The Impact of Information Integration in a Simulation of Future Submarine Command and Control

Steph Michailovs, **Stephen Pond**, The University of Western Australia, Crawley, **Megan Schmitt**, **Jessica Irons**, Defence Science and Technology Group, Canberra, Australia, **Matthew Stoker**, Ergonomie, Sydney, Australia, **Troy A. W. Visser**, The University of Western Australia, Crawley, **Samuel Huf**, Defence Science and Technology Group, Canberra, Australia, and **Shayne Loft**, The University of Western Australia, Crawley

Objective: Examine the extent to which increasing information integration across displays in a simulated submarine command and control room can reduce operator workload, improve operator situation awareness, and improve team performance.

Background: In control rooms, the volume and number of sources of information are increasing, with the potential to overwhelm operator cognitive capacity. It is proposed that by distributing information to maximize relevance to each operator role (increasing information integration), it is possible to not only reduce operator workload but also improve situation awareness and team performance.

Method: Sixteen teams of six novice participants were trained to work together to combine data from multiple sensor displays to build a tactical picture of surrounding contacts at sea. The extent that data from one display were available to operators at other displays was manipulated (information integration) between teams. Team performance was assessed as the accuracy of the generated tactical picture.

Results: Teams built a more accurate tactical picture, and individual team members had better situation awareness and lower workload, when provided with high compared with low information integration.

Conclusion: A human-centered design approach to integrating information in command and control settings can result in lower workload, and enhanced situation awareness and team performance.

Application: The design of modern command and control rooms, in which operators must fuse increasing volumes of complex data from displays, may benefit from higher information integration based on a human-centered design philosophy, and a fundamental understanding of the cognitive work that is carried out by operators.

Keywords: human–machine interface, information integration, performance, workload, situation awareness

Address correspondence to Shayne Loft, School of Psychological Science, University of Western Australia, Crawley, WA 6009, Australia; e-mail: shayne.loft@uwa.edu.au

HUMAN FACTORS
2023, Vol. 65(7) 1473–1490
DOI:10.1177/00187208211045872
Article reuse guidelines: sagepub.com/journals-permissions

The capabilities of command and control (C2) rooms are changing dramatically due to rapid technological advances. New technologies for C2 rooms are resulting in greater volumes of data, necessitating that innovative information displays and automated systems be developed to assist humans. Such advances have the potential to improve the capability of C2 rooms to adapt to changing requirements of operation. However, for this to be achieved, it is imperative that information is presented in a manner that maximizes relevance to each operator role and to the operational context, so that each operator in a team can effectively use this information to do their jobs.

Historically, many C2 settings have used designs in which the information displayed to each operator is determined by the system's physical architecture (e.g., the source of the information) and the individual operator task analysis and resulting input to their work console (i.e., the individual human–machine pair). In contrast, more contemporary C2 designs aim to increase "information integration"—that is, to visually share on the display of each operator in a team the information that they potentially require to complete their tasks, regardless of the source of that information, or whom in the team generated that information (Boehm-Davis et al., 2015).

The aim of the current study was to examine the impact of increased information integration on operator workload, operator situation awareness (SA), and team performance in a task in which teams of participants performed tactical picture compilation in a simulation of a future submarine C2 room. Although we used a submarine simulation, our results provide an evidence base for design principles applicable to other C2 work domains that require teams to fuse large volumes of information from various sensors in order to develop a tactical picture (e.g., air traffic control, uninhabited vehicle control, air battle management).

DEVELOPING THE TACTICAL PICTURE

In a submarine C2 room, submariners interpret sensor and other data to build an understanding of the operational situation (Roberts et al., 2018). The C2 team comprises a number of departments including navigation, communications, and sensor and tactical systems. The current study concerned how effectively the C2 team (in our simulation, comprised of only sensor and tactical system operators) could compile a tactical picture of the position of other vessels (contacts) in relation to their own vessel (Ownship). An accurate tactical picture is critical to maintain safety, stealth, and mission success. However, picture compilation takes time, and the compiled tactical picture will always vary in the degree to which it matches the "ground truth" because of, for example, the unpredictability of the ocean, and third-party actions (hostile deception; Loft et al., 2016).

There are several roles that contribute to reducing uncertainty in the tactical picture. The Track Motion Analyst (TMA) estimates a *solution* (bearing, range, course, and speed) for each contact, based on its relative motion derived from bearing cuts provided by the Sonar operator, and visual information provided by the Periscope operator. The relative motion of each contact is displayed on a time-bearing plot to the Track Manager, allowing them to assess the accuracy of the TMA solution, monitor information derived from sensors, and resolve conflicting information from team members to improve contact solutions. Supervised by a Watch Leader, the Track Manager assesses and fuses the sensor data with other geospatial and environmental data to build a tactical picture, and to recommend submarine maneuvers required to ensure safety, stealth, and mission success.

INFORMATION INTEGRATION

Submarine C2 rooms have traditionally used stove-piped human–machine interface (HMI) designs, where information is only shown to an operator if the source of that information is from sensor(s) connected to that operator console architecture (role), and the operator can then choose to verbally share it with other operators. Similarly, outcomes from information processed by an operator largely remain on that display on which it is generated, and are shared verbally by the operator (we refer to this work design as low integration).

Prototype high-integration HMI designs have emerged from cognitive work analyses and human-centered design workshops conducted by the Australian Defence Science and Technology Group (e.g., Schmitt et al., 2013). For the current study, the HMI prototypes were networked together in the Control Room Use Simulation Environment (CRUSE), where data were integrated across consoles. The high integration design displays to each operator the information that they required to complete their task goals, regardless of the source of that information, or whom in the team generated it. The concept of information integration is similar to that of a common operating picture (e.g., Hwang & Yoon, 2020), but does not involve presenting a common display, but rather information specific to roles.

On the face of it, highly integrated C2 rooms may have several advantages. High integration provides each operator with a complete set of information regarding their own tasks, and timely task updates from team members completing interdependent tasks. This work design emulates how effective teammates anticipate the information needs of others and "push" information in a timely fashion (transactive memory; Hollingshead, 1998; Wegner, 1987). Higher integration reduces the need for the operator to wait for information because it is available on their display. By comparison, with lower integration, information must be verbally requested (i.e., "pulled"; McNeese et al., 2018), and the operator must then wait for the information to be reported. We therefore expected that the number of verbal emissions per team to be higher in low compared with high integration teams due to the increased need to verbalize information. Additionally, because teams using low integration rely more on verbal exchange, operators could be more vulnerable to team miscommunication (e.g., being mis-heard, heard but forgotten, or not heard at all), and given that communication requires time/cognitive resources, the increased need for communication may produce a "communication overhead" (MacMillan et al., 2004) that increases workload and reduces team performance. On the basis of this theoretical analyses, we would predict more verbal communication in low compared with high integration teams, and that the increase in immediately available (pushed)/reliable information on highly integrated displays should improve operator SA by facilitating their understanding of the current and likely future state of the task environment (Vu & Chiappe, 2015), and improve team performance.

