Effects of Cueing and Knowledge of Results on Workload and Boredom in Sustained Attention

Edward M. Hitchcock, William N. Dember, Joel S. Warm, Brian W. Moroney, University of Cincinnati, Cincinnati, Ohio, and **Judi E. See,** Logicon Technical Services, Inc., Albuquerque, New Mexico

Two models of recently reported high workload associated with vigilance tasks are the direct-cost and indirect-cost views. The former attributes high workload to the need for continuous observation in discriminating signals from neutral events; the latter attributes it to efforts to combat the boredom associated with monotonous vigilance tasks. These opposing views were tested by providing observers with reliable cueing, which rendered observation necessary only when low-probability critical signals were imminent, or with knowledge of results (KR) regarding performance efficiency. On the basis of cue and KR differences in elicited observation activity and motivational value, the direct-cost model led to the anticipation that cueing would result in a high-boredom, low-workload profile and a greater reduction in workload than KR. The indirect-cost model led to the prediction that cueing would result in a high-boredom, high-workload profile and a lesser reduction in workload than KR. The results clearly supported the direct-cost view that the workload of vigilance is task-induced. Consequently, efforts to combat high workload in complex automated systems requiring substantial monitoring by operators should focus specifically upon task-related determinants.

INTRODUCTION

Vigilance (sustained attention) tasks require prolonged monitoring of repeated stimulus events for infrequently and unpredictably occurring critical signals. Such tasks characterize many human-machine interactions in automated systems (Howell, 1993; Nickerson, 1992). By virtue of their repetitiveness and simplicity, vigilance tasks seem tedious and cognitively undemanding (e.g., Heilman, 1995). However, recent studies using the NASA-Task Load Index (NASA-TLX; Hart & Staveland, 1988) have shown that the mental workload of vigilance tasks is substantial (Deaton & Parasuraman, 1993; Scerbo, Greenwald, & Sawin, 1992; Warm, Dember, & Hancock, 1996; Warm, Dember, & Parasuraman, 1991). In these experiments workload scores usually fell at the upper level of the NASA-TLX scale and exceeded those typical of other tasks, such as memory search, choice reaction time, mental arithmetic, and grammatical reasoning.

The objective for the present investigation was to test two competing models used to explain the surprisingly high workload found in vigilance tasks. The first, the direct-cost model (Warm, Dember, & Hancock, 1996), maintains that the elevated workload is an immediate consequence of the high rate of observation and decision making demanded of observers in their efforts to discriminate critical signals from neutral events. The second, the indirect-cost model (Scerbo, 1998; Scerbo, Greenwald, & Sawin, 1992), maintains that the elevated workload scores arise not from task requirements, but indirectly from observers' efforts to combat the tedium that typifies vigilance tasks.

One way to address the direct- and indirect-cost models would be to employ psychophysical

Address correspondence to Joel S. Warm, Department of Psychology, University of Cincinnati, Cincinnati, OH 45221-0376; joel.warm@uc.edu. **HUMAN FACTORS,** Vol. 41, No. 3, September 1999, pp. 365–372.

manipulations (e.g., event rate, signal salience) that vary the observing and decision-making difficulty while preserving the tedium of vigilance tasks. In support of the direct-cost view, studies have uniformly found that increases in task demand both degrade performance efficiency and increase the overall workload of vigilance tasks (Warm, Dember, & Hancock, 1996). In contrast to the workload results, such psychophysical manipulations do not affect boredom (Scerbo & Holcomb, 1993).

