Using Situation Awareness and Workload to Predict Performance in Submarine Track Management: A Multilevel Approach

Shayne Loft, **Lisa Jooste**, **Yanqi Ryan Li**, University of Western Australia, Perth, **Timothy Ballard**, University of Queensland, Brisbane, Australia, **Samuel Huf**, Defence Science and Technology Group, Australia, **Ottmar V. Lipp**, Curtin University, Perth, and **Troy A. W. Visser**, University of Western Australia

Objective: Examine the extent to which subjective workload and situation awareness (SA) can predict variance in performance at the between- and within-person levels of analysis in a simulated submarine track management task.

Background: SA and workload are crucial constructs in human factors that are conceptualized as states that change within individuals over time. Thus, a change in an individual's subjective workload or SA over the course of performing a task should be predictive of their subsequent performance (within-person effects). However, there is little empirical evidence for this.

Method: Participants monitored displays to track the behaviors of contacts in relationship to their own ship (Ownship) and landmarks. The Situational Awareness Global Assessment Technique measured SA, and the Air Traffic Workload Input Technique measured subjective workload.

Results: When a participant's subjective workload rating increased, their subsequent performance decreased, but there was no evidence for within-person effects of SA on performance. We replicated prior between-person level effects of SA; participants with higher SA performed better than those with lower SA.

Conclusion: Change in an individual's subjective workload rating (but not SA) was predictive of their subsequent performance. Because an increase in SA should increase the extent to which operators hold the knowledge required to perform subsequent tasks, further research is required to examine SA effects on performance at the within-person level.

Application: Adapting automation is more likely to produce optimal outcomes if based on measurement of operator states that predict future task performance, such as workload.

Keywords: situation awareness, workload, submarine track management, multilevel modeling

Address correspondence to Shayne Loft, The University of Western Australia, 35 Stirling Highway, Crawley, Perth, Western Australia 6009, Australia; e-mail: shayne.loft@uwa.edu.au.

HUMAN FACTORS
Vol. 60, No. 7, November 2018, pp. 978–991
DOI: 10.1177/0018720818784803

The evaluation of operator mental workload and situation awareness (SA) has been the subject of extensive scientific inquiry and has greatly informed our understanding of human performance in complex work systems as diverse as air traffic control, driving, unmanned vehicle control, and piloting (for reviews, see Vu & Chiappe, 2015; Young, Brookhuis, Wickens, & Hancock, 2015). *Workload* is the term used to describe the relationship between task demands and available operator mental capacity (Parasuraman, Sheridan, & Wickens, 2008). SA refers to an operator's understanding of the relevant elements of his or her task environment and how these elements might change as a consequence of external factors or control actions (Durso & Dattel, 2004; Endsley, 1995b). Workload and SA are arguably the two most important constructs in the human factors literature for understanding human performance (Vidulich & Tsang, 2012).

The efficient and safe operation of complex transportation, military, and production systems would not be possible without effective training programs, well-designed displays, and task automation. Prior studies examining the impact of such interventions on workload, SA, and performance have used cross-sectional designs that aggregate repeated measurements from individuals and conduct between-person level analyses. This level of analysis is entirely appropriate for assessing the impact of design interventions, for example, demonstrating that individuals who undertake training have higher SA, reduced workload, and improved performance compared to individuals provided alternative or no training.

However, much less work has examined whether a change in an individual's workload or

SA predicts their subsequent performance (within-person effects). Thus, a significant research gap exists in that the prevailing between-person level analytical techniques aggregate repeated measurements from individuals, while the constructs of workload and SA are conceptualized as operator states that change within individuals over time (Funke, Knott, Salas, Pavlas, & Strang, 2012; Helton, Funke, & Knott, 2014). Further, it is the within-person level of analysis that enables change in an operator's performance to be predicted, and it is often crucial to do so. For example, while task automation often improves system performance, it can lead to impaired operator SA and return-to-manual performance deficits (Chen, Visser, Huf, & Loft, 2017; Onnasch, Wickens, Li, & Manzey, 2014). Such costs from automation use may be overcome by more selectively engaging automation when it is most required. The schedule for engaging automation or other task adaptations made in real time are more likely to be optimal if based on measurement of operator states (SA, workload) that can predict future task performance (Hancock et al., 2013).

In the current study, we used a multilevel approach to examine the extent to which workload and SA could predict variance in task performance at the between- and within-person levels of analysis. We used a simulated submarine track management task that required individuals to interpret a tactical picture of the position and behavior of contacts in relation to their own ship (Ownship) and landmarks. The task is similar to a range of work contexts that involve displaying abstract features of dynamic situations occurring outside of the operator's physical perceptual experience, such as air traffic control and unmanned vehicle control. In these work settings, there is an increasing need to be able to predict variations in future operator performance to facilitate human-machine integration.

PREDICTING BETWEEN-PERSON VARIATION IN PERFORMANCE

From a cognitive-energetic perspective, an individual's resource capacity can be conceptualized as a finite quantity of processing units available for performance (Gopher & Donchin, 1986). Subjective workload ratings reflect the relationship between perceived task demands and the operator's self-appraisal of available resources (Helton, Matthews, & Warm, 2009). More broadly, analysis of workload at the between-person level represents differences between individuals in resource capacity and the ability to self-regulate these resources as a function of task demands (Hockey, 1997; Humphreys & Revelle, 1984). Analysis of SA at the between-person level represents differences between individuals in their current understanding of the task environment and how it might change.

The relationship between workload, SA, and performance is complex and shaped by a variety of exogenous and endogenous factors. Higher average subjective workload can indicate that, compared to others, an individual's capacity is exceeded, which can degrade performance. In other circumstances, high workload can predict better performance (e.g., due to increased effort), or low workload can predict better performance (e.g., easy tasks). High SA can improve performance, but under some circumstances (e.g., easy tasks), there may be no relationship. These points notwithstanding, in our prior submarine track management studies, we have found that participants with higher workload or lower SA perform more poorly than those with lower workload or higher SA (Loft et al., 2015, 2016; Loft, Morrell, & Huf, 2013).