On the other hand, an argument can be made that the increase in display complexity (Donderi, 2006) associated with the larger volume of immediately available (pushed) information on highly integrated displays may make it challenging for operators to locate relevant information (display clutter; Moacdieh & Sarter, 2015), or lead to information overload (Marusich et al., 2016), due to the cognitive limitations that constrain the amount of information humans can process (Gigerenzer & Selten, 2002). To the extent this is the case, higher integration may not improve SA, and may increase operator workload, and possibly not lead to improved team performance. Higher integration could also lead to greater operator autonomy, in that operators may work more within their own information bubble, creating potential temporal dissonance between operators, and undermining team process. The reduced communication from higher integration may also come at a cost. The need to verbalize more information (as expected in the lower integration condition) may help synchronize the efforts of the C2 team by increasing awareness of which contacts are being worked on and conflicting information (closed-loop communication, Salas et al., 2005) facilitating a common team understanding (interactive team cognition; Cooke et al., 2013).

With these uncertainties in mind, the goal was to investigate whether increased information integration would indeed reduce the number of verbal emissions, impact operator workload, and improve operator SA or team performance. Teams of six participants were trained to perform tactical picture compilation. Participants were trained to perform one of six roles (Sonar x 2, TMA x 2, Periscope, Track Manager), and information integration was manipulated between-subjects. Operator SA was measured by presenting operators with queries about the scenarios. Participants rated their own workload every 5 min, and after each scenario. Team performance was assessed by examining tactical picture error (the difference in contact location between the simulation truth and the solution derived by teams). Communication between operators was recorded.

METHOD

Participants

Sixteen teams of six participants were recruited through word of mouth, the UWA community participation website, and from local sporting and yacht clubs. There were 38 males and 58 females with an average age of 29.14 years. There was no a priori constraint on sample size, but we did find it difficult to recruit teams. Participants were paid AUD 15 per hour (total for 9 hr = AUD 135). Participants were randomly assigned to either a high integration (eight teams) or low integration (eight teams) team, and within those teams, participants were randomly assigned a role. This research complied with the National Statement of Ethical Conduct in Human Research (code of conduct) and was approved by the Human Research Ethics Office at the University of Western Australia. Informed consent was obtained from each participant.

Control Room Use Simulation Environment (CRUSE)

For this experiment, CRUSE was configured with six consoles (each with two displays): two Sonar consoles, two TMA consoles, one Periscope console and a console showing the tactical picture compilation which was operated

Figure 1. Participants undertaking their tasks within the CRUSE. From left to right the displays are Periscope, Sonar 2, Sonar 1, Track Manager, TMA 1, and TMA 2. CRUSE = Control Room Use Simulation Environment; TMA = Track Motion Analyst.

by the Track Manager. The team's goal was to detect, locate, classify, and estimate the range, course, and speed of all contacts within sonar range. The team then reported this solution to the Watch Leader, who was one of the researchers. Figure 1 shows a team completing the task (in starboard-facing line configuration).

Two scenarios were developed that were the 180º degree mirror (on Ownship heading axis) of each other and thus had comparable task load. This meant that the number of contacts, contact classification, and timing of contact detection both visually and on sonar were identical, but that contacts were detected on a different bearing. For example, as shown in Figure 2, if Scenario 1 included a Merchant contact that could be detected at the 10 -min mark, was traveling at 12 knots at a bearing of 315, and on a course of 225, then in Scenario 2 this contact would be a Merchant contact detected at the 10 -min mark, traveling at 12 knots, at a bearing of 045, and on a course of 135 (assuming Ownship heading of 000).

Each scenario had 14 contacts introduced gradually over 60 min. For each contact, Sonar operators initially detected and reported it. The Track Manager then prioritized each contact based on its classification, range, behavior, and solution accuracy in relation to received sonar data, and assigned each contact to one of the TMAs. The Track Manager was instructed to attempt to assign equal task-load across TMAs. The Track Manager used the information on their display to track the task load of each TMA, the relative priority of contacts, and the progress of TMA contact solutions. The Periscope

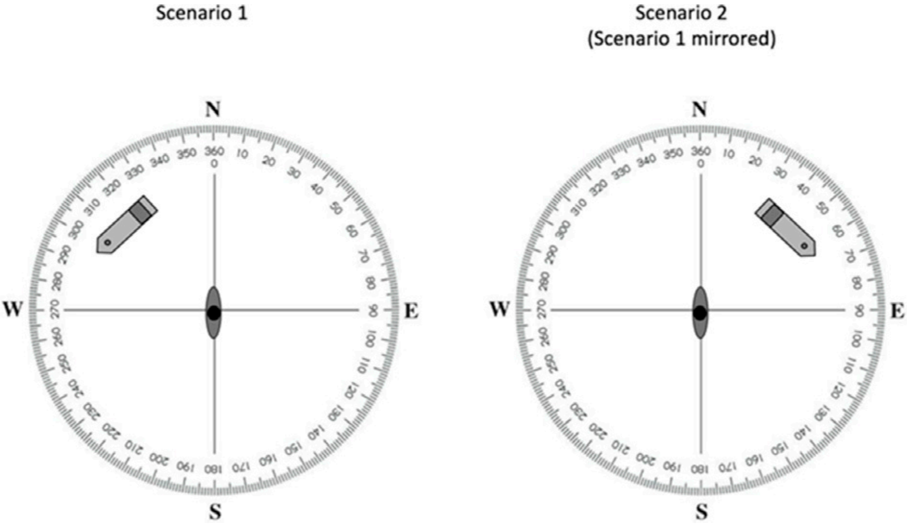

Figure 2. The running example of how a scenario was mirrored so that workload was kept comparable whilst ensuring minimal transfer of learning.

operator looked for each contact and centered it on their display, so the range-finding tool (adjustable reticules over the contact image) could be applied to estimate contact range. The Periscope operator could also estimate the aspect of the contact to derive the angle to the bow (ATB). The TMAs use the contact's ATB (which the system converts to a course) and the estimated range from the periscope operator, as well as estimated contact speed (derived from classification) to find a solution that would best fit the sonar data for that contact. TMA did this using on-screen tools, which included a virtual "line of best fit" ladder that could be adjusted until the TMA was satisfied the solution was correct (best matched the sonar information). The solution was verbally reported to the Watch Leader before being entered into the system by the TMA.

Information Integration Conditions

Cognitive work analyses (utilizing the Cognitive Work Analysis framework; for example, Rasmussen, 1986) were conducted by the Australian Defence Science and Technology Group to identify the information processing requirements of critical roles in the submarine control room (e.g., Schmitt et al., 2013). This was followed by human-centered design workshops that applied Ecological Interface Design principles (e.g., Burns & Hajdukiewicz, 2004) to design the HMIs to support operator behavior across the range of activities performed in each role, and that applied the Proximity Compatibility Principle (Wickens & Carswell, 1995) to visually co-locate information common to specific tasks or required mental operations. These analyses/workshops resulted in the generation of highly integrated HMI designs to support the cognitive work likely undertaken in future submarine C2 rooms, whilst the low integration design represented a more conventional stove-piped approach.

There were a number of differences between the high- and low-integration conditions (for a detailed list, see Table 1). It is critical to note that, as illustrated by Table 1, the information provided in the high integration displays was equivalent to the information presented in the low integration displays, and what differed was that the information was visually available across more positional displays (consoles) in the high integration condition.

