However, response rate did not significantly change across blocks or across time bins – there was no main effect of blocks, F(2, 30) = .73, p = .49, n_p² = .05 in Experiment 1, F(2, 30) = 1.52, p = .23, n_p² = .14 in Experiment 3, the main effect of time bins, F(3, 29) = 1.21, p = .33, n_p² = .11 in Experiment 1, F(3, 29) = .90, p = .45, n_p² = .09 in Experiment 3, or interaction, F(6, 26) = 1.11, p = .39, n_p² = .20 in Experiment 1, F(6, 26) = 1.26, p = .31, n_p² = .33 in Experiment 3. Similarly, in Experiment 4 where perfect probability matching should yield a 50% response rate, participants under-matched, making 45% of responses, t(31) = -9.66, p = 7.27×10⁻¹¹, Cohen’s d = -1.71. Nonetheless, response rate did not significantly change across blocks, F(2, 30) = .29, p = .75, n_p² = .02, or across time bins, F(3, 29) = .92, p = .44, n_p² = .09, and there was no interaction between blocks and time bins, F(6, 26) = 1.75, p = .15, n_p² = .29. An adjustment of response rate over time was seen only in Experiment 2, where perfect probability matching should yield a 10% response rate. Here, participants started out making more responses — the response to target ratio exceeded 1 at the beginning of Block 1. They then downward shifted their response
rate. The overall response rate of 9.4% was significantly lower than perfect matching, $t(31) = -5.30$, $p = 9.00 \times 10^{-6}$, Cohen’s $d = -.94$. The downward adjustment in Experiment 2 yielded significant main effects of time bins, $F(3, 29) = 5.03$, $p = 6.29 \times 10^{-3}$, $n_p^2 = .34$, blocks, $F(2, 30) = 3.40$, $p = .047$, $n_p^2 = .19$, and a marginal interaction, $F(6, 26) = 2.30$, $p = .065$, $n_p^2 = .35$, due to the adjustment occurring primarily in block 1.

The analysis above suggests that although participants tend to probability match their responses to the target rate, this factor is unlikely to account for the vigilance decrements we have demonstrated. Downward adjustment of response rate was only notable in Experiment 2, where the response rate was very low (10%). Yet this was the same experiment that failed to reveal a significant performance decrement. Response rate was relatively constant in Experiments 1, 3, and 4 — all experiments where the vigilance decrement was notable.

### Table 1. Proportion of trials receiving a response. Values in parentheses indicate the standard error of the mean.

<table>
<thead>
<tr>
<th></th>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bin 1</td>
<td>Bin 2</td>
<td>Bin 3</td>
</tr>
<tr>
<td>Exp. 1</td>
<td>.89 (.01)</td>
<td>.88 (.01)</td>
<td>.89 (.01)</td>
</tr>
<tr>
<td>Exp. 2</td>
<td>.10 (&lt;.01)</td>
<td>.10 (&lt;.01)</td>
<td>.09 (&lt;.01)</td>
</tr>
<tr>
<td>Exp. 3</td>
<td>.86 (&lt;.01)</td>
<td>.87 (.01)</td>
<td>.87 (.01)</td>
</tr>
<tr>
<td>Exp. 4</td>
<td>.45 (.01)</td>
<td>.45 (.01)</td>
<td>.44 (.01)</td>
</tr>
</tbody>
</table>

### Stimulus novelty

The visual system rapidly adapts to stimulus repetition (Grill-Spector, Henson, & Martin, 2006). Given the prominent involvement of the occipital lobe in the global index of sustained attention (Rosenberg et al., 2016a), one might expect that the novelty of the task stimuli would influence performance on the gradCPT. In Experiment 3, the repetitive period involved just 12 city images. Over the course of 4 blocks each city image was shown 88 times. This contrasted with the novel period, during which only novel city images were presented. The repetition had some effects on behavior: On the surprise memory test given after the gradCPT task, participants recognized the repeated city images much better than the unrepeated ones. Yet it did not influence $A'$ on the CPT task or the time course of the performance decrement across blocks.

The lack of a novelty effect may hint at a broader feature of gradCPT – its insensitivity to the content of distractor stimuli. Consistent with this idea, Rosenberg et al. (2013) tested two groups of participants using faces as stimuli. One group viewed faces against scenes, another viewed faces against scrambled noise. Despite the greater semantic incongruency of the stimuli presented to the first group, performance was similar between the two groups.

