Vigilance and Automation Dependence in Operation of Multiple Unmanned Aerial Systems (UAS): A Simulation Study

Ryan W. Wohleber, **Gerald Matthews**, **Jinchao Lin**, **James L. Szalma**, University of Central Florida, Orlando, USA, **Gloria L. Calhoun**, **Gregory J. Funke**, Air Force Research Laboratory, Wright-Patterson AFB, OH, USA, **C.-Y. Peter Chiu**, University of Cincinnati, USA, and **Heath A. Ruff**, Infoscitex, Dayton, OH, USA

Objective: This simulation study investigated factors influencing sustained performance and fatigue during operation of multiple Unmanned Aerial Systems (UAS). The study tested effects of time-on-task and automation reliability on accuracy in surveillance tasks and dependence on automation. It also investigated the role of trait and state individual difference factors.

Background: Warm's resource model of vigilance has been highly influential in human factors, but further tests of its applicability to complex, real-world tasks requiring sustained attention are necessary. Multi-UAS operation differs from standard vigilance paradigms in that the operator must switch attention between multiple subtasks, with support from automation.

Method: 131 participants performed surveillance tasks requiring signal discrimination and symbol counting with a multi-UAS simulation configured to impose low cognitive demands, for 2 hr. Automation reliability was manipulated between-groups. Five Factor Model personality traits were measured prior to performance. Subjective states were assessed with the Dundee Stress State Questionnaire.

Results: Performance accuracy on the more demanding surveillance task showed a vigilance decrement, especially when automation reliability was low. Dependence on automation on this task declined over time. State but not trait factors predicted performance. High distress was associated with poorer performance in more demanding task conditions.

Conclusions: Vigilance decrement may be an operational issue for multi-UAS surveillance missions. Warm's resource theory may require modification to incorporate changes in information processing and task strategy associated with multitasking in low-workload, fatiguing environments.

Application: Interface design and operator evaluation for multi-UAS operations should address issues including motivation, stress, and sustaining attention to automation.

Keywords: vigilance, multitasking, automation, fatigue, stress, Unmanned Aerial Systems

Address correspondence to Gerald Matthews, Institute for Simulation & Training, University of Central Florida, 3100 Technology Pkwy., Orlando, FL 32826, USA; e-mail: gmatthews@ist.ucf.edu.

HUMAN FACTORS
Vol. 61, No. 3, May 2019, pp. 488–505
DOI: 10.1177/0018720818799468
Article reuse guidelines: sagepub.com/journals-permissions

Joel Warm's priceless legacy to human factors includes the resource theory of vigilance (Warm, Dember, & Hancock, 1996), acclaimed as the leading account of sustained attention (e.g., Proctor & Vu, 2010). It also includes more than 50 years of meticulous experimental work on vigilance based on his expertise in psychophysics, experimental design, workload assessment, neuroergonomics, and human factors engineering (Warm & Dember, 1998; Warm & Jerison, 1984; Warm, Matthews, & Finomore, 2008; Warm, Parasuraman, & Matthews, 2008). The present article addresses two of the many facets of vigilance that Warm's work addressed: vigilance decrement in complex, real-world operational settings and the role of individual differences. We report on a study that examined sustained attention in the context of a simulation of operating multiple Unmanned Aerial Systems (UAS) supported by automation, and the influence of trait and state individual difference factors on performance, workload, and stress.

ATTENTIONAL RESOURCES, VIGILANCE, AND UAS OPERATION

The resource theory of vigilance combines three principal propositions. First, contrary to earlier conceptions of vigilance as mentally undemanding, sustaining focused attention often imposes high cognitive demands and workload (Warm et al., 1996). Second, sustained allocation of resources to attentional processes leads to depletion of the resource pool, an assumption supported by Warm's groundbreaking work on hemodynamic markers for vigilance (Warm, Tripp, Matthews, & Helton, 2012). Third, high-workload vigilance tasks are commonly resource limited (Norman & Bobrow, 1975), so that resource depletion is directly expressed as vigilance decrement. A meta-analysis (See, Howe, Warm, & Dember, 1995) showing that perceptual sensitivity declines increase with task difficulty substantiates this proposition.

Theoretical accounts of vigilance including resource theory and its competitors continue to be debated (e.g., Fraulini, Hancock, Neigel, Claypoole, & Szalma, 2017; Thomson, Besner, & Smilek, 2016). The aim of the present study was to test the applicability of the resource model to a complex task environment (multi-UAS control). The generalizability of results from standard laboratory paradigms for vigilance to real-world tasks is a longstanding issue. Skepticism over generalizability has been fueled by observations that personnel tasked with monitoring often experience breaks from the task that may help to reduce vigilance decrement (Casner & Schooler, 2015; Mackie, 1987). Hancock (2013) attributes the decrement to design features of typical laboratory displays that elicit boredom and loss of personal agency, features not necessarily present in operational environments. A contrasting perspective advanced by Warm and his colleagues (e.g., Warm et al., 1996) is that regardless of the quality of the display design, automation of systems increasingly shifts the operator's role from active control to supervisory monitoring, enhancing the likelihood of vigilance decrement.

Military UAS missions are often monotonous, so that operators become fatigued, bored, distractible, and potentially vulnerable to loss of vigilance (Cummings, Mastracchio, Thornburg, & Mkrtchyan, 2013; Mouloua, Gilson, & Hancock, 2003; Tvaryanas et al., 2006). Indeed, a UAS study conducted by Warm and colleagues (Gunn et al., 2005) showed a vigilance decrement in detection of threat warnings presented in the context of a 32.4 min simulated target acquisition task.

However, it is unclear whether vigilance decrement in detection tasks would occur with more complex task configurations, in which the UAS operator timeshares monitoring of displays with performing additional subtasks (Ouma, Chappelle, & Salinas, 2011). Real-world monitoring tasks often require multiple, heterogeneous forms of processing (Donald, 2008). If the operator frequently switches between subtasks requiring different multiple resources, depleted resources may be able to recover; for example, a visual resource might recover during periods when the operator is performing an auditory subtask. In addition, task complexity may counter operator passivity and disengagement. In a follow-up to Gunn et al. (2005), Parsons, Warm, Nelson, Matthews, and Riley (2007) showed that vigilance decrement was eliminated in a detection-action scenario, in which target detection allowed the participant to use the mouse to destroy the threat to the UAS. Needs for strategic management of multiple task elements may increase engagement and counter the decrement.

Vigilance may also be influenced by the increasing automation of operator functions in the UAS. The current research focuses on automation to support multi-aircraft control (MAC), where a single operator controls multiple vehicles (Calhoun, Goodrich, Dougherty, & Adams, 2016). Automation is essential to manage multiple sources of demand including information retrieval, flight control, navigation, mission and payload management, and communication. Implications for vigilance are equivocal. On the one hand, automation allows the operator to manage multiple subtasks that would otherwise be allocated to different personnel, for example, sensor operation and piloting. The consequent increase in subtask diversity may reduce resource depletion and hence mitigate vigilance decrement. On the other hand, automation that relegates the operator to a passive monitoring role could exacerbate vigilance decrement (Warm, Parasuraman, & Matthews, 2008). Design factors such as the reliability and level of automation (LOA) also moderate the impact of the automation on attention (Parasuraman & Manzey, 2010; Parasuraman & Wickens, 2008). For example, higher LOAs can encourage operators to delegate too much responsibility to the automation for target detection (Calhoun, Ruff, Draper, & Wright, 2011).

The introduction of automation also raises the question of how dependence on automation may change as the operator becomes less vigilant. Successful MAC requires that operators calibrate dependence on automation appropriately, avoiding the twin hazards of under- and overdependence (Parasuraman & Riley, 1997). Parasuraman and Manzey's (2010) theory of automation use provides a resource theory perspective. It links two forms of overdependence on automation—automation complacency and automation bias—to attentional processes. It emphasizes attentional overload as a source of automation misuse, especially with more reliable, higher LOA systems. Thus, automation that successfully mitigates workload should minimize overdependence and counter vigilance decrements.

A contrasting view comes from studies of effort-regulation in fatigue states (Hockey, 1997; Sauer, Wastell, Robert, Hockey, & Earle, 2003). Fatigued operators reduce performance standards as an effort-conservation strategy, potentially leading to increasing dependence on automation, coupled with loss of vigilance. Similarly, Desmond and Hancock (2001) describe a passive fatigue state induced by boredom and underload in which operators tend to disengage from the task. In simulated driving studies, automated driving impairs alertness following manual takeover (Matthews, Neubauer, Saxby, Wohleber, & Lin, 2018; Saxby, Matthews, Warm, Hitchcock, & Neubauer, 2013), and fatigued drivers are more likely to initiate automation (Neubauer, Matthews, Langheim, & Saxby, 2012).