PREDICTING WITHIN-PERSON VARIATION IN PERFORMANCE

Changes in an individual's subjective workload or SA should predict subsequent performance. Even an operator who has low overall workload and high SA (compared to other operators) will exhibit temporal variation in their self-regulation (Hockey, 1986, 1997), and therefore, their workload, SA, and performance should vary over time. Changes in subjective workload and SA should be indicative of an operator's capacity to cope with task demands and thus should be predictive of that operator's subsequent performance.

To our knowledge, only one study (Mracek, Arsenault, Day, Hardy, & Terry, 2014) has examined the relationship between workload and performance at the within-person level. Participants

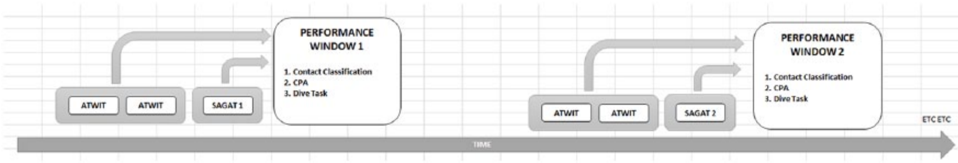

Figure 1. Conceptual illustration of the experimental design.

completed a command and control peacekeeping task. At the within-person level, Mracek et al. (2014) found some evidence that if an individual's subjective workload increased, their performance subsequently decreased, but this relationship was not consistently observed. In addition, to our knowledge, only one prior study has examined the relationship between SA and performance at the within-person level. Endsley (1990) found that increased SA, as measured by the Situational Awareness Global Assessment Technique (SAGAT), positively predicted expert pilots' subsequent performance. The lack of other research surprised us given the extensive number of SA papers published. Theories of SA, whether they assume that SA reflects a stored situation model (Endsley, 1995b) or knowledge of where goal-relevant information is located (Chiappe et al. 2016), predict that variations in the extent to which an operator has SA should influence whether that operator has the knowledge to make subsequent task decisions.

THE CURRENT STUDY

The objective was to examine whether subjective workload and SA predicted performance at the between-person and within-person levels of analysis. The submarine track management task required participants to monitor two displays. The left display presented a tactical plot of landmarks, contacts, and Ownship. The right display presented a time-bearing plot, representing the directional bearing of each contact on the tactical display in relation to Ownship and how those bearings change with time. The classification task required participants to judge how long each contact spent inside certain geospatial boundaries, which defined contacts as friendly, trawler, and so forth. The closest point of approach (CPA) task required participants to mark the closest point of approach of contacts to Ownship. The dive task required participants to "dive" the submarine when all contacts on the tactical display were heading in the same direction and at least one contact was heading toward Ownship.

SA was measured using SAGAT (Endsley, 1995a), which is the SA measure most predictive of track management performance at the between-persons level in previous studies (e.g., Loft et al., 2015). SAGAT involved pausing and blanking the task displays six times in each scenario in order to query participants about their SA for the information related to making subsequent classification, CPA, and dive task decisions. The Air Traffic Workload Input Technique (ATWIT; Stein, 1985) measured subjective workload by prompting participants to rate their workload between 1 and 10 once a minute. After each scenario, participants completed the National Aeronautics and Space Administration's Task Load Index (NASA-TLX; Hart & Staveland, 1988). We expected positive correlations between ATWIT and the NASA-TLX, providing evidence for convergent validity.

Figure 1 conceptually illustrates how our design enabled us to examine whether subjective workload and SA predicted performance at the within- and between-person levels. A performance window consisted of a number of contacts that required a classification, CPA, or dive task decision, and there were several performance windows per scenario. Each SAGAT freeze queried information about specific contacts, relating to task decisions that had to be made in the subsequent "performance window." We analyzed the two ATWIT probes preceding each performance window. This design allowed us to determine whether an increase in subjective workload or a decrease in SA for a participant at a specific time (i.e., an increase/decrease in workload/SA relative to their own average workload/SA across all of the performance windows) was associated with a decrease in performance for that participant in the subsequent performance window (i.e., a decrease in performance relative to their own average performance across all of the performance windows).

At the between-person level, we aggregated the repeated measurements of workload, SA, and performance across all the performance windows and examined whether participants with higher average workload or lower average SA had poorer performance compared to those with lower average workload or higher average SA.

METHOD

Participants

Participants were 59 (33 females) psychology students (M = 23.2 years, SD = 4.48) who volunteered for course credit or were reimbursed $40. This research complied with the American Psychological Association Code of Ethics and was approved by the Human Research Ethics Office at the University of Western Australia. Informed consent was obtained from each participant.

Simulated Submarine Track Management Task

The simulated submarine track management task is presented in Figure 2. The number of contacts increased (peaking at eight) and decreased (plateauing at two) three times during each 27.5-minute scenario.

Contact classification. Contacts could be classified after they had spent more than two continuous minutes within a specific area of the tactical display. A contact was a *merchant* if it spent more than 2 min within the "shipping lane" denoted by the two parallel white lines. A contact was a *friendly* if it spent more than 2 min within the sectors bounded by blue lines. A contact was a *trawler* if it spent more than 2 min within "shallow" dark blue areas. A contact was a *hostile* if it spent less than one continuous minute in any classification area over 4 min. Participants could place horizontal lines on the time-bearing display when the contact entered an area. Once that line reached the relevant time indicator on the time-bearing display, the contact could be classified. Once classified, the contact and their sound track changed from yellow to green triangles (friendly), white triangles (merchant), blue squares (trawler), or red diamonds (hostile). The 24 contacts presented in a scenario could be classified once each.

Closest point of approach. A CPA was defined as the point at which a contact was at its closest to Ownship. This could be the point where a contact heading toward Ownship turned away from Ownship or when a contact traveled past Ownship. Participants indicated the time that the CPA occurred by marking the sound track of that contact on the time-bearing display. The 24 contacts presented in each scenario had one CPA each.

Dive. Participants were instructed to "dive" the submarine when all contacts were heading in the same direction and at least one contact was heading directly toward Ownship. There were nine dive windows in each scenario, and each window lasted between 10 s and 30 s. Participants clicked the Dive button to signal the Ownship to dive and received feedback, namely, "Dive successful" or "Dive unnecessary."