If our manipulation of stimulus novelty had little discernible effect on CPT performance, what other stimulus features might interact with the gradCPT? Previous vigilance research has classified tasks according to the mode of presentation (successive discrimination versus...
simultaneous comparison), rate of presentation (greater or less than 24 stimuli per minute), and information content (cognitive versus sensory; Ballard, 1996). According to this classification, both the novel and repetitive stimuli used in our study fall into the same classification: the gradual transition required successive discrimination, the presentation rate was high, and the stimuli were largely sensory. Changing the scene stimuli to more semantic stimuli, such as alphanumeric characters, thereby manipulating the cognitive versus sensory nature of the information, may influence the results.

Response inhibition

To understand the role of response inhibition in the gradCPT, we focused on two effects – the effect of the required response rates on overall performance, and the performance decrement over time. High response rates substantially lowered the overall sensitivity \( A' \) – an effect with a large effect size (Cohen’s \( f \) exceeded .50 when comparing the 90% and 10% response rates of Experiments 1 and 2). The overall \( A' \) was inversely related to the required motor response rate. Two of these experiments – Experiments 1 and 2, used identical stimuli and included the same requirement for perceptual discrimination. Yet the overall \( A' \) dropped substantially when the task-required response rate was higher (90% in Experiment 1). When the response rate was 50% in Experiment 4, the overall sensitivity was intermediate – it was higher than that in Experiment 1 (90% response rate) but lower than that in Experiment 2 (10% response rate). This finding supports the idea that response inhibition is an important component of the gradCPT. An analysis of how the two types of errors (commission and omission) changed over time corroborated this claim. In Experiment 1, in which response inhibition was strongly required, commission error rates increased over time but omission error rates did not. And when the required response rate was substantially reduced, in Experiment 2, neither type of errors increased with time on task.

The finding that response inhibition contributes to the gradCPT is consistent with more recent work by Rosenberg and colleagues (2016b), who showed that a whole-brain functional connectivity network originally based on the gradCPT predicted performance on the stop signal task that assesses inhibitory control. It is also in line with recent whole brain vowel-wise neuroimaging findings of Fortenbaugh et al. (2018) – in a large sample of 140 Veteran participants – that revealed a cluster of activation in bilateral somatomotor cortex preceding commission errors on the gradCPT, suggesting that "greater activity in the task-relevant motor system or fine motor preparation signals could lead to errors of commission" (pp. 157–158). Although the gradCPT has been instrumental in indexing a sustained attention network, the function of this network goes beyond sustaining attention over time; it also includes one’s ability to exert inhibitory control over responses.

Target prevalence

Frequent alteration of targets and nontargets may be expected to increase the task difficulty, relative to when the targets are infrequent. Yet we found higher \( A' \) when the target to nontarget ratio was 1:1, than when the ratio was 1:9. This finding again supports the idea that response inhibition is a major component of the gradCPT task.

The rare target effect is commonly found in standard visual search tasks, manifested as high miss rates for rarely occurring targets (Wolfe et al., 2007). However, several findings suggest that the target prevalence effect differs from vigilance decrement. Wolfe et al. (2007) showed that, with the exception of one experiment, the target prevalence effect was comparable between the first half and the second half of the experiment, even though vigilance was
expected to decline in the second half. In addition, whereas target prevalence primarily influences response criterion, vigilance decrement manifests as a decline in sensitivity. In this regard, the task-required response rate in our study affected performance differently than target prevalence rate. When the rate of task-required responding was high (because targets occurred frequently and required a response), participants did not simply increase their overall number of responses. If they had done that, it would have raised hits as well as false alarms, leading to a criterion rather than a sensitivity change. Thus, whereas the target prevalence effect in visual search primarily reflects a change in decision threshold (Wolfe & Van Wert, 2010), increasing the response inhibition demands of a continuous performance task adds to the difficulty of the task and impairs sensitivity.

**Individual differences**

The gradCPT yields some of the most robust individual differences among attention tasks. In our study, the reliability of this task across blocks was frequently above .80. Rosenberg et al. (2016a) reported even higher reliability in the behavioral data of their MRI participants – a stunning reliability of 0.975 as measured by Spearman-Brown-corrected split-half correlation between even- and odd-numbered trials. In contrast, many commonly used attention tasks lack such reliability. For example, interference in the Eriksen flanker task was unreliable in a split-half analysis (MacLeod et al., 2010). Even the standard CPT tasks are associated with a moderate degree of reliability, in the range of 0.40-0.50 in a split-half analysis. The high and continuously present demand on multiple aspects of cognitive processing – from visual perception to motor responses and inhibitory control – may be the reason why the gradCPT is so sensitive to individual differences.