INDIVIDUAL DIFFERENCES

Warm's research also investigated the roles of trait and state individual-difference factors in vigilance. A review (Finomore, Matthews, Shaw, & Warm, 2009) concluded that measures of temporary stress states were stronger predictors of vigilance than were stable personality traits, such as the Five Factor Model (FFM; McCrae, 2009). Extraversion is modestly negatively associated with vigilance in some studies (Finomore et al., 2009), whereas neuroticism and low conscientiousness may be associated with stress response to vigilance, though not with performance (Shaw et al., 2010).

A key state dimension is task engagement, which contrasts energy, motivation, and alertness with fatigue symptoms such as tiredness, apathy, and distractibility (Matthews et al., 2002). This dimension correlates with perceptual sensitivity and the hemodynamic response to task demands, and may index resource availability (Matthews et al., 2010; Matthews, Warm, & Smith, 2017). Thus, to the extent that vigilance decrement occurs in UAS operation, operators who can maintain task engagement should perform better. If fatigue leads to increased automation-dependence (Neubauer et al., 2012), low engagement might also be associated with dependence.

Individual-difference factors might play a somewhat different role in UAS operation than they do in conventional laboratory vigilance, although evidence is limited. Szalma and Taylor (2011) found that FFM conscientiousness and neuroticism were associated with performance accuracy in a simulated unmanned ground vehicle (UGV) scenario, but relationships varied across task conditions. In Szalma and Taylor's study, neuroticism was the only trait to predict operator agreement with automation. By contrast, Kidwell, Calhoun, Ruff, and Parasuraman (2012) found that extraversion was associated with usage of adaptable automation. In addition, the optimal operator state for multitasking may differ from the high engagement state ideal for conventional vigilance. Distress was a better predictor than task engagement of attention in studies of simulated UGV (Matthews, Reinerman-Jones, Abich, & Kustubayeva, 2017) and UAS performance (Lin et al., 2018). In multitasking, distress may disrupt executive control of multiple processing components (Matthews & Campbell, 2010; Matthews, Reinerman-Jones, et al., 2017).

AIMS AND HYPOTHESES

The current study utilized the ALOA (Adaptive Levels of Autonomy) multi-UAS research simulation (see Calhoun et al., 2011). ALOA supports multiple subtasks including routing, surveillance, health checks, and communication, with automation at different LOAs. In a previous study (Lin et al., 2018) we configured the platform to be highly cognitively demanding and showed that overload tended to impair accuracy on surveillance tasks, increase neglect of these tasks, and encourage disuse of automation. The surveillance tasks were designed to correspond to sensory and cognitive vigilance tasks, respectively (see See et al., 1995). A Weapon Release (WR) task required the participant to view a still image and distinguish friendly versus enemy tanks that differed subtly in appearance. The use of perceptually confusable stimuli builds on Warm's demonstrations that perceptually demanding vigilance tasks may elicit rapid performance decrement (e.g., Temple et al., 2000). By contrast, the Image Analysis (IA) task required the participant to count green diamonds overlaying a map display, ignoring other geometric shapes. Stimuli were easily discriminated, so demands were cognitive rather than sensory in nature; typically, cognitive vigilance tasks are less prone to vigilance decrement than are sensory tasks (See et al., 1995). Lin et al. (2018) confirmed that accuracy was significantly lower for WR than for IA.

In the present study, the simulation was configured to impose low cognitive demands, so as to encourage passive fatigue. A manipulation of automation reliability was included. Trial duration was extended to 2 hr to provide ample time for vigilance decrement to emerge. Aims and hypotheses were formulated to test generalizability of resource theory predictions to the MAC context, as follows.

Detection accuracy, stress, and workload during UAS operation. High-demand vigilance tasks have several signature characteristics (Warm, Parasuraman, & Matthews, 2008). These include high workload, increased distress, and lower task engagement following performance. We hypothesized that these subjective responses would be observed for the current task (H1). In addition, signal detection declines over time (vigilance decrement), especially as attentional demands increase. We hypothesized that vigilance decrement would be greater for the more difficult WR task and when automation was less reliable, placing more demands on attention (H2).

Temporal change in automation dependence. There is little previous research on changes in dependence on automation during vigilant monitoring, and so this aspect of the research was exploratory. The Parasuraman and Manzey (2010) theory predicts greater dependence with more reliable automation (H3), but does not make strong predictions for temporal change. The effort-minimization perspective (Sauer et al., 2003) implies that fatigue should produce greater reliance on automation, and hence automation-dependence should increase with time on task (H4).

Individual differences in stress response and performance. Previous studies of vigilance (Shaw et al., 2010) and unmanned vehicle operation (Szalma & Taylor, 2011) broadly predict that low neuroticism and high conscientiousness should predict more adaptive states during performance, including higher task engagement and lower distress (H5). Vigilance studies identify task engagement with resource availability, implying that task engagement should correlate with detection accuracy (H6), especially as task demands increase (WR task, low reliability). An effort-minimization perspective predicts that task engagement should be negatively associated with automation dependence (H7).

METHOD

A previous short paper (Wohleber et al., 2016) reported results from this study pertaining to the automation reliability manipulation. The present article provides a comprehensive description of this study's results and specifically addresses vigilance effects. It adds analyses of individual differences and further post hoc analysis and regression modeling relevant to our current aims. The study utilized a mixed-model design with automation reliability (high vs. low) as a between-groups variable and time on task (eight 15-min intervals) as a repeated-measures variable.

Participants

Participants were 131 University of Central Florida undergraduate psychology students (50 women, 81 men, M_{age} = 19.86, age range: 18–31 years), who received class credit. Populations vulnerable to stress manipulations due to mental disorder or unable to perform required tasking (due to uncorrected vision impairment, lack of fluency in English, physical disability) were excluded. The study was approved by the University of Central Florida Institutional Review Board.

Subjective Measures

Demographics questionnaire. The demographics questionnaire asked about biographical information including education, computer use, and expertise on several categories of video games (rated on 7-point scales).

Workload. The NASA-Task Load Index (NASA-TLX: Hart & Staveland, 1988) requires 0 to 100 ratings of six workload components: performance, mental demand, physical demand, temporal demand, effort, and frustration. Overall workload was calculated as an unweighted mean of ratings, given that the weighting procedure sometimes used does not improve the psychometric properties of the scale (Hendy, Hamilton, & Landry, 1993).

Stress state. The short, 21-item Dundee Stress State Questionnaire (DSSQ: Matthews et al., 2002) was administered to gauge symptoms of passive fatigue. Participants completed a pretask DSSQ prior to training, and after the experimental trial, a posttask DSSQ on their state in the final 10 min of the trial. The DSSQ assesses three higher order dimensions of subjective state in a task-performance context: task engagement, distress, and worry.

Personality. FFM extraversion, agreeableness, openness, conscientiousness, and neuroticism traits were assessed using Saucier's (2002) Mini-Markers. Participants rated the extent to which 40 different adjectives described their personalities using a 9-point Likert scale.

Simulator

The ALOA research testbed (Johnson, Leen, & Goldberg, 2007) provided a UAS control environment for eight tasks that mimicked projected cognitive demands of controlling four UAS. Simulator configuration for this study was based on insights from Lin et al. (2018) and pilot testing. Table 1 lists the tasks, their priorities (1 = highest), the frequency of stimuli for the whole trial, and the roles assigned to the automation and the participant. Figure 1 shows task locations. Each UAS was allocated to a route automatically, and the participant could follow the progress of the UAS toward the end of its route on the map display. Overall task frequency was set to around 2 per min, trial duration was 120 min, and various tasks were automated to elicit passive fatigue (see details below). This report focuses on the two surveillance tasks: IA and WR.

Surveillance tasks. IA and WR tasking shared top priority (Figure 1a, left side of panel). As the four vehicles passed target locations, their sensor payloads collected image data that participants then had to analyze. These IA and WR tasks populated respective queues (Figure 1a) for participants to select and complete one at a time. Clicking on the row of data in the queue windows displayed the IA or WR image in the surveillance task window. Tasks in each queue timed out after 30 s (IA) or 20 s (WR) if not selected and completed. Sample stimuli for the tasks are shown in Figure 2. In the IA task, 19 to 26 green symbols (diamonds, squares, circles, and triangles) overlaid the imagery, of which 1 to 8 were diamonds. The eight options for the correct number of diamonds were presented below the image, one of which was highlighted by the automation. The participant either confirmed the automation's recommended option or clicked on another option. For the WR task, participants had to ensure that the automation had correctly identified enemy tanks (marked with red squares). The terrain photo was overlaid with a total of 2 to 7 tank images. Of these, 0 to 5 were friendly, and 0 to 4 were hostile. The two types of tank differed slightly in their dimensions and gun barrel length. For each image, the automated target recognition marked 1 to 5 tanks with the red box indicating they should be targeted, including 0 to 4 friendly tanks and 0 to 4 enemy tanks. Participants either confirmed that the recommended weapon strike option ("Authorize" or "Do Not Authorize") was correct or clicked on the alternate option.