Measures

Performance windows. The six performance windows in each scenario consisted of a number of contacts that required a classification, CPA, or dive decision. The average duration of each performance window was 102.7 s (SD = 14.2 s, minimum = 82 s, maximum = 129 s), and over the duration of each performance window, there were between four and eight contacts on the display. There were three or four contact classifications, two or three CPAs, and either one or no dive decisions to be made during each performance window.

All contact classifications were timed to occur during performance windows. In contrast, 10 or 11 CPAs in each scenario (out of the total of 24 CPAs presented) occurred outside performance windows, and four dive task conditions met in each scenario (out of the total nine presented) were presented outside performance windows. Timing some task events to occur outside performance windows ensured that participants expected their performance was being continually assessed. When using multilevel

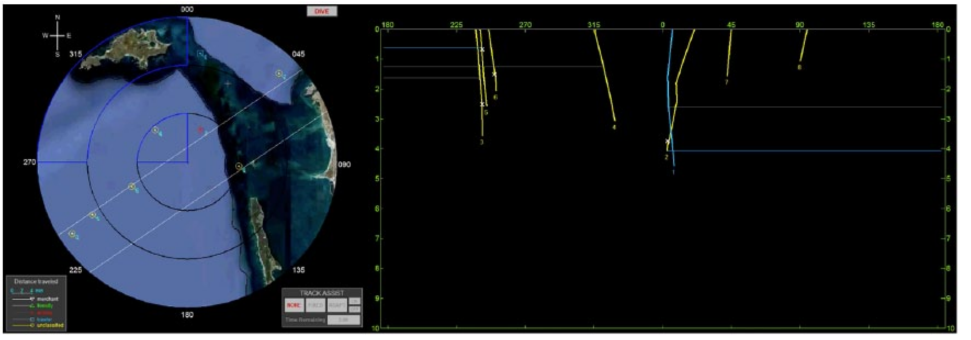

Figure 2. The submarine track management task. The left tactical display represents the area of operations, with Ownship located at the center. On the tactical display, the concentric range rings indicate the distance from Ownship. The range rings extend in 5 km increments. The parallel white lines indicate a shipping lane. The two friendly sectors are bounded by the blues lines. The fishing areas are darker in color, which depicts shallow waters. Note that the other two Australian coastal maps used different physical arrangements of these strategic zones. Eight contacts are displayed in Figure 2. The leader lines projecting from the center of each contact indicate the contact heading, and the contacts are numbered. The red diamond icon for Contact 2 indicates a contact that has been classified as an enemy. Contact 1 is a blue square because it has been classified as a trawler. Contacts represented by yellow circles are unclassified. The screen on the right is the sonar time-bearing display, which is a time-bearing plot of the history of each contact shown on the tactical display. The time-bearing display provides the bearing of contacts (horizontal axis) in relation to Ownship and indicated how those bearings changed with time (vertical axis). As shown, participants could place blue horizontal lines on the waterfall display when the contact entered an area of interest. Contacts 2, 3, 5, and 6 have had a closest point of approach (CPA) marked on the time-bearing display. Participants clicked the Dive button to signal the Ownship to dive. When a contact abrupted out (was no longer detected/displayed), the track for that contact terminated from both of the displays.

modeling to analyze the between- and within-person relationships between workload, SA, and performance, we only used the performance data from within the performance windows.

Situation awareness (SAGAT). A SAGAT freeze was timed to occur before each of the six performance windows. The average time between a SAGAT and the subsequent performance window was 17.4 s (SD = 11.3 s, minimum = 6 s, maximum = 36 s). During each SAGAT freeze, contacts were removed from the tactical display, and the time-bearing display was blanked. Seven SAGAT queries were then sequentially presented on the tactical display, targeting the knowledge required for the classification, CPA, and dive task decisions that needed to be made in the next performance window. This allowed us to assess whether the quality of SA for contact behavior at one point in time predicted how well participants made decisions relating to those same contacts a short time later. The full list of SAGAT queries used and their relationship to the three tasks is presented in Table 1.

Subjective workload. ATWIT was presented on the tactical display, and participants had 10 seconds to rate their workload between 1 and 10, described as *very low* (1–2), *moderate* (3–5), *relatively high* (6–8), and *very high* (9–10). Participants were provided with anchors; very low = *can accomplish everything easily*; moderate = *can accomplish everything, but takes some effort*; relatively high = *can accomplish everything, but is difficult, and takes some effort*; very high = *extremely difficult to accomplish everything*. An ATWIT probe was presented every minute. The simulation did not pause during ATWIT administration. In analyzing the

TABLE 1: The Situation Awareness (SAGAT) Queries Presented for Each Freeze During the Three Scenarios