All information that was unavailable visually in the low integration condition was available verbally. For Periscope, Sonar, and Track

TABLE 1: How Information Required for Contact Solutions Was Shared Visually Between Displays in Low and High Integration

	Periscope		Sonar		Track Manager		TMA	
Information Displayed	LO	HI	LO	HI	LO	HI	LO	HI
Bearing – from Sonar	✗	✓	✓	✓	✓	✓	✓	✓
Bearing – from TMA	✓	✓	✓	✓	✓	✓	✓	✓
Bearing rate – from Sonar	✗	✗	✓	✓	✗	✗	✗	✓
Classification – from Sonar & Periscope	✓	✓	✗	✓	✓	✓	✗	✓
Range – from TMA	✗	✗	✗	✗	✗	✓	✓	✓
Course – from TMA	✗	✓	✗	✗	✗	✓	✓	✓
Speed – from TMA	✗	✓	✗	✗	✗	✓	✓	✓
Annotations – from Track manager	✗	✗	✗	✓	✓	✓	✗	✗
Solution on Geoplot - TMA	✗	✓	✗	✓	✓	✓	✓	✓

Note. TM = Track Manager; TMA = Track Motion Analyst. The Sonar information *bearing* and *bearing rate* are known by the participants to be certain (the simulation truth). All information from TMA is from their estimated solution. The bearing from Sonar is represented visually as both a three-figure bearing, or a "sonar tracker" icon on a bearing strip.

Manager, the bearing and the TMA range, course and speed could be attained by listening to the scripted "mandatory" verbal solution reports from TMA to the Watch Leader, or requesting it as needed. The bearing was most important to Periscope to know where to look (and could also be asked of Sonar), and the range, course, and speed useful for the Track manager to assess whether it was consistent with that expected for the class of contact and sensor information. Track manager annotations confirmed verbal reports. The bearing rate, most important to TMA, was reported by Sonar on detection and requested when required.

Figures 3–6 show the low and high information integration displays for each role (Sonar, Periscope, Track Manager, and TMA) with the accompanying figure captions explaining the differences in each display for each role as a function of integration condition.

We also orthogonally manipulated team configuration upon the request of the Defence research sponsor (to assess the *physical use* of control room space). Teams were in a starboard-facing line-type configuration for one scenario (Figure 1), and forward-facing classroom-type configuration for the other. We did not expect configuration to influence the results as there was no need for egocentric spatial awareness because the submarine did not maneuver. Indeed, the analyses indicated that this manipulation had no effect on any outcome variable, so analyses are collapsed across this variable and are not discussed further.

Measures

Situation awareness. The Situation Present Awareness Method (SPAM; Durso & Dattel, 2004) delivers SA queries in real time without pausing scenarios or blanking displays. An initial prompt asks operators whether they are ready for an SA query. The time the operator takes to accept the query is referred to as SPAM accept time, and often correlates with subjective workload (Vu et al., 2012) and task load (Loft et al., 2015). The actual SA query is then posed after the operator indicates that they are ready to receive the query. Operators with better SA should know, or know where to find, information to answer the question, and thus should respond faster and more accurately to the SA queries.

Figure 3. The Sonar display as presented to the low integration (left) and high integration (right) conditions. For high integration, the Sonar and the TMA solution bearings were both presented on the bearing strip as Sonar trackers and circle icons respectively. Additionally, the bearings of the sonar trackers were presented in the contact information table. The team's estimate of the contact location relative to Ownship (TMA solution) was presented on the Geoplot (a bird's eye depiction of the immediate area around the submarine) via the contact icons. The bearing rate was presented in the contact information table. The contact classification, derived from the Time-Frequency plot, was indicated by the icon color. For Sonar in low integration, the Geoplot, contact classification by icon color, and the annotations from the Track Manager were not presented. TMA = Track Motion Analyst.

In each scenario, the participants were presented with "Are you ready for a question?" prompt at roughly 5 -min intervals (13 occasions), followed by a SPAM query. Both the prompt and the queries were delivered audibly through headphones, and simultaneously via text on displays, with participants answering by clicking on "yes" on the screen for the prompt, and one of four possible answers for the query. The SPAM prompt appeared at the same time for all participants. Table 2 presents the SA queries delivered. All participants in the team received the same SA query at the same time. The SPAM queries and their order were preserved for the second scenario.

Mean accuracy on SPAM queries, and response time (RT) for correct SPAM responses, were calculated for each participant based on the 26 SPAM queries delivered over the two scenarios. The dependent variable for SA was the participant mean RT divided by the proportion of correct responses (inverse efficiency score), thereby providing single index of SA that reflected both response accuracy and speed (Bruyer & Brysbaert, 2011; Townsend & Ashby, 1983). The inverse efficiency scores were averaged within each team.

Subjective workload. The Air Traffic Workload Input Technique (ATWIT, Stein, 1985) required participants to rate their workload on a 10-point scale (1 = very low, 10 = very high). ATWIT was presented for 10 s at 5 -min intervals during each scenario (12 times per scenario) at the same time for all participants. The simulation continued in the background while the ATWIT query was displayed.

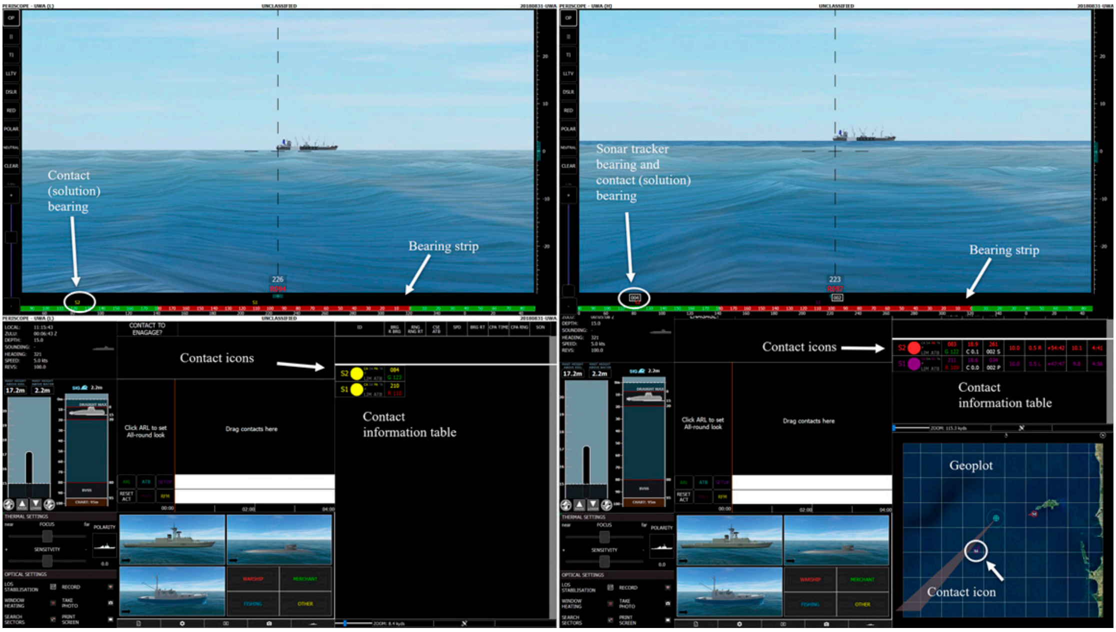

Figure 4. The Periscope display as presented to the low integration (left) and high integration (right) conditions. For high integration, the TMA solution bearings, sonar tracker bearings, contact classification (color of contact icon), the solution range, course and speed information, and Geoplot were presented. For low integration, only the contact (solution) bearings were presented. TMA = Track Motion Analyst.

A mean ATWIT score was calculated for each participant based on the 24 ATWIT probes delivered over the two scenarios. The ATWIT scores were averaged within each team.

The NASA Taskload Index (NASA-TLX); Hart and Staveland (1987), a multidimensional subjective workload scale, was administered after each scenario. Participants rated their workload on a 20-point scale on dimensions of workload: Mental Demands, Physical Demands, Temporal Demands, Performance, Effort, and Frustration. The NASA-TLX was administered and scored as per the procedures outlined by Hart and Staveland (1987). NASA-TLX scores were averaged within each team.