For both tasks, participants had to change or confirm the automation's recommendation before the task timed out (management-by-consent LOA). Recommendations were 60% accurate in the low-reliability condition and 86.7% in the high-reliability condition, for both tasks, with minor variation across time blocks due to scenario constraints. These reliability values aimed to establish a strong contrast between conditions, based on Wickens and Dixon's (2007) finding from 20 studies that a reliability of 70% was the "crossover point" below which

TABLE 1: Summary of ALOA Simulation Tasks

Task	Priority	Frequency	Automation (A) and Operator (O) Roles
1. Allocation & rerouting	N/A	20 each	A: Full (100% reliable) O: None
2. Image analysis	1	60 (time out 30 s)	A: Recommends option (60 or 86.7% reliable) O: Select option with correct number of diamonds
3. Weapon release	1	60 (time out 20 s)	A: Recommends option (60 or 86.7% reliable) O: Select correct strike authorization option
4. Unidentified aircraft	2	48 (time out 10 s)	A: None O: Click red plane symbol when appears on map
5. Digit pairs	3	16 (time out 10 s)	A: None O: Respond (true/false) whether two digits meet two criteria (for difference & sum)
6. Audio chatter	3	16 (time out 15 s)	A: None O: If call sign prompted, enter designated color/number in chat. Otherwise, ignore
7. Health/status	3	16 (time out 15 s)	A: None O: Click lighted (yellow or red) indicator
8. Chat questions	3	8 response prompts 16 "noise" prompts	A: None O: Type answer in chat using information in vehicle status windows. Ignore "noise" prompts

unreliable automation was worse than no automation at all. Pilot studies confirmed that participants were aware of the reliability difference.

Three performance measures were assessed for each 15-min interval. The interval duration was chosen to be short enough to track vigilance effects, but long enough to provide an adequate sampling of responses for each task type. In that the multi-UAS scenario simulates an envisioned operational mission, the timing of the IA and WR tasks was driven by the progression of the four UAS on the routes with respect to the location of targets in the mission environment. Thus, the number of IA and WR tasks slightly varied in each time interval (7–9 of each surveillance task type per interval). *Accuracy* was defined as the overall percentage of correct decisions. *Automation dependence* was measured as the percentage of trials on which the participant confirmed the automation's decision, following Barg-Walkow and Rogers's (2016) definition. *Neglect* was defined as the percentage of images appearing in the queues that the participant failed to select, so that the image was not processed before time expired.

Procedure

Each participant was allocated at random to the high-reliability ($n = 67$) or low-reliability ($n = 64$) condition. After informed consent, participants completed the demographics questionnaire, Mini-Marker scale, and pretask DSSQ. Next, training (lasting approximately 60 min) was conducted with the participant's assigned reliability level in effect. Training consisted of a PowerPoint orientation detailing the test environment and steps for completing all tasks, followed by practice trials containing all tasking. Participants were provided with the task priorities listed in Table 1. A researcher monitored participants, answered questions, and ensured that they were aware of task priorities and completed each task component according to the rules provided. When the researcher was satisfied with the participant's competence, the main 2-hr trial was initiated. Participants ended the 4-hr test session by completing the NASA-TLX and posttrial DSSQ scales.

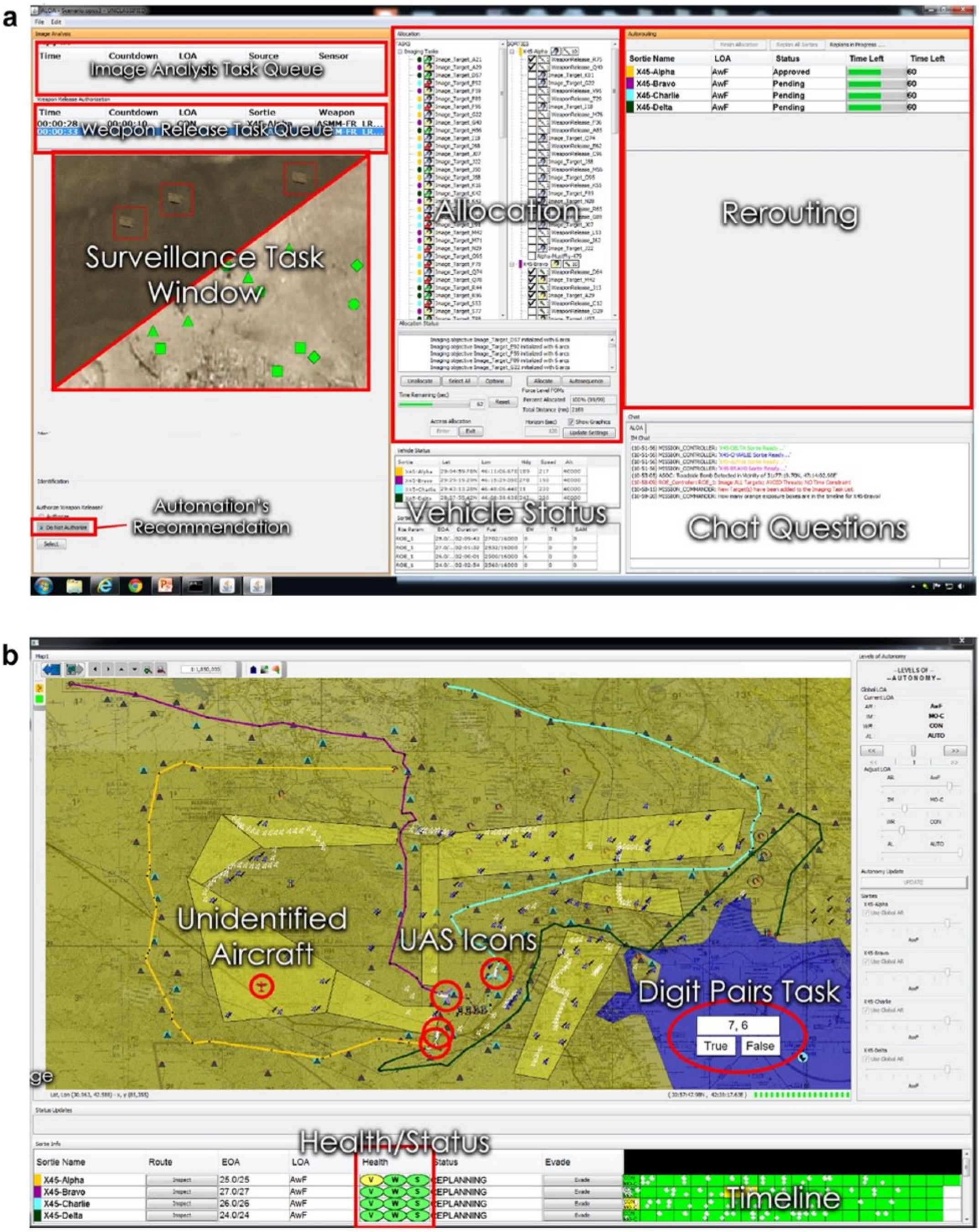

Figure 1. ALOA Simulator left (a) and right (b) monitors.

RESULTS

For all analyses, Box's corrections were used to correct for violations of sphericity, and Bonferroni corrections were used to correct for post hoc family-wise error.

Subjective Outcomes

Three 2 (pre vs. posttask) × 2 (low vs. high reliability) mixed-model analyses of variance (ANOVAs), run for each stress state factor (Figure 3), revealed that the 2-hr trial significantly

Figure 2. Sample images for Image Analysis (upper) and Weapon Release (lower) tasks. The Weapon Release image shows the automation correctly marking two enemy tanks on the left, and incorrectly marking two friendly tanks on the right. Additionally, the enemy tank at the center of the image should be marked.

reduced task engagement, $F(1, 129) = 197.64$, $p < .001$, $\eta^2_p = .605$, $d = 1.31$. There was also a small but significant interaction effect for worry, $F(1, 129) = 5.27$, $p = .023$, $\eta^2_p = .039$; post hoc contrasts showed a significant increase with low-reliability automation ($p < .001$, $d = .48$), but no change with high-reliability automation ($p = .570$, $d = .07$). A *t* test revealed that the low-reliability condition elicited higher NASA-TLX workload ($M = 39.65$, $SD = 16.64$) than did the high-reliability condition ($M = 32.26$, $SD = 16.82$), $t(129) = 2.53$, $p = .013$, $d = .44$. There were no significant main or interactive effects of task factors on distress.

Performance-Based Outcomes

A series of 2 (low vs. high reliability) × 8 (time on task: 15 min intervals) mixed-model ANOVAs were run for accuracy, automation-dependence, and neglect performance measures.