	Scenario		
Block	Cockatoo Island	Exmouth	Regnard
1	Place vessel 1 on the screen[a,b]	Place vessel 2 on the screen[a]	Place vessel 2 on the screen[a]
	Is vessel 3 moving closer towards you?[b,c]	Is vessel 5 moving closer towards you?[b,c]	Which vessel near bearing 315 is currently heading further away from you?[b,c]
	Which vessel near bearing 000 is currently in an unclassified zone?[a]	Is vessel 1 currently in a trawler zone?[a]	Which vessel is currently in a trawler zone?[a]
	Is vessel 3 currently in a merchant zone?[a]	Which vessel is currently in a merchant zone?[a]	Is vessel 2 currently in an unclassified zone?[a]
	Is vessel 4 currently in a trawler zone?[a]	Which vessel has been in a trawler zone for less than 1 minute?[a]	Is vessel 3 currently in a trawler zone?[a]
	Is the vessel at bearing 45 headed away from you?[b,c]	Is the vessel at bearing 225 headed away from you?[b,c]	Which vessel at bearing 270 is currently heading closer towards you?[b,c]
	Which vessel is currently facing directly towards you?[c]	Which vessel is currently facing directly towards you?[b,c]	How many vessels are currently facing in the same direction?[c]
2	Place vessel 7 on the screen[a]	Place vessel 7 on the screen[a]	Place vessel 7 on the screen[a,b]
	Which vessel at bearing 90 is in a friendly zone?[a]	Which vessel has been in the friendly zone for the shortest time?[a]	Is vessel 5 currently in a friendly zone?[a]
	Are any vessels facing directly towards you?[b]	Are any vessels facing directly towards you?[b]	How many vessels are currently facing in the same direction?[c]
	Is vessel 6 moving further away from you?[b]	Is vessel 4 moving further away from you?[b]	Is vessel 7 heading further away from you?[b,c]
	Is vessel 8 currently in a trawler zone?[a]	Is vessel 6 currently in a trawler zone?[a]	Is vessel 8 currently in a friendly zone?[a]
	Is vessel 6 currently in a merchant zone?[a]	Is vessel 5 currently in a merchant zone?[a]	Which vessel is currently in a merchant zone?[a]
	Which vessel near bearing 90 is heading closer towards you?[b]	Is vessel 6 moving further away from you?[b]	Is vessel 5 currently heading closer towards you?[b,c]
3	Place vessel 10 on the screen[a]	Place vessel 11 on the screen[a]	Place vessel 12 on the screen[a]
	Which vessel is currently in a friendly zone?[a]	Which vessel is currently in a friendly zone?[a]	Which vessel is currently in a merchant zone?[a]
	Is vessel 13 moving further away from you?[b,c]	Is vessel 10 moving further away from you?[b,c]	Is vessel 14 heading closer towards you?[b,c]
	How many vessels are currently facing the same direction?[c]	How many vessels are currently facing the same direction?[c]	Which vessel is currently facing directly towards you?[b,c]
	Identify any vessel which is currently in a merchant zone[a]	Has any vessel been in a merchant zone for more than 1 minute?[a]	Identify any vessel which is currently in a friendly zone[a]
	Which vessel near bearing 125 is in a merchant zone?[a]	Identify any vessel which is currently in an unclassified zone[a]	Identify any vessel which is currently in a trawler zone[a]
	Which vessel at bearing 000 is heading closer towards you?[b,c]	Which vessel at near bearing 225 is heading towards you?[b,c]	Is vessel 13 moving further away from you?[b,c]

(continued)

TABLE 1: (continued)

Block	Cockatoo Island	Exmouth	Regnard
		Scenario	
4	Place vessel 15 on the screen[a]	Place vessel 14 on the screen[a]	Place vessel 14 on the screen[a]
	How many vessels are currently in an unclassified zone?[a]	How many vessels are currently in an unclassified zone?[a]	How many vessels are currently in an unclassified zone?[a]
	How many vessels are currently facing towards you?[b,c]	How many vessels are currently facing towards you?[b,c]	How many vessels are currently facing towards you?[b]
	Is vessel 14 moving further away from you?[b,c]	Is vessel 15 moving further away from you?[b,c]	Is vessel 10 moving closer towards you?[b]
	Is vessel 13 currently in an unclassified zone?[a]	Identify any vessel which is currently in a trawler zone[a]	Identify any vessel which is currently in an unclassified zone[a]
	Which vessel near bearing 180 is in a merchant zone?[a]	Which vessel has been in a friendly zone for the shortest time?[a]	Is vessel 16 currently in a friendly zone?[a]
	Is vessel 16 moving towards you?[b,c]	Is vessel 13 moving towards you?[b,c]	Is vessel 12 currently moving further away from you?[b]
5	Place vessel 18 on the screen[a,b]	Place vessel 17 on the screen[a]	Place vessel 18 on the screen[a]
	Which vessel is currently in a trawler zone?[a]	How many vessels are currently in a friendly zone?[a]	Identify any vessel which is currently in a trawler zone[a]
	Is vessel 20 currently in an unclassified zone?[a]	Is vessel 20 currently in an unclassified zone?[a]	Is vessel 20 currently in an unclassified zone?[a]
	Is vessel 19 moving closer towards you?[b,c]	Is vessel 22 moving closer towards you?[b,c]	Is vessel 22 moving closer towards you?[b,c]
	Which vessel is facing directly towards you?[b,c]	Are any vessels are currently facing directly towards you?[b,c]	Which vessel is facing directly towards you?[b,c]
	How many vessels are currently in a friendly zone?[a]	How many vessels are currently in a merchant zone?[a]	Is vessel 19 currently in a friendly zone?[a]
	Is vessel 18 moving further away from you?[b,c]	Is vessel 20 moving further away from you?[b,c]	Is vessel 19 moving further away from you?[b,c]
6	Place vessel 24 on the screen[a]	Place vessel 20 on the screen[a]	Place vessel 21 on the screen[a,b]
	Are there any vessels currently facing the same direction?[c]	Are there any vessels currently facing the same direction?[c]	Are any vessels currently facing the same direction?[c]
	Is vessel 21 currently moving further away from you?[b,c]	Is vessel 21 currently moving further away from you?[b,c]	Is vessel 21 currently moving further away from you?[b,c]
	Is vessel 22 currently in a trawling zone?[a]	Is vessel 22 currently in a trawling zone?[a]	Which vessel is currently in a merchant zone?[a]
	Which vessel at bearing 90 is currently moving closer towards you?[b,c]	Which vessel at bearing 270 is currently moving closer towards you?[b,c]	Is vessel 23 currently moving closer towards you?[b,c]
	Which vessel is currently in a merchant zone?[a]	Identify any vessel which is currently in a friendly zone.[a]	Is vessel 23 currently in an unclassified zone?[a]
	Is vessel 24 currently in an unclassified zone?[a]	Which vessel is currently in a trawler zone?[a]	Is vessel 24 currently in a friendly zone?[a]

Note. Note that there were no dive windows presented during Cockatoo Block 2, Exmouth Block 2, or Regnard Block 4. SAGAT = Situational Awareness Global Assessment Technique.
[a]Classification task.
[b]Closest point of approach task.
[c]Dive task.

between- and within-person effects of ATWIT in our multilevel analysis, we used the two ATWIT probes presented immediately before each performance window (the last one being presented 10 seconds before the SAGAT freeze and the other 1 minute earlier than that). This meant that the first ATWIT was presented on average 87.4 s before the subsequent performance window and the second 27.4 s before the subsequent performance window.

