Effects of Workload and Workload Transitions on Attention Allocation in a Dual-Task Environment: Evidence From Eye Tracking Metrics

Nadine Marie Moacdieh, American University of Beirut, Lebanon, **Shannon P. Devlin**, University of Virginia, Charlottesville, VA, USA, **Hussein Jundi**, American University of Beirut, Lebanon, and **Sara Lu Riggs**, University of Virginia, Charlottesville, VA, USA

High mental workload, in addition to changes in workload, can negatively affect operators, but it is not clear how sudden versus gradual workload transitions influence performance and visual attention allocation. This knowledge is important as sudden shifts in workload are common in multitasking domains. The objective of this study was to investigate, using performance and eye tracking metrics, how constant versus variable levels of workload affect operators in the context of a dual-task paradigm. An unmanned aerial vehicle command and control simulation varied task load between low, high, gradually transitioning from low to high, and suddenly transitioning from low to high. Performance on a primary and secondary task and several eye tracking measures were calculated. There was no significant difference between sudden and gradual workload transitions in terms of performance or attention allocation overall; however, both sudden and gradual workload transitions changed participants' strategy in dealing with the primary and secondary task as compared to low/high workload. Also, eye tracking metrics that are not frequently used, such as transition rate and stationary entropy, provided more insight into performance differences. These metrics can potentially be used to better understand operators' strategies and could form the basis of an adaptive display.

Keywords: workload, topics, eye movements, attention

Address correspondence to Nadine Marie Moacdieh, Department of Industrial Engineering and Management, American University of Beirut, Beirut, Lebanon, nm102@aub.edu.lb

Journal of Cognitive Engineering and Decision Making
Vol. 00, No. 0, Month XXXX, pp. 1
DOI:10.1177/1555343419892184
Article reuse guidelines: sagepub.com/journals-permissions

INTRODUCTION

High mental workload can cause significant performance decrements and breakdowns in attention allocation (Rouse et al., 1993). Preventing problems related to high workload is critical in complex, safety-sensitive domains, such as aviation, process control, driving, and medicine (e.g., Dixon et al., 2005, Van Benthem et al., 2015), where delays and errors due to high workload can have life-threatening consequences. The 1978 United Airlines DC-8 airplane crash in Portland, OR, for example, was linked to pilots' high workload as they were trying to address issues with the landing gear (National Transportation Safety Board, 1978).

Workload is a multidimensional construct for which there are several definitions in the literature; we have adopted Wickens' (1992) definition of workload as the gap between one's attentional resources and the cognitive demands placed on the user. One common way to influence the cognitive demands of a person is by manipulating their task load (Hancock et al., 1995). Task load can be defined as the number of items that one has to attend to in order to successfully complete a task (e.g., Veltman & Gaillard, 1996). While research on high workload has fully established its detrimental effects on performance (e.g., Brookings et al., 1996; Dixon & Wickens, 2006; Matthews et al., 2015), studies tend to assess the effects of workload in discrete and separate time intervals. However, this does not always typify the work of operators in real-world situations, where workload can fluctuate over time. This has led to research on what has been termed workload transitions (Huey & Wickens, 1993; Prytz & Scerbo, 2015), workload history (Cox-Fuenzalida, 2007), or hysteresis (Morgan & Hancock, 2011). The literature has shown that workload transitions negatively affect operator performance, suggesting the concern about fluctuating workload levels is well founded (Cox-Fuenzalida et al., 2006; 2007; Cumming & Croft, 1973; Goldberg & Stewart, 1980).

However, despite these efforts to study the effects of workload transitions, there are still three major gaps in the literature on this topic. The first gap is the lack of emphasis on the difference between gradual and sudden workload changes, as opposed to just analyzing the difference between low-to-high versus high-to-low workload transitions (e.g., Cumming & Croft, 1973; Goldberg & Stewart, 1980). An air traffic control operator, for example, may have to deal with a sudden increase in the number of airplanes to attend to; this could affect the operator's performance differently than if the number of planes increased gradually over time. In other words, it could be that the element of surprise in sudden workload transitions could lead to additional decrements in performance (Kochan et al., 2004; Wickens, 2001).

The second gap is the absence of workload transition studies occurring in a more realistic environment, especially ones that involve multitasking. For instance, Cox-Fuenzalida (2007) investigated the effects of workload transitions by presenting participants with various number strings and asking them to identify particular sequences of digits. The study showed that, contrary to previous research (e.g., Matthews, 1986), a sudden increase in workload did lead to performance decrements. While such studies are valuable for understanding workload transitions, they do not necessarily generalize to more complex environments, such as air traffic control operations or military environments, where operators may have more than one task and several different areas of the screen to attend to.