Accuracy. Accuracy was improved with high-reliability automation for both IA, $F(1, 129) = 9.52$, $p = .002$, $\eta^2_p = .069$, and WR tasks, $F(1, 129) = 11.05$, $p = .001$, $\eta^2_p = .079$. Accuracy also changed with time for both tasks: $F(5.57, 718.39) = 18.70$, $p < .001$, $\eta^2_p = .127$, and $F(5.15, 663.75) = 27.52$, $p < .001$, $\eta^2_p = .176$, respectively. Finally, results revealed significant interactions between reliability and time for IA, $F(5.57, 718.39) = 2.41$, $p = .030$, $\eta^2_p = .018$, and for WR, $F(5.15, 663.75) = 17.51$, $p < .001$, $\eta^2_p = .120$.

Temporal trends for accuracy are shown in Figure 4. To clarify time on task effects, Table 2 shows effect sizes (*d*s) for changes in accuracy from the initial baseline, that is, block 1, calculated separately for each group using Morris and DeShon's (2002) procedure for single-group repeated-measures designs. Conventionally, values of .2, .5, and .8 are considered small, medium, and large effect sizes (Cohen, 1988). Significance levels for change from baseline were derived from one-factor repeated-measures ANOVAs for each group; the main effect of time block was significant at $p < .001$ in each case. Simple contrasts with block 1 were calculated by the SPSS GLM procedure. Table 2 suggests that IA accuracy was unstable during the first hour of the performance, with performance declines in blocks 2 and 4 in both reliability conditions. However, performance returned to baseline levels in the second hour of the task (blocks 5–8) in the high-reliability condition, and showed only modest tendencies toward decrement in blocks 5 and 7 when reliability was low. WR accuracy, in the high-reliability condition, after initial instability showed a 30 min interval of impairment in blocks 5 and 6, prior to recovery. In the low-reliability condition, WR accuracy showed a more sustained 1-hr period (blocks 4–7) of impairment relative to baseline, before recovering in the final block.

Automation dependence. The high-reliability automation evoked substantially higher dependence in both IA, $F(1, 129) = 393.35$, $p < .001$, $\eta^2_p = .753$, and WR tasks, $F(1, 129) = 205.66$, $p < .001$, $\eta^2_p = .615$. Main effects of time were significant for both IA, $F(5.44, 701.92) = 40.48$, $p < .001$, $\eta^2_p = .239$, and WR, $F(5.14, 663.40) =$

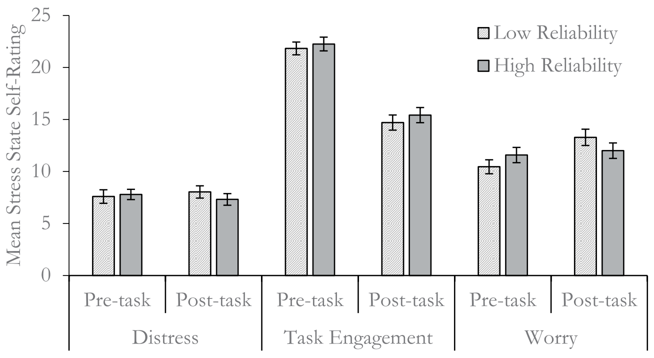

Figure 3. Self-reported stress states (DSSQ). Error bars represent standard errors.

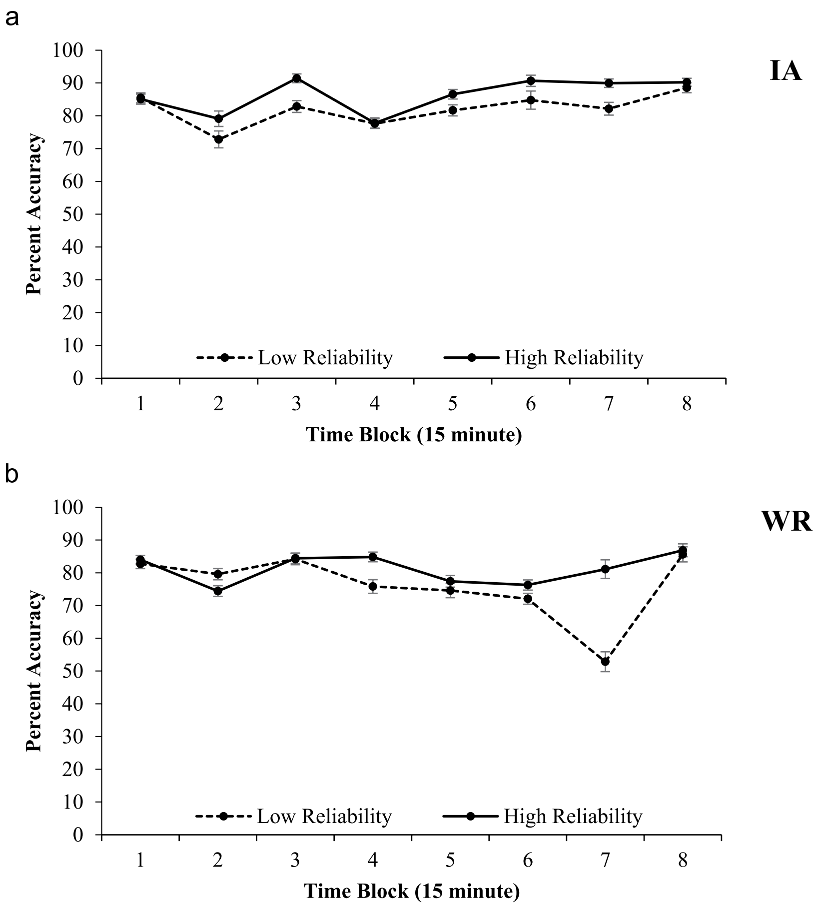

Figure 4. Accuracy for Image Analysis (a) and Weapon Release (b) tasks. Error bars represent standard errors.

41.99, $p < .001$, $\eta^2_p = .246$. Time on task effects were significantly moderated by reliability of automation for WR, $F(5.44, 701.92) = 19.41$, $p < .001$, $\eta^2_p = .131$, and IA, $F(5.14, 663.40) = 12.46$, $p < .001$, $\eta^2_p = .088$. For IA, the trend is difficult to interpret (Figure 5a), but dependence on the high-reliability automation appeared more consistent than dependence on the low-reliability automation. For WR, dependence generally declined with time, but dependence on

TABLE 2: Effect sizes (d) for accuracy changes from initial baseline (time block 1)

		Time Block						
Task	Reliability	2	3	4	5	6	7	8
IA	High	−.49**	.37**	−.81**	−.07	.20	.18	.28*
IA	Low	−.95**	−.28	−1.05**	−.40**	−.10	−.39*	.11
WR	High	−.79**	−.01	−.09	−.47**	−1.02**	−.37	.12
WR	Low	−.29	.05	−.62**	−.53**	−1.04**	−1.99**	.17

*p < .05. **p < .01.

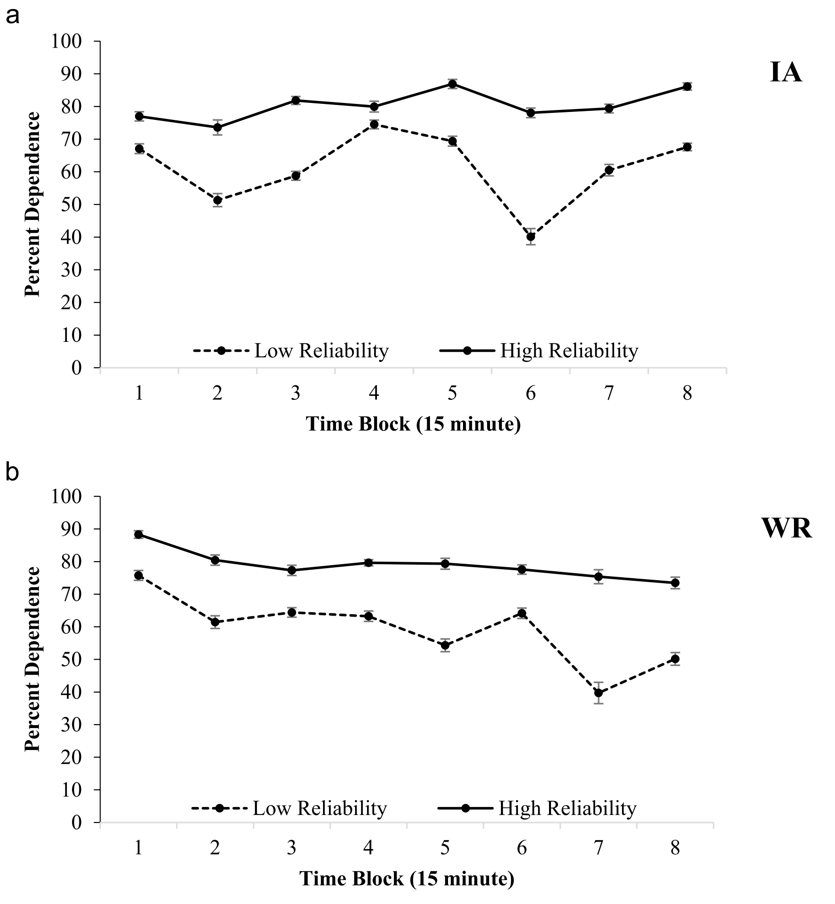

Figure 5. Dependence on automation in (a) Image Analysis and (b) Weapon Release tasks. Error bars represent standard errors.

low-reliability automation saw a more pronounced decline (and was more erratic) than dependence on high-reliability automation.