Team performance. We quantified team performance in terms of Tactical Picture Error (TPE), which was defined as the average difference in location between the simulation truth and the solution entered by the team for each contact, weighted by the priority of the contact (Figure 7). These difference scores were calculated at 20-s intervals across the scenario. After a solution is submitted, position error will often increase over time, particularly if a solution is poor, which will impact on the TPE metric. For example, if course and speed are inaccurate, the solution will likely drift further away from the ground truth, until the solution can be refined. In this way, the TPE is sensitive to errors in range, course and speed, as well as solution update speed.

Individual contact position errors were weighted by their priority to capture the fact that errors in some contacts (e.g., a close warship) carry more serious consequences than others (e.g., a distant merchant). Contacts were assigned a weight between 1 and 3 on each of four dimensions listed in Table 3, in accordance with task instructions. A contact's total weight was the sum of all four weights (e.g., a Merchant at 4000 yards closing on a steady course would have a weighting of 1 + 3 + 3 +1 = 8).

At a given 20 -s time-point, the position error (distance in yards) between ground truth of contacts held on Sonar and the corresponding

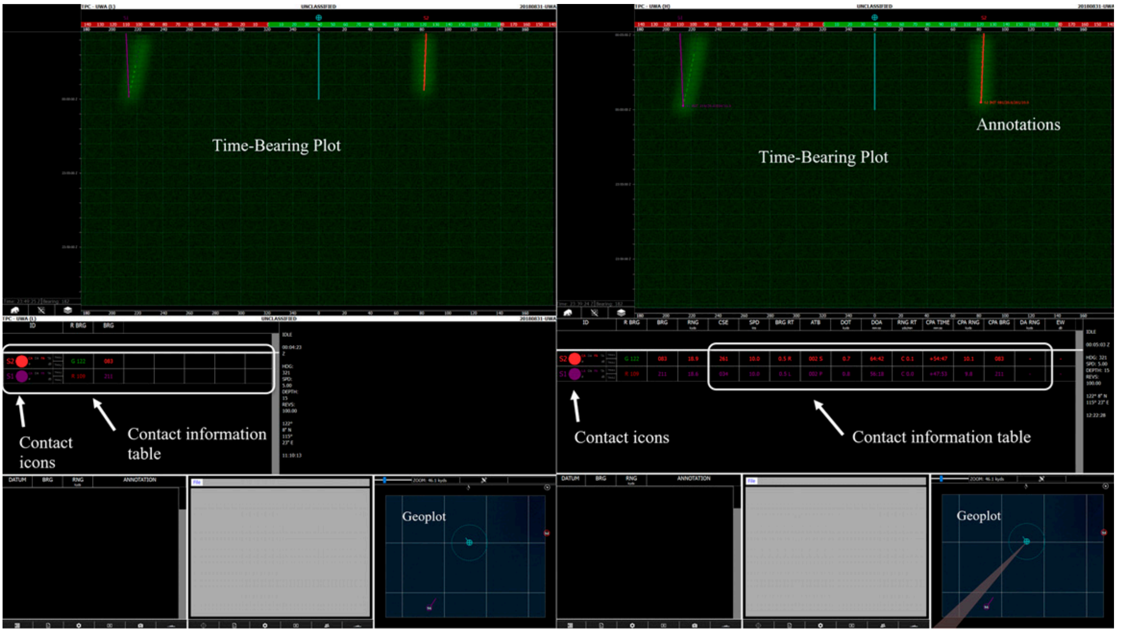

Figure 5. The Track Manager displays as presented to the low integration (left) and high integration (right) conditions. The difference between integration conditions was the presentation of additional contact information in the contact information table, such as the TMA solution range, course, speed and bearing rate in high integration but not in low integration, as well as time-bearing plot annotations from TMA and Sonar. TMA = Track Motion Analyst.

contact solution was calculated. To reduce the impact of outliers, extreme position errors were winsorized to cap them at a maximum of 17,500 yards (corresponding to approximately the 95th percentile of all individual position errors). If a ground truth contact had not yet been given a TMA solution, it was also given a position error of 17,500 yards, thereby penalizing slow first solutions. Next, the priority weighting of each contact at that time-point was calculated and expressed as a percentage of the sum of all current contact weights. Each position error was then multiplied by the percent weighting and summed together to provide a total weighted average TPE.

Team communication. We assessed the frequency of communication (number of verbal emissions per team). Examples of a single operator emissions include TMA saying to Periscope " Can I get a range on Sierra1," and Sonar saying to the team "All positions, Sonar 2, updating report on Sierra2, classified as a possible warship."

Procedure

Teams attended one 9 -h testing session (0800 to 1700). The first 3 hr comprised training via video modules and question-and-answer tutorial sessions. Each participant was then randomly assigned to a role, and viewed a 12–14-min tutorial specific to their role. This was followed by seven practice scenarios ranging from 15 to 30 min in duration. During the practice scenarios, researchers coached participants how to complete their tasks, including instructions about which team member had the information they would need to perform their tasks. They were also instructed on how to ask for and report information. The voice protocol was based on that used in the Australian submarines, which is designed for brevity and clarity so as

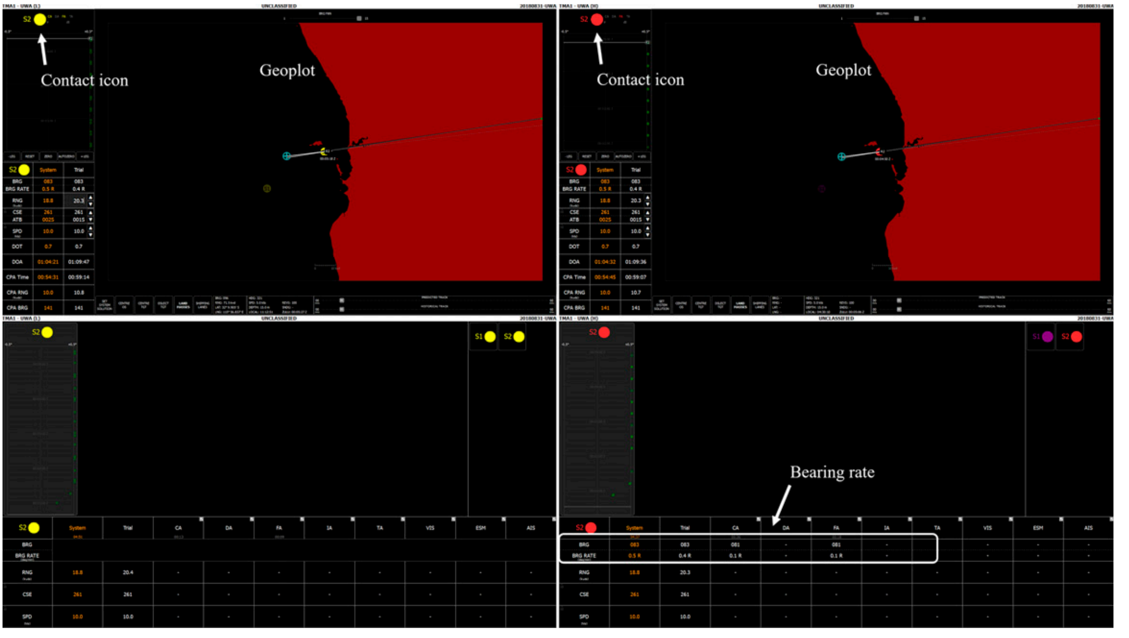

Figure 6. The Track Motion Analyst (TMA) displays as presented to the low integration (left) and high integration (right) conditions. The bearing rate and classification colors for contact icons were available in high integration but not in low integration.

not to "clog" communication channels and minimize miscommunication. For example, "Sonar, TMA" from TMA meant TMA wanted Sonar's attention, and the Sonar operator would respond with "Sonar" so that TMA knew they could then pose their request. Terms such as "standby" (if busy) or "disregard" (if current reporting was wrong or needed updating mid-reporting) were also trained. Participants communicated via a "push to talk" protocol via the keyboard (control button) and headsets fitted with microphones. After training, the team of participants completed two high- or two low-integration test scenarios. Each scenario lasted approximately 1 hr, and the two scenarios were presented in counterbalanced order.

