View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by CORE

Wright State University

CORE Scholar

International Symposium on Aviation Psychology - 2011

International Symposium on Aviation Psychology

2011

The SART Task Does Not Promote Mindlessness in Vigilance Performance

Michael Dillard

David B. Boles

Matthew Funke

Gregory Funke

Victor Finomore

See next page for additional authors

Follow this and additional works at: https://corescholar.libraries.wright.edu/isap_2011

Part of the Other Psychiatry and Psychology Commons

Repository Citation

Dillard, M., Boles, D. B., Funke, M., Funke, G., Finomore, V., Dukes, A., Warm, J. S., Knott, B. A., Matthews, G., & Parasuraman, R. (2011). The SART Task Does Not Promote Mindlessness in Vigilance Performance. *16th International Symposium on Aviation Psychology*, 309-314.
https://corescholar.libraries.wright.edu/isap_2011/63

This Article is brought to you for free and open access by the International Symposium on Aviation Psychology at CORE Scholar. It has been accepted for inclusion in International Symposium on Aviation Psychology - 2011 by an authorized administrator of CORE Scholar. For more information, please contact library-corescholar@wright.edu.

Authors
Michael Dillard, David B. Boles, Matthew Funke, Gregory Funke, Victor Finomore, Allen Dukes, Joel S. Warm, Benjamin A. Knott, Gerald Matthews, and Raja Parasuraman

This article is available at CORE Scholar: https://corescholar.libraries.wright.edu/isap_2011/63

THE SART TASK DOES NOT PROMOTE MINDLESSNESS IN VIGILANCE PERFORMANCE

Michael Dillard, David B. Boles

University of Alabama

Matthew Funke, Gregory Funke, Victor Finomore, Allen Dukes, Joel S. Warm, Benjamin A. Knott

Air Force Research Laboratory Wright-Patterson AFB

Gerald Matthews

University of Cincinnati

Raja Parasuraman

George Mason University

Vigilance tasks typically require observers to respond to critical signals on their monitored displays and withhold responding to neutral events. The Sustained Attention to Response Task (SART) features the opposite response requirements which supposedly lead it to promote a mindless, non-thoughtful approach to the vigilance task that lacks attentional focus. To test that possibility, this study compared the SART to the standard vigilance task in terms of perceived mental workload – indexed by the Multiple Resource Questionnaire (MRQ) – and eye tracking activity – reflected via the Nearest Neighbor Index (NNI) – in the performance of a simulated air-traffic control assignment. Observers with both types of tasks identified a subset of identical MRQ dimensions as being highly involved in their monitoring assignment. The NNI scores indicated that observers with both types of tasks experienced higher workload than controls who viewed the display without a work imperative. Evidently, the SART does not promote mindlessness in vigilance performance.

Vigilance or sustained attention tasks require observers to monitor displays for extended periods of time and detect the appearance of critical signals. The signals, which occur infrequently, are embedded in a background of neutral or non-signal events. Observers are typically instructed to make an overt response, such as a button press, 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 the aviation community because of the critical role that vigilance plays in military surveillance, supervisory control of unmanned systems, air traffic control, and airport and border security (Vidulich, Wickens, Tsang, & Flach, 2010; Warm, Parasuraman, & Matthews, 2008).

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). As described by Warm et al. (2008), 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 associated with performing a task (Wickens & Hollands, 2000), from studies showing that vigilance performance is poorer in tasks that require the use of working memory to distinguish signals from non-signals than in tasks in which signal detection does not involve a working memory component, from neuroimaging studies of resource demand using Transcranial Doppler sonography, 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, Baddley, & Yiend, 1997). The model was prompted by the suggestion that when confronted with repetitive tasks in which signals are separated by long 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). 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 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 task, observers are asked to respond to the more frequent neutral events and to withhold responding in the presence the less frequent critical signals. In support of the mindlessness model, research with the SART has shown that failures to detect signals are proceeded by periods of increased routinization and decreased effort, and by findings that absent minded observers do more poorly than non-absent minded observers (Langer, 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 task as a means for promoting mindlessness. That validity has been questioned by Grier and her associates (Grier et al., 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) who have shown that with both types of task formats, observers are able to detect subtle patterns in the temporal structure of critical signal appearances. High workload and the detection of subtle changes in task elements do not seem to be consistent with the mindlessness perspective. The present study was designed to investigate the SART-validity issue further in terms of workload and eye scanning behavior. Toward that end, the study featured a vigilance task that required spatial discriminations and the need to visually explore the display in order to locate critical signals (see Figure 1 below).