Neglect: For both surveillance tasks, neglect (failure to process tasks) did not differ by reliability level but did change with time on task, $F(5.05, 631.25) = 4.57, p < .001, \eta^2_p = .035$, and $F(5.03, 618.22) = 17.94, p < .001, \eta^2_p = .127$. Whereas time trends for both were irregular, each showed an increase in neglect after 45 min

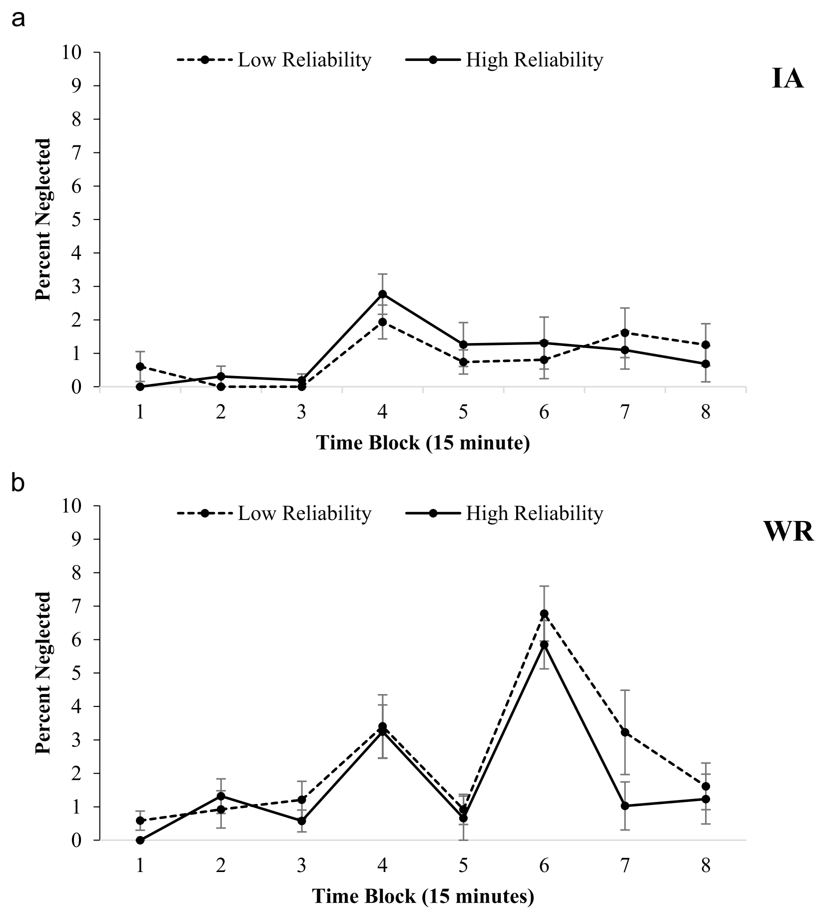

Figure 6. Neglect in (a) Image Analysis and (b) Weapon Release tasks. Error bars represent standard errors.

(Figure 6). Neither task showed a reliability × time interaction.

Individual Differences

Trait correlates. There were several significant correlations between FFM traits and posttask DSSQ states. Extraversion was negatively correlated with task engagement ($r = -.19, p < .05$), Neuroticism was associated with higher distress ($r = .29, p < .01$) and worry ($r = .33, p < .01$), and Conscientiousness predicted lower distress ($r = -.35, p < .01$) and worry ($r = -.22, p < .05$).

Correlations were computed between the FFM and the principal performance measures for WR and IA, that is, accuracy, dependence on automation, and neglect of surveillance tasks. Significant correlations did not exceed chance levels, and regression analyses did not show any interactive effects of FFM traits and reliability on performance.

State correlates. A multiple regression approach was used to investigate performance correlates of posttrial DSSQ states. For these analyses, we focused on the second hour of performance, during which it was likely that participants were experiencing fatigue. That is, dependent measures were the performance variables averaged across time blocks 5 through 8. Six regressions were computed, corresponding to three performance variables × two surveillance tasks. Four-step hierarchical regressions were computed. Step 1 included two control variables found to be predictive of stress and performance in a previous study (Lin et al., 2018): gender and self-rated expertise in "first person shooter" (FPS) video games. Lin et al. (2018) found that FPS expertise predicted higher WR accuracy;

TABLE 3: Regressions of Selected Performance Measures on Stress States, Reliability, and Control Variables: Summary Statistics

	Accuracy: Weapon Release		Accuracy: Imaging		Neglect: Weapon Release		Neglect: Imaging		
Step	*R*	ΔR^2	*R*	ΔR^2	*R*	ΔR^2	*R*	ΔR^2	*df*(Δ)
1.Control variables	.04	.00	.05	.00	.09	.00	.11	.01	2,128
2. Reliability	.36**	.13**	.26*	.06**	.17	.02	.11	.00	1,127
3. States: Linear	.39**	.02	.43**	.12**	.35*	.10**	.35*	.11**	3,124
4. States: Interactions	.50**	.10**	.47**	.03	.36	.01	.39*	.03	3,121
Significant at final step	Reliability (β = .36**), Distress × Reliability (β = .32**)		Reliability (β = .22*), Worry (β = −.17*), Task Engagement (β = −.16, *p* = .068)		Distress (β = −.25*)		Distress (β = −.21*), Task Engagement (β = −.24*)		

$^*p < .05$. $^{**}p < .01$.

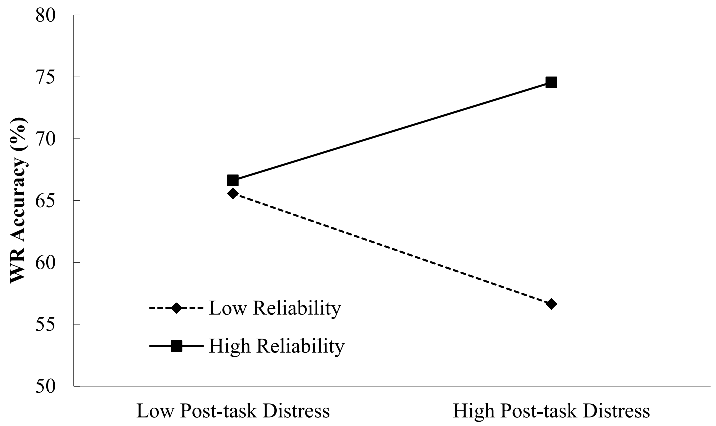

Figure 7. Association between posttask distress and accuracy moderated by automation reliability for Weapon Release tasks: Regression model.

both gaming and UAS operation may require multitasking skills. Step 2 controlled for reliability, coded as 1 (high) or −1 (low). Step 3 comprised the three centered post-DSSQ scales, and Step 4 added the DSSQ × Reliability product terms.

Table 3 shows summary statistics for regressions showing significant contributions from state measures. For both accuracy measures, higher reliability was positively associated with performance, consistent with the ANOVAs. For WR, the distress × reliability interaction was significant. As shown in Figure 7, distress tended to be positively correlated with accuracy in the high-reliability condition, but the regression was negative at low reliability. Imaging accuracy was linearly associated with lower worry and higher task engagement (trend only). Higher distress was associated with greater neglect for both tasks; task engagement was associated with lower neglect on IA. For the two automation-dependence regressions, neither linear nor interactive stress state terms contributed significantly to the equations. Similar to the ANOVAs, reliability had a significant effect in both; βs at Step 4 were .74 ($p < .01$) for WR and .86 ($p < .01$) for IA.

DISCUSSION

UAS operators commonly report that prolonged missions lead to fatigue and difficulties in sustaining attention (Chappelle, Salinas, & McDonald, 2011; Cummings et al., 2013; Tvaryanas & Macpherson, 2009). The present multi-UAS simulation study elicited comparable symptoms of large-magnitude loss of task engagement and vigilance decrement on the more demanding of two surveillance tasks. Individual differences in task engagement and distress were associated with different performance indices. An unexpected temporal decline in automation-dependence was also observed. We will discuss the central issue for this research: Is this vigilance as we know it from laboratory studies and the Warm et al. (1996) resource theory? We will also consider the relationship between vigilance decrement and changes in automation dependence, as well as practical implications and study limitations.

