The Sustained Attention to Response Task (SART) Does Not Promote Mindlessness During Vigilance Performance

Michael B. Dillard, National Research Council, Washington, D.C., **Joel S. Warm** and **Gregory J. Funke**, Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio, **Matthew E. Funke**, Naval Medical Research Unit–Dayton, Wright-Patterson Air Force Base, Ohio, **Victor S. Finomore Jr.**, Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio, **Gerald Matthews**, University of Central Florida, Orlando, and **Tyler H. Shaw** and **Raja Parasuraman**, George Mason University, Fairfax, Virginia

Objective: In this study, we evaluated the validity of the Sustained Attention to Response Task (SART) as a means for promoting mindlessness in vigilance performance.

Background: Vigilance tasks typically require observers to respond to critical signals and to withhold responding to neutral events. The SART features the opposite response requirements, which supposedly leads it to promote a mindless, nonthoughtful approach to the vigilance task. To test that notion, we compared the SART to the traditional vigilance format (TVF) in terms of diagnostic accuracy assessed through decision theory measures of positive and negative predictive power (PPP and NPP), perceived mental workload indexed by the Multiple Resource Questionnaire, and oculomotor activity reflected in the Nearest Neighbor Index and fixation dwell times.

Method: Observers in TVF and SART conditions monitored a video display for collision flight paths in a simulated air traffic control task.

Results: Diagnostic accuracy in terms of NPP was high in both format conditions. While PPP was poorer in the SART than in the TVF, that result could be accounted for by a loss of motor control rather than a lack of mindfulness. Identical high levels of workload were generated by the TVF and SART tasks, and observers in both conditions showed similar dynamic scanning of the visual scene.

Conclusion: The data indicate that the SART is not an engine of mindlessness.

Application: The results challenge the widespread use of the SART to support a model in which mindlessness is considered to be the principal root of detection failures in vigilance.

Keywords: vigilance, SART, mental workload, mindlessness model, resource model, MRQ, oculometrics, gaze control, positive predictive power, negative predictive power

---

Address correspondence to Gregory J. Funke, Air Force Research Laboratory, 2510 Fifth Street, Wright-Patterson Air Force Base, OH 45433-7951, USA; e-mail: Gregory.Funke.1@us.af.mil.

HUMAN FACTORS
Vol. 56, No. 8, December 2014, pp. 1364–1379
DOI: 10.1177/0018720814537521

INTRODUCTION

Vigilance or sustained attention tasks require observers to monitor displays for extended periods and detect the appearance of critical signals. The signals, which occur infrequently, are often embedded in a background of neutral or nonsignal events. Observers are typically instructed to make an overt response to the critical signals and to make no response to the more frequent neutral events. Thus, vigilance tasks can be described as “go/no-go” attentional assignments in which the frequency of “no-go” events outweighs that of “go” events. These assignments are of interest to human factors/ergonomic specialists because of the critical role that vigilance plays in a wide array of automated human–machine systems in aviation, industrial process and quality control, medical monitoring and screening, and airport and border security. Signal detection failures in these situations have led to unfortunate consequences. Thus, it is important to understand the origin of such failures on the part of human observers (Hancock, 2013; Vidulich, Wickens, Tsang, & Flach, 2010; Warm, Parasuraman, & Matthews, 2008; Wickens, Hollands, Banbury, & Parasuraman, 2013).

At present, there are two competing models to account for failure of signal detection in vigilance tasks. One of these is the resource model, in which the need to make continuous signal/noise discriminations is held to deplete observers’ information-processing assets over time, leading to missed signals (Davies & Parasuraman, 1982; Parasuraman, 1979). As described in several sources (e.g., Helton & Russell, 2013; Shaw et al., 2013; Warm et al., 2008; Warm, Finomore, Vidulich, & Funke, in press; Warm, Tripp, Matthews, & Helton, 2012; Wickens et al., 2013), support for the resource model comes from studies indicating that vigilance tasks impose a substantial mental burden on observers as reflected in high scores on the NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988), a major instrument for measuring the perceived mental workload in performing a task; from studies showing that vigilance performance is poorer in tasks that require the use of working memory to distinguish signals from nonsignals than in tasks in which signal detection does not involve a working memory component; from neuroimaging studies of resource demand using cerebral blood flow velocity measures; from the susceptibility of vigilance tasks to interference from concurrent tasks; and from investigations featuring physiological and subjective report measures indicating that vigilance tasks induce stress in observers that is linked to task demand.

An alternative view of detection failures in vigilance is the mindlessness model proposed by Robertson and associates (Manly, Robertson, Galloway, & Hawkins, 1999; Robertson, Manly, Andrade, Baddeley, & Yiend, 1997). The model was prompted by the suggestion that when confronted with repetitive tasks in which signals are separated by long time intervals, as in the case of vigilance, a supervisory attention system loses its potency and observers cease to focus their awareness on the task at hand (Shallice, 1988; Stuss, Shallice, Alexander, & Picton, 1995). With this in mind, Robertson and colleagues have asserted that the repetitive nature of vigilance tasks leads to a mindless lack of attentional focus and thence to failures of signal detection. Support for the mindlessness model, or as Head and Helton (2013) have termed it, the perceptual decoupling model, comes from studies using the Sustained Attention to Response Task (SART; Robertson et al., 1997), which was designed to promote mindlessness in vigilance by inverting the "go/no-go" ratio. With this format, observers are asked to respond to the more frequent neutral events and to withhold responding in the presence of the less frequent critical signals. The SART has been widely used to measure sustained attention (Smallwood, McSpadden, & Schooler, 2007; Smilek, Carriere, & Cheyne, 2010). In support of the mindlessness model, research with the SART has shown that failures to detect signals are preceded by periods of increased routinization and decreased effort and that absentminded observers do more poorly than nonabsentminded observers (Langner, Willmes, Chatterjee, Eickhoff, & Sturm, 2010).

A key issue in dealing with the competing resource and mindlessness models of detection failures in vigilance is the validity of the SART as a means for promoting mindlessness. That validity has been questioned by Grier and associates (2003), who have reported that, as is the case with the traditional vigilance task format (TVF), perceived workload on the SART falls at the upper range of the NASA-TLX, and by Helton and colleagues (Helton et al., 2005; Helton, Head, & Russell, 2011; Helton, Weil, Middlemiss, & Sawers, 2010), who have shown that with both types of task formats, observers are responsive to subtle patterns in the temporal and spatial structure of critical signal appearances and to the reliability of warning signals. Further, observers have been found to report more task-related thoughts in performing the SART than in performing a traditional vigilance task (Carter, Russell, & Helton, 2013) and to frequently be aware of their commission errors, that is, failures to withhold responding to signal events (Seli, Cheyne, & Smilek, 2012). High workload, the detection of subtle changes in task elements, elevated levels of task-related thoughts, and awareness of failures to act appropriately in the presence of signal events do not seem to be consistent with the mindlessness perspective the SART is supposed to promote.

The present study was designed to provide further evaluation of the SART as an engine of mindlessness by examining the accuracy of decision processes in determining the presence or absence of critical signals, by carrying out an additional exploration of task-induced workload, and by studying oculomotor activity as observers searched a complex display for critical signals.

Decision Process Accuracy

Given the importance of veridical decisions in operational vigilance assignments, the diagnostic accuracy of observers' reports about the presence or absence of critical signals is vital; that is, it is essential that when an observer indicates that a signal is present, it is actually there, and when the observer indicates a signal is absent, it is really not there. These aspects of performance are inherent in the decision theory measures of positive predictive power (PPP), the proportion of an observer's "signal present" responses that are actually correct, and negative predictive power (NPP), the proportion of an observer's "signal absent" responses that are actually correct. The computational formula for PPP is (H) / (H + FA), where H = number of correct detections (hits) and FA = number of false alarms. The comparable formula for NPP is (CR) / (CR + M), where CR = number of correct rejections and M = the number of signals missed.