The third and final gap is the type of approach used to measure the effects of workload transitions, which currently do not capture changes in visual attention allocation in detail. Eye tracking technology is a promising tool in this regard. Several eye tracking metrics have been shown to be sensitive to differences between low and high workload (see Coral, 2016 for a review), but eye tracking has not been used to date to examine the changes in attention allocation that result from sudden and gradual workload transitions in comparison to constant low and high workload.

Objectives

The aim of this research study was to analyze how performance and attention allocation, as evidenced by eye movements, are affected by gradual and sudden workload transitions as compared to constant low or high workload in a multitasking environment. Our expectations were that sudden workload transitions would result in more performance detriments compared to gradual ones, based on the reported detrimental effects of sudden workload transitions (e.g., Cox-Fuenzalida, 2007; Kochan et al., 2004). It was also expected that performance on both the sudden and gradual conditions would be midway between low workload (best) and high workload (worst). The eye tracking measures, though, were the main focus of this study, and it was expected that these measures would help provide insight into how the performance effects came about. More specific hypotheses are provided at the end of the "Background" section, after the eye tracking metrics are defined.

To this end, a realistic simulation environment was used, with variations in task load used as a means of modulating mental workload. The focus in this study was on the types of tasks that require monitoring several spatially dispersed areas of a screen and searching for certain targets, such as what could be found in military applications, air traffic control operations, and security system monitoring. The selected application domain was unmanned aerial vehicle (UAV) operations, an example of a multitasking domain where workload fluctuates.

BACKGROUND

Workload and Workload Transitions

Several theories have attempted to explain what happens during workload transitions. However, the well-known theories that have been posited—Cumming and Croft's (1973) error acceptance theory, Goldberg and Stewart's (1980) short-term memory theory, Matthews' (1986) task overworking theory, and

Cox-Fuenzalida's (2007) adaptation model—focus more on why high to low workload transitions tend to be more detrimental than the inverse. There is no satisfactory explanation or focus in these studies on the differences between sudden and gradual changes in workload levels, least of all at the level of attention allocation.

Approaches that have usually been adopted to examine workload include subjective ratings, performance measures, and physiological measures (Cain, 2007). Subjective ratings include the widely used NASA-Task Load Index (TLX; Hart & Staveland, 1988) and the Subjective Workload Assessment Technique (Reid & Nygren, 1988). The challenge with subjective measures is that they cannot provide detailed information regarding attention allocation. Likewise, while performance measures, such as response time and error rate (e.g., Bliss & Dunn, 2000; Boyer et al., 2015; Dixon & Wickens, 2006), provide a good impression of the effects of low and high workload, they are also not well suited for tracing changes in attention allocation. Physiological measures, on the other hand, are better suited for that purpose. These include measures such as electroencephalography (EEG; Berka et al., 2007), electrocardiogram recordings (ECG; e.g., Solovey et al., 2014), heart rate variability (e.g., Hoover et al., 2012; Schulz et al., 2011), galvanic skin response (GSR; e.g., Solovey et al., 2014), and eye tracking—the focus of this study.

Eye Tracking

Eye tracking is a technique used to trace where a user is looking on a display, typically using infrared light (Poole & Ball, 2005). Eye tracking can be used to track approximate gaze location, which in turn indicates overt attention allocation or the shifting of the eyes toward a stimulus (Bergen & Julesz, 1983; Treisman & Gelade, 1980). This is in contrast to covert attention, which relates to mentally focusing on a certain stimulus. A person could be looking at an item and thinking about something completely different, but the eye tracker cannot recognize this discrepancy and would only indicate where the person is looking. In addition, it is important to note that eye tracking output only provides information about foveal vision—that is, the high acuity central vision that is used to visually process items in detail—but less about peripheral vision (Rosenholtz, 2016) or the useful field of view (Wolfe et al., 2017), which are both relevant to visual perception. This means that there could be information that participants are obtaining that is not directly reflected in the eye tracking metrics. This limitation of eye trackers, however, has not prevented the approach being widely used in human factors research as a means to understand visual attention allocation (Duchowski, 2007).

In contrast to other measures of assessing workload, such as EEG, ECG, and GSR, eye tracking is usually nonobtrusive, meaning that the user is not physically tethered or attached to anything. Moreover, eye tracking can be used to assess a person's situation awareness (SA), which is defined as people's perception of the items around them, their understanding of these items, and their estimate of their state in the near future (Endsley, 1995). However, assessing SA is a challenge because current SA measures that have high reliability and validity require frequently interrupting the participant to ask questions/probes, for example, Situation Awareness Rating Technique (Taylor, 1990) and Situation Awareness Global Assessment Techniques (Endsley, 1988). Eye tracking, on the other hand, would not disturb the operator during a task.