Recently, Boles and his associates (Boles, Bursk, Phillips, & Perdelwitz, 2007) have introduced a new workload scale, the Multiple Resources Questionnaire (MRQ), which characterizes workload with respect to multiple mental processes based upon a combination of dimensions drawn from Wickens’ multiple resource theory (Wickens & Holands, 2000) and factor-analytic studies carried out by Boles and colleagues. The instrument consists of the 17 resource dimensions listed in Table 1. Fifteen of the dimensions reflect encoding and central processing resources; the remaining two are response resources. Using a scale of 0 (no usage) to 100 (extreme usage; Finomore et al., 2008) observers are asked to rate the extent to which a task they just performed utilized 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., 2008). If the SART does indeed promote a mindless, non-thoughtful approach to vigilance performance, one would anticipate that it would engage a more limited subset of resources and employ them at significantly lower level than the standard vigilance format. One goal for present study was to test these possibilities.

Eye-tracking has been used as a tool in aviation psychology since the field’s earliest days and a considerable amount of data is available to indicate a close coupling between eye-movements and attention (McCarley & Kramer, 2007; Wright & Ward, 2008). Oculomotoractivity offers an additional medium to assess the degree to which the SART promotes mindlessness in vigilance performance. A recently developed eye-tracking metric known as the Nearest Neighbor Index (NNI) measures the spatial dispersion produced by a pattern of fixations, or more specifically, the ratio of the average minimum distance between observed fixations to the average distance between a hypothetical set of randomly distributed points. Previous research has demonstrated the sensitivity of NNI values to variations in task difficulty such that demanding tasks led to NNI values approaching 1 (i.e., wider fixation distributions), and less demanding tasks led to values approaching zero (i.e., clustered fixation distributions; Di Nocera, Terenzi, & Camilli, 2006). To the extent that the SART promotes mindlessness, one would anticipate less task demand, and therefore lower NNI values compared to the TVF. Moreover, the SART task should result in an NNI that is similar to a control condition with no task imperative.

Method

Observers assumed the role of air traffic controllers monitoring the flight pattern of a squadron of jet fighters on a circular display divided into four quadrants. Within each quadrant was a triangular jet icon. In all conditions, the jet fighter icons would appear to travel on either a clockwise or counterclockwise course (defined by the noses of the plane) throughout the vigil. In the TVF and SART formats, the task of the observers was to look for cases in which one of the jet icons appeared to be flying in an opposite direction relative to the other three aircraft.

Thirty right handed observers (15 men and 15 women) were assigned at random to each format condition. An additional 15 observers served as passive controls who viewed the flight display without an information-processing imperative, and were instructed to simply gaze at the display until the session ended. All participants served in a 40-min vigil divided into four continuous 10-min periods of watch. In all conditions, the display was updated 30 times/min with a dwell time of 1000 msec. Observers in the TVF and SART conditions were allowed 1200 msec from the onset of the signal to indicate a response. Twelve critical signals occurred during each period of watch (three in each display quadrant). In the TVF, observers were instructed to press the spacebar in response to critical signals – which are illustrated in the image on the right in Figure 1 below – and to make no response to non-critical signals – illustrated in the image on the left in Figure 1 below. In the SART condition, opposite instructions were given; participants were directed to press the spacebar for every occurrence of a non-critical signal and to withhold a response upon the occurrence of a critical signal.

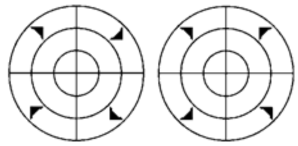

Figure 1. Examples of critical and non-critical signals in the flight path display (after Funke et al. 2010).