Another way to attack the direct-cost versus indirect-cost issue experimentally – the approach adopted in this study – is to provide supplementary information to aid in monitoring efficiency. This can be accomplished by providing observers with a consistent, reliable cue about the imminent arrival of a critical signal. Consistent cueing of this sort is an effective way to improve signal detection (e.g., Aiken & Lau, 1967; Annett & Patterson, 1967; Weiner & Attwood, 1968). Because cued participants in such a situation would need to observe the display only after having been prompted about the arrival of a signal, cueing should substantially reduce observing and decision-making demands placed upon them. At the same time, participants would still find themselves in a monotonous environment with little to do; hence cueing should not reduce the tedium of vigilance. Thus the direct- and indirect-cost views lead to differential predictions regarding the effects of cueing on workload and boredom in a sustained-attention task. On the basis of the direct-cost model, one would expect a high-boredom, low-workload profile in the context of cueing, whereas the indirect-cost view leads to the expectation of a high-boredom, high-workload profile.

A second form of supplementary information that has been used to bolster vigilance performance is the provision of information feedback or knowledge of results (KR), which has been shown to enhance the speed and accuracy of signal detections (Davies & Parasuraman, 1982; Dittmar, Warm, & Dember, 1985; Szalma, Hitchcock, Miller, Warm, & Dember, 1999). Moreover, a recent study has shown that KR can reduce workload in vigilance tasks (Becker, Warm, Dember, & Hancock, 1995). When operating under KR, however, observers are not relieved of the need for continuous observation and decision making, as they are in the case of consistent cueing. Observers supplied with KR must still observe continually in order to obtain positive evaluations and avoid negative ones. Hence the direct-cost model leads to the prediction that cueing will foster a greater reduction in workload than will KR. From the indirect-cost perspective, however, the well-known motivational effects of KR (Mackworth, 1970; Warm & Jerison, 1984), by promoting greater engagement and interest in the task, could be seen as a vehicle for reducing the monotony of the vigil, thereby reducing workload to a greater extent than would cueing, which should not alleviate the monotony of sustained attention. Thus the differential effects of cueing and KR on the workload of sustained attention provide an additional vehicle for testing the direct-cost and indirect-cost views of workload in vigilance.

METHODS

A total of 108 students, 54 men and 54 women, from the University of Cincinnati participated as observers to fulfill a course requirement. They ranged in age from 18 to 24 years, with a mean age of 19.6 years. All the students had normal or corrected-to-normal vision and were free of any known hearing impairment. Three experimental groups were employed. One of the groups was cued as to the imminent arrival of critical signals, another was given KR regarding performance efficiency, and the third served as a no-supplementary-information control. We assigned thirty-six students at random to each group with the restriction that the groups be equated for gender.

All observers participated in a 40-minute vigil using the simulated air-traffic control display shown in Figure 1.

The display was presented on a color video display terminal (VDT). It consisted of a "city" (a solid red circle 10.5 mm in diameter) banded by a white circle (0.75 mm wide × 12 mm in diameter) surrounded by three circular "outer markers" (0.75 mm wide; 28 mm, 53 mm, and 83 mm in diameter, respectively), and two "jet aircraft" represented by two 1