The output from an eye tracker is a series of gaze points that allow researchers to assess when and for how long users were looking at screen elements. The coordinates are used to determine eye *fixations*, or spatially stable gaze points during which visual processing takes place, and *saccades,* which are ballistic movements of the eye between two fixations (Holmqvist et al., 2011).

The most frequently used eye tracking measures for workload evaluation are pupillometry metrics such as pupil diameter (Hampson et al., 2010) and eye blink frequency and duration (Hwang et al., 2008; Veltman & Gaillard, 1996). A full discussion and meta-analysis of these metrics can be seen in Coral (2016). It has been found that these measures are positively correlated with workload; however, such measures are also sensitive to other factors, such as the amount of light (Monfort et al., 2016). Avoiding this problem would require keeping the amount of light constant at all times, incorporating ambient light into the estimate of workload, or incorporating some other measure of workload, such as EEG (e.g., Rozado & Dunser, 2015).

Alternatives to these pupillometry metrics do exist and were the focus of this research study. The selected eye tracking metrics, together with their definitions and our hypotheses regarding each metric, are summarized in Table 1. Mean fixation duration and saccade amplitude have both been used in workload research (De Rivecourt et al., 2008; Di Stasi et al., 2013), as has the nearest neighbor index (NNI; Di Nocera et al., 2007). NNI has been used to determine how spread out or concentrated fixations are and has been shown to be sensitive to workload in various domains that include aviation (Di Nocera et al., 2007) and health care (Moacdieh & Sarter, 2015a). Other measures that have been shown to be useful are stationary and transition entropy (Krejtz et al., 2014). These entropy measures provide an estimate of fixation sequence randomness and have been used in previous studies to estimate workload (e.g., Monfort et al., 2016).

The metrics in Table 1 have been categorized by the *spread* (where are users looking?), *directness* (how efficiently are users scanning?), and *duration* (how long are users looking at a certain area?) of eye gaze points, as classified in Moacdieh and Sarter (2015b). All of these metrics have been used to a much lesser extent than pupillometry metrics, but they offer two major advantages over the latter. First, the metrics in Table 1 are not as sensitive to light as the pupillometry metrics (controlling for light variation can attenuate this problem, although that could be hard to do in a realistic environment). Second, the metrics provide a more reliable and comprehensive assessment of how workload affects scan patterns than just measuring pupil size or blink frequency. The metrics have been used to provide useful insight into the effects of display clutter (Moacdieh & Sarter, 2015a); however, spread metrics have rarely been explored in the context of workload (e.g., Rantanen & Goldberg, 1999) and only a small selection of directness metrics, such as mean saccade amplitude (e.g., Savage et al., 2013), have been explored in this context. In particular, directness metrics, which focus on the efficiency of users' scan patterns, have the potential to reveal more insights about a person's use of a display in divided attention scenarios such as the ones in this study (as detailed in the "Methods" section). Since participants will be asked to deal with a primary task and a number of secondary tasks, with each task located in a different area of the screen, larger saccade amplitudes and more transitioning between areas of the screen would suggest uncertainty about the location of the target or about the task to focus on. For example, questions that directness measures can answer include: *How frequently did operators transition between multiple areas of interest? Did they use the shortest path to reach their goal? Was there a lot of inefficient back-and-forth scanning*? By answering such questions, directness metrics can provide insight into the efficiency of users' scan patterns. Efficiency of eye movement scanning is related to how much screen distance the user's eyes had to travel in order to complete the task, with more efficient users being ones who cover less distance.

METHODS

Participants

Twenty-one students participated in this study (13 men and 8 women; mean age = 20.9, *SD* = 1.5). Participants had self-reported normal or corrected-to-normal vision. Participants were compensated $10/hr for their participation. Participants gave informed consent and the study was conducted in accordance with the tenets of the Declaration of Helsinki. The study was approved by the Clemson University Institutional Review Board (IRB2015-217).