Ocular activity data were collected using a Seeing Machines Inc. faceLAB eye tracker, which recorded eye movements at a rate of 60 Hz. Using corneal reflectance to track the movement of the eyes, two desk-mounted infrared cameras recorded information regarding the location of observers' fixations on the visual display throughout the task. Subjective workload was measured using the MRQ immediately following the conclusion of the vigil.

Results

Performance Efficiency. Given the importance of veridical decisions in operational vigilance assignments, the diagnostic accuracy of observers' decisions about the presence or absence of critical signals is vital. Diagnosticity was measured in the present study in terms 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. As described by Szalma et al. (2006), PPP= (number of correct detections)/(correct detections + false alarms) while NPP = (number of correct rejections)/ (correct rejections + misses). PPP and NPP scores of 1.0 indicate a perfectly accurate observer, scores of 0 indicate no correct decisions about signal presence/absence and no diagnosticity. Means and standard errors of the PPP and NPP scores for the TVF and the SART conditions are displayed in Figures 2 and 3, respectively. A 2 (conditions) × 4 (periods of watch) mixed-model ANOVA of the PPP scores showed a significant main effect for condition, $F(1,28) = 17.41$, $p < .05$, $\eta_p^2 = .38$. It is evident in Figure 2 that the mean PPP score in the TVF (89.94) was at the upper level of the PPP range while the mean in the SART format (67.44) was much lower. All other sources of variance in this analysis were not significant, $p > .05$. A similar ANOVA of the NPP scores revealed a significant main effect for periods of watch, $F(2.82, 78.83) = 3.70$, $p < .05$, $\eta_p^2 = .12$. It is evident in the figure that there was an overall temporal decline in the NPP scores but that the degree of decline was quite limited in magnitude. All other sources of variance in this analysis were not significant, $p > .05$. In these and all subsequent ANOVAs, the Box correction was employed to compensate for violations of the sphericity assumption (Field, 2009).

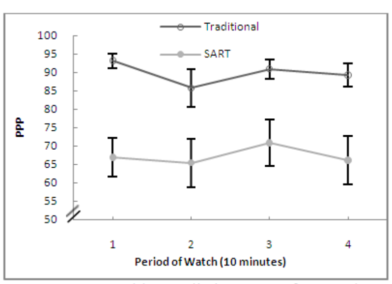

Figure 2. Mean positive predictive power of TVF and SART conditions as a function of periods of watch. Error bars are standard errors.

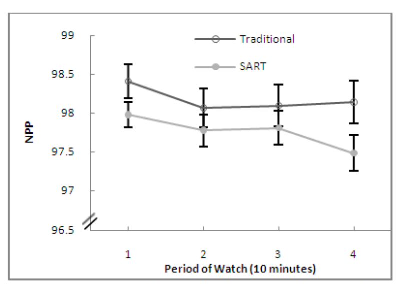

Figure 3. Mean negative predictive power of TVF and SART conditions as a function of periods of watch. Error bars are standard errors.

Workload. Means and standard deviations of the ratings for each of the 17 MRQ dimensions are presented for the TVF and SART format conditions in Table 1. In order to determine a resource profile for each format condition, it was necessary to identify the resources that met a "greater than zero usage" standard within each condition. Toward that end, Bonferroni corrected one-tail *t*-tests with alpha set at .05 were employed in each condition to determine the resource dimensions in which usage ratings were significantly greater than zero. The dimensions that met the usage standard in each format condition are starred in the table. It is evident in the table that the same eight dimensions met the usage standard in each condition. Moreover, the mean levels of engagement for the eight resources were 56.46 and 56.54, for the TVF and SART formats, respectively; values that were above the midpoint of the scale indicating a substantial level of workload. A 2 (conditions) × 8 (dimensions) mixed ANOVA did not reveal a significant main effect for condition or a conditions × dimension interaction, $p > .05$ in each case. In sum, the two format conditions showed identical multidimensional resource profiles and identical high levels of workload in performing the vigilance task.