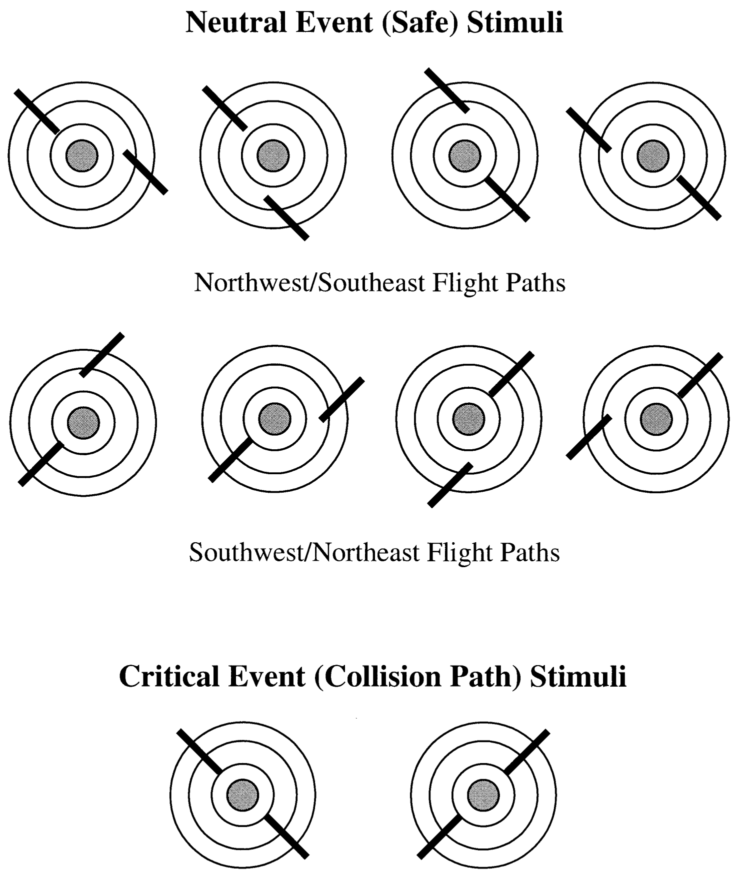

Figure 1. The simulated air-traffic control display employed in this study. All possible neutral and critical events are illustrated.

mm × 25 mm black lines, all of which were presented on a gray background. The aircraft were equidistant from the city (each reached the innermost marker) and approached the city from opposite headings, either from northwest to southeast or from southwest to northeast. Within each of the two headings, one of the aircraft was always directed toward the center of the city. The flight path of the other aircraft was parallel to its cohort, but displaced to the left or the right of it so that it would pass on a tangent to the innermost marker. The eight permutations of flight headings (see Figure 1) were presented randomly throughout the vigil and constituted neutral (safe) situations requiring no overt responses from observers. Critical signals for detection (emergency events) were cases in which the two aircraft, in either the northwest/southeast or the southwest/northeast headings (see Figure 1), were aligned on a collision path over the center of the city. In all experimental conditions, the display was updated 30 times/min with a dwell time of 300 ms. Ten critical signals occurred at random intervals within each of four continuous 10-min periods of watch in all conditions (signal probability/period = .033).

Participants in the cueing group received oral prompts that a critical signal (emergency event) would occur in one of the five display updates immediately following the prompts. The cue consisted of the word *Look* provided through a digitized male voice. Cued participants were informed that the prompt was perfectly reliable and that emergency events would appear only during these cued intervals. Monitors in the KR group received composite information regarding correct detections (hits),

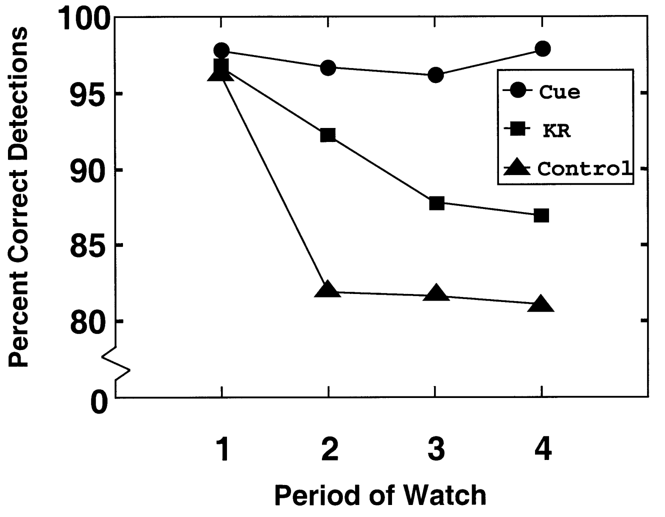

Figure 2. Mean percentages of correct detections for the cue, KR, and control groups.

errors of commission (false alarms), and detection failures (misses) given in a male voice that announced "*Yes*," "*No*," and "*Miss*," respectively, after each response, or lack thereof. Control monitors received acknowledgment after each response in the form of the statement *Logged* spoken by the same male voice to control for accessory auditory stimulation. All verbalizations were presented binaurally at an intensity of 60 dB(A) at the participant's ear. Stimulus presentations and verbalizations were orchestrated by a MacIntosh IIci (Apple: Cupertino, CA) personal computer with a built-in speaker and VDT. Programming was achieved by means of SuperLab (v1.5; Cedrus, Phoenix, AZ) and Microsoft Excel (v3.0; Microsoft, Seattle, WA) software.