Experimental Setup

The simulation was developed using the Unity game development platform and was based on the "Vigilant Spirit Control Station" used by the Air Force to develop interfaces to control multiple UAVs (Feitshans et al., 2008). The simulation was displayed on the full screen of a desktop computer with a 32″ monitor (2,560 × 1,600 screen resolution). A Fovio eye

TABLE 1: Eye Tracking Metrics Investigated as Part of This Study, Together with the Definition of Each and Our Hypotheses for How They Will Change in the Experiment

Metric	Definition and Calculation	Hypotheses
Spread Metrics		
Convex hull area (pixels^2)	The minimum convex area which contains the fixation points (Goldberg & Kotval, 1999). This is calculated using the Matlab function convHull, with the X and Y positions of the fixation points as input. The maximum area of the screen is 2,560 × 1,600 = 4.096 ×10^6 pixels^2 A larger convex hull area indicates more spread of gaze points and larger cognitive load as the user attempts to sample all the information available within the display (Di Nocera et al., 2007)	H1: As performance deteriorates, these metrics will exhibit increased spread as participants distribute their visual attention to many and wide-ranging areas of the display
Spatial density	The number of grid cells containing gaze points divided by the total number of cells (Goldberg & Kotval, 1999). A 20 × 20 evenly divided grid (128 × 80 pixels per cell) was created to cover the full screen dimensions. Similar to convex hull area, a higher spatial density would indicate a larger dispersion of attention	
Stationary entropy	Stationary entropy indicates how equally distributed a person's attention is, with larger values indicating more evenly spread attention across areas of interest and lower values indicating more narrowed attention (Krejtz et al., 2014). Stationary entropy is calculated using the following equation $H_s = -\sum_{i \in \mathrm{AOIs}} \pi_i log \pi_i$ where π_i represents the long-run fraction of time the chain spends in the *i*th state. For our purposes, it represents the long-run fraction of time a participant spent in the *i*th AOI (the AOIs are as defined in Figure 1). Assuming the properties of a first-level Markov chain hold, π_i can be determined by solving the system of equations $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$, where $\boldsymbol{\pi}$ is stationary distribution of the chain (i.e., every π_i in the chain) and $\boldsymbol{P}$ is a one-step transition matrix of the chain	
Directness Metrics		
Mean saccade amplitude (pixels)	The average amplitude of saccades. Higher mean saccade amplitude indicates lower scanning efficiency	

(Continued)

TABLE 1 (Continued)

Metric	Definition and Calculation	Hypotheses
Scanpath length per second (pixels/s)	The sum of all the saccade lengths divided by the total time. Similar to mean saccade amplitude, a larger scanpath length indicates less efficiency	H2: As performance deteriorates, these metrics will exhibit less directness and efficiency as participants switch their visual attention more frequently and randomly among areas of the display
Backtrack rate (s^{-1})	A backtrack is defined as an angle between two saccades that is greater than 90° (Goldberg & Kotval, 1999), indicating a change in direction. A higher backtrack rate indicates lower efficiency	
Transition rate (s^{-1})	The rate of transitions between equal grid cells (Goldberg & Kotval, 1999). A higher rate of transitions indicates lower efficiency. The same grid cells used for spatial density were used here	
Transition entropy	The transition entropy represents the randomness and complexity of a person's eye movements, with higher values indicating more randomness and lower efficiency (Krejtz et al., 2014). The transition entropy was calculated based on the following formula: $H_t = -\sum_{i \in AOIs} \pi_i \sum_{j \in AOIs} p_{ij} \log p_{ij}$ where π_i is calculated as for stationary entropy, and p_{ij} is the probability of transitioning from state *i* to state *j* in one unit of time. Assuming the assumption for a first-order Markov chain holds, this was calculated by counting the number of transitions from *i* to *j* and then dividing by the total number of transitions from *i*. This was done for each pairing of AOIs (the AOIs are as defined in Figure 1)	
Duration Metric		
Mean fixation duration (ms)	A lower mean fixation duration suggests that the user is quickly moving from one focus to the next	H3: As performance deteriorates, there will be increased mean fixation duration as participants struggle to discriminate information from the display

Note that any mention of attention here refers to overt attention allocation.