Sources of Vigilance Decrement in Multi-UAS Operation

Warm's resource theory (Warm et al., 1996; Warm, Finomore, Vidulich, & Funke, 2015; Warm, Parasuraman, & Matthews, 2008) describes several signature features of the demanding vigilance tasks that typically show performance decrements (See et al., 1995). The task elicits high workload, elevated distress, and reduced task engagement (H1), and declines in detection rate are moderated by task demands (H2). The current data provide partial support for H1, although distress did not increase. Patterns of temporal change in accuracy suggested some instability in performance in the initial time blocks. In the second hour of the task, temporal decline in accuracy was more evident for the more demanding of the two surveillance tasks (WR), and for this task, decrement was more pronounced with low-automation reliability. Low reliability produced higher workload and might elicit greater sensitivity of performance to loss of resources. Thus, H2 was supported for WR, but performance on the IA task was better sustained over time.

However, the time course for loss of vigilance on WR differed from standard vigilance decrement. Typically, most decrement occurs during the first 20 to 30 min of the session (See et al., 1995; Warm, Parasuraman, & Matthews, 2008). By contrast, in the low-reliability condition, WR accuracy was fairly stable for the first 45 min, declined progressively for the next hour, and recovered sharply in the final 15 min. In the high-reliability condition, there was a less pronounced curvilinear trend of this kind. Recovery of vigilance toward the end of the task or "end effect" (Bergum & Lehr, 1963) is seen when participants can anticipate the end of the task, resulting from motivational factors (Oken, Salinsky, & Elsas, 2006). In ALOA, participants can track the progress of vehicles to an end point on the map display, which may have encouraged a burst of effort in the final block. However, the delayed onset of vigilance decrement is more unusual and suggests that over shorter durations, UAS surveillance tasks may be less vulnerable to decrement than laboratory studies would suggest.

Subjective state data also differed somewhat from the typical response to vigilance. A large-magnitude decline in task engagement was observed, characteristic of vigilance (Matthews Szalma, Panganiban, Neubauer, & Warm, 2013) and observed in another UAS simulation study (Guznov, Matthews, Warm, & Pfahler, 2017). However, the NASA-TLX workload of 35.8 was well below the scale midpoint of 50, similar to Gunn et al.'s (2005) study requiring detection of warning signals in a UAS context. The stability of distress across the trial was also unexpected, given previous studies on the stress of vigilance (Warm, Matthews, & Finomore, 2008). For example, a vigilance study requiring detection of changes in symbology on a map display (Matthews, Warm, Shaw, & Finomore, 2014) found that distress increased by nearly 1 standard deviation during a 1-hr vigil.

The relatively low workload of the present tasks may have mitigated distress, producing a passive fatigue state similar to those observed in studies of automated car driving (Matthews et al., 2018; Saxby et al., 2013). The multi-UAS task required greater cognitive activity than the vehicle driving tasks, due in part to the multiple task demands. However, the declines in task engagement were of similar or greater magnitude to those observed in the automated driving studies (e.g., Saxby et al., 2013), without any concomitant increase in distress. Thus, passive fatigue may be an operational issue even when operators are required to perform routine task activities, given sufficient task duration.

The modest workload of the task raises the question of why any vigilance decrement should be found, given that the resource model (Warm et al., 1996) attributes resource-depletion to high, prolonged workload. A possible explanation for both the workload finding and delayed onset of vigilance decrement is that although workload was generally low, it was still sufficient to slowly deplete resources. Despite the low event rate for the surveillance tasks, the impact of low reliability on workload implies that processing these tasks contributed to workload, and resources may have depleted faster under low reliability. After an hour or so, resources were sufficiently low for more resource-demanding tasks (i.e., WR) to show performance deficits. In low-workload, passive fatigue states, variation in task-directed effort may influence performance separately from resource variation (Matthews & Desmond, 2002). The increasing neglect of tasks in the second half of the trial also suggests loss of motivation, which may have accentuated resource-depletion effects on WR during the period from 45 to 105 min. Increased motivation associated with the end effect was sufficient to compensate for the resource shortfall in the final task period.

Dependence on Automation

System reliability influenced automation-dependence as expected (H3). However, the anticipated increase in dependence as operators became more fatigued, and therefore effort-minimizing (Sauer et al., 2003) was not observed, contrary to H4. There was no clear trend for the IA task, and WR dependence decreased systematically over time, especially in the low-reliability condition. Relative to automation accuracy, participants tended to be overdependent initially, and underdependent by the end of the task. Decreasing automation dependence may have contributed to the observed WR vigilance decrement, but the two temporal plots do not closely correspond.

In Parasuraman and Manzey's (2010) account, high reliability promotes dependence, so it is surprising that WR dependence also declined in this condition. Relying on the automation to handle this difficult task would seem an effective strategy for energy conservation. A possible explanation is that evaluating the performance of the automation is itself seen as an additional task. As fatigue develops, participants sought to reduce demands by increasingly ignoring the automation, consistent with evidence on task-shedding under fatigue (Russo et al., 2004), even though this strategy was counterproductive.

Individual Differences in Stress Response and Performance

Neuroticism and low conscientiousness predicted distress, similar to Shaw et al. (2010) and other studies (Matthews et al., 2013), providing partial support for H5. Contrary to the hypothesis, neither trait was associated with task engagement. Lin et al. (2018) found that conscientiousness was associated with engagement only under high workload, which may moderate personality effects. Extraversion, which correlates with boredom-proneness (Hunter, Abraham, Hunter, Goldberg, & Eastwood, 2016), was the only trait predictor of (low) task engagement. However, there were no significant associations between traits and performance.

Task engagement is a reliable correlate of detection accuracy in standard vigilance paradigms (Matthews et al., 2010, 2014). Here, engagement was associated only with the easier of the two tasks (Imaging), contrary to prediction from resource theory (H6). Engagement was also associated with lower task neglect, as in a previous study (Lin et al., 2018), suggesting a motivational mechanism. However, state factors were unrelated to automation-dependence, contrary to H7. The regression analysis showed a positive association between distress and WR detection accuracy under high reliability, but a negative one when reliability was low. Distress was also associated with greater neglect, irrespective of reliability. Negative relationships between distress and attention have been observed in other multitasking environments (Matthews & Campbell, 2010; Matthews, Reinerman-Jones, et al., 2017) and may reflect disruptive effects of the state on the executive processing necessary to coordinate multiple tasks. Similarly, Eysenck and Derakshan's (2011) Attentional Control Theory acknowledges that anxiety and stress may enhance performance on easy tasks due to compensatory effort.

Practical Implications

In their review of vigilance for *Human Factors*, Warm, Parasuraman, and Matthews (2008, p. 438) identified applications in "the design of work environments involving vigilance functions and in the evaluation of those who carry out such functions." From a design standpoint, contemporary vigilance research highlights workload mitigation, but the current findings implicate the need to support operator motivation during extended missions. Hancock (2013) discusses design principles that support operator engagement (see also Szalma, 2014); interface design might also be individualized according to operator personality (Szalma, 2009). This study also identified a specific issue of progressive underutilization of automation, implying that interface design should increase the salience of recommendations, along with training.

The findings also highlight applications for operator evaluation. They do not support selecting operators on the basis of personality, although they identify possible attrition risks for individuals who find the task especially distressing or fatiguing. The link between high distress and poor performance in demanding task configurations is significant because distress also correlated with lower accuracy in a high-workload version of ALOA (Lin et al., 2018). Diagnostic monitoring for stress and fatigue that could drive adaptive automation may be of value (Kidwell et al., 2012), although valid psychological markers for operator stress need to be determined (Kamzanova, Kustubayeva, & Matthews, 2014).

Limitations

An obvious limitation is the use of untrained civilian participants rather than Air Force personnel, which was necessary for an adequate sample size. Similarly, a simulation cannot fully capture operational conditions and demands, although in this case simulation is necessary because of the absence of fielded multivehicle systems. Future work should test whether the current findings generalize across different task configurations, LOAs, and mission demands. It is also challenging to test resource theory predictions using complex tasks such as ALOA, which include multiple subtasks, imposing somewhat differing cognitive demands. It is possible that vigilance effects on surveillance tasks would be influenced by the nature of concurrent multitasking. The automation used here was also limited in nature and does not afford investigation of key factors such as transparency (Lyons, 2013). Due to scenario constraints, there was some variability in the frequency of events and automation reliability from block to block, which may have added statistical noise to the data. In addition, the present dependence metric is not informative about how participants arrived at their response decision on each trial (Rice & Geels, 2010). Further theoretical and empirical efforts are necessary to understand sources of changes in dependence on automation. Finally, while the current focus was on task-induced fatigue, operational settings introduce additional stressors, including sleep quality and long hours of work (Ouma et al., 2011), which may impinge differently on performance.