RESULTS

The data from six participants (from three low integration teams, and three high integration teams) were removed due to a data-recording failure on the Sonar 2 position.

Descriptives and Correlations

The descriptives and inter-correlations at the team level between the three workload measures (ATWIT, NASA-TLX, and SPAM query accept time), SA, operator emissions, solutions per contact, and team performance (TPE) are shown in Table 4. The two subjective workload measures (ATWIT, NASA-TLX) were highly positively correlated, as was ATWIT and SPAM query accept time, providing evidence of convergent validity. SA (higher scores = poorer SA) was strong positively correlated with ATWIT, and to a lesser extent NASA-TLX. SPAM query accept time was strongly positively associated with SA. Team performance was strongly negatively associated with number of solutions per contact.

Situation Awareness

Participants in the high integration condition responded to 57.52 % (SD = 5.23%) of SPAM queries correctly and took on average 12.66

TABLE 2: Situation Awareness Questions

Item	Time	Question
1	3.30	Where is the most recently detected contact?
2	6.00	Where will the Merchant B ship be in 10 min?
3	11.00	Where is the closest merchant ship?
4	16.30	Where is the most recently detected contact?
5	21.00	Which contact will pass behind us in about 10 min time?
6	23.30	How many merchant contacts are we currently tracking?
7	29.00	Where will the merchant contact at around Green 25 be in 10 min?
8	34.00	Where is the most recently detected contact?
9	38.00	Where will the fishing contact at around Red 15 be in 10 min?
10	43.00	What type of contact is the closest?
11	47.00	Where is the most recently detected contact?
12	53.00	Where will the warship contact at around Red 20 be in 10 min?
13	58.30	Where will the merchant contact at around Red 130 be in 10 min?

s (SD = 2.32 s) to make correct responses. Participants in the low integration condition responded to 52.33% (SD = 9.36%) of SPAM queries correctly and took on average 15.09 s (SD = 3.85 s) to make correct responses. The SA inverse efficiency score (lower score = better SA) indicated that participants in the high integration condition (M = 24.14, SD = 4.71) had better SA than participants in the low integration condition (M = 30.64, SD = 6.39), $t(14) = 2.32$, $p = .036$, $d = 1.16$.

We examined SA inverse efficiency score differences between conditions at the console level. SA was improved in the high compared to low integration condition for all positions,

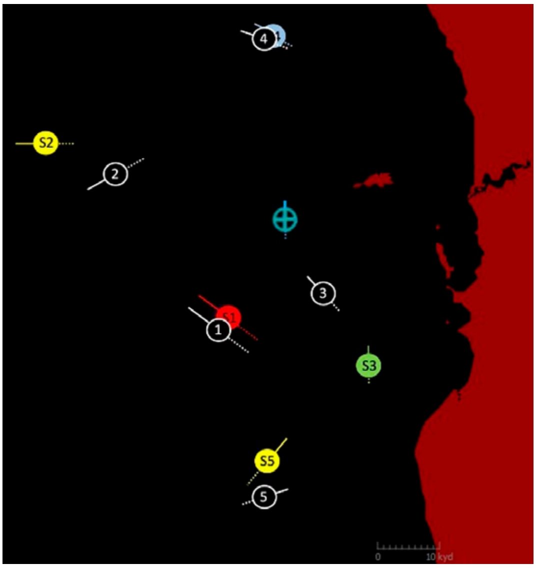

Figure 7. Example of a tactical picture error (TPE). Each numbered icon represents a contact, its direction of travel (solid line), and its wake (dashed line). The middle circle with the cross depicts Ownship. The unfilled circles represent truth (programmed scenario contacts), and the filled circles represent the solutions entered by the team (the "S" denotes the sensor the contact is held on: Sonar). If the solution for a contact was perfect, the team solutions would match the truth. Better solutions can be seen on contacts 1 and 4. Contacts 2, 3, and 5, however, have poorer solutions.

TABLE 3: Priority Weightings Used for the Tactical Picture Error Algorithm

	Weight		
Priority Weight	1	2	3
Classification	Merchant	Fishing	Warship
Range	>10,000 yards	<10,000 yards	<5000 yards
Course	Opening		Closing
Zigging	On a steady course		Frequent course changes
Total	*Sum of all weights*		

indicating a high consistency of the impact of integration on SA across all positions. Periscope

TABLE 4: Means, Standard Deviations, and Correlations Calculated Across Both Integration Conditions for Workload, Situation Awareness, Communication Frequency, and Team Performance (N = 16 Teams)

Variable	*M*	SD	1	2	3	4	5	6
1. ATWIT	5.68	0.84						
2. NASA-TLX	59.72	9.35	.73**					
3. SPAM-AT	3.16s	0.49s	.55*	.12				
4. SA	27.39	6.38	.63**	.40	.68**			
5. Emissions	959.62	150.66	−.12	−.25	.36	.17		
6 . N. Solutions	2.47	0.74	−.33	−.21	−.21	−.06	.42	
7. Performance	8850	1681	.25	.16	.33	.31	.001	−.65**

Note. ATWIT = Air Traffic Workload Input Technique; SA = situation awareness (Situation Present Awareness Method inverse efficiency score); SPAM-AT = Situation Present Awareness Method query accept time. Emissions = average number of verbal emissions; *N*. Solutions = number solutions per contact; Performance = Tactical Picture Error (TPE). *M* and SD are used to represent mean and standard deviation, respectively.

operators had the greatest improvement in SA, but no differences in SA as a function of integration at the position level were statistically significant after using a Bonferroni-corrected alpha level of p = .0125 (smallest p = .02).

Workload

ATWIT scores were significantly lower for participants in the high (M = 5.23, SD = .65) compared with low (M = 6.13, SD = .79) integration condition, $t(14)$ = 2.49, p = .026, d = 1.25, indicating subjective workload was lower with higher integration.

With respect to the SPAM workload indicator, participants in the high integration condition (M = 2.85 s, SD = .21s) were faster to accept SPAM queries than participants in the low integration condition (M = 3.47 s, SD = .50s), $t(14)$ = 3.25, p = .006, d = 1.63, also indicating that workload was lower in the high integration condition.

NASA-TLX scores indicated no difference between participants in the high (M = 57.38, SD = 9.59) compared with low (M = 62.05, SD = 9.10) integration condition, $t(14)$ = 1.01, p = .334. We examined subscale scores as a function of information integration (using a Bonferroni-corrected alpha level of p = .008) but found no significant differences (smallest p = .03 for the Effort subscale which was higher in the low integration condition).

In sum, two of the three workload measures indicated that participants in the high integration condition had lower workload than those in the low integration condition.