Despite the multifaceted nature of the gradCPT, it did not significantly correlate with any of the other measures we assessed, including complex working memory span, mindfulness, cognitive errors in everyday activities, and multimedia use. The lack of correlation was not due to the unreliability of the individual measures. For example, mindfulness scores were highly inversely correlated with daily cognitive errors, with an average correlation of -.55. The lack of correlation between performance on the gradCPT and the individual differences measures suggests that sharing a cognitive component is not sufficient for cross-task correlations. Consider the complex working memory span task and the gradCPT. At a conceptual level, both involve cognitive control of some sort – manipulating information, selecting relevant information, ignoring irrelevant input, and so on. Though such shared processing can be the basis for significant correlations in task performance, each is implemented in different ways. The working memory task involves mathematical calculation, discarding that information, and mentally rehearsing letters. The gradCPT involves selecting rapidly presented stimuli and finely tuning responses to the right moment. It is conceivable that though some common regions of the brain are involved in the two tasks, how they communicate with other regions of the brain differs. On this view, while it is possible to de-construct a cognitive task into individual components, the operation of each component is not independent of other components — in time or space (cf. Calhoun, Miller, Pearlson, & Adali, 2014; Mattar, Betzel, & Bassett, 2016). The degree of independence varies across domains or tasks. This puts constraints on the utility of an individual-differences approach. While significant correlations suggest that two tasks share common processes, a lack of correlation is not sufficient ground for inferring their independence. Future research should further investigate the relationship between the gradCPT and other cognitive functions using tasks that tap into cognitive control, including more direct measures of individual differences in response inhibition, such as the stop-signal task (Logan,
1994; Rosenberg et al., 2016b) or the Go/No-go task (Leimkuhler, & Mesulam, 1985), and within-participant manipulations of the density of targets requiring a response.

**Conclusion**

The present study examined key characteristics of the gradCPT. Vigilance decrement over time was observed in most experiments, but this decrement was not significantly modulated by either the repetitive nature of the visual stimuli or by the prevalence of target events. The frequency of task-required motor responses, however, was inversely related to overall performance, and exacerbated the vigilance decrement. Though the effect size of the vigilance decrement was moderate, the rapidity with which it emerges suggests that sustained attention quickly wanes in continuous performance tasks.
Appendix

Table A1. Two types of errors over four 2-min time bins across the first three blocks, including the proportion of false alarms (Panel A) and the proportion of misses (Panel B). Values in parentheses indicate the standard error of the mean.

(A) Commission errors, or P(“Go response” | “No-go stimulus”)