Observers indicated their detection of critical signals by pressing the spacebar on the computer's keyboard. A response occurring within 1500 ms after the appearance of a critical signal was automatically recorded by the computer as a correct detection. All other responses were recorded as false alarms. The 1500 ms cutoff ensured that responses to any particular event could not overlap the appearance of the next event. Pilot work indicated that if monitors were going to respond to a critical signal, they would do so within this 1500 ms window.

Participants were tested individually in a 1.95 × 1.90 × 1.88 m Industrial Acoustics (Industrial Acoustics, New York, NY) sound chamber. They were seated in front of the VDT, which was mounted on a table at eye level at a viewing distance of 70 cm. Ambient illumination in the chamber was 0.74 cd/m^2. It was provided by a 25-watt incandescent bulb housed in a parabolic reflector positioned above and behind the seated observer so as to minimize glare on the display. After reporting for the experiment, participants completed an informed consent form, removed their watches, and received two 5-min practice trials on the vigilance task they would later encounter. Prompts, KR, and response acknowledgment were not presented during practice. In order to remain in the study, participants were required to detect at least 70% of all critical signal appearances during practice with no more than 10% false alarms. Three participants failed to meet this criterion and were replaced. Participants were unaware of the length of the vigil, other than that it would not exceed 90 min.

Immediately following the main vigil, half the male and female participants in each experimental group completed a computerized version of the NASA-TLX, a well-regarded instrument for assessing workload (Nygren, 1991). The remaining participants completed the Task-related Boredom Scale (TBS) that has been used to measure the boredom of vigilance (Scerbo, 1998; Scerbo, et al., 1992). The TBS assesses the situational characteristics that lead to boredom and the feelings often associated with boredom. The workload and boredom scales were administered to separate subgroups within each experimental condition to avoid the possibility that completing one scale would contaminate responses to the other.

RESULTS

Performance Efficiency

Mean percentages of correct detections for the control, KR, and cue groups are plotted as a function of time on task in Figure 2.

It is evident in Figure 2 that the detection scores for all groups exceeded 95% at the outset of the vigil. The performance of the cueing group remained at that high level throughout the watch. In contrast, detection probability in the control group declined by 15 % over the course of the vigil. Performance efficiency also declined over time for observers who were afforded KR, though the decline was not as pronounced in this group as in the control group. These impressions were confirmed by an analysis of variance (ANOVA) based on an arcsine transformation of the detection data, which revealed significant main effects for groups, $F(2, 105) = 8.34$, $p < .001$, and periods $F(3, 315) = 14.20$, $p < .001$, and a significant Groups × Periods interaction, $F(6, 315) = 4.14$, $p < .001$. False alarms were negligible (< 1 %) throughout the vigil in all three experimental groups.

Workload

As measured by the NASA-TLX, overall workload is indexed on a 0–100 interval scale. Mean overall workload scores and associated standard errors on the NASA-TLX for the three experimental groups are displayed in Figure 3. Also shown is the mean workload score on a 104-s card-sorting task in which observers (9 women and 9 men sampled from the same population of participants as used in the vigil) sorted for suit while paced by an auditory indicator at the rate of one card/s. Including the card-sorting task ensured that any absence of workload differences among the three experimental groups could not be attributed to lack of sensitivity to task-relevant workload on the part of the NASA-TLX.