After each scenario, participants completed the NASA-TLX, which was administered and scored as per the procedures outlined by Hart and Staveland (1988).

Procedure

On Day 1, participants were trained for 2 hr. The training began with a 40-min audio-visual presentation. This was followed by a 15-min narrated video, including a prerecorded demonstration of tasks being completed and completion of a 27.5-min practice scenario. On Day 2, participants first viewed a 15-min refresher presentation and then completed the three 27.5-min scenarios, each of which used a different coastal map in counterbalanced order.

RESULTS

Descriptive Statistics and Between-Person Correlations

The classification hit rate was the proportion of contacts correctly classified. The CPA hit rate was the proportion of CPAs correctly marked. The CPA was scored as correct if the cross the participant made on the time-bearing display was placed within a 3 mm radius of the actual CPA point. Otherwise, the cross was recorded as a CPA false alarm. Participants made several CPA false alarms (M = 10.6, SD = 6.6 per scenario) for which we needed to account. The exact number of contacts to which participants could have potentially made a CPA false alarm was indeterminable. We followed the Chen et al. (2017) reasoning that CPA false alarms would be more likely in response to a contact course change. The false alarm rate was estimated to be the number of false alarms divided by the total number of course changes (minus the actual number of CPAs). CPA performance was then calculated by subtracting the CPA false alarm rate from the CPA hit rate. Dive task false alarms were rare (<1 made per scenario), and we therefore used dive hit rate. Response times (RTs) for the classification, CPA, and dive task are based on correct decisions only.

The descriptive statistics and between-person correlations are presented in Table 2. For completeness and to allow comparison to prior work (Chen et al., 2017; Loft et al., 2015), the analysis in Table 2 includes all the data, including the CPAs and dives that occurred outside the performance windows, and all of the ATWIT ratings. As shown in Table 2, ATWIT was positively correlated with the NASA-TLX (convergent validity), and ATWIT was negatively correlated with SAGAT. SAGAT and the NASA-TLX were not related. There were several significant positive correlations between SAGAT and performance but no significant correlations between ATWIT and performance.

Multilevel Modeling

When using multilevel modeling to analyze the between-person and within-person relationships between workload/SA and performance, we only used the performance data from within each performance window, the SAGATs (which all preceded a performance window), and the two ATWIT ratings that preceded each performance window, thereby ensuring we used the same data set for the between- and within-person levels of analysis. We used multilevel modeling to control for the nested structure of the data. All the analyses (except for dive task accuracy) were conducted in R by running a series of linear mixed effects models via the *lme* function within the multilevel package (Bliese, 2006). Dive task accuracy within each performance window was binary: 0 or 1. We therefore conducted a logistic version of the multilevel regression by using a linear mixed effects model via the *glmer* function. The distribution of the residuals for the accuracy and RT data were normal except for dive task RT, which was positively skewed. We corrected dive task RT residuals by transforming dive RT using a log function.

The nested data had a two-level structure. The higher level was "participants" (between-person

TABLE 2: Descriptive Statistics and Between-Person Correlation Matrix for Performance, Situation Awareness, and Workload for the Entire Scenarios

Variable	*M*	*SD*	1	2	3	4	5	6	7	8
1. Classification accuracy	0.59	0.23								
2. Classification response time	35.83	12.38	−.28*							
3. CPA hit – false alarm rate	0.28	0.23	.43**	−.49**						
4. CPA response time	20.35	11.95	−.29*	.27*	−.45**					
5. Dive hit rate	0.77	0.19	.32*	−.39**	.52**	−.17				
6. Dive response time	8.98	3.80	−.32*	.36**	−.58**	.41**	−.77**			
7. SAGAT (situation awareness)	0.51	0.12	.31*	−.31*	.47**	−.09	.29*	−.25		
8. ATWIT (workload)	4.66	1.12	.01	.22	−.08	.07	−.23	.21	−.33**	
9. NASA-TLX	66.28	12.61	−.21	−.01	.09	−.18	−.04	.02	−.05	.28*

Note. Response time is in seconds. CPA = closest point of approach task; SAGAT = Situation Awareness Global Assessment Technique; ATWIT = Air Traffic Workload Input Technique; NASA-TLX = NASA Task Load Index.
*$p < .05$. **$p < .01$ (two-tailed).

level). Nested within participants was the "observation" (within-person) level. As a first step, we ran a "null model" for each variable to examine the proportion of the total variance in each variable that existed at each of the two levels. We specified the relevant variable of interest as the dependent variable and included the random effect of participant. In these models, the random effect of participant represents the variance at the between-person level, and the residual variance represents the variance at the within-person level. The percentage of between-person variance in ATWIT, SAGAT, and performance on the three tasks ranged from 8% to 51%, whereas the within-person variance ranged from 50% to 92%. The relatively high percentage of variance (i.e., >5%) at the between-person level means that ordinary least squares regression was inappropriate because the assumption of independence of observations was likely to be violated. Our multilevel analysis was therefore justified. We ran six multilevel models, one for each dependent variable (classification accuracy and RT, CPA accuracy and RT, and dive accuracy and RT).

Subjective workload. As indicated in Table 3, at the between-person level of analysis, subjective workload, as measured by ATWIT, was not a significant predictor of CPA or dive task accuracy or RT for any of the three tasks. However, there was a significant positive relationship between workload and classification accuracy: Unexpectedly, we found that participants with higher average subjective workload ratings made more accurate classification decisions than participants with lower average subjective workload ratings.

At the within-person level, subjective workload was a significant predictor of the accuracy of classification and dive task decisions. When a participant's workload increased (relative to their own average workload across all of the performance windows), their classification accuracy and dive task accuracy in the subsequent performance window decreased, and their dive task RT in the subsequent performance window increased (relative to their own average classification and dive task performance across all of the performance windows). However, we also found that when an individual's workload ratings increased, their classification task RT in the subsequent performance window decreased.