PPP and NPP scores of 1.0 indicate a perfectly accurate observer, whereas scores of 0.0 indicate no correct decisions about signal presence/absence and no diagnosticity. These measures, which have been frequently employed to evaluate decision making in medicine (see Linton, 1996) and in the evaluation of human performance with automated alarms (Parasuraman, Hancock, & Olofinboba, 1997), were introduced into evaluating vigilance performance in a study by Szalma, Hancock, Warm, Dember, and Parsons (2006) in which vigilance training effectiveness was assessed. In the present study, we made use of the PPP and NPP measures to provide the initial portrait of diagnostic efficiency in operators working within the TVF and SART formats. To the extent that the SART promotes a lack of attentional focus, it would be anticipated that diagnostic decisions about the presence and absence of critical signals with this format would be poorer than those with the TVF. One goal for this study was to test that expectation.

Mental Workload

Recently, Boles and his associates (Boles, Bursk, Phillips, & Perdelwitz, 2007) have introduced a new subjective workload scale, the Multiple Resources Questionnaire (MRQ), which characterizes workload with respect to multiple mental processes based upon a combination of dimensions drawn largely from factor-analytic studies carried out by Boles and colleagues (Boles, 1998; Boles & Law, 1998). The instrument consists of the 17 resource dimensions listed in Table 1. Fourteen of the dimensions reflect encoding/central processing resources; the remaining three are response resources. Using a scale of 0 (*no usage*) to 100 (*extreme usage*; Boles & Dillard, in press; Finomore, Shaw, Warm, Matthews, & Boles, 2013), observers are asked to rate the extent to which a task they just performed required the employment of each dimension. Research with the MRQ has shown that the instrument is able to uncover different key resource dimensions in tasks involving dissimilar skills, such as reading bar graphs, determining the spatial position of a line, word interpretation, medical imaging, and of critical importance for the present study, vigilance (Boles et al., 2007; Finomore et al., 2013; Klein et al., 2012).

In the case of vigilance, Finomore and his associates (2013) performed two experiments in which observers were asked to monitor visual displays involving spatial discriminations. They found that, like the NASA-TLX, the MRQ revealed that the workload of the vigilance tasks was substantial and that the MRQ was equivalent to the NASA-TLX in reflecting different degrees of workload in the tasks involved. In addition, in their responses to the MRQ, the observers in the Finomore et al. study indicated that they utilized a number of spatial resources in signal detection, a result consistent with the fact that spatial discriminations were a key feature of the tasks they performed. As will be seen later, the present study also made use of a task in which spatial discriminations were integral—critical signals could occur unpredictably in one of four locations in the visual display and were defined by a difference in the spatial orientation of one of four elements relative to the others. If the SART does indeed promote a mindless, nonthoughtful approach to vigilance performance, one would anticipate that observers working with this task in the SART format would engage a more limited subset of spatial resources and do so at a significantly lower level than those working with the task in the TVF. A second goal for present study was to test these expectations.

Oculomotor Activity

Oculomotor activity offers an additional medium for assessing the degree to which the

TABLE 1: Definition of Multiple Resource Questionnaire (MRQ) Dimensions

MRQ Dimension	Definition
Encoding/central processing	
Auditory emotional process	Required judgments of emotion (e.g., tone of voice of musical mood) presented through the sense of hearing
Auditory linguistic process	Required recognition of words, syllables, or other verbal parts of speech presented through the sense of hearing
Facial figural process	Required recognition of faces, or of the emotions shown on faces, presented through the sense of vision
Short-term memory process	Required remembering of information for a period ranging from a couple of seconds to half a minute
Spatial attentive process	Required focusing of attention on a location, using the sense of vision
Spatial categorical process	Required judgment of simple left-versus-right or up-versus-down relationships, without consideration of precise location, using the sense of vision
Spatial concentrative process	Required judgment of how tightly spaced are numerous visual objects or forms
Spatial emergent process	Required "picking out" of a form or object from highly cluttered or confusing background, using the sense of vision
Spatial positional process	Required recognition of a precise location as differing from other locations, using the sense of vision
Spatial quantitative process	Required judgment of numerical quantity based on a nonverbal, nondigital representation (e.g., bar graphs or small clusters of items), using the sense of vision
Tactile figural process	Required recognition or judgment of shapes (figures), using the sense of touch
Visual lexical process	Required recognition of words, letters, or digits, using the sense of vision
Visual phonetic process	Required detailed analysis of the sound of words, letters, or digits, presented using the sense of vision
Visual temporal process	Required judgment of time intervals, or of the timing of events, using the sense of vision
Response resources	
Facial motive process	Required movement of one's own face muscles, unconnected to speech or the expressing of emotion
Manual process	Required movement of the arms, hands, and/or fingers
Vocal process	Required use of one's voice

SART promotes mindlessness in vigilance performance. Several studies have shown that detection failures in vigilance are accompanied by oculomotor changes indicative of reductions in scanning efficiency and increased fatigue as reflected in declines in the frequency of saccades and the number and accuracy of fixations on target objects and increases in blink frequency and blink duration (Funke et al., 2012; Lavine, Sibert, Gokturk, & Dickens, 2002; Morris & Miller, 1996; Schroder & Holland, 1968; Stern, Boyer, Schroder, Touchstone, & Stoliarov, 1994). Along that line, we utilized an eye tracking metric known as the Nearest Neighbor Index (NNI) and fixation dwell time to measure oculomotor activity in observers performing this study's vigilance task in either the TVF or SART format and in control observers who viewed the vigilance display for an equal period but without a work imperative (perhaps the maximum condition for the induction of mindlessness). This study was the initial attempt to utilize the NNI metric in the context of a vigilance task because of the utility of the NNI for testing predictions about differences in the manner in which observers in the TVF, SART, and control formats viewed the vigilance display.

The NNI measures the spatial dispersion produced by a pattern of fixations, or more specifically, the ratio of the average minimum distance between actual observed fixations to the average distance between a hypothetical set of randomly distributed points (Clark & Evans, 1954; Di Nocera, Camilli, & Terenzi, 2007; Wickens et al., 2103). Previous research has demonstrated the sensitivity of NNI values to variations in task difficulty due to temporal or visuospatial demand. In this research, demanding tasks led to NNI values approaching 1 (wide fixation distributions), and less demanding tasks led to values approaching 0 (clustered fixation distributions; Camilli, Terenzi, & Di Nocera, 2007; Di Nocera et al., 2007; Di Nocera & Bolia, 2007; Di Nocera, Terenzi, & Camilli, 2006).

As described by Di Nocera et al. (2007), a possible explanation for this result is that when task demands are high, it becomes necessary to monitor display elements in the shortest time frame without pausing over noninformative regions. Thus, the dispersion of eye fixations reflected in NNI scores and the associated shortening of dwell times can be utilized as an index of how rapidly an observer scanned the entirety of a dynamic display, such as the one employed in this study, which, in addition to involving spatial discriminations, also required those discriminations to be made within a limited time frame. Consequently, if the SART does indeed promote a mindless, nonthoughtful approach to vigilance performance, one would anticipate that in scanning this study's visual display, the NNI scores and dwell times for observers in the SART format condition would more closely approximate those of passive control observers who view the display in the absence of a work imperative compared to those of observers performing the task in the TVF. Specifically, among the SART and passive control observers, the NNI scores would be expected to be less dispersed and the dwell time scores to be longer than those obtained for observers in the TVF. A final goal for the present study was to test these oculomotor predictions.