It is clear in Figure 3 that the workload means for the control and KR groups were almost identical, both falling at the upper level of the TLX scale. Conversely, perceived workload in the cued group was considerably lower than that in the other two groups. The mean for this group was approximately half as great as the means for the control and KR groups and was only about 10 points higher than that of the card-sorting task. An ANOVA revealed a statistically significant groups effect in the

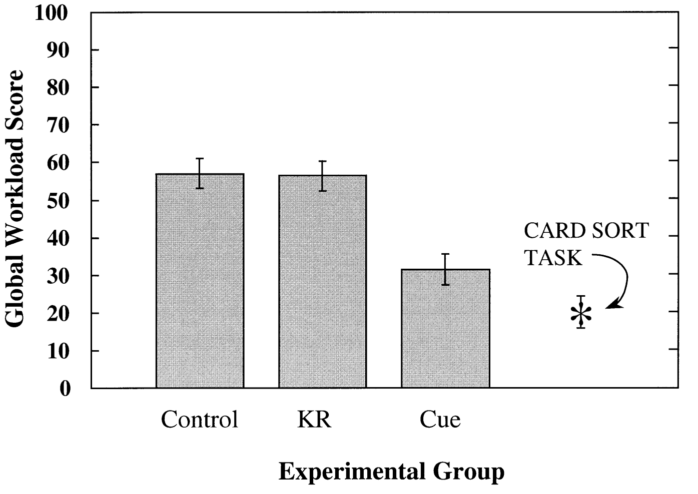

Figure 3. Mean overall workload scores (and associated standard error bars) on the NASA-TLX for the control, KR, and cue groups. The results for a card-sorting task are provided for comparison.

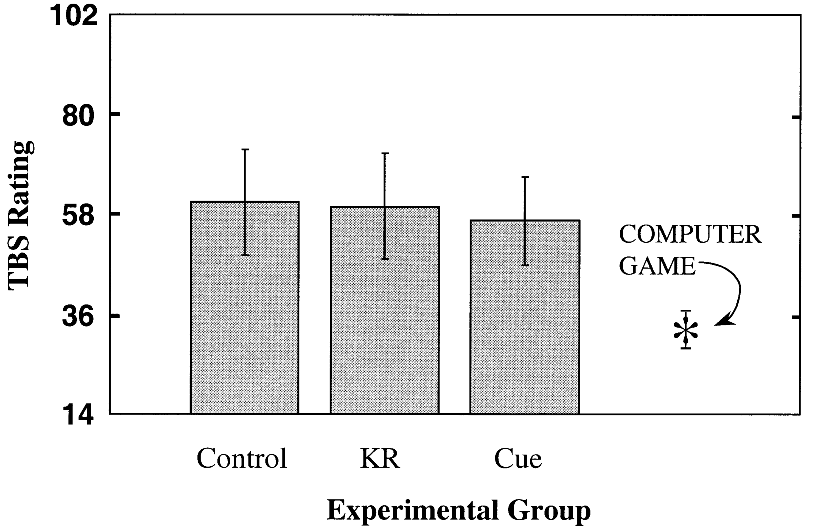

Figure 4. Mean boredom scores (and associated standard error bars) on the TBS for the cue, KR, and control groups. The results for a computer game are provided for comparison.

overall workload data, $F(2, 51) = 11.21$, $p < .001$. Scores for the card-sorting task were not included in the analysis.

Boredom

As measured by the TBS, boredom scores could range from 14 to 102. Mean boredom scores and associated standard errors on the TBS are presented in Figure 4. Also shown in the figure is the mean boredom score on a 20-m computer game known as “Asteroids,” (ATARI, Sunnyvale, CA) in which players (9 women and 9 men drawn from the same participant population) used the computer space bar to fire missiles at floating asteroids to avoid collisions with their spaceship. Including the game ensured that potential absence of boredom differences among the three experimental conditions could not be attributed to a lack of sensitivity on the part of the TBS.

Figure 4 reveals that boredom scores for the three experimental groups were essentially the same, falling within midlevel on the boredom scale. Moreover, the mean for each group was approximately twice as high as that for participants who engaged in the computer game. An ANOVA of the TBS scores for the three experimental groups indicated that the groups did not differ significantly in regard to their boredom levels, $F(2, 51) < 1$.