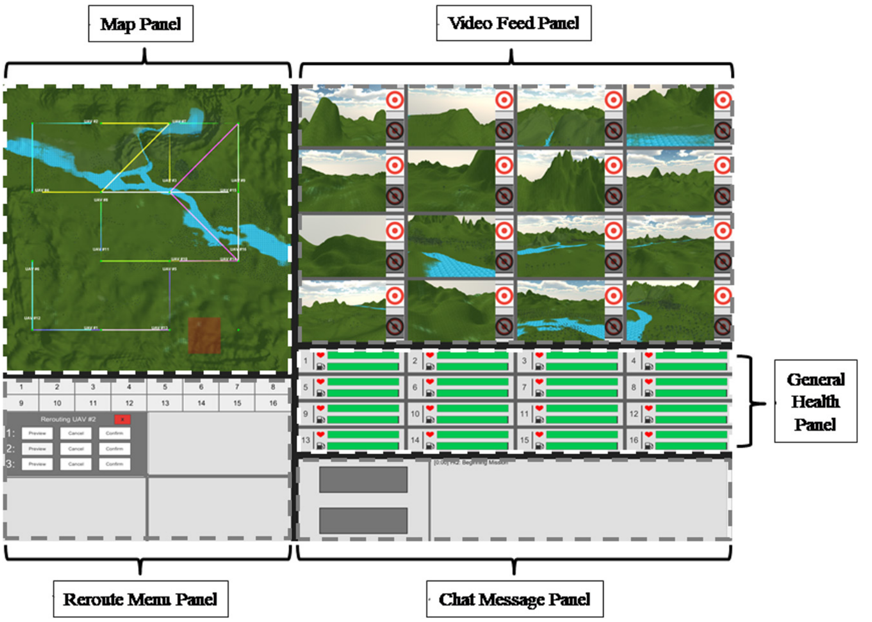


Figure 1. Screenshot of the UAV simulation with labeled panels. Each of these panels also constituted the AOIs for the calculation of the entropy measures.

tracker, a desktop-mounted eye tracker with a sampling rate of 60 Hz, was used to collect eye tracking data. The eye tracker was placed right below the monitor. Participants sat 28″–31″ from the monitor and used a standard mouse to input responses. The average degree of error for this eye tracker is 0.78 degrees (*SD* = .59; (Eyetracking, 2011).

UAV Control Simulation and Tasks

Participants were responsible for simultaneously controlling and managing 16 UAVs under four different task load conditions that mapped to four different scenarios (see "Task Load Modulation" section). Each scenario was 15 min in duration. This time frame was selected as it would be long enough to include multiple transitions from low and high workload, while not being too long where significant vigilance decrements could occur (typically considered around 15 min, although several factors can play a role; Teichner, 1974). Figure 1 depicts the simulation interface with task-specific areas of interest, referred to as panels, in dotted lines. The four tasks included one primary task (target detection) and three secondary tasks (reroute, fuel leak, and chat message task). For all four scenarios, the rate at which the primary task occurred was varied (see "Task Load Modulation" section) and one secondary task occurred every 20 s.

Target detection task (primary task). Target detection was the primary task and participants were instructed that this task had the highest priority. At the most fundamental level, the primary task was a visual search task for targets (transparent cubes on the video feed panel; Figure 2). There were up to 16 UAVs (i.e., 16 video feeds) at any given time. When a UAV approached a waypoint, the corresponding UAV video feed would become highlighted. When a UAV video feed was highlighted and a target was present, the participants were instructed to press the "target" button to indicate a target was present. Otherwise they were instructed to leave the default "no target" button selected. UAV video feeds were active for 10 s and targets could appear any time during this period. Of the UAV feeds that were active, a target was present

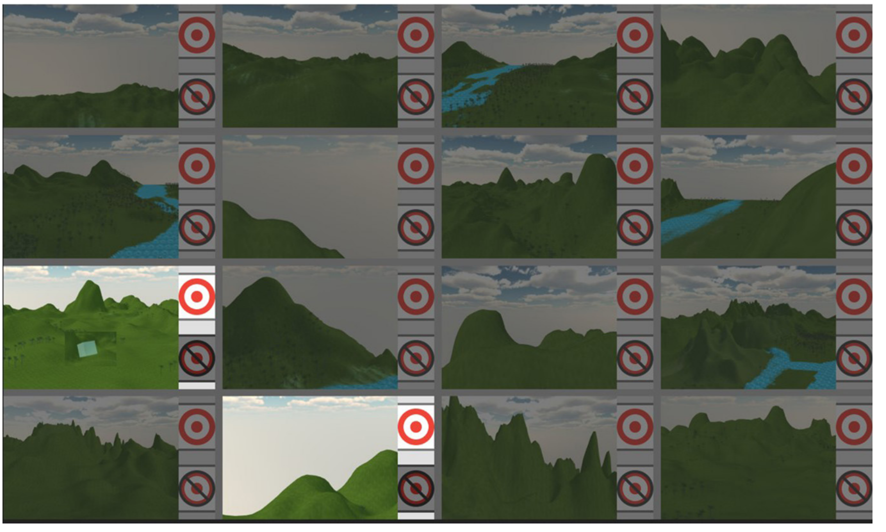


Figure 2. Screenshot of the video feed panel. UAVs 9 and 14 are highlighted (numbering from left to right, top to bottom). UAV 9 has a target (transparent cube, circled here for illustration purposes).

in 30% of those UAVs, on average. There could only be one target per UAV per 10 s interval. The specific UAVs that were concurrently highlighted at any point in time and the UAVs' arrival time to each waypoint were pseudorandomly set by the experimenter.