METHOD

Participants

Forty-five individuals (23 men, 22 women; mean age = 22.7 years) recruited from the Dayton, Ohio, area served as observers for a single payment of $30. All observers had normal or corrected-to-normal vision and normal hearing. The experiment was conducted under conditions approved by the Wright-Patterson Air Force Base Institutional Review Board.

Design

A 3 (task format) × 4 (periods of watch) split-plot experimental design was employed. Fifteen participants, approximately equated for sex, were assigned at random to one of three task format conditions: TVF, SART, and passive control. All observers participated in a 40-min session divided into four continuous 10-min periods.

Vigilance Tasks

Active observers assumed the role of controllers monitoring the flight pattern of a squadron of four unmanned aerial vehicles (UAVs) projected in the center of a 43.18-cm visual display terminal (VDT) as shown in Figure 1. The display, adapted from Shaw et al. (2013) and Funke et al. (2011, 2012), contained a single circular viewing field, 10.19 cm in diameter, that was presented on a gray background (transluminance = 42 cd/m^2). The viewing field consisted of three concentric circles. The diameters of the small and middle circles were 2.54 cm and 6.35 cm, respectively. The largest circle formed the exterior black border of the viewing field, which was divided into four equal 90° quadrants defined by black lines. In all cases, the lines defining the viewing field were 0.32 cm thick, their transluminance was 37 cd/m^2, and their contrast with the gray background based on the Michaelson contrast ratio (maximum luminance – minimum luminance /

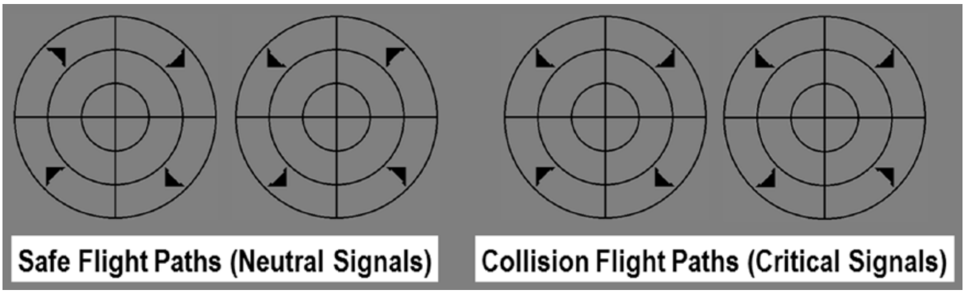

Figure 1. Examples of neutral events and critical signals in the display (adapted from Funke et al., 2011, 2012; Shaw et al., 2013).

maximum luminance + minimum luminance; Coren, Ward, & Enns, 1999) was 6.33%.

Normally, the quadrants of the viewing field were blank. When activated, each quadrant of the display contained a black triangular icon (base = 1.35 cm, altitude = 0.95 cm, transluminance = 37 cd/m^2, contrast with the gray background = 6.33%), which represented a UAV. In all conditions, the squadron of UAVs flew in either a clockwise or a counterclockwise direction (defined by the "noses" of the UAVs), but not both, throughout the vigil. The flight directions were determined at random for each observer with the restriction that they occurred nearly equally across the 15 observers in each condition. Critical signals for detection were cases in which one of the UAVs was flying in an inappropriate direction relative to the others so that a collision could occur. Neutral events and critical signals in the clockwise and counterclockwise flight paths are illustrated in Figure 1.

In the TVF, observers were instructed to indicate the detection of a critical signal by pressing the space bar on a computer keyboard and to withhold responses for the more frequent neutral event presentations. In the SART response format, observers were given the opposite instructions; they were tasked with pressing the space bar for each neutral event and withholding response to the infrequent critical signal presentations. In both task formats, observers were instructed to make the required "go" response as quickly as possible when an appropriate stimulus event appeared on their display. The vigilance task was programmed so that only one response was recorded per key press. Therefore, observers in the TFV and SART conditions could not secure high signal detection rates or high nonsignal rejection rates, respectively, by merely keeping the space bar depressed throughout the vigilance session.

In all conditions, the display was updated 30 times per minute (one stimulus event every 2,000 ms) with an exposure time of 1,000 ms. For each observer, 12 critical signals occurred at random intervals during each 10-min period of watch (three in each display quadrant, overall signal probability = 4%). As in the Funke et al. (2011, 2012) studies in which the task employed herein was first used, observers in the TVF condition were allowed 1,000 ms from the onset of a critical signal to its offset (the critical signal response window) to produce a "go" response and be credited with a correct detection (hit). Failures to respond within the window were counted as errors of omission (misses). "Go" responses to nonsignal events within 1,000 ms of event onset were considered errors of commission (false alarms), whereas withholding responses within that window (i.e., the response of not responding) were considered as correct rejections. In the SART, observers were credited with a correct detection if they *did not* respond to a critical stimulus within the 1,000-ms window between the onset of the stimulus and its offset, whereas "go" responses within that window were considered as errors of commission (misses). Failures to respond to nonsignal events within the 1,000 ms between event onset and offset were considered errors of omission (false alarms), whereas responses within that window were considered as correct rejections.

Prior to participating in the experimental vigil, observers assigned to the TVF and SART conditions completed a 10-min practice session to familiarize them with the task. During the practice, a computerized female voice provided feedback about correct detections, misses, and false alarms. Observers were required to detect at least 7 of 12 presented critical signals and make no more than 10 false alarms in this phase of the study to be considered for inclusion in the final analysis. All observers in the TVF and SART conditions met this dual criterion. During the experimental vigil, audio feedback was removed and observers completed the vigil in silence. Following completion of the experimental vigil, observers in both active vigilance formats (TVF and SART) completed a computerized version of the MRQ. Stimulus presentations, vigilance response recording, and MRQ presentation/response were controlled by a Dell PC running Windows XP.

Observers in the passive control condition viewed the flight display without an information-processing imperative. These observers were not provided with a definition of critical and neutral events, nor were they given any information about pressing keys on the keyboard. They were instructed to simply gaze at the display until the session ended. Since they did not need to actively interact with the display, they did not require task instruction or initial performance feedback.

All observers were tested individually in a 1.78 × 2.41 × 2.67-m windowless laboratory room. The VDT was mounted on a table 99.10 cm directly in front of the seated observer (visual $\text{angle}_{\text{VDT}} = 23.54°$; visual $\text{angle}_{\text{stimulus display}} = 5.89°$). Ambient illumination in the testing room was 5 cd/m^2, provided by a single 50-watt incandescent bulb, dimmed to half power, and positioned above and behind the seated observer in order to minimize glare on the VDT. To curtail distraction, observers were separated from computer instrumentation by a cubicle wall dividing the width of the room in half.

Eye movement data were collected using a Seeing Machines Inc. faceLAB 4.0 eye tracker. The desktop-mounted eye tracking system, which was located immediately under the VDT, consisted of two infrared cameras and a group of infrared light–emitting diodes. The cameras recorded eye movements at a rate of 60 Hz, using corneal reflectance to determine the point of gaze during fixations. Fixations were determined from the recorded eye movements in this experiment using the I-DT method described by Salvucci and Goldberg (2000).

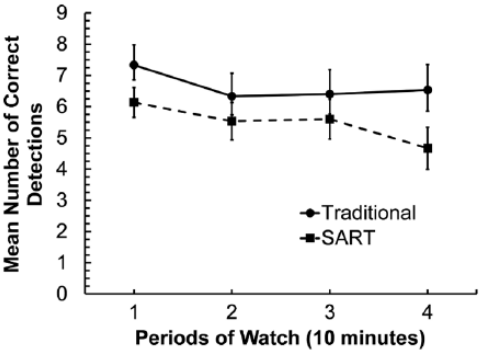

Figure 2. Mean number of correct detections (hits) in the traditional vigilance format and Sustained Attention to Response Task (SART) conditions as a function of periods of watch. Error bars are standard errors. The maximum number of correct detections per period of watch was 12.