DISCUSSION

This study tested predictions from two competing explanations of the high workload found in vigilance tasks. One, the direct-cost view, maintains that the workload of vigilance originates in characteristics of the task itself. The alternate, the indirect-cost view suggested by Scerbo and his associates (Scerbo, 1998; Scerbo, et al., 1992), maintains that the elevated workload scores in vigilance arise not from specific task requirements but from efforts to combat the “fog of boredom” that is associated with these tasks. From the direct-cost perspective, it was anticipated that cueing would result in a high-boredom, low workload profile, whereas a high-boredom, high-workload profile was anticipated on the basis of the indirect-cost model. The results with regard to the cueing manipulation clearly supported the direct-cost model; cueing was associated with a high level of boredom and low workload.

These seemingly clear results might be challenged by reference to two possible artifacts. One stems from the notion that the low workload of the cue group simply reflected those observers’ disengagement from the vigilance task as a consequence of excessive boredom. That argument is countered, however, by the fact that the cue group not only produced lower workload ratings than its control and KR counterparts but also showed superior performance in comparison with the other two groups. Detection probability in the cue group exceeded 95%, and the typical vigilance decrement, evident in the control and KR groups, was eliminated by the cueing procedure. Thus the low workload in the cue group cannot be attributed to an artifact of disengagement from the vigilance task.

A second potential artifact centers on the sensitivity of the TBS to variations in experimental conditions. The similarity in the boredom scores of the three experimental groups could be attributable to an insensitive instrument. But the computer game used to check on the sensitivity issue produced considerably lower boredom scores than the three vigilance conditions. Although the results of this experiment support the direct-cost view of workload in vigilance, it should be emphasized that they in no way minimize the importance of boredom in prolonged monitoring situations. Clearly, high levels of task-induced workload and the boredom engendered by monotony and lack of control need to be considered in enhancing job satisfaction and performance in complex automated systems that require substantial monitoring by operators (Nickerson, 1992).

As a further test of the direct-cost and indirect-cost models of the workload of sustained attention, the present study determined which of two supplementary information procedures – cueing or KR – was more effective in reducing the workload of vigilance. Because participants receiving KR were not relieved of the need for continual observation and decision making, as was the case with cueing, the direct-cost model led to the expectation that workload would be greater in the KR group than in the cue group. An opposite expectation was derived from the indirect-cost model, based on the assumption that its motivational value could permit KR to attenuate the monotony of the vigil to a greater degree than could cueing. Accordingly KR would be more effective than cueing, which has little inherent motivational value, in reducing the subsequent boredom-induced workload of the task.

As with the results of the boredom-workload comparison vis-à-vis cueing, the results of the cueing-KR comparison clearly confirmed the expectation generated by the direct-cost model. Thus both the cueing and the KR manipulations provided evidence of the boredom-workload dissociation required to discontinue the indirect-cost model. If suitable manipulations could be developed, this case could be strengthened by showing the same dissociation, but with workload remaining constant over conditions and boredom varying.

An unexpected finding was the failure of KR to reduce workload. Though Becker et al. (1995) showed that KR can have such an effect, its occurrence is not inevitable and must be dependent on as yet unspecified conditions of the vigil – for example, the nature of the display. The display used here was different from that used by Becker et al.; their display required observers to discriminate line length. In any case, the effect of KR on workload was not central to the objective of this experiment, though it does pose an interesting issue for future investigation.

REFERENCES

Aiken, E. G., & Lau, A. W. (1967). Response prompting and response confirmation: A review of recent literature. *Psychological Bulletin, 68,* 330–341.

Annett, J., & Patterson, L. (1967). Training for auditory detection. *Acta Psychologica, 27,* 420–426.

Asteroids [Computer Software] (1997). Sunnyvale, CA: ATARI.

Becker, A. B., Warm. J. S., Dember, W. N., & Hancock, P. A. (1995). Effects of jet engine noise and performance feedback on perceived workload in a monitoring task. *International Journal of Aviation Psychology, 5,* 49–62.