<table>
<thead>
<tr>
<th></th>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp. 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bin 1</td>
<td>.20 (.03)</td>
<td>.29 (.03)</td>
<td>.29 (.03)</td>
<td>.27 (.03)</td>
</tr>
<tr>
<td>Bin 2</td>
<td>.24 (.02)</td>
<td>.29 (.04)</td>
<td>.29 (.04)</td>
<td>.28 (.03)</td>
</tr>
<tr>
<td>Bin 3</td>
<td>.25 (.02)</td>
<td>.26 (.03)</td>
<td>.26 (.03)</td>
<td>.29 (.03)</td>
</tr>
<tr>
<td>Bin 4</td>
<td>.26 (.03)</td>
<td>.29 (.04)</td>
<td>.29 (.04)</td>
<td>.36 (.03)</td>
</tr>
<tr>
<td>Exp. 2</td>
<td>.003 (&lt;.01)</td>
<td>.003 (&lt;.01)</td>
<td>.003 (&lt;.01)</td>
<td>.003 (&lt;.01)</td>
</tr>
<tr>
<td>Bin 1</td>
<td>.003 (&lt;.01)</td>
<td>.002 (&lt;.01)</td>
<td>.004 (&lt;.01)</td>
<td>.002 (&lt;.01)</td>
</tr>
<tr>
<td>Bin 2</td>
<td>.003 (&lt;.01)</td>
<td>.002 (&lt;.01)</td>
<td>.003 (&lt;.01)</td>
<td>.003 (&lt;.01)</td>
</tr>
<tr>
<td>Bin 3</td>
<td>.001 (&lt;.01)</td>
<td>.002 (&lt;.01)</td>
<td>.004 (&lt;.01)</td>
<td>.003 (&lt;.01)</td>
</tr>
<tr>
<td>Bin 4</td>
<td>.001 (&lt;.01)</td>
<td>.002 (&lt;.01)</td>
<td>.003 (&lt;.01)</td>
<td>.003 (&lt;.01)</td>
</tr>
<tr>
<td>Exp. 3</td>
<td>.23 (.02)</td>
<td>.26 (.03)</td>
<td>.30 (.03)</td>
<td>.28 (.03)</td>
</tr>
<tr>
<td>Bin 1</td>
<td>.23 (.02)</td>
<td>.26 (.03)</td>
<td>.30 (.03)</td>
<td>.28 (.03)</td>
</tr>
<tr>
<td>Bin 2</td>
<td>.25 (.02)</td>
<td>.26 (.03)</td>
<td>.30 (.03)</td>
<td>.35 (.02)</td>
</tr>
<tr>
<td>Bin 3</td>
<td>.25 (.02)</td>
<td>.26 (.03)</td>
<td>.30 (.03)</td>
<td>.36 (.04)</td>
</tr>
<tr>
<td>Bin 4</td>
<td>.25 (.02)</td>
<td>.26 (.03)</td>
<td>.30 (.03)</td>
<td>.30 (.04)</td>
</tr>
<tr>
<td>Exp. 4</td>
<td>.02 (&lt;.01)</td>
<td>.04 (&lt;.01)</td>
<td>.04 (&lt;.01)</td>
<td>.04 (&lt;.01)</td>
</tr>
<tr>
<td>Bin 1</td>
<td>.03 (&lt;.01)</td>
<td>.04 (&lt;.01)</td>
<td>.04 (&lt;.01)</td>
<td>.04 (&lt;.01)</td>
</tr>
<tr>
<td>Bin 2</td>
<td>.03 (&lt;.01)</td>
<td>.04 (&lt;.01)</td>
<td>.04 (&lt;.01)</td>
<td>.05 (&lt;.01)</td>
</tr>
<tr>
<td>Bin 3</td>
<td>.03 (&lt;.01)</td>
<td>.04 (&lt;.01)</td>
<td>.04 (&lt;.01)</td>
<td>.04 (&lt;.01)</td>
</tr>
<tr>
<td>Bin 4</td>
<td>.03 (&lt;.01)</td>
<td>.04 (&lt;.01)</td>
<td>.04 (&lt;.01)</td>
<td>.04 (&lt;.01)</td>
</tr>
</tbody>
</table>

Commission error rates increased significantly across blocks in Experiment 1, where 90% of the stimuli required a response, $F(2, 30) = 7.63, p = 2.10 \times 10^{-3}, \eta_p^2 = .34$, in Experiment 3 with 90% required responses, $F(3, 29) = 8.13, p = 4.44 \times 10^{-4}, \eta_p^2 = .46$, and in Experiment 4 with 50% required responses, $F(2, 30) = 6.75, p = 3.79 \times 10^{-3}, \eta_p^2 = .31$; they did not change across blocks in Experiment 2, with only 10% of targets requiring a response, $F(2, 30) = .49, p = .62, \eta_p^2 = .03$. 

(B) Omission errors, or P(“No-go response” | “Go stimulus”)