RESULTS

Performance Efficiency

Means and standard errors of the number of correct detections (hits) and the number of false alarms in the TVF and SART conditions are presented in Figures 2 and 3, respectively. In examining these figures, it is important to keep in mind that a hit in the TVF condition involved pressing the computer space bar to the occurrence of a critical signal, whereas refraining from pressing the space bar to the occurrence of a critical signal defined a hit in the SART condition. Pressing the space bar to the occurrence of a nonsignal event constituted a false alarm in the TVF condition, whereas refraining from pressing the space bar to the occurrence of a nonsignal event defined a false alarm in the SART condition.

The data for both performance metrics were tested for statistical significance by means of 2 (TVF, SART) × 4 (periods of watch) mixed-model analyses of variance (ANOVAs). In these and all subsequent ANOVAs, the Box correction was employed when necessary to compensate

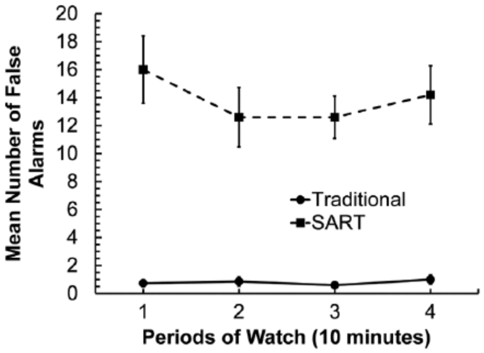

Figure 3. Mean number of false alarms in the traditional vigilance format and Sustained Attention to Response Task (SART) conditions as a function of periods of watch. Error bars are standard errors. The maximum number of false alarms per period of watch was 288.

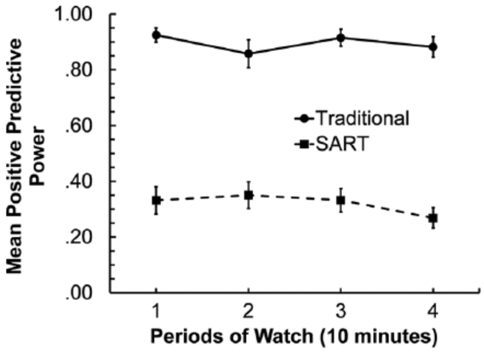

Figure 4. Mean positive predictive power in the traditional vigilance format and Sustained Attention to Response Task (SART) conditions as a function of periods of watch. Error bars are standard errors.

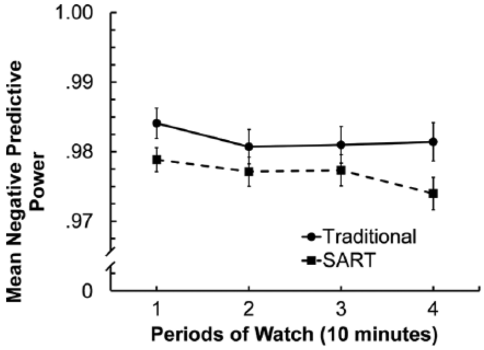

Figure 5. Mean negative predictive power in the traditional vigilance format and Sustained Attention to Response Task (SART) conditions as a function of periods of watch. Error bars are standard errors.

for violations of the sphericity assumption (Field, 2009). In regard to hits, there was a statistically significant main effect for periods of watch, $F(2.80, 78.50) = 3.75$, $p = .016$, $\eta_p^2 = .12$. The main effect for task format and the Task × Period interaction were not significant, both p values > .18, both η_p^2 values < .13. Inspection of Figure 2 will reveal that the overall frequency of hits declined over time; means for Periods 1 through 4 were 6.73, 5.93, 6.00, and 5.60, respectively. The ANOVA of the false alarm data indicated that although false alarms were rare in both task conditions, false alarms were significantly more frequent in the SART ($M = 13.85$) than in the TVF ($M = .80$) condition, $F(1, 28) = 60.66$, $p < .001$, $\eta_p^2 = .68$. The main effect for periods and the Task × Period interaction were not significant, both p values > .26, both η_p^2 values < .06.

Means and standard errors of the PPP and NPP scores for the TVF and the SART conditions are displayed in Figures 4 and 5, respectively. A 2 (TVF, SART) × 4 (periods of watch) mixed-model ANOVA of the PPP scores indicated a statistically significant main effect for task format, $F(1, 28) = 163.12$, $p < .001$, $\eta_p^2 = .85$. It is evident in Figure 4 that the mean PPP score in the TVF ($M = .90$) was near the upper level of the PPP range, whereas the mean in the SART format ($M = .32$) was considerably below that. The main effect for periods and the Task × Period interaction were not significant, both p values > .26, both η_p^2 values < .04.

It is evident in Figure 5 that the NPP scores for the TVF and SART conditions closely approached the upper limit of the NPP range. A mixed-model ANOVA of the NPP data revealed a statistically significant main effect for periods of watch, $F(2.82, 78.84) = 3.40$, $p = .024$, $\eta_p^2 = .11$. The main effect for task format and the Task × Period interaction were not significant, both p values > .09, both η_p^2 values < .10. Examination of the figure will show that although statistically significant, the overall temporal decline in the NPP scores was quite limited in magnitude (means for Periods 1 through 4 were .981, .979, .979, and .978, respectively).

TABLE 2: Mean Multiple Resource Questionnaire Scores for the Resource Dimensions Meeting the Inclusion Criterion in the Traditional Vigilance Format (TVF) and Sustained Attention to Response Task (SART Conditions)

	Condition			
Dimension	TVF		SART	
Manual process	29.33	(7.14)	52.67	(8.53)
Short term memory process	46.00	(10.07)	38.33	(9.36)
Spatial attentive process	91.33	(3.43)	81.00	(6.85)
Spatial categorical process	76.33	(6.73)	83.33	(4.57)
Spatial concentrative process	41.33	(10.64)	25.67	(7.59)
Spatial emergent process	58.33	(9.43)	51.00	(9.07)
Spatial positional process	59.67	(7.21)	57.67	(9.07)
Visual temporal process	49.33	(9.31)	62.67	(9.93)
Global score (means and standard errors of dimensions meeting inclusion criterion)	56.46	(7.01)	56.54	(6.92)

Note. Standard errors are in parentheses.

Mental Workload

The MRQ provides two indices of workload: (a) a global workload score defined as the mean of the observer's ratings across all resource dimensions and (b) a profile of resource contributions to workload defined by the absolute value of the rating given to each resource dimension (Boles et al., 2007). As discussed by Boles and his associates (Boles et al., 2007; Boles & Dillard, in press) including resources that are rated as having "no usage" (a rating of zero) when calculating a global score can distort the global workload picture by masking the utilization magnitude of resources involved in the task. One approach, employed by Klein et al. (2012), is to retain only those items with utilization ratings significantly greater than zero across raters. Toward that end, Bonferroni corrected one-tail t tests with alpha set at .05 were employed in each condition to determine the resource dimensions for which usage ratings were significantly greater than zero. The means and standard errors of the dimensions that met the usage standard in each format condition are presented in Table 2.

It is evident in the table that the same eight dimensions met the usage standard in each condition. Specifically, the manual processing, short term memory, spatial attentive, spatial categorical, spatial concentrative, spatial emergent, spatial positional, and visual temporal dimensions were utilized across both format conditions though not necessarily to the same degree; the scores for the TVF are greater than those for the SART in some cases and lower than those for the SART in others. Moreover, the table reveals that the mean ratings across the eight resource dimensions meeting the usage standard in each format condition, which compose the global MRQ workload scores for those conditions, were 56.46 and 56.54 for the TVF and SART formats, respectively. These values are above the midpoint of the scale, indicating a substantial level of workload.