CONCLUSION

The current study confirmed that a 2-hr simulation of multi-UAS operation produced vigilance decrement and loss of task engagement, together with increasing underuse of automation. The Warm et al. (1996) resource theory does not fully explain the effects observed, and future work in this domain should focus on changes in sustained attention during multicomponent task performance, as well as on motivational and individual difference factors. Improved understanding of temporal change in cognitive, motivational, and affective processes will support better strategies for interface design and operator support.

ACKNOWLEDGMENT

This research was sponsored by AFOSR A9550-13-1-0016 and 13RH05COR. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of AFOSR or the U.S. Government.

KEY POINTS

- Operators of future multiple Unmanned Aerial Systems (UAS) may be vulnerable to fatigue and loss of vigilance.
- A multi-UAS simulation study was conducted that induced fatigue and tested for changes in vigilance and dependence on automation.
- Vigilance decrement was observed during performance of surveillance tasks, accompanied by decreasing dependence on automation; these findings cannot be fully explained by the resource theory of Joel Warm and others.
- The participant's level of distress correlated with poorer performance, especially when automation reliability was low.

REFERENCES

Barg-Walkow, L. H., & Rogers, W. A. (2016). The effect of incorrect reliability information on expectations, perceptions, and use of automation. *Human Factors*, *58*(2), 242–260.

Bergum, B. O., & Lehr, D. J. (1963). End spurt in vigilance. *Journal of Experimental Psychology*, *66*(4), 383-385.

Calhoun, G. L., Goodrich, M. A., Dougherty, J. R., & Adams, J. A. (2016). Human-autonomy collaboration and coordination toward multi-RPA missions. In N. J. Cooke, L. J. Rowe, W. Bennett, & D. Q. Joralmon (Eds.), *Remotely piloted aircraft systems: A human systems integration perspective* (pp. 109–136). Chichester, UK: Wiley.

Calhoun, G. L., Ruff, H. A., Draper, M. H., & Wright, E. J. (2011). Automation-level transference effects in simulated multiple unmanned aerial vehicle control. *Journal of Cognitive Engineering and Decision Making*, *5*(1), 55–82. https://doi.org/10.1177/1555343411399069

Casner, S. M., & Schooler, J. W. (2015). Vigilance impossible: Diligence, distraction, and daydreaming all lead to failures in a practical monitoring task. *Consciousness and Cognition*, *35*, 33–41. https://doi.org/10.1016/j.concog.2015.04.019

Chappelle, W. L., Salinas, A., & McDonald, K. (2011). *Psychological health screening of remotely piloted aircraft (RPA) operators and supporting units* (No. ADA582856). Wright-Patterson Air Force Base, OH: USAF School of Aerospace Medicine Department of Neuropsychiatry. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA582856

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Hillsdale, NJ: Erlbaum.

Cummings, M. L., Mastracchio, C., Thornburg, K. M., & Mkrtchyan, A. (2013). Boredom and distraction in multiple unmanned vehicle supervisory control. *Interacting with Computers*, *25*(1), 34–47.

Desmond, P. A., & Hancock, P. A. (2001). Active and passive fatigue states. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload, and fatigue* (pp. 455–465). Mahwah, NJ: Erlbaum.

Donald, F. M. (2008). The classification of vigilance tasks in the real world. *Ergonomics*, *51*(11), 1643–1655. https://doi.org/10.1080/00140130802327219

Eysenck, M. W., & Derakshan, N. (2011). New perspectives in attentional control theory. *Personality and Individual Differences*, *50*(7), 955–960. https://doi.org/10.1016/j.paid.2010.08.019

Finomore, V., Matthews, G., Shaw, T., & Warm, J. (2009). Predicting vigilance: A fresh look at an old problem. *Ergonomics*, *52*(7), 791–808. https://doi.org/10.1080/00140130802641627

Fraulini, N. W., Hancock, G. M., Neigel, A. R., Claypoole, V. L., & Szalma, J. L. (2017). A critical examination of the research and theoretical underpinnings discussed in Thomson, Besner, and Smilek (2016). *Psychological Review*, *124*(4), 525–531. https://doi.org/10.1037/rev0000066

Gunn, D. V., Warm, J. S., Nelson, W. T., Bolia, R. S., Schumsky, D. A., & Corcoran, K. J. (2005). Target acquisition with UAVs: Vigilance displays and advanced cuing interfaces. *Human Factors*, *47*(3), 488–497.

Guznov, S., Matthews, G., Warm, J. S., & Pfahler, M. (2017). Training techniques for visual search in complex task environments. *Human Factors*, *59*(7), 1139–1152.

Hancock, P. A. (2013). In search of vigilance: The problem of iatrogenically created psychological phenomena. *American Psychologist*, *68*(2), 97–109. https://doi.org/10.1037/a0030214

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in psychology* (Vol. 52, pp. 139–183). Amsterdam: Elsevier.

Hendy, K. C., Hamilton, K. M., & Landry, L. N. (1993). Measuring subjective workload: When is one scale better than many? *Human Factors*, *35*(4), 579–601. https://doi.org/10.1177/001872089303500401

Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*, *45*, 73–93. https://doi.org/10.1016/S0301-0511(96)05223-4

Hunter, J. A., Abraham, E. H., Hunter, A. G., Goldberg, L. C., & Eastwood, J. D. (2016). Personality and boredom proneness in the prediction of creativity and curiosity. *Thinking Skills and Creativity*, *22*, 48–57. https://doi.org/10.1016/j.tsc.2016.08.002

Johnson, R., Leen, M., & Goldberg, D. (2007). *Testing adaptive levels of automation (ALOA) for UAV supervisory control*. Whittier, CA: OR Concepts Applied. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA470315

Kamzanova, A. T., Kustubayeva, A. M., & Matthews, G. (2014). Use of EEG workload indices for diagnostic monitoring of vigilance decrement. *Human Factors*, *56*(6), 1136–1149.

Kidwell, B., Calhoun, G. L., Ruff, H. A., & Parasuraman, R. (2012). Adaptable and adaptive automation for supervisory control of multiple autonomous vehicles. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 428–432. https://doi.org/10.1177/1071181312561096

Lin, J., Matthews, G., Wohleber, R. W., Szalma, J. L., Ruff, H. A., Calhoun, G. L., & Funke, G. J. (2018). *Cognitive overload, stress and automation utilization: A simulation study of multiple Unmanned Aerial Systems (UAS) operation*. Manuscript submitted for publication.

Lyons, J. B. (2013). Being transparent about transparency: A model for human-robot interaction. In *2013 AAAI Spring Symposium Series* (pp. 48–53). Retrieved from http://www.aaai.org/ocs/index.php/SSS/SSS13/paper/view/5712

Mackie, R. R. (1987). Vigilance research—are we ready for countermeasures? *Human Factors*, *29*(6), 707–723.

Matthews, G., & Campbell, S. E. (2010). Dynamic relationships between stress states and working memory. *Cognition & Emotion*, *24*(2), 357–373. https://doi.org/10.1080/02699930903378719

Matthews, G., Campbell, S. E., Falconer, S., Joyner, L. A., Huggins, J., Gilliland, K., . . . Warm, J. S. (2002). Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. *Emotion*, *2*(4), 315–340. https://doi.org/10.1037//1528-3542.2.4.315

Matthews, G., & Desmond, P. A. (2002). Task-induced fatigue states and simulated driving performance. *The Quarterly Journal of Experimental Psychology Section A*, *55*(2), 659–686. https://doi.org/10.1080/02724980143000505

Matthews, G., Neubauer, C., Saxby, D. J., Wohleber, R. W., & Lin, J. (2018). Dangerous intersections? A review of studies of fatigue and distraction in the automated vehicle. *Accident Analysis and Prevention*. Online publication: https://doi.org/10.1016/j.aap.2018.04.004

Matthews, G., Reinerman-Jones, L., Abich, J., & Kustubayeva, A. (2017). Metrics for individual differences in EEG response to cognitive workload: Optimizing performance prediction. *Personality and Individual Differences*, *118*, 22–28. https://doi.org/10.1016/j.paid.2017.03.002

Matthews, G., Szalma, J. L., Panganiban, A. R., Neubauer, C., & Warm, J. S. (2013). Profiling task stress with the Dundee Stress State Questionnaire. In L. Cavalcanti & S. Azevedo (Eds.), *Psychology of stress: New research* (pp. 49–91). Hauppauge, NY: Nova Science Publishers.