<table>
<thead>
<tr>
<th></th>
<th>Block 1</th>
<th></th>
<th></th>
<th>Block 2</th>
<th></th>
<th></th>
<th>Block 3</th>
<th></th>
<th></th>
<th>Block 3</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bin 1</td>
<td>Bin 2</td>
<td>Bin 3</td>
<td>Bin 4</td>
<td>Bin 1</td>
<td>Bin 2</td>
<td>Bin 3</td>
<td>Bin 4</td>
<td>Bin 1</td>
<td>Bin 2</td>
<td>Bin 3</td>
<td>Bin 4</td>
</tr>
<tr>
<td>Exp. 1</td>
<td>.04 (.01)</td>
<td>.04 (.01)</td>
<td>.04 (.01)</td>
<td>.06 (.01)</td>
<td>.05 (.01)</td>
<td>.05 (.01)</td>
<td>.06 (.02)</td>
<td>.08 (.02)</td>
<td>.04 (.01)</td>
<td>.06 (.01)</td>
<td>.05 (.01)</td>
<td>.05 (.01)</td>
</tr>
<tr>
<td>Exp. 2</td>
<td>.05 (.01)</td>
<td>.08 (.02)</td>
<td>.09 (.02)</td>
<td>.11 (.02)</td>
<td>.08 (.01)</td>
<td>.08 (.02)</td>
<td>.10 (.03)</td>
<td>.12 (.03)</td>
<td>.12 (.03)</td>
<td>.14 (.03)</td>
<td>.13 (.03)</td>
<td>.11 (.03)</td>
</tr>
<tr>
<td>Exp. 3</td>
<td>.05 (.01)</td>
<td>.04 (.01)</td>
<td>.05 (.01)</td>
<td>.06 (.01)</td>
<td>.06 (.01)</td>
<td>.08 (.02)</td>
<td>.08 (.02)</td>
<td>.07 (.02)</td>
<td>.06 (.01)</td>
<td>.08 (.01)</td>
<td>.08 (.01)</td>
<td>.08 (.02)</td>
</tr>
<tr>
<td>Exp. 4</td>
<td>.12 (.01)</td>
<td>.13 (.01)</td>
<td>.14 (.01)</td>
<td>.14 (.01)</td>
<td>.14 (.01)</td>
<td>.16 (.02)</td>
<td>.14 (.03)</td>
<td>.14 (.01)</td>
<td>.14 (.01)</td>
<td>.16 (.01)</td>
<td>.16 (.02)</td>
<td>.14 (.01)</td>
</tr>
</tbody>
</table>

Omission errors did not significantly increase across blocks in Experiment 1, $F(2, 30) = 1.76, p = .19, \eta_p^2 = .11$, in Experiment 2, $F(2, 30) = 2.39, p = .11, \eta_p^2 = .14$, or in Experiment 4 $F(2, 30) = 3.08, p = .06, \eta_p^2 = .17$; they increased significantly across blocks in Experiment 3, $F(3, 29) = 4.45, p = 1.09 \times 10^{-2}, \eta_p^2 = .32$. 
Table A2. Signal detection measures over four 2-min time bins across the first three blocks, including sensitivity ($d'$, Panel A) and response bias ($c$, Panel B). Values in parentheses indicate the standard error of the mean.

(A) Sensitivity ($d'$)

<table>
<thead>
<tr>
<th></th>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bin 1</td>
<td>Bin 2</td>
<td>Bin 3</td>
</tr>
<tr>
<td>Exp. 1</td>
<td>2.89 (.17)</td>
<td>2.72 (.13)</td>
<td>2.65 (.14)</td>
</tr>
<tr>
<td>Exp. 2</td>
<td>4.59 (.14)</td>
<td>4.36 (.16)</td>
<td>4.34 (.16)</td>
</tr>
<tr>
<td>Exp. 3</td>
<td>2.56 (.11)</td>
<td>2.66 (.12)</td>
<td>2.49 (.12)</td>
</tr>
<tr>
<td>Exp. 4</td>
<td>3.33 (.08)</td>
<td>3.21 (.09)</td>
<td>3.12 (.08)</td>
</tr>
</tbody>
</table>

(B) Response criterion ($c$). A positive value indicates a response bias towards “go” responding and a negative value indicates a bias towards “no-go” responses.

<table>
<thead>
<tr>
<th></th>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bin 1</td>
<td>Bin 2</td>
<td>Bin 3</td>
</tr>
<tr>
<td>Exp. 1</td>
<td>.42 (.06)</td>
<td>.53 (.07)</td>
<td>.55 (.05)</td>
</tr>
<tr>
<td>Exp. 2</td>
<td>-.22 (.06)</td>
<td>-.32 (.07)</td>
<td>-.37 (.07)</td>
</tr>
<tr>
<td>Exp. 3</td>
<td>.47 (.03)</td>
<td>.49 (.05)</td>
<td>.51 (.05)</td>
</tr>
<tr>
<td>Exp. 4</td>
<td>-.44 (.04)</td>
<td>-.43 (.03)</td>
<td>-.44 (.04)</td>
</tr>
</tbody>
</table>

Statistical results for these measures can be obtained from the authors.
References