A 2 (task format) × 8 (resource dimensions) mixed-model ANOVA revealed that although there was a statistically significant main effect for resource dimensions, $F(5.23, 146.54) = 14.63$, $p < .001$, $\eta_p^2 = .34$, the main effect for task format lacked significance, as did the Format × Dimension interaction, both p values $> .09$, both $\eta_p^2 < .07$. In sum, the two format conditions showed identical high levels of workload, and in the absence of a significant Format × Dimension interaction, it appears that they also evidenced similar multidimensional resource profiles in performance of the vigilance task. With regard to the main effect for resource dimensions, post hoc Tukey tests with alpha set at .05 indicated the spatial attentive and spatial categorical resources were the most heavily utilized; their ratings differed significantly from

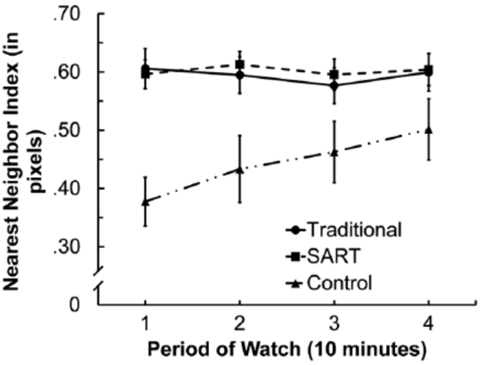

Figure 6. Mean Nearest Neighbor Index scores for the traditional vigilance format and Sustained Attention to Response Task (SART), and control conditions. Error bars are standard errors.

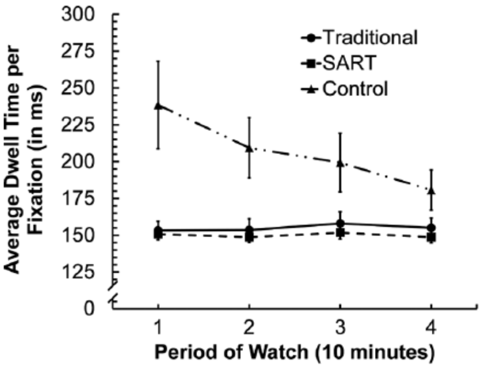

Figure 7. Mean dwell time per fixation for the traditional vigilance format and Sustained Attention to Response Task (SART), and control conditions. Error bars are standard errors.

those of each the other six resource dimensions but not from each other. No significant differences were noted among the remaining units in the resource profile.

Oculomotor Activity

Eye movement data were used to calculate NNI scores and mean fixation dwell times across the four periods of watch for each active format condition and the control condition. Means and standard errors of the NNI scores for each of the three conditions are plotted as a function of periods in Figure 6. It is evident in the figure that the NNI scores for the TVF and SART conditions were nearly identical and higher than those for the control condition, indicating greater dispersion of fixations in the TVF and SART conditions in comparison to the control. It is also evident in the figure that the scores for the TVF and SART conditions showed a stable pattern of scanning over time, whereas those for the control condition became increasingly dispersed over time. These impressions were confirmed by a 3 (task format) × 4 (periods of watch) mixed-model ANOVA, which revealed statistically significant main effects for task format, $F(2, 42) = 6.20$, $p < .001$, $\eta_p^2 = .23$, and periods of watch, $F(2.29, 96.15) = 4.10$, $p = .015$, $\eta_p^2 = .09$, and a significant Task Format × Period interaction, $F(4.58, 96.15) = 4.87$, $p < .001$, $\eta_p^2 = .19$. Subsequent Bonferroni corrected *t* tests with alpha set at .05 indicated that the mean NNI scores in the TVF ($M = .59$) and SART ($M = .60$) conditions did not differ significantly from each other ($p = .83$), but they both were significantly greater than the mean for the control condition ($M = .44$), TVF versus control, $t(28) = 2.61$, $p = .014$, $d = .99$; SART versus control, $t(28) = 2.91$, $p = .007$, $d = 1.10$. Simple effects tests for periods were significant only in the case of the control condition, $F_{\text{control}}(2.42, 33.90) = 9.02$, $p < .001$, $\eta_p^2 = .39$ (p values for periods in both the TVF and SART conditions > .40, both η_p^2 values < .06).

Means and standard errors of the fixation dwell times for each of the three conditions are plotted as a function of periods in Figure 7. Dwell times for the TVF and SART conditions were alike and lower than those for the control condition. It is also evident in the figure that the dwell time scores for the TVF and SART conditions showed a stable pattern of visual processing over time, whereas those for the control condition decreased over the vigil. These impressions were confirmed by a 3 (task format) × 4 (periods of watch) mixed-model ANOVA of the data of Figure 7, which revealed statistically significant main effects for task format, $F(2, 42) = 6.37$, $p = .004$, $\eta_p^2 = .23$, and periods of watch, $F(1.60, 67.22) = 6.26$, $p = .006$, $\eta_p^2 = .13$, and a significant Task Format × Period interaction, $F(3.20, 67.22) = 6.64$, $p < .001$, $\eta_p^2 = .24$. Subsequent Bonferroni corrected *t* tests with alpha set at .05 indicated that the mean dwell time per fixation in the TVF (154.95 ms) and SART

(149.98 ms) conditions did not differ significantly from each other (p = .52), but they both were significantly less than the mean for the control condition (206.94 ms), TVF versus control, $t(28) = 2.43$, $p = .022$, $d = .92$; SART versus control, $t(28) = 2.77$, $p = .010$, $d = 1.05$. Simple effects tests for periods were significant only in the case of the control condition, $F_{\text{control}}(1.47, 20.56) = 7.35$, $p = .007$, $\eta_p^2 = .34$ (p values for periods in both the TVF and SART conditions > .44, both η_p^2 values < .06).

DISCUSSION

This study was designed to assess the legitimacy of the claim made by Robertson et al. (1997) that the SART promotes a mindless, nonthoughtful approach to vigilance performance. That view led to several predictions in regard to the diagnostic accuracy of observers' decisions regarding the presence and absence of critical signals, their perceptions of task-induced workload, and the manner in which they visually inspected the monitored display in search of critical signals.

Manipulation Check

The predictions featured in this study were based on comparisons of the SART with a TVF task. Consequently, prior to discussing the outcomes with regard to the predictions, it is important to examine the suitability of the TVF task employed in this study to be used in testing them. The task appears to meet the suitability criterion for several reasons. As is typical in vigilance studies in general (Davies & Parasuraman, 1982; Finomore et al., 2013; Warm et al., 2008) and has been found in prior studies with the TVF task used herein (Funke et al., 2011, 2012; Shaw et al., 2013), performance efficiency showed a significant decline over time in terms of the frequency of signal detections and a high level of perceived mental workload. Moreover, although the present study was the first to employ the NNI measure in vigilance, the scores for the TVF task (M = .59) were comparable to those noted by Di Nocera and his colleagues (Di Nocera et al., 2007) in individuals engaged in high-workload activities, viz., pilots during descent and landing phases of flight (M = ~.62), and the dwell times in the TVF task of this study (M = 152.95 ms) were considerably shorter than the 200 to 300 ms typically obtained in normal scene viewing (McCarley & Kramer, 2008) but almost identical to the dwell times found in an earlier study with the TVF task (M = 155.20; Funke et al., 2012).