Table 1. *Means and standard errors that met the inclusion criterion for the TVF and SART conditions.*

Dimensions	**Conditions**	
	TVF	**SART**
	Met Inclusion Criteria	
Auditory Emotional Process	-	-
Auditory Linguistic Process	-	-
Facial Figural Process	-	-
Facial Motive Process	-	-
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)
Spatial Quantitative Process	-	-
Tactile Figural Process	-	-
Visual Lexical Process	-	-
Visual Phonetic Process	-	-
Visual Temporal Process*	49.33 (9.31)	62.67 (9.93)
Vocal Process	-	-

Ocular Activity. Fixation point data collected from the TVF, SART, and control conditions were used to calculate NNI scores across the four periods of watch for each 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 4. It is evident in the figure that the NNI scores for the TVF and SART conditions were higher than those for the control condition, indicating a greater task demand in the TVF and SART conditions in comparison to the control. It is also evident in the figure that the scores for the TVF, SART, and control conditions showed a declining pattern of scanning over time. These impressions were confirmed by A 2 (conditions) × 4 (periods of watch) mixed-ANOVA of the data of Figure 4 which revealed significant main effects for conditions, $F(1, 42) = 6.59$, $p < .05$, $\eta_p^2 = .24$, and periods, $F(2.52, 105.93) = 9.98$, $p < .05$, $\eta_p^2 = .19$. Subsequent Bonferroni corrected *t*-tests with alpha set at .05 indicated that that the mean NNI scores in the TVF (.64) and SART (.68) conditions did not differ significantly from each other ($p > .05$) but they both were significantly greater than the mean for the control condition ($M = .54$), $p < .05$ in each case.

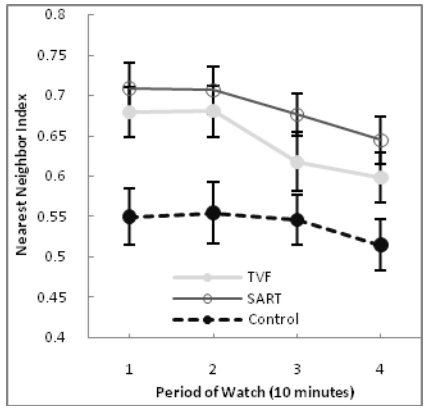

Figure 4. Mean NNI scores for the TVF, SART, and control conditions. Error bars are standard errors.

Discussion

As noted above, a key element in vigilance performance is the diagnostic accuracy of observers' decisions about the presence or absence of critical signals. In terms of signal presence, the PPP data showed that diagnostic acumen was significantly greater when observers performed the vigilance task in the TVF than in the SART format; signal present decisions were 90% accurate in the former condition but only 67% accurate in the latter. A result of this sort could be accounted for in terms of the view that the SART promotes a mindless lack of attentional focus in observers. However, the results of the MRQ workload and ocular workload measures indicate that a conclusion of that sort would be inappropriate.

The principal purpose for the use of the workload and ocular activity measures was to examine the validity of the notion that the SART fosters mindlessness. Toward that end, it was expected that the SART would engage a more limited set of resources on the MRQ than would the TVF and that it would employ them at a significantly lower level than that of the TVF. Neither of these expectations were borne out.

With respect to the MRQ, the two vigilance formats engaged identical ensembles of eight resource dimension many of which involved spatial elements befitting the need of observers to monitor a spatially dynamic visual display. In addition, the mean levels of engagement for the eight resources were above the midpoint of the MRQ scale for both vigilance formats confirming earlier findings by Grier et al. (2003) with the NASA-TLX that both formats induce a substantial level of workload in observers. These outcomes were complimented by the NNI results in which the pattern of eye-scanning was indicative of a higher level of workload in observers across the duration of the vigil than control observers who viewed the display without a work imperative. Rather than depicting the SART as promoting a mindless, non-thoughtful approach to vigilance performance, the results of the present study indicate that under both the TVF and the SART conditions, observers adopt a cognitively active approach in performing a vigilance task.