Situation awareness. At the between-person level, SA, as measured by SAGAT, was a significant predictor of classification, CPA and dive task accuracy, and correct classification RT. This indicates that participants with higher average SA performed more accurately on all three tasks and made faster classification decisions than participants with lower average SA.

TABLE 3: Hierarchical Linear Model: Between- and Within-Person Effects of Situation Awareness and Subjective Workload on Classification, CPA, and Dive Task Performance

	Classification		CPA		Dive	
	Accuracy	RT	Accuracy	RT	Accuracy	RT
Between-person effect						
SAGAT (SA)	0.63** (0.25)	–38.93** (15.17)	0.78** (0.21)	–25.33 (22.76)	3.39* (1.61)	–0.73 (0.58)
ATWIT (workload)	0.05* (0.02)	0.08 (1.43)	0.03 (0.02)	0.79 (2.17)	0.27 (0.16)	–0.06 (0.06)
Within-person effect						
SAGAT (SA)	0.04 (0.04)	–5.81 (3.76)	–0.02 (0.06)	2.11 (7.95)	–0.67 (0.55)	–0.17 (0.22)
ATWIT (workload)	–0.02** (0.005)	–1.60** (0.46)	–0.008 (0.007)	–1.08 (0.95)	–0.26** (0.07)	0.09** (0.03)

Note. Values represent the standardized parameter estimates, and parenthetical values indicate standard errors. CPA = closest point of approach task; SAGAT = Situation Awareness Global Assessment Technique; ATWIT = Air Traffic Workload Input Technique; RT = response time.
*$p < .05$. **$p < .01$ (two-tailed).

At the within-person level, we found no evidence that SA predicted either the accuracy or RT on any of the three tasks (smallest $p = .12$), indicating that variation in a participant's SA as measured by SAGAT was not predictive of their subsequent performance.

Our multilevel model used composite SAGAT and performance measures. This use of composites does not take into account the possibility that participants may have traded off the information that they attended to across tasks or across contacts (Endsley, 2000). For example, a participant could have had better SA for information relating to classification than for CPA or at a contact-by-contact level could have had better SA for Contact 3 than Contact 5. To examine these possibilities, we separated SAGAT queries by the task that they targeted (as specified in Table 1) and to the extent possible, also examined the relationship between SA and performance on a contact-by-contact basis. The latter analysis conceptually replicates Endsley (1990), who reported that participants were more likely to "kill an aircraft" when they had previously correctly reported the location of that aircraft.

A SAGAT query was deemed related to classification if the query asked for information about a contact that was related to classification (contact location, contact location relative to area, time spent in area) and that same contact had to be classified in the next performance window (see Table 1). The pattern of multilevel statistical significance reported in Table 3 for SA and workload did not change when we selectively regressed the classification-related SAGAT queries (instead of using the SAGAT composite measure) on classification accuracy and RT. On a contact-by-contact basis, classification accuracy and RT for contacts did not differ as a function of whether participants had previously answered the classification SAGAT query for that same specific contact correctly or not, ts < 1.

A SAGAT query was deemed related to CPA if the query asked for information about a contact that was related to a CPA decision (contact location, contact heading relative to Ownship) and that same contact had a CPA in the next performance window. The pattern of multilevel statistical significance reported in Table 3 for SA and workload did not change when we selectively regressed CPA-related SAGAT queries (instead of using the SAGAT composite measure) on CPA accuracy and RT. It was not possible to calculate a CPA false alarm rate on a contact-by-contact basis, so we used the CPA hit rate. The CPA hit rate ($p = .12$, but with the effect in opposite direction) and CPA RT ($t < 1$) for contacts did not differ as a function of whether participants had previously answered the CPA

SAGAT query for that same specific contact correctly or not.

A SAGAT query was marked as related to dive if the query asked for information that was related to dive decisions (contact heading relative to Ownship, relative headings of contacts) in the next performance window. The pattern of multilevel statistical significance reported in Table 3 for SA and workload did not change if we selectively regressed dive-related SAGAT queries (instead of using the SAGAT composite measure) on dive accuracy and RT. Dive decisions involved all contacts on the display and thus could not be statistically linked to SA for a specific contact. Nonetheless, there was no difference in dive accuracy or dive RT as a function of whether participants had previously correctly answered dive task SAGAT queries correct or not, ts < 1.

DISCUSSION

SA and workload are both conceptualized as operator states that change over time. Yet to our knowledge, only one study has examined the relationship between workload and performance (Mracek et al., 2014) or between SA and performance (Endsley, 1990) at the within-person level. In the current study, we report evidence that an increase/decrease in subjective workload within an individual predicted their subsequent performance. However, while we replicated previously reported between-person effects of SA on performance, there were no significant within-person effects of SA on performance.

Subjective Workload

Mracek et al. (2014) reported preliminary evidence of a within-person association between subjective workload and performance, but their results were inconsistent. In contrast, we found more robust evidence that subjective workload can predict subsequent performance within individuals. After a participant's subjective workload ratings increased (relative to their own average workload rating), their subsequent classification accuracy and dive task accuracy decreased, and their dive task RT increased (relative to their own average performance). However, workload was negatively associated with classification RT at the within-person level. The fact that increased workload within an individual led to faster subsequent classification decisions may reflect a perceived need to make faster (but as a result, less accurate) decisions because of perceived time pressure.

In contrast to the negative relationship between workload and classification accuracy found at the within-person level, we found a positive relationship between workload and classification accuracy at the between-person level. Thus, participants with higher average subjective workload made more accurate classifications compared to participants with lower average subjective workload. However, as previously noted, we also found that if a participant's subjective workload increased, the accuracy of their subsequent classifications decreased. Thus, although individuals who reported higher average workload compared to others made more accurate classifications, increased workload within an individual actually impaired their subsequent classification performance. This finding demonstrates that the relationship between workload and performance can change according to the level at which workload is measured (Yeo & Neal, 2004).