Predicted Outcomes

Diagnostic efficiency. As noted in the introduction to this paper, the accuracy of observers' decisions about signal presence/absence is a key aspect of vigilance performance. The current study provides the initial examination of comparative diagnostic accuracy in the TVF and SART formats. If the SART does promote mindlessness or perceptual decoupling, we anticipated that diagnostic accuracy in the SART condition would be poorer than in the TVF condition. That expectation was confirmed in terms of the "signal present" decisions reflected in the PPP data, which showed that diagnostic acumen was significantly poorer when observers performed the vigilance task in the SART format (M = .32) than in the TVF (M = .90). Along this line, it is important to note that while the two types of vigilance tasks did not differ significantly in performance efficiency as reflected by the number of correct detections, they did differ significantly in regard to the PPP measure that integrated correct detections with the frequency of false alarms, which was significantly greater in the case of the SART than the TVF. A result of that sort underscores the ability of the PPP measure to reflect the veridicality of observers' reports about signal presence.

In regard to the accuracy of "signal absent" decisions, the results did not confirm the expectation that NPP scores would be poorer in the SART as compared to the TVF. The NPP scores in both vigilance formats were similar to each other and fell close to the upper limit of the NPP range. At first glance, the fact that the NPP scores in the SART fell at the upper bound of the NPP range would in itself suggest that the SART did not induce perceptual decoupling. However, the absolute magnitude of the NPP scores needs to be considered carefully. As Szalma et al. (2006) have noted, the number of neutral events in a vigilance task far exceeds the number of critical signals, and therefore NPP scores would be expected to be very high in this type of task. To be sure, an observer in either the TVF or SART conditions of the present study who correctly rejected all nonsignals in a given period of watch ($n = 288$) but missed all critical signals in that period ($n = 12$) would have an NPP score of .96 for that period. Nevertheless, as Szalma and his associates have shown, NPP scores, although high, can still differ significantly between experimental conditions and be meaningful. Although a difference in NPP scores was not present in this study in regard to variations in task format, a small but statistically significant overall difference in the temporal course of the NPP scores was present in the data. Consequently, it would appear that the absence of task format differences in this study is not necessarily the result of a ceiling effect in the NPP measure.

In terms of diagnostic accuracy, the evidence obtained in this study in support of the claim that the SART induces perceptual decoupling is mixed; the PPP results are affirmative whereas the NPP results are not. This conundrum can be resolved by noting that the PPP results can be accounted for in a manner that does not involve the view that the SART is an engine of mindlessness. Rather than perceptual decoupling, Helton and his associates (Head & Helton, 2013, in press; Helton et al., 2011) have argued that motor decoupling may be the root of performance failures in the SART.

In considering this alternative explanation in regard to the PPP data of the present study, it is helpful to keep four points in mind. (a) The PPP index reflects the ratio of correct responses to the sum of correct responses and false alarms. Consequently, for any frequency of correct responses, increases in false alarms will suppress the observer's diagnosticity for the presence of critical signals. (b) Grier and her colleagues (2003) have shown that although the SART does not differ significantly from the TVF in terms of the frequency of correct detections, the SART is susceptible to a higher false alarm rate than the TVF, results that were also observed in the present investigation. Although there were no statistically significant task differences in regard to the number of correct detections in the current study, there was a significant task difference in regard to false alarms in which the overall mean per period for the SART (13.85) was more than 17 times greater than that for the TVF (0.80). (c) In the case of the SART, a false alarm is defined as an error of omission, that is, the failure to execute a motor response in the presence of a neutral or noncritical stimulus event. (d) Observers in the SART format were required to make a motor response to neutral events that occurred frequently at the rate of 288 per 10-min period.

With these points in mind, the argument advanced by Helton and his associates (2011) that errors of omission in the SART are due to loss of motor control in the form of momentary tactical rest stops or "taking a breather" from the need for a high level of continuous responding becomes critical. Such "rest stops" could lead to an increase in false alarms and a consequent reduction in PPP. Thus, instead of lapses of attention, the poor PPP in the SART condition was more likely the result of difficulty in continuously initiating motor responses to an arduous flow of neutral events with a consequent increase in the false alarm rate. An explanation of this sort is consistent with Reason's (1984) findings that well-rehearsed motor routines can conflict with attentional control systems, thereby leading to poorer performance in the SART condition through a loss of motor control.

With an account of this sort, one might wonder how the "rest stops" were distributed over the course of the continuous sequences of neutral events that SART observers encountered in performing the vigilance task; that is, were the "rest" occurrences clustered or bunched together, or were they isolated incidents? If the SART indeed induces episodes of perceptual decoupling during which observers cease to focus their awareness on the task at hand, errors of omission might be expected to be clustered in their occurrence. On the other hand, if the errors of omission were committed by perceptually focused observers who needed occasional relief from continuous motor responding, the errors of omission might be anticipated to be more isolated in occurrence. To explore these possibilities, we examined the false alarm data in the SART condition for clustering, defining a cluster as two or more consecutive errors of omission (false alarms). Less than half (47%) of the SART observers exhibited clustered errors of omission throughout the experiment. Of those who did, the mean number of clusters was 1.57 and the average duration per cluster was 5.27 s, a period encompassing just two consecutive neutral events. Evidently, the errors of omission or false alarms in the SART did not emanate from extensive periods of perceptual decoupling.

Additional evidence that false alarms in the SART condition are not rooted in mindlessness comes from a study by Parasuraman and colleagues (2009). Observers in that study were asked to monitor videos of hand-object movements and detect infrequent target threats (grasping a gun in order to shoot it) in the presence of more frequent and different types of nontargets (grasping a hair dryer or moving a gun). Observers responded either in the TVF or SART format. Parasuraman et al. found that false alarms across nontarget categories were not distributed equally in the SART task, as would be predicted if observers were mindless. Instead, they made more false alarms for nontargets that shared intentional features with the target (moving a gun rather than grasping a hair dryer), indicating that they were actively processing targets and nontargets.

Workload and oculomotor measures. The results with the perceived workload and oculomotor measures also challenge a conclusion that the SART promotes mindlessness in vigilance performance. With regard to workload, it was expected that if the SART fosters mindlessness, it would lead observers to engage a more limited subset of spatial resources in the task employed herein and do so at a significantly lower level than the TVF. It was also expected that, as revealed by oculomotor measures, the manner in which observers in the SART condition scanned the multi-element flight pattern display would be more similar to that adopted by passive control observers than to that of observers in the TVF condition. None of these expectations were borne out.

With respect to the MRQ, the two vigilance formats engaged similar ensembles of eight resource dimensions, many of which involved spatial elements befitting the need of observers to monitor a spatially dynamic visual display. In addition, short-term memory, a dimension needed to keep information active and available (Medin, Ross, & Markman, 2005), was also included in the panoply of resource dimensions utilized in the TVF and the SART conditions, and the mean level of overall workload across the eight resources was above the midpoint of the MRQ scale (56.46 and 56.54 for the TVF and SART conditions, respectively) and almost identical for both vigilance formats. The high level of workload for the SART observed in this study with the MRQ confirms a similar effect noted with the NASA-TLX by Grier et al. (2003)—high workload and the engagement of similar task-relevant resource dimensions in the TVF and SART conditions do not align with the view that the SART induces perceptual decoupling in observers.

The workload outcomes were supplemented by NNI and dwell time results showing that observers with both vigilance formats evidenced more dynamic scanning of the visual scene than control observers who viewed the display without a work imperative. Observers in both the TVF and SART conditions showed a wider dispersion of fixations and shorter dwell times than control observers. Thus, rather than depicting the SART as promoting a mindless, nonthoughtful approach to vigilance performance, the visual scanning results of this study indicate that observers adopted a cognitively active approach in performing their vigilance assignments in both the TVF and SART conditions. Moreover, it is possible that cognitive activation may even have played a role in the behavior of the passive control observers. The finding of a temporal increase in fixation dispersions and a temporal reduction in dwell times for these observers suggests that in the absence of a work imperative, the control observers may have pursued more cognitive stimulation in viewing the vigilance display as time on task continued; that is, these observers may have sought mindfulness even in a situation geared for the maximum case of mindlessness.