Given that the SART is not an engine for the promotion of mindlessness, what can account for the finding that PPP is significantly poorer when observers perform a vigilance task in the SART format as compared to the TVF? In seeking an account for this effect, it is helpful to keep four major points in mind. (1) 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. (2) Grier and her colleagues (2003) have shown that the SART is susceptible to a higher false alarm rate than the TVF, a result also observed in this study in which the mean false alarm rates for the SART and the TVF conditions were 1.15% and 0.26%, respectively. (3) In the case of the SART, a false alarm is defined as an error of omission, i.e., the failure to execute a motor response in the presence of a neutral or non-critical stimulus event. (4) In this study, 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 (Helton, Head, & Russell, in press) that errors of omission in the SART are due to loss of motor control in the form of tactical forced rest stops or “taking a breather” from the need for a high level of continuous responding becomes critical. Instead of lapses of attention, the poor diagnosticity 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.

In sum, the results of this study challenge the validity of the proposition that detection failures in the SART emanate from a withdrawal of attentional effort. In so doing, they also challenge the viability of the mindlessness model of vigilance which draws its major support from research with the SART. In addition to implications for theories of vigilance, the present results also have potential meaning at an operational level. Advocates of a mindlessness account of detection failures in vigilance may advocate that steps be taken at the operational level to remove the factors that lead to loss of attentional focus. As Helton et al. (in press) have noted, to the extent to which mindlessness theory represents a misunderstanding of the cause of attentional lapses in vigilance performance, such steps could lead to the adoption of inappropriate solutions to the failure of signal detection in operational vigilance tasks.

References

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.

Davies, D.R., & Parasuraman, R. (1982).*The psychology of vigilance.* New York: 1982.

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.*

Di Nocera, F. Terenzi, M. & Camilli, M. (2006). Another look at scanpath: distance to nearest neighbor as a measure of mental workload. In D. de Waard, K.A. Brookhuis, and A. Toffetti (Eds.) (2006), *Developments in Human Factors in Transportation, Design, and Evaluation* (pp.1-9). Maastricht, the Netherlands: Shaker Publishing.

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

Finomore, V.S., Shaw, T.H., Warm, J.S., Matthews, G., Riley, M.A., Boles, D.B., & Weldon, D. (2008). Measuring the workload of sustained attention: Further evaluation of the Multiple Resources Questionnaire. *Proceedings of the Human Factors and Ergonomics Society, 52,* 1209-1213.

Funke, M., Warm, J.S., Matthews, G., Finomore, V., Vidulich, M., Knott, B., et al. (2010). Static and dynamic discriminations in vigilance: Effect on cerebral hemodynamics and workload. In T. Marek, W. Karwowski, & V. Rice (Eds.), *Advances in understanding human performance:Neuroergonomics,human factors design, and special populations* (pp. 80-90). Boca Raton, FL: CRC Press.

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.

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

Helton, W.S., Head, J., & Russell, P.N. (in press). Reliable and unreliable warning cues in the Sustained Attention to Response Task (SART). *Experimental Brain Research.*

Helton, W.S., Hollander, T.D., Warm, J.S., Matthews, G., Dember, W.N., Wallart, M., Beauchamp, G., Parasuraman, R., & Hancock, P.A. (2005). Signal regularity and the mindlessness model of vigilance. *British Journal of Psychology*, *96¸* 249-261.

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.*

Manley, T., Robertson, I.H., Galloway, M., & Hawkins, K. (1999). The absent-mind: Further investigations of sustained attention to response. *Neuropsychologica*, 37, 661-670.

McCarley, J.S., & Kramer, A.F. (2007). Eye movements as a window on perception and cognition. In R. Parasuraman & M. Rizzo (Eds.), *Neuroergonomics: The brain at work*: New York: Oxford.

Robertson, Manly, T., Andrade, J., Baddeley, B.T., & Yiend, J. (1997). “Opps!”: Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologica*, 35, *747-758.*

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

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.

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: Elsevier.

Warm, J.S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors*, 50, *433-441*.

Wickens, C.D., & Hollands, J.G. (2003). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall,

Wright, R.D., & Ward, L.M., (2008). *Orienting of attention*. New York: Oxford.