We have shown that changes in an individual's subjective workload provided statistically reliable indications of their subsequent performance. If an individual had a high workload (relative to their own average workload), their subsequent performance decreased (relative to their own average performance). Therefore, subjective workload is a potentially useful predictor of an operator's capacity to cope with future task demands and could be used as an automation trigger to proactively mitigate against operator performance risks. However, a more crucial prediction concerns when task demands will exceed an operator's capability to *adequately respond* to task demands. This requires measurement of minimum acceptable performance. Subjective workload in this study was not used to predict minimum acceptable performance but rather only the direction of change in an individual's performance at specific time points (relative to their own average performance across the entire task). In command and control settings, it can be difficult to derive measures of "operator effectiveness" (Hancock, 2007). This requires specification not just of the operator's mental processes but also of the environment and its interaction with these mental processes (Brunswick, 1943; Simon, 1956). For example, the accuracy of responses made by experts is often "satisficed" rather than optimized, particularly as a function of increased workload (Loft, Bolland, Humphreys, & Neal, 2009).

Note that a drawback of subjective workload is that it can only be measured at discrete times and can be disruptive to the operator. Future research could examine whether secondary task performance, which provides a continuous measure of residual capacity, can also predict performance at the within-person level.

Situation Awareness

The between-person effects of SA on performance replicate prior work in submarine track management (Loft et al., 2015) and other task domains (Jones & Endlsey, 2004; Manning, Mills, Fox, Pfleiderer, & Mogilka, 2002). However, we found no evidence that an increase or decrease in an individual's SA was related to their subsequent performance. This result is surprising given we found large between-person effects of SA on performance and that 76% to 77% of the variance that we observed in SA and performance, respectively, was situated within-persons (rather than between-persons).

A reviewer raised the possibility that by referencing contact numbers our SAGAT queries may have induced a memory demand unrelated to the SA required for performance. In air traffic control (ATC) for example, Endsley and Rodgers (1996) found that air traffic controllers have poor SA for aircraft call-signs. Unlike ATC, however, in track management, it is likely more important to be aware of contact numbers to make classification and CPA decisions. The common factor across the two displays that could be used to integrate information was contact number (see Figure 1). For example, for classification, participants placed lines on the time-bearing plot for the contact number that had entered the area of interest on the tactical display and then needed to keep track of that contact on both the tactical display (ensuring it did not leave the area) and the time-bearing plot (checking time spent in the area). Nonetheless, the impact of memory demands, particularly for the dive task in which awareness of contact number was arguably less critical, deserves further investigation.

Although there are many approaches to defining and measuring SA, all assume that SA reflects the degree to which an individual understands the relevant elements of the task environment and how these elements might change in the future (Vu & Chiappe, 2015). Therefore, an increase in SA within an individual should increase the extent to which that individual holds the knowledge required to make subsequent decisions. Endsley (1990) found some evidence for this, but it is difficult to draw comparisons between Endsley's aircraft tactical task and track management, and we will not attempt to do so here. Clearly, fluctuations in SA over time should predict fluctuations in subsequent performance, and it is critical for future research to demonstrate this.

That said, the current findings are by no means definitive, and it would not be appropriate to conclude that SA (measured by SAGAT or other means) cannot predict within-person variation in performance in track management (or other tasks). Future research is required; for example, we measured SA shortly before performance windows, and it may be the case that SA measured further away from performance windows could predict subsequent classification performance because it would reflect earlier and therefore more temporally accurate awareness of when contacts had entered areas.

CONCLUSION

The vast majority of the human factors literature has focused on the between-person effects of SA and workload. We replicated the between-person effects of SA and demonstrated that subjective workload could predict within-person variation in performance. The current research highlights that it is important for human factors researchers and practitioners to understand conceptually at which level SA or workload measures are applied.

ACKNOWLEDGMENTS

This research was supported by Research Grant RA6146 awarded to Loft and Visser from the Defence

Science and Technology Group (Australia) and Discovery Grant DP160100575 from the Australian Research Council awarded to Loft. We thank Janelle Braithwaite for her programming of the submarine simulation.

KEY POINTS

- Situation awareness (SA) and workload are crucial constructs in human factors that are conceptualized as states that change within individuals over time. An increase/decrease in SA or subjective workload within individuals should be predictive of their subsequent performance (within-person effects).
- Changes in subjective workload predicted subsequent performance (within-person effects) in a submarine track management task; when a participant's subjective workload rating increased (relative to their own average workload), their subsequent performance decreased (relative to their own average performance).
- We found a positive relationship between workload and contact classification accuracy at the between-person level. Although individuals who experienced higher average levels of workload compared to others made more accurate classification decisions compared to others, increased workload relative to an individual's own average workload impaired subsequent performance for that individual.
- We replicated prior between-person level effects of SA; participants with higher average SA performed better on all three tasks compared to participants with lower SA, but we found no evidence of within-person effects of SA on performance. Increased SA should increase the extent to which operators hold the knowledge required to perform subsequent tasks, so further research is required.
- It is crucial for human factors researchers and practitioners to understand conceptually at which level SA or workload measures are applied.

REFERENCES

Bliese, P. D. (2006). *Multilevel modeling in R (2.2): A brief introduction to R, the multilevel package and the nlme package*. Washington, DC: Walter Reed Army Institute of Research.

Brunswick, E. (1943). Organismic achievement and environmental probabilities. *Psychological Review*, *50*, 255–272.

Chaippe, D., Morgan, C. A., Kraut, J., Ziccardi, J., Sturre, L., Strybel, T. Z., & Vu, K. P. L. (2016). Evaluating probe techniques and a situated theory of situation awarness. *Journal of Experimental Psychology: Applied*, *22*, 436–454.

Chen, S., Visser, T. A. W, Huf, S., & Loft, S. (2017). Optimizing the balance between task automation and human manual control in simulated submarine track management. *Journal of Experimental Psychology: Applied*, *23*, 240–262.

Durso, F. T., & Dattel, A. R. (2004). SPAM: The real-time assessment of SA. In S. Banbury & S. Tremblay (Eds.), *A cognitive approach to situation awareness* (pp. 137–154). Farnham, UK: Ashgate.