CONCLUSION

In conclusion, the results of this study are concordant with the findings of other researchers (Carter et al., 2013; Grier et al., 2003; Head & Helton, 2013, in press; Helton et al., 2005, 2010, 2011; Parasuraman et al., 2009) that the SART is not an engine of mindlessness in vigilance. Therefore, use of the SART for the experimental validation of the claim that mindlessness is the principal root of detection failures in vigilance is likely to be inappropriate.

ACKNOWLEDGMENTS

This research was performed while the first author held a National Research Council Research Associateship Award at the Air Force Research Laboratory, 711th Human Performance Wing. The authors thank Samantha Epling for assisting with data collection.

KEY POINTS

- The present results challenge the use of the Sustained Attention to Response Task (SART) to support the view that mindlessness is the principal root of detection failures in vigilance.
- When compared to performance in the traditional vigilance format (TVF), the SART engages a similar ensemble of mental resources at a comparably high level.
- As is the case with the TVF, observers working with the SART employ dynamic scanning of the visual scene.
- As is the case with the TVF, observers' diagnostic efficiency in determining signal absence (negative predictive power score) with the SART is high.
- Observers are less efficient in diagnosing signal presence (positive predictive power score) when working with the SART than with the TVF. However, that result can be explained in terms of a loss of motor control rather than an absence of mind.
- The SART is not an engine of mindlessness in vigilance performance.

REFERENCES

Boles, D. B. (1998). Relationships among multiple task asymmetries: II. A large-sample factor analysis. *Brain and Cognition*, *36*, 268–289. doi:10.1006/brcg.1998.0995

Boles, D. B., Bursk, J. H., Phillips, J. B., & Perdelwitz, J. R. (2007). Predicting dual-task performance with the Multiple Resources Questionnaire (MRQ). *Human Factors*, *49*, 32–45. doi:10.1518/001872007779598073

Boles, D. B., & Dillard, M. B. (in press). The measurement of perceptual resources. In R. R. Hoffman, P. A. Hancock, R. Parasuraman, J. L. Szalma, & M. Scerbo (Eds.), *The handbook of applied perception research*. New York, NY: Cambridge University Press.

Boles, D. B., & Law, M. B. (1998). A simultaneous task comparison of differentiated and undifferentiated hemispheric resource theories. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 204–215. doi:10.1037/0096-1523.24.1.204

Camilli, M., Terenzi, M., & Di Nocera, F. (2007). Concurrent validity of an ocular measure of mental workload. In D. de Waard, G. R. J. Hockey, P. Nickel, & K. A. Brookhuis (Eds.), *Human factors issues in complex system performance* (pp. 117–129). Maastricht, Netherlands: Shaker.

Carter, L., Russell, P. N., & Helton, W. S. (2013). Target predictability, sustained attention, and response inhibition. *Brain and Cognition*, *82*, 35–42. doi:10.1016/j.bandc.2013.02.002

Clark, P. J., & Evans, F. C. (1954). Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology*, *35*, 445–453. doi:10.2307/1931034

Coren, S., Ward, L. M., & Enns, J. T. (1999). *Sensation and perception* (5th ed.). Fort Worth, TX: Harcourt-Brace.

Davies, D. R., & Parasuraman, R. (1982). *The psychology of vigilance*. London, UK: Academic Press.

Di Nocera, F., & Bolia, R. S. (2007). PERT networks as a method for analyzing the visual scanning strategies of aircraft pilots. *Proceedings of the International Symposium on Aviation Psychology*, *14*, 165–169.

Di Nocera, F., Camilli, M., & Terenzi, M. (2007). A random glance to the flight deck: Pilot's scanning strategies and real-time assessment of mental workload. *Journal of Cognitive Engineering and Decision Making*, *3*, 271–285. doi:10.1518/155534307X255627

Di Nocera, F., Terenzi, M., & Camilli, M. (2006). Another look at scanpath: Distance to nearest neighbour as a measure of mental workload. In D. de Waard, K. A. Brookhuis, & A. Toffetti (Eds.), *Developments in human factors in transportation, design, and evaluation* (pp. 295–303). Maastricht, Netherlands: Shaker.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Los Angeles, CA: Sage.

Finomore, V. S., Jr., Shaw, T. H., Warm, J. S., Matthews, G., & Boles, D. B. (2013). Viewing the workload of vigilance through the lenses of the NASA-TLX and the MRQ. *Human Factors*, *55*, 1044–1063. doi:10.1177/0018720813484498

Funke, M. E., Warm, J. S., Matthews, G., Finomore, V., Vidulich, M. A., Knott, B. A., Helton, W. S., Shaw, T. H., & Parasuraman, R. (2011). Static and dynamic discriminations in vigilance: Effects on cerebral hemodynamics and workload. In T. Marek, W. Karwowski, & V. Rice (Eds.), *Advances in understanding human performance* (pp. 80–90). Boca Raton, FL: CRC Press.

Funke, M. E., Warm, J. S., Matthews, G., Funke, G. J., Chiu, P., & Riley, M. (2012, August). *Neuroergonomic dynamics associated with spatial uncertainty during vigilance task performance*. Paper presented at the meeting of the American Psychological Association, Orlando, FL.

Grier, R. A., Warm, J. S., Dember, W. N., Matthews, G., Galinski, T. L., Szalma, J. A., & Parasuraman, R. (2003). The vigilance decrement reflects limitations in effortful attention, not mindlessness. *Human Factors*, *45*, 349–359. doi:10.1518/hfes.45.3.349.27253

Hancock, P. A. (2013). In search of vigilance. The problem of iatrogenically created psychological phenomena. *American Psychologist*, *68*, 97–109.

Hart, S. G., & Staveland, L. E. (1988). Development of a multidimensional workload scale: Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Amsterdam, Netherlands: North-Holland.

Head, J., & Helton, W. S. (2013). Perceptual decoupling or motor decoupling? *Consciousness and Cognition*, *22*, 913–919. doi:10.1016/j.concog.2013.06.003

Head, J., & Helton, W. S. (in press). Practice does not make perfect in a modified sustained attention to response task. *Experimental Brain Research*. doi:10.1007/s00221-013-3765-0.

Helton, W. S., Head, J., & Russell, P. N. (2011). Reliable and unreliable warning cues in the sustained attention to response task (SART). *Experimental Brain Research*, *209*, 401–407. doi:10.1007/s00221-011-2563-9

Helton, W. S., Hollander, T. D., Warm, J. S., Matthews, G., Dember, W. N., Wallart, M., . . . Hancock, P. A. (2005). Signal regularity and the mindlessness model of vigilance. *British Journal of Psychology*, *96*, 249–261. doi:10.1348/000712605X38369

Helton, W. S., & Russell, P. N. (2013). Visuospatial and verbal working memory load: Effects on visuospatial vigilance. *Experimental Brain Research*, *224*, 429–436. doi:10.1007/s00221-012-3322-2

Helton, W. S., Weil, L., Middlemiss, A., & Sawers, A. (2010). Global interference and spatial uncertainty in the Sustained Attention to Response Task (SART). *Consciousness and Cognition*, *19*, 77–85. doi:10.1016/j.concog.2010.01.006

Klein, T. A., Warm, J. S., Riley, M. A., Matthews, G., Doarn, C., Donovan, J. F., & Gaitonde, K. (2012). Mental workload and stress perceived by novice operators in the laparoscopic and robotic minimally invasive surgical interfaces. *Journal of Endourology*, *26*, 1089–1094. doi:10.1089/end.2011.0641

Langner, R., Willmes, K., Chatterjee, A., Eickhoff, S. B., & Sturm, W. (2010). Energetic effects of stimulus intensity on prolonged simple reaction-time performance. *Psychological Research*, *74*, 499–512. doi:10.1007/s00426-010-0275-6

Lavine, R. A., Sibert, J. L., Gokturk, M., & Dickens, B. (2002). Eye-tracking measures and human performance in a vigilance task. *Aviation, Space, and Environmental Medicine*, *73*, 367–372.