We examined workload differences between conditions at the console level using ATWIT, NASA, and SPAM-AT measures. For each of the three workload measures, workload was greater in the low integration condition for all positions, indicating a high consistency of the impact of integration on workload across all positions. However, no differences in workload as a function of integration at the position level were statistically significant after using a Bonferroni-corrected alpha level of p = .0125 (smallest p = .02).

Team Performance

Tactical picture error. The TPE as a function of information integration is plotted in Figure 8. The initial ceiling TPE reflects the time until the first TMA solution was inputted, during which time the contact was assigned maximum position error of 17,500 yards.

Examination of Figure 8 indicates that TPE began to diverge between the low and high integration conditions after 10 min. This is likely because, even with the provision of high integration, it takes some time for information to be estimated (e.g., for Periscope to estimate range/

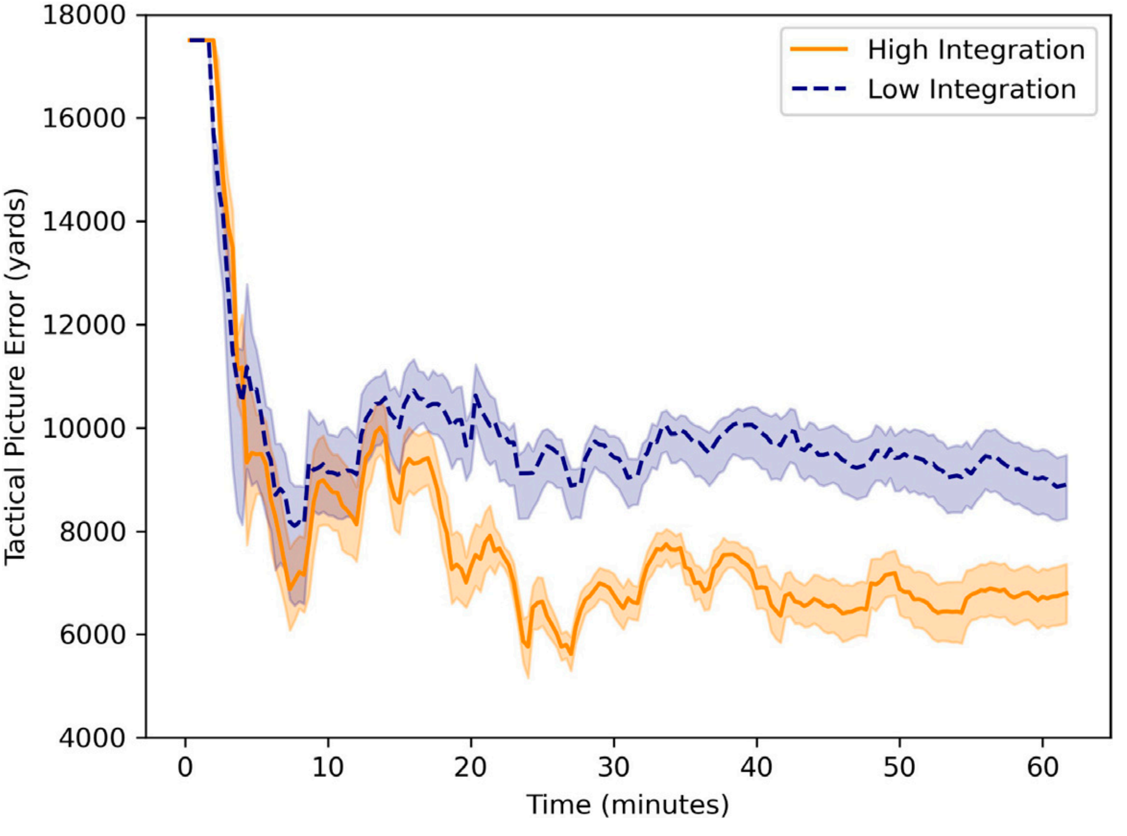

Figure 8. Tactical picture error over time for high and low integration. Error bands represent the standard error of the mean.

course or for Sonar to classify) and share this with TMA to refine the solution.

To compare team performance statistically, the two scenarios were split into three 20 -min blocks and TPE was averaged for each block (Figure 9). A 3 (block) x 2 (integration) mixed ANOVA revealed a significant main effect of block, $F(2, 28) = 21.99$, $p < .001$, $\eta_p^2 = 2.61$, with team performance improving over time. There was also a significant main effect of integration, $F(1,14) = 9.43$, $p = .008$, $\eta_p^2 = 2.40$, and a significant interaction between integration and block ($F(2,28) = 3.35$, $p = .005$, $\eta_p^2 = 2.21$). Follow-up simple effects (Bonferroni-corrected alpha level of $p = .017$) analyses revealed that while the conditions did not differ in the first block, $t < 1$, the high integration condition had a significantly lower TPE than the low integration condition in the second block, $t(14) = 4.33$, $p < .001$, $d = 2.16$, and the third block, $t(14) = 3.35$, $p = .005$, $d = 1.68$. In summary, teams were able to create a more accurate tactical picture when information was more highly integrated.

Solutions per contact. One of the possible reasons for the lower TPE with higher information integration is that it may have allowed teams more time to work on contact solutions, and thus to enter more solutions per contact. Further, in general, we found each new solution on a contact was more likely to be closer to the truth than the previous solution. Teams using high integration entered more solutions per contact ($M = 2.93$, SD = .55) than teams using low integration ($M = 2.01$, SD = .63), $t(14) = 3.11$, $p = .008$, $d = 1.56$. The negative correlation between the number of solutions per contact and TPE (Table 4) indicated that 42% of

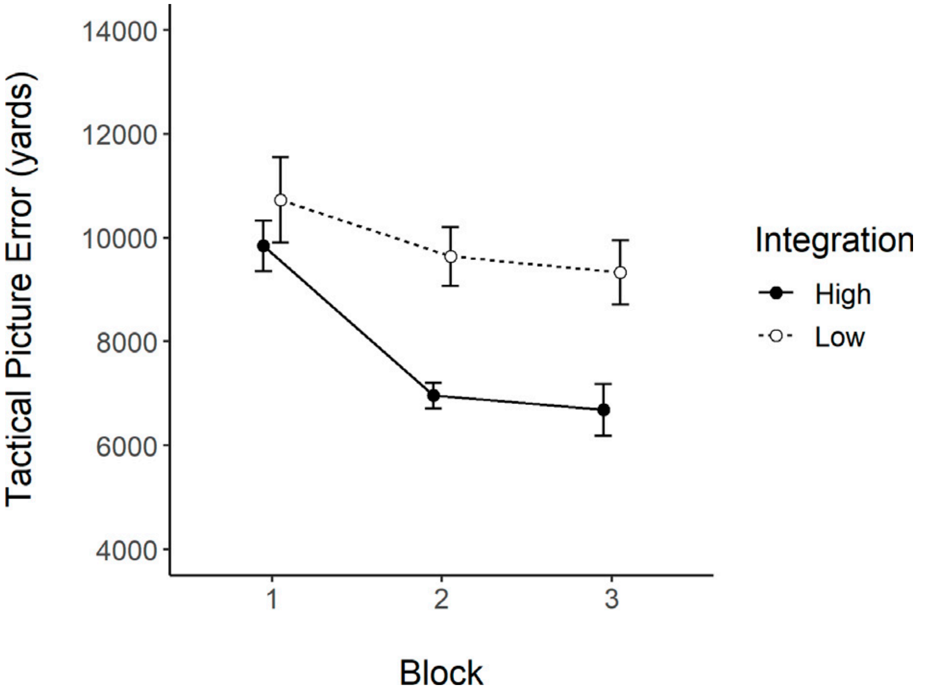

Figure 9. TPE for high and low integration teams (collapsed across configuration conditions) split into 20 -min blocks. Error bars depict standard error of the mean.

the variance in TPE was related to the number of contact solutions entered.