Reroute task (secondary task #1). The secondary tasks consisted of monitoring tasks for changes in the status of the UAVs. Participants were tasked to reroute a UAV when it entered a no-fly zone (i.e., red square on the map panel in Figure 3). To reroute a UAV, a participant clicked on the respective UAV's numbered square in the reroute menu panel (Figure 3). Participants had the option of selecting "Preview" to see three suggested routes, "Confirm" to reroute the UAV to a new suggested route, or "Cancel" to exit from the window to allow the UAV to continue on its original route. When a UAV was not rerouted in time and entered a no-fly zone, it became nonoperational for the remainder of the scenario (i.e., it could not participate in any of the target detection, reroute, or fuel leak tasks). Only three UAVs could enter a no-fly zone per scenario, meaning that even in the worst case scenario, if a participant failed to reroute all three UAVs, the number of UAVs available would still be within the range of active UAV needs in the high task load condition. The rerouting task occurred 18 times in each scenario.

Fuel leak task (secondary task #2). Participants were also tasked with maintaining the overall health of each UAV using the general health panel (Figure 4). Participants were asked to monitor for fuel leaks. When a fuel leak occurred, the color of the health status bar (top bar denoted with a heart) changed from green to yellow with a "FIX LEAK" warning. To stop a fuel leak, participants had 10 s to click on the yellow bar. If the leak was not stopped in time, the health status bar would change from yellow to orange and read "FATAL FUEL LEAK," meaning that the UAV could not participate in the fuel leak task again for that scenario. However, the UAV could still attend to the primary task, that is, the target detection task. A fuel leak occurred 14 times in each scenario.

Chat message task (secondary task #3). Participants were tasked with responding to chat messages by selecting between the two

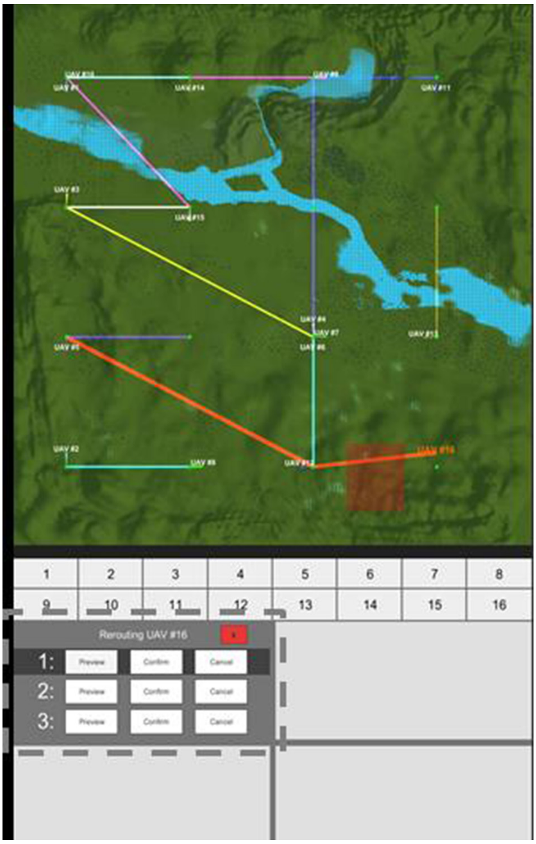


Figure 3. Screenshot of the map panel (top half) and reroute menu panel (bottom half). After clicking on the respective UAV's number (buttons numbered 1–16), a menu of route options was presented (dotted area with options 1–3 where "Preview," "Confirm," or "Cancel" could be selected).

options on the left-hand side of the chat message panel (Figure 5). Participants could respond to a chat message by clicking on one of two options until another message showed up. There were 19 chat messages in each scenario.

Response time and accuracy for the primary task were calculated for only the correct detection of targets. Cases where participants did not respond to the occurrence of a target were not considered in the calculation of response time. For all secondary tasks, response times were calculated from the onset of the event to when the participant responded. Accuracy was calculated as the percentage of correct responses (i.e., the percentage of correctly detected targets) within the time limit for each task.