Endsley, M. R. (1990). Predictive utility of an objective measure of situation awareness. *Proceedings of the 34th Annual Meeting of f the Human Factors and Ergonomics Society* (pp. 41–45). Santa Monica, CA: Human Factors and Ergonomics Society.

Endsley, M. R. (1995a). Measurement of situation awareness in dynamic systems. *Human Factors*, *37*, 65–84.

Endsley, M. R. (1995b). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*, 32–64.

Endsley, M. R. (2000). Direct measurement of situation awareness: Validity and use of SAGAT. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 147–173). Mahwah, NJ: Lawrence Erlbaum Associate.

Endsley, M. R., & Rodgers, M. D. (1996). Attention distribution and situation awareness in air traffic control. *Proceedings of the 40th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 82–85). Santa Monica, CA: Human Factors and Ergonomics Society.

Funke, G. J., Knott, B. A., Salas, E., Pavlas, D., & Strang, A. J. (2012). Conceptualization and measurement of team workload: A critical need. *Human Factors*, *54*, 36–51.

Gopher, D., & Donchin, E. (1986). Workload: An examination of the concept. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 1–49). New York, NY: Wiley.

Hancock, P. A. (2007). On the process of automation transition in multitask human-machine systems. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, *37*, 586–598.

Hancock, P. A., Jagacinski, R. J., Parasuraman, R., Wickens, C. D., Wilson, G. F., & Kaber, D. B. (2013). Human-automation interaction research: Past, present, and future. *Ergonomics in Design: The Quarterly of Human Factors Applications*, *21*, 9–14.

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

Helton, W. S., Funke, G. J., & Knott, B. A. (2014). Measuring workload in collaborative contexts trait versus state perspectives. *Human Factors*, *56*, 322–332.

Helton, W. S., Matthews, G., & Warm, J. S. (2009). Stress state mediation between environmental variables and performance: The case of noise and vigilance. *Acta Psychologica*, *130*, 204–213.

Hockey, G. R. J. (1986). A state control theory of adaptation and individual differences in stress management. In G. R. J. Hockey, A. W. K Gaillard, & M. G. H. Coles (Eds.), *Energetics and human information processing* (pp. 285–298). Leiden, Netherlands: Martinus Nijhoff.

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.

Humphreys, M. S., & Revelle, W. (1984). Personality, motivation, and performance: A theory of the relationship between individual differences and information processing. *Psychological Review*, *91*, 153–184.

Jones, D. G., & Endsley, M. R. (2004). Use of real-time probes for measuring situation awareness. *The International Journal of Aviation Psychology*, *14*, 343–367.

Loft, S., Bolland, S., Humphreys, M. S., & Neal, A. (2009). A theory and model of conflict detection in air traffic control: Incorporating environmental constraints. *Journal of Experimental Psychology: Applied*, *15*, 106–124.

Loft, S., Bowden, V., Braithwaite, J., Morrell, D. B., Huf, S., & Durso, F. T. (2015). Situation awareness measures for simulated submarine track management. *Human Factors*, *57*, 298–310.

Loft, S., Morrell, D. B., & Huf, S. (2013). Using the situation present assessment method to measure situation awareness in simulated submarine track management. *International Journal of Human Factors and Ergonomics*, *2*, 33–48.

Loft, S., Morrell, D., Ponton, K., Braithwaite, J, Bowden, V., & Huf, S. (2016). The impact of uncertain contact location on situation awareness and performance in simulated submarine track management. *Human Factors*, *58*, 1052–1068.

Manning, C., Mills, S., Fox, C., Pfleiderer, E., & Mogilka, H. (2002). *Using air traffic control taskload measures and communication events to predict subjective workload* (No. DOT/FAA/AM-02/4). Washington, DC: Federal Aviation Administration Office of Aviation Medicine.

Mracek, D. L., Arsenault, M. L., Day, E. A., Hardy, J. H., & Terry, R. A. (2014). A multilevel approach to relating subjective workload to performance after shifts in task demand. *Human Factors*, *56*, 1401–1413.

Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2014). Human performance consequences of stages and levels of automation: An integrated meta-analysis. *Human Factors*, *56*, 476–488.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making*, *2*, 140–160.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*, 129–138.

Stein, E. S. (1985). *Air traffic controller workload: An examination of workload probe* (DOT/FAA/CT-TN84/24, Trans.). Atlantic City, NJ: Federal Aviation Administration Technical Centre.

Vidulich, M. A., & Tsang, P. S. (2012). Mental workload and situation awareness. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 243–273). Hoboken, NJ: John Wiley & Sons, Inc.

Vu, K. P. L., & Chiappe, D. (2015). Situation awareness in human systems integration. In D. A. Boehm-Davis, F. T. Durso, & J. D. Lee (Eds.), *APA handbook of human systems integration* (pp. 293–308). Washington, DC: American Psychological Society.

Yeo, G., & Neal, A. (2004). A multilevel analysis of the relationship between effort and performance: Effects of ability, conscientiousness and goal orientation. *Journal of Applied Psychology*, *89*, 231–247.

Young M. S., Brookhuis K, A., Wickens, C. D., & Hancock P. A. (2015). State of science: Mental workload in ergonomics. *Ergonomics*, *58*, 1–17.

Shayne Loft is an associate professor at The University of Western Australia. He received his PhD in psychology in 2004 from The University of Queensland.

Lisa Jooste is a master's student enrolled in the industrial and organization psychology program at the University of Western Australia.

Yanqi Ryan Li is a research fellow at The University of Western Australia. He received his PhD in psychology in 2015 from The University of Western Australia.

Timothy Ballard is a research fellow at The University of Queensland. He received his PhD in psychology in 2017 from The University of Western Australia.

Samuel Huf is a team leader of Submarine Human Systems and Analysis, Maritime Division at the Defence Science and Technology Organisation, Australia. He received his PhD in psychology in 2002 from James Cook University.

Ottmar V. Lipp is a professor at Curtin University. He received his PhD in psychology in 1989 from the University of Giessen, Germany.

Troy A. W. Visser is an associate professor at The University of Western Australia. He received his PhD in cognitive systems in 2001 from the University of British Columbia.

Date received: September 22, 2017
Date accepted: May 25, 2018