Matthews, G., Warm, J. S., Reinerman-Jones, L. E., Langheim, L. K., Washburn, D. A., & Tripp, L. (2010). Task engagement, cerebral blood flow velocity, and diagnostic monitoring for sustained attention. *Journal of Experimental Psychology: Applied*, *16*(2), 187–203. https://doi.org/10.1037/a0019572

Matthews, G., Warm, J. S., Shaw, T. H., & Finomore, V. S. (2014). Predicting battlefield vigilance: a multivariate approach to assessment of attentional resources. *Ergonomics*, *57*(6), 856–875. https://doi.org/10.1080/00140139.2014.899630

Matthews, G., Warm, J. S., & Smith, A. P. (2017). Task engagement and attentional resources: Multivariate models for individual differences and stress factors in vigilance. *Human Factors*, *59*(1), 44–61.

McCrae, R. R. (2009). The Five-Factor-Model of personality traits. In P. J. Corr & G. Matthews (Eds.), *The Cambridge handbook of personality psychology* (pp. 148–161). Cambridge: Cambridge University Press.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*(1), 105–125. https://doi.org/10.1037//1082-989X.7.1.105

Mouloua, M., Gilson, R., & Hancock, P. A. (2003). Human-centered design of unmanned aerial vehicles. *Ergonomics in Design*, *11*(1), 6–11.

Neubauer, C., Matthews, G., Langheim, L., & Saxby, D. (2012). Fatigue and voluntary utilization of automation in simulated driving. *Human Factors*, *54*(5), 734–746. https://doi.org/10.1177/0018720811423261

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, *7*(1), 44–64.

Oken, B. S., Salinsky, M. C., & Elsas, S. M. (2006). Vigilance, alertness, or sustained attention: physiological basis and measurement. *Clinical Neurophysiology*, *117*(9), 1885–1901. https://doi.org/10.1016/j.clinph.2006.01.017

Ouma, J. A., Chappelle, W. L., & Salinas, A. (2011). *Facets of occupational burnout among US Air Force active duty and National Guard/reserve MQ-1 predator and MQ-9 reaper operators*. DTIC Document. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA548103

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, *52*(3), 381–410. https://doi.org/10.1177/0018720810376055

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, *39*(2), 230–253.

Parasuraman, R., & Wickens, C. D. (2008). Humans: Still vital after all these years of automation. *Human Factors*, *50*(3), 511–520. https://doi.org/10.1518/001872008X312198

Parsons, K. S., Warm, J. S., Nelson, W. T., Matthews, G., & Riley, M. A. (2007). Detection-action linkage in vigilance: Effects on workload and stress. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *51*, 1291–1295.

Proctor, R. W., & Vu, K.-P. L. (2010). Cumulative knowledge and progress in human factors. *Annual Review of Psychology*, *61*(1), 623–651. https://doi.org/10.1146/annurev.psych.093008.100325

Rice, S., & Geels, K. (2010). Using system-wide trust theory to make predictions about dependence on four diagnostic aids. *The Journal of General Psychology*, *137*(4), 362–375. https://doi.org/10.1080/00221309.2010.499397

Russo, M. B., Sing, H., Santiago, S., Kendall, A. P., Johnson, D., Thorne, D., . . . Redmond, D. (2004). Visual neglect: occurrence and patterns in pilots in a simulated overnight flight. *Aviation, Space, and Environmental Medicine*, *75*(4), 323–332.

Saucier, G. (2002). Orthogonal markers for orthogonal factors: The case of the Big Five. *Journal of Research in Personality*, *36*(1), 1–31. https://doi.org/10.1006/jrpe.2001.2335

Sauer, J., Wastell, D. G., Robert, G., Hockey, G. R. J., & Earle, F. (2003). Performance in a complex multiple-task environment during a laboratory-based simulation of occasional night work. *Human Factors*, *45*(4), 657–669.

Saxby, D. J., Matthews, G., Warm, J. S., Hitchcock, E. M., & Neubauer, C. (2013). Active and passive fatigue in simulated driving: Discriminating styles of workload regulation and their safety impacts. *Journal of Experimental Psychology: Applied*, *19*(4), 287–300. https://doi.org/10.1037/a0034386

See, J. E., Howe, S. R., Warm, J. S., & Dember, W. N. (1995). Meta-analysis of the sensitivity decrement in vigilance. *Psychological Bulletin*, *117*(2), 230–249.

Shaw, T. H., Matthews, G., Warm, J. S., Finomore, V. S., Silverman, L., & Costa, P. T. (2010). Individual differences in vigilance: Personality, ability and states of stress. *Journal of Research in Personality*, *44*(3), 297–308. https://doi.org/10.1016/j.jrp.2010.02.007

Szalma, J. L. (2009). Individual differences in human–technology interaction: incorporating variation in human characteristics into human factors and ergonomics research and design.

Theoretical Issues in Ergonomics Science, *10*(5), 381–397. https://doi.org/10.1080/14639220902893613

Szalma, J. L. (2014). On the application of motivation theory to human factors/ergonomics: Motivational design principles for human–technology interaction. *Human Factors*, *56*(8), 1453–1471.

Szalma, J. L., & Taylor, G. S. (2011). Individual differences in response to automation: The five factor model of personality. *Journal of Experimental Psychology: Applied*, *17*(2), 71–96. https://doi.org/10.1037/a0024170

Temple, J. G., Warm, J. S., Dember, W. N., Jones, K. S., LaGrange, C. M., & Matthews, G. (2000). The effects of signal salience and caffeine on performance, workload, and stress in an abbreviated vigilance task. *Human Factors*, *42*(2), 183–194.

Thomson, D. R., Besner, D., & Smilek, D. (2016). A critical examination of the evidence for sensitivity loss in modern vigilance tasks. *Psychological Review*, *123*(1), 70–83. https://doi.org/10.1037/rev0000021

Tvaryanas, A. P., Lopez, N., Hickey, P., DaLuz, C., Thompson, W. T., & Caldwell, J. L. (2006). *Effects of shift work and sustained operations: Operator performance in remotely piloted aircraft (OP-REPAIR)* (Final No. HSW-PE-BR-TR-2006-0001). Brooks City-Base. TX: United States Air Force 331st Human Systems Wing. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA443145

Tvaryanas, A. P., & Macpherson, G. D. (2009). Fatigue in pilots of remotely piloted aircraft before and after shift work adjustment. *Aviation, Space, and Environmental Medicine*, *80*(5), 454–461. https://doi.org/10.3357/ASEM.2455.2009

Warm, J. S., & Dember, W. N. (1998). Tests of vigilance taxonomy. In R. R. Hoffman, M. F. Sherrick, & J. S. Warm (Eds.), *Viewing psychology as a whole: The integrative science of William N. Dember* (pp. 87–112). Washington, DC: American Psychological Association. Retrieved from http://content.apa.org/books/10290-004

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., Finomore, V. S., Vidulich, M. A., & Funke, M. E. (2015). Vigilance: A perceptual challenge. In R. R. Hoffman (Ed.), *The Cambridge handbook of applied perception research* (Vol. 1, pp. 241–283). New York, NY: Cambridge University Press.

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, UK: Wiley.

Warm, J. S., Matthews, G., & Finomore, V. S. (2008). Vigilance, workload, and stress. In P. A. Hancock & J. L. Szalma (Eds.), *Performance under stress* (pp. 115–141). Burlington, VT: Ashgate.

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors*, *50*(3), 433–441. https://doi.org/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., & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science*, *8*(3), 201–212.

Wohleber, R. W., Calhoun, G. L., Funke, G. J., Ruff, H., Chiu, C.-Y. P., Lin, J., & Matthews, G. (2016). The impact of automation reliability and operator fatigue on performance and reliance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*(1), 211–215. https://doi.org/10.1177/1541931213601047

Ryan W. Wohleber is a research associate in human factors psychology at the Institute for Simulation and Training, University of Central Florida. He received his MA in experimental psychology from University of Cincinnati and his PhD in modeling and simulation from University of Central Florida.

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

Jinchao Lin is a postdoctoral researcher at the Institute of Simulation and Training, University of Central Florida. He received his PhD in modeling and simulation from the University of Central Florida.

James L. Szalma is an associate professor of psychology at the University of Central Florida. He earned his PhD from the University of Cincinnati in 1999.

Gloria L. Calhoun is a senior engineering research psychologist with the Air Force Research Laboratory's 711th Human Performance Wing. She received an MA in human performance from Wright State University in 1984.

Gregory J. Funke is a research psychologist in the Applied Neuroscience Branch of the U.S. Air Force Research Laboratory's Human Effectiveness Directorate at Wright-Patterson Air Force Base in Dayton, Ohio. He received his PhD in human factors psychology from the University of Cincinnati in 2008.

C.-Y. Peter Chiu is an Associate Professor in Psychology at the University of Cincinnati. He received his PhD in cognitive neuroscience from Harvard University in 1997.

Heath A. Ruff is a Human Factors Specialist at Infoscitex. He received his MS in engineering from Wright State University in 2000.

Date received: February 16, 2018
Date accepted: July 18, 2018