TASK LOAD MODULATION

Task load (low, high, gradual, sudden) was the only independent variable and was varied fully within-subjects. Participants had to go through one scenario for each of the four task load conditions. Task load was manipulated by varying the number of active UAVs (the number of highlighted video feeds) in the target detection task. The four task load conditions include the following:

1. **Low task load.** There were three to five UAVs active for the entirety of the scenario. These values were determined and validated based on pilot testing data using both performance and NASA-TLX measures.
2. **High task load.** There were 13–16 UAVs active for the entirety of the scenario. These values were also determined based on pilot tests.
3. **Gradual task load.** The number of active UAVs increased gradually (Figure 6). The scenario started at low task load for 20 s, and one active UAV was added every 10 s until high task load was reached, that is, 13–16 active UAVs. The scenario would remain at high task load for 2 min, before returning to low task load. This low to high task load cycle repeated five times for this scenario.
4. **Sudden task load.** The number of active UAVs increased instantaneously (see Figure 6). One minute of low task load (3-5 UAVs) was followed by a jump to high task load (13-16 UAVs) for two minutes. This cycle repeated five times.

Note that the number of areas of the screen that participants had to monitor and attend to did not change based on task load condition because participants had to monitor all AOIs in order to complete all tasks.

Procedure

The experiment took place over two consecutive days around the same time of day (morning, afternoon, or evening). On the first day, participants signed a consent form and were briefed about the study's goals and expectations. Participants then completed a 5 min training session that included all tasks and task load conditions. By the end of the training session, participants had to demonstrate proficiency by having a minimum accuracy of 70% across all

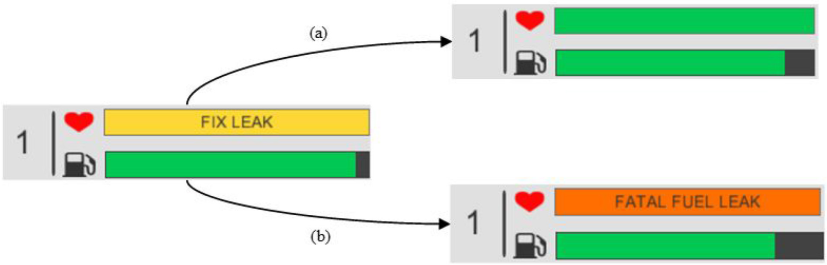


Figure 4. Screenshot of UAV 1's health status bar. Participants were tasked to press the yellow "FIX LEAK" button when a fuel leak occurred. (a) When a fuel leak was fixed in time, the yellow health status bar changed from yellow to green and the "FIX LEAK" warning disappeared. (b) When a fuel leak was not fixed in time, the yellow "FIX LEAK" changed to an orange "FATAL FUEL LEAK" warning.

training tasks, both primary and secondary. If they did not, they had another chance to complete the 5 min training session to achieve the desired proficiency. If they did not meet the proficiency requirements the second time, they were excused from the study. Only four participants had to repeat the training session, and all were successful the second time.

The eye tracker was calibrated using a 5-point grid. The calibration procedure was done at the start of each scenario. Participants completed two of the four scenarios on the first day and the other two on the second day, with the order counterbalanced across subjects. There was a 10 min break between scenarios. The entire study across 2 days lasted about 2 hr.

RESULTS

Results were analyzed using a one-way repeated measures ANOVA, with task load (four levels) as the variable of interest. Bonferroni corrections were applied for all post hoc tests. In all cases, Epsilon (ε) was calculated according to Greenhouse and Geisser (1959) and used to correct the one-way repeated measures ANOVA. For all graphs, error bars indicate the standard error of the mean and asterisks denote significant differences between conditions.

Figure 5. Screenshot of the chat message panel with response buttons on the left ("yes"/"no") and timestamped chat window on the right.

The initial gaze data were screened to meet data quality requirements as outlined in ISO/TS 15007-2:2014-09, which states that at most 15% data loss is acceptable for good quality data. Following this guideline, the eye tracking and corresponding performance data of five participants was not used in any of the analyses. The mean data loss of the included participants was 7.07%. The gaze points from the eye tracker were used to calculate fixations and saccades (the eye tracker automatically filters out blinks and any fixations outside the screen were discarded). Goldberg and Kotval (1999) fixation algorithm was applied: A cluster of gaze points was classified as a fixation if the points within the cluster were within 75 pixels of each other, and there was a minimum number of six gaze points within this fixation cluster. This made for a minimum fixation duration of approximately 100 ms. The first gaze point outside the 75-pixel limit was considered to be not part of the fixation; the gaze point just before would be the endpoint of the fixation. Any gaze points that were not part of fixations were assumed to be saccades. The calculated fixations were then used to calculate the eye tracking metrics described in Table 1.