Communication quantity. There was no difference in the number of verbal emissions between teams provided low (M = 937, SD = 145) compared with high (M = 982, SD = 162) integration, $t < 1$. We discuss this unexpected finding in the Discussion section below.

DISCUSSION

The aim of the present study was to examine the impact of increased information integration (the degree to which relevant information was shared between operator displays) on operator SA, operator workload, number of verbal emissions, and team performance. The SPAM data indicated that participants had better SA when provided high compared with low information integration. The workload measures indicated that workload decreased with higher integration. Solutions were more accurate for teams provided with high compared to low integration, partly because TMAs in the high integration condition entered more solutions, with each iteratively closer to ground truth. There was no difference in the number of verbal emissions between the low and high integration conditions.

The SA findings indicate that participants in the high-integration condition had better understanding of the status, and current/future location of contacts (including in relation to Ownship), than participants in the low integration condition. It is important to note, however, that this finding does not necessarily mean that participants provided higher integration had a better memory-based situational model (Endsley, 1995). The amount and complexity of information in the task may have prevented participants from committing many display details to memory. Instead, we suggest the SA findings reflect the greater ability of operators using the more highly integrated HMI to find SA-relevant information quickly on their displays (Fiore & Wiltshire, 2016; Kirschenbaum et al., 2014). For example, in the high integration condition, the Periscope operator had the sonar tracker bearing for each contact, and the Track Manager had annotations from Sonar. All displays also had each contact color coded with the contact's classification, allowing the Periscope, Sonar, and Track Manager to use their bearing strip to identify the contact being queried.

The two self-report workload measures (ATWIT, NASA-TLX) were significantly positively correlated (convergent validity). In addition, SPAM query accept time correlated with subjective workload (ATWIT). Contrary to our concerns that the increased and immediately available (pushed) information on highly integrated displays may have made it challenging for operators to locate or process information (Marusich et al., 2016; Moacdieh & Sarter, 2015), two (ATWIT, SPAM query accept time) of the three workload measures provided evidence that higher integration reduced workload. The ATWIT was likely more sensitive to the impact of information integration than the NASA-TLX because it measured real-time subjective workload. In contrast, the NASA-TLX is a post-scenario measure that required participants to remember and average what they had experienced over the proceeding 60 min. Nonetheless, the patterns of means in the NASA-TLX data as a function of information integration were consistent with the two other workload measures, although clearly not reaching statistical significance.

Teams provided with higher information integration also built a more accurate tactical picture, and entered more solutions per contact. More solutions per contact decreased tactical picture error (42% of variance explained in the TPE by the number solutions per contact). The number of solutions a TMA would submit for each contact is likely driven by factors such as how efficiently the Periscope and Sonar operators could compile, process, and relay information to the TMA, the extent to which the TMA had the information required to compile solutions, and the quality of guidance provided from the information analysis conducted by the Track Manager. These processes could have been facilitated by the higher information integration that provided each operator with more immediate visual access (information "pushing") to the information required for their tasks. For example, the Periscope operator could see the solution bearing of each contact and thus could visually locate contacts more quickly; the TMA had the bearing rate displayed on their display so they did not have to ask for it; and the Track Manager could assign a color to each track denoting their classification, which enabled the team to more quickly identify high priority contacts (e.g., warships). The required information in the high integration displays was also proximal (co-located) to other information sources in terms of the extent to which the information was used as part of a required cognitive task, in line with the Proximity Compatibility Principle (Wickens & Carswell, 1995; also see Burns, 2000). For example, for TMA, the bearing rate (from Sonar) had to be processed in combination with the classification of the contact (indicated by color from the Track manager), and the range of the contact (from Periscope).

These benefits are likely direct outcomes of the cognitive work analysis/human-centered design (including the application of Ecological Interface Design principles and Proximity Compatibility Principle) conducted by Defence (Schmitt et al., 2013) to support the cognitive work anticipated in future submarine C2 rooms. To our knowledge, this is the first empirical demonstration in the literature that more highly integrated HMIs in a simulated submarine C2 room can improve SA, workload, and team performance. These results are also potentially applicable to other C2 work domains that require teams to interpret large volumes of information from sensors to develop a tactical picture, such as air traffic control and uninhabited vehicle control. However, each task domain will have unique challenges for designing highly integrated displays in a manner that minimizes the potential costs of increased visual complexity (for review, see Donderi, 2006).

Given that the higher integration was designed to reduce the need for verbal emissions (communication), we were surprised to find no difference in communication quantity as function of information integration, and thus no evidence to suggest a communication overhead for low integration (MacMillan et al., 2004). While it is beyond the scope of this brief report to analyze the content of communication, the fact that the teams using high integration submitted more solutions may have contributed to their verbal emissions (the fact that the number of solutions and verbal emissions were moderately positively correlated in Table 4 indicates these variables relate). By having basic information immediately available, the high integration condition may have been able to focus their communication more on solution refinement. Given that verbal emissions and team performance are uncorrelated (Table 4), the evidence suggests that the number of solutions drove the performance advantage associated with high integration rather than the quantity of communication.

It is important however to acknowledge limitations of this study and interpret the results of the current work with appropriate caution. Although the information provided in the highly integrated displays was equivalent to the information presented in low integration, in that what differed was only that the information was visually available across more positional displays with high integration, there was *more* information presented visually to each operator. For this reason, follow-up work is needed to discriminate between the potential effect of information *access* and information *integration*. The scenarios were purposely designed to be routine and unclassified, we used novice participants, and ultimately there was limited consequence to participants from their performance, unlike control rooms in complex high-risk domains. It is critical that follow-up work is conducted in higher fidelity environments with expert teams. Third, it is critical to be clear that we certainty do not purport to have measured *team* SA or *team* workload, as the SA and workload of a team is not simply the aggregation of individual SA and individual workload (see Bedwell et al., 2014; Salmon et al., 2008). While beyond the scope of this paper, analysis of the content of the communication data could prove useful for building task, social, and information networks (Stanton & Roberts, 2019) to assess constructs such as team (distributed) SA, team (distributed) workload, and shared mental models.

ACKNOWLEDGMENTS

This research was supported by Research Agreement RA7057 with Loft and Visser from the Defence Science and Technology Group (Australia). The authors would like to thank Justin Hill (Royal Australian Navy Reserve) for his subject matter expertise, David Munro-Ford (Total Technology Partners) for his simulation programming, and Graeme Muller (elmTEK) for his work on the situation awareness measure, simulation programming, and technical and infrastructure support.

KEY POINTS

- The increased volume of information in modern command and control rooms needs to be appropriately integrated across operator roles for full benefits to be realized.
- We examined the extent to which increasing information integration across displays in a simulated future submarine command and control room could improve operator situation awareness, reduce operator workload, reduce the need for team communication, and improve team performance.
- Teams of novice participants were trained to work together to combine data from multiple simulated submarine sensors to build a tactical picture of surrounding contacts at sea. The extent that information was shared across operator displays was manipulated (information integration).
- Participants had better situation awareness, lower workload, and teams built a more accurate tactical picture, when provided with high compared with low information integration. There was no difference in the quantity of communication.
- The design of modern command and control rooms may benefit from higher information integration, identified through a design philosophy that adequately considers the cognitive requirements of the human operators and actively engages them in the design process.