Linton, C. S. (1996). General internal medicine. In U. B. S. Prakash (Ed.), *Mayo Internal Medicine Board review 1996-1997* (pp. 333–352). Rochester, MN: Mayo Foundation for Medical Education and Research.

Manly, T., Robertson, I. H., Galloway, M., & Hawkins, K. (1999). The absent-mind: Further investigations of sustained attention to response. *Neuropsychologia*, *37*, 661–670. doi:10.1016/S0028-3932(98)00127-4

McCarley, J. S., & Kramer, A. F. (2008). Eye movements as a window on perception and cognition. In R. Parasuraman & M. Rizzo (Eds.,), *Neuroergonomics: The brain at work* (pp. 95–112). New York, NY: Oxford University Press.

Medin, D. L., Ross, B., & Markman, A. (2005). *Cognitive psychology* (4th ed.). New York, NY: John Wiley & Sons.

Morris, T. L., & Miller, J. C. (1996). Electrooculographic and performance indices of fatigue during simulated flight. *Biological Psychology*, *42*, 343–360. doi:10.1016/0301-0511(95)05166-X

Parasuraman, R. (1979). Memory load and event rate control sensitivity decrements in sustained attention. *Science*, *205*, 924–927. doi:10.1126/science.472714

Parasuraman, R., De Visser, E., Clarke, E., McGarry, W. R., Hussey, E., Shaw, T., & Thompson, J. (2009). Detecting threat-related intentional actions of others: Effects of image quality, response mode, and target cueing on vigilance. *Journal of Experimental Psychology: Applied*, *15*, 275–290. doi:10.1037/a0017132

Parasuraman, R., Hancock, P. A., & Olofinboba, O. (1997). Alarm effectiveness in driver-centered collision-warning systems. *Ergonomics*, *40*, 390–399. doi:10.1080/001401397188224

Reason, J. (1984). Lapses in attention in everyday life. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 515–549). Orlando, FL: Academic Press.

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

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In A. T. Duchowski (Ed.), *Proceedings of the Eye Tracking Research and Application Symposium* (pp. 71–78). New York, NY: ACM Press.

Schroder, S. R., & Holland, J. G. (1968). Operant control of eye movements. *Journal of Applied Behavior Analysis*, *1*, 161–166. doi:10.1901/jaba.1968.1-161

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

Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge, UK: Cambridge University Press.

Shaw, T. H., Funke, M. E., Dillard, M., Funke, G. J., Warm, J. S., & Parasuraman, R. (2013). Event-related cerebral hemodynamics reveal target-specific resource allocation for both "go" and "no-go" response-based vigilance tasks. *Brain and Cognition*, *82*, 265–273. doi:10.1016/j.bandc.2013.05.003

Smallwood, J., McSpadden, M., & Schooler, J. W. (2007). The lights are on but no one's home: Meta-awareness and the decoupling of attention when the mind wanders. *Psychonomic Bulletin and Review*, *14*, 527–533. doi:10.3758/BF03194102

Smilek, D., Carriere, J. S. A., & Cheyne, J. A. (2010). Failures of sustained attention in life, lab, and brain: Ecological validity of the SART. *Neuropsychologia*, *48*, 2564–2570. doi:10.1016/j.neuropsychologia.2010.05.002

Stern, J. A., Boyer, D., Schroder, D., Touchstone, M., & Stoliarov, N. (1994). *Blinks, saccades, and fixation pauses during vigilance task performance: I. Time on task* (Tech. Rep. DOT-FAA-AM-94-26). Washington, DC: Office of Aviation Medicine, Federal Aviation Administration.

Stuss, D. T., Shallice, T., Alexander, M. P., & Picton, T. W. (1995). A multidisciplinary approach to anterior attentional functions. *Annals of the New York Academy of Sciences*, *769*, 191–211. doi:10.1111/j.1749-6632.1995.tb38140.x

Szalma, J. L., Hancock, P. A., Warm, J. S., Dember, W. N., & Parsons, K. S. (2006). Training for vigilance: Using predictive power to evaluate feedback effectiveness. *Human Factors*, *48*, 682–692. doi:10.1518/001872006779166343

Vidulich, M. A., Wickens, C. D., Tsang, P. A., & Flach, J. M. (2010). Information processing in aviation. In E. Salas & D. Maurino (Eds.), *Human factors in aviation* (2nd ed., pp. 175–215). New York, NY: Elsevier.

Warm, J. S., Finomore, V. S., Vidulich, M. A., & Funke, M. E. (in press). Vigilance: A perceptual challenge. In R. R. Hoffman, P. A. Hancock, R. Parasuraman, J. L. Szalma, & M. Scerbo (Eds.), *The handbook of applied perception research*. New York, NY: Cambridge University Press.

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors*, *50*, 433–441. doi:10.1518/001872008X312152

Warm, J. S., Tripp, L. D., Matthews, G., & Helton, W. S. (2012). Cerebral hemodynamic indices of operator fatigue in vigilance. In G. Matthews, P. A. Desmond, C. Neubauer, & P. A. Hancock (Eds.), *The handbook of operator fatigue* (pp. 197–207). Burlington, VT: Ashgate.

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013). *Engineering psychology and human performance* (4th ed.). Upper Saddle River, NJ: Pearson.

Michael B. Dillard is a National Research Council postdoctoral fellow at the Air Force Research Laboratory, Wright-Patterson Air Force Base. He received his PhD in experimental psychology from the University of Alabama in 2012.

Joel S. Warm is a senior scientist at the Air Force Research Laboratory, Wright-Patterson Air Force Base, Distinguished Researcher in the Human Factors Group at the University of Dayton Research Institute, and professor emeritus of psychology at the University of Cincinnati. He received his PhD in experimental psychology from the University of Alabama in 1966.

Gregory J. Funke is an engineering research psychologist at the Air Force Research Laboratory, Wright-Patterson Air Force Base. He received his PhD in experimental psychology/human factors from the University of Cincinnati in 2007.

Matthew E. Funke is a research psychologist at the Naval Medical Research Unit–Dayton. He received his PhD in experimental psychology/human factors from the University of Cincinnati in 2011.

Victor S. Finomore Jr. is an engineering research psychologist at the Air Force Research Laboratory, Wright-Patterson Air Force Base. He received his PhD in experimental psychology/human factors from the University of Cincinnati in 2008.

Gerald Matthews is a research professor at the Institute of Simulation and Training, University of Central Florida. He received his PhD in psychology from the University of Cambridge in the United Kingdom in 1984.

Tyler H. Shaw is an assistant professor of psychology at George Mason University. He received his PhD in experimental psychology/human factors from the University of Cincinnati in 2008.

Raja Parasuraman is a university professor and director of the Center of Excellence in Neuroergonomics, Technology, and Cognition (CENTEC) at George Mason University. He received his PhD in psychology from Aston University in the United Kingdom in 1976.

Date received: June 14, 2013
Date accepted: April 15, 2014