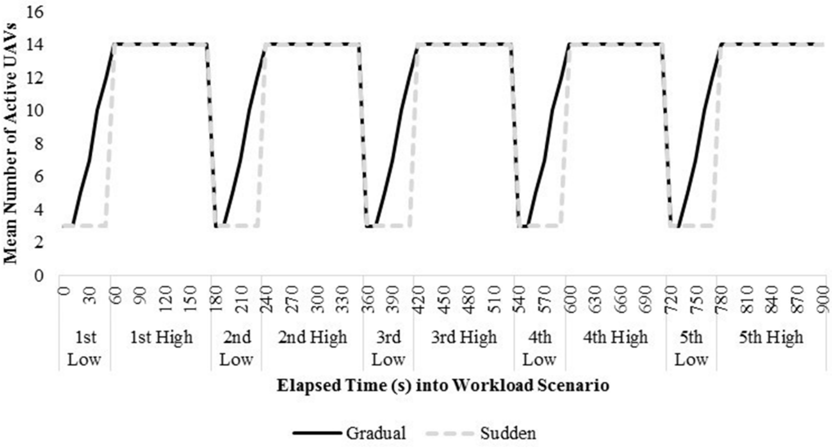


Figure 6. The number of active UAVs throughout the gradual and sudden task load scenarios.

Performance Results

Primary task. There was a significant effect of task load on response time (F[2.36,33.48] =161.51, $p < .001$, $\eta_p^2 = .95$, $\varepsilon = .94$), with means (standard error of the mean [*SE*]) equal to 2.37 s (*SE* = .04), 3.42 s (*SE* = .04), 2.86 s (*SE* = .04), and 2.86 s (*SE* = .03) in the low, high, gradual, and sudden conditions, respectively. Linear contrasts showed that the low task load scenario elicited significantly faster response times than all other scenarios (all $p < .001$). The high task load scenario had significantly slower participant response times than all other scenarios (all $p < .001$).

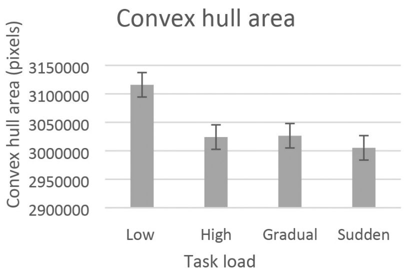
For primary task accuracy, there was a significant effect of task load on accuracy (F(2.61,39.25) = 101.42, $p < .001$, $\eta_p^2 = .87$, $\varepsilon = 1.00$), with means equal to 83.52% (*SE* = 1.20), 58.20% (*SE* = 1.50), 69.68% (*SE* = 1.90), and 72.09% (*SE* = 1.80) in the low, high, gradual, and sudden conditions, respectively. The low task load scenario led to significantly higher accuracy than all other scenarios (all $p < .001$). Participant accuracy in the high task load scenario was significantly lower than all other scenarios (all $p < .001$).

Secondary task. There was a significant effect of task load on secondary task response time (F(1.75,26.24) = 16.05, $p < .001$, $\eta_p^2 = .51$, $\varepsilon = .58$), with means of 5.09 s (*SE* = .28), 5.03 s (*SE* = .26), 3.88 s (*SE* = .10), and 3.70 s (*SE* = .15) in the low, high, gradual, and sudden conditions, respectively. Participant response time in the low task load scenario was significantly slower than the gradual and sudden task load scenarios (all $p < .001$). The high task load scenario was significantly slower than in the gradual and sudden task load scenarios (all $p < .001$).

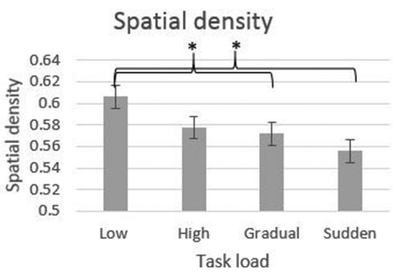
There was a significant effect of task load on secondary task accuracy (F(2.02,30.40) = 7.15, $p = .003$, $\eta_p^2 = .32$, $\varepsilon = .67$), with means of 88.41% (*SE* = 2.80), 89.94% (*SE* = 2.40), 96.71% (*SE* = 1.00), and 96.08% (*SE* = 1.10) in the low, high, gradual, and sudden conditions, respectively. In the low task load scenarios, participants were significantly less accurate than in the gradual ($p = .006$) and sudden task load scenarios ($p = .003$). Participant accuracy in the high task load scenarios was also significantly lower than in