ORCID iDs

Steph Michailovs https://orcid.org/0000-0002-6767-6692
Troy A. W. Visser https://orcid.org/0000-0003-3960-2263
Shayne Loft https://orcid.org/0000-0002-5434-0348

REFERENCES

Bedwell, W. L., Salas, E., Funke, G. J., & Knott, B. A. (2014). Team workload: A multilevel perspective. *Organizational Psychology Review*, *4*, 99–123.

Boehm-Davis, D. A., Durso, F. T., & Lee, J. D. (2015). *APA handbook of human systems integration*. American Psychological Association.
Bruyer, R., & Brysbaert, M. (2011). Combining speed and accuracy in cognitive psychology: Is the inverse efficiency score (IES) a better dependent variable than the mean reaction time (RT) and the percentage of errors (PE)? *Psychologica Belgica*, *51*, 5–13. https://doi.org/10.5334/pb-51-1-5
Burns, C. M. (2000). Putting it all together: Improving display integration in ecological displays. *Human Factors*, *42*, 226–241. https://doi.org/10.1518/001872000779656471
Burns, C. M., & Hajdukiewicz, J. (2004). *Ecological interface design*. CRC Press.
Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive team cognition. *Cognitive Science*, *37*, 255–285. https://doi.org/10.1111/cogs.12009
Donderi, D. C. (2006). Visual complexity: A review. *Psychological Bulletin*, *132*, 73–97. https://doi.org/10.1037/0033-2909.132.1.73
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). Routledge.
Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*, 32–64. https://doi.org/10.1518/001872095779049543
Fiore, S. M., & Wiltshire, T. J. (2016). Technology as teammate: Examining the role of external cognition in support of team cognitive processes. *Frontiers in Psychology*, *7*, 1531. https://doi.org/10.3389/fpsyg.2016.01531
Gigerenzer, G., & Selten, R. (2002). *Bounded rationality: The adaptive toolbox*. MIT press.
Hart, S. G., & Staveland, L. E. (1987). Development of NASA-TLX: Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Elsevier.
Hollingshead, A. B. (1998). Communication, learning, and retrieval in transactive memory systems. *Journal of Experimental Social Psychology*, *34*, 423–442. https://doi.org/10.1006/jesp.1998.1358
Hwang, G. H., & Yoon, W. C. (2020). A new approach to requirement development for a common operational picture to support distributed situation awareness. *Safety Science*, *125*, 104569. https://doi.org/10.1016/j.ssci.2019.104569
Kirschenbaum, S. S., Trafton, J. G., Schunn, C. D., & Trickett, S. B. (2014). Visualizing uncertainty: The impact on performance. *Human Factors*, *56*, 509–520. https://doi.org/10.1177/0018720813498093
Loft, S., Morrell, D. B., 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. https://doi.org/10.1177/0018720816652754
Loft, S., Sadler, A., Braithwaite, J., & Huf, S. (2015). The chronic detrimental impact of interruptions in a simulated submarine track management task. *Human Factors*, *57*, 1417–1426. https://doi.org/10.1177/0018720815599518
MacMillan, J., Entin, E. E., & Serfaty, D. (2004). Communication overhead: The hidden cost of team cognition. In E. Salas & S. Fiore (Eds.), *Team cognition: Understanding the factors that drive process and performance* (pp. 61–83). APA Books.
Marusich, L. R., Bakdash, J. Z., Onal, E., Yu, M. S., Schaffer, J., O'Donovan, J., Höllerer, T., Buchler, N., & Gonzalez, C. (2016). Effects of information availability on command-and-control decision making: Performance, trust, and situation awareness. *Human Factors*, *58*, 301–321. https://doi.org/10.1177/0018720815619515
McNeese, N. J., Demir, M., Cooke, N. J., & Myers, C. (2018). Teaming with a synthetic teammate: Insights into human-autonomy teaming. *Human Factors*, *60*, 262–273. https://doi.org/10.1177/0018720817743223
Moacdieh, N., & Sarter, N. (2015). Display clutter: A review of definitions and measurement techniques. *Human Factors*, *57*, 61–100. https://doi.org/10.1177/0018720814541145
Rasmussen, J. (1986). *Information processing and human machine interaction: An approach to cognitive engineering*. Elsevier Science.
Roberts, A. P. J., Stanton, N. A., & Fay, D. T. (2018). Go deeper, go deeper: Understanding submarine command and control during the completion of dived tracking operations. *Applied Ergonomics*, *69*, 162–175. https://doi.org/10.1016/j.apergo.2018.02.003
Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "Big Five" in teamwork? *Small Group Research*, *36*, 555–599. https://doi.org/10.1177/1046496405277134
Salmon, P. M., Stanton, N. A., Walker, G. H., Baber, C., Jenkins, D. P., McMaster, R., & Young, M. S. (2008). What really is going on? Review of situation awareness models for individuals and teams. *Theoretical Issues in Ergonomics Science*, *9*, 297–323. https://doi.org/10.1080/14639220701561775
Schmitt, M. L., Parera, D. L., Henley, P. J., & Myer, A. (2013). *A cognitive work analysis-based user interface design process: A conceptual user interface for a submarine periscope system* [Paper presentation]. Paper presented at the 10th Industrial and Organisational Psychology Conference, Perth, Western Australia.
Stanton, N. A., & Roberts, A. P. (2019). Better together? Investigating new control room configurations and reduced crew size in submarine command and control. *Ergonomics*, 1–17.
Stein, E. S. (1985). *Air traffic controller workload: An examination of workload probe* (DOT/FAA/CT-TN84/24). Federal Aviation Administration Technical Center.
Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. CUP Archive.
Vu, K.-P. L., Strybel, T. Z., Battiste, V., Lachter, J., Dao, A.-Q. V., Brandt, S., Ligda, S., & Johnson, W. (2012). Pilot performance in trajectory-based operations under concepts of operation that vary separation responsibility across pilots, air traffic controllers, and automation. *International Journal of Human-Computer Interaction*, *28*, 107–118. https://doi.org/10.1080/10447318.2012.634761
Wegner, D. M. (1987). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior* (pp.pp. 185–208). Springer.
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). American Psychological Society.
Wickens, C. D., & Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors*, *37*, 473–494. https://doi.org/10.1518/001872095779049408

Stephanie Michailovs is a research associate at the University of Western Australia. She received her PhD in psychology in 2019 from the University of Western Australia.

Stephen Pond is the Human Factors CRUSE Laboratory Manager at the University of Western Australia. He received his Master of Science from the University of Guelph in 2015.

Megan Schmitt is a psychologist at Defence Science and Technology Group, Australia. She received her Bachelor of Arts (psychology) (Honors) in 2001 from the University of Western Australia.

Jessica Irons is a behavioral scientist at Defence Science and Technology Group, Australia. She received her PhD in psychology in 2013 from the University of Queensland.

Matthew Stoker is a Chartered Human Factors Specialist and a Certified Professional Ergonomist and is a contractor at the Defence Science and Technology Group, Australia. He received an Honors degree in human factors and ergonomics from Loughborough University, UK.

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.

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

Shayne Loft is a professor at the University of Western Australia. He received his PhD in psychology in 2004 from the University of Queensland.

Date received: June 11, 2020
Date accepted: August 21, 2021