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

Deaton, J. E., & Parasuraman, R. (1993). Sensory and cognitive vigilance: Effects of age on performance and mental workload. *Human Performance, 6,* 71–97.

Dittmar, M. L., Warm, J. S., & Dember, W. N. (1985). Effects of knowledge of results on performance in successive and simultaneous vigilance tasks: A signal detection analysis. In R. E. Eberts & C. G. Eberts (Eds.), *Trends in ergonomics/human factors II* (pp. 195–202). Amsterdam: North Holland.

Hart, S., & Staveland, L. (1988). Development of the NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Amsterdsam: North-Holland.

Heilman, K. M. (1995). Attentional asymmetries. In R. J. Davidson & K. Hugdahl (Eds.), *Brain asymmetry* (pp. 217–234). Cambridge: MIT Press.

Howell, W. C. (1993). Engineering psychology in a changing world. *Annual Review of Psychology, 44,* 231–263.

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

Nickerson, R. S. (1992). *Looking ahead: Human factors challenges in a changing world.* Mahwah, NJ: Erlbaum.

Nygren, T. E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors, 33,* 17–33.

Scerbo, M. W. (1998). What's so boring about vigilance? In R. R. Hoffman, M. F. Sherrick, & J. S. Warm (Eds.), *Viewing psychology as a whole: The integrative science of William N. Dember* (pp. 145–166). Washington, DC: American Psychological Association.

Scerbo, M. W., Greenwald, C. Q., & Sawin, D. A. (1992). Vigilance: It's boring, it's difficult, and I can't do anything about it. In *Proceedings of the Human Factors Society 36th Annual Meeting* (pp. 1508–1512). Santa Monica, CA: Human Factors and Ergonomics Society.

Scerbo, M. W., & Holcomb, J. R. (1993, February). *Effects of signal conspicuity and time on boredom in vigilance*. Paper presented at the first Mid-Atlantic Human Factors conference, Reston, VA.

Szalma, J. L., Hitchcock, E. M., Miller, L. C., Warm, J. S., & Dember, W. N. (1999). Intraclass and interclass transfer of training for vigilance. In. M. W. Scerbo & M. Mouloua (Eds.), *Automation technology and human performance: Current research and trends* (pp. 183–187). Mahwah, NJ: Erlbaum.

Warm, J. S., Dember, W. N., & Hancock, P. A. (1996). Vigilance and workload in automated systems. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 183–200) Mahwah, NJ: Erlbaum.

Warm, J. S., Dember, W. N., & Parasuraman, R. (1991). Effects of olfactory stimulation on performance and stress in a visual sustained attention task. *Journal of the Society of Cosmetic Chemists, 42,* 199–210.

Warm, J. S., & Jerison, H. J. (1984). The psychophysics of vigilance. In J. S. Warm (Ed.), *Sustained attention in human performance* (pp. 15–59). Chichester, England: Wiley.

Weiner, E. L., & Attwood, D. A. (1968). Training for vigilance: Combined cueing and knowledge of results. *Journal of Applied Psychology, 52,* 474–479.

Edward M. Hitchcock received his M.S. from the University of Cincinnati in 1996. He is a doctoral candidate in experimental psychology/human factors at the university.

William N. Dember received his Ph.D. in experimental psychology from the University of Michigan in 1955. He is Professor Emeritus of Psychology at the University of Cincinnati.

Joel S. Warm received his Ph.D. in experimental psychology from the University of Alabama in 1966. He is a professor of Psychology at the University of Cincinnati.

Brian W. Moroney received his Ph.D. in experimental psychology/human factors from the University of Cincinnati in 1999. He is a human factors engineer with IBM in Boulder, CO.

Judi E. See received her Ph.D. in experimental psychology/human factors from the University of Cincinnati in 1994. She is a human factors engineer at Logicon Technical Services, Inc., Albuquerque, NM.

Date Received: August 10, 1998
Date Accepted: January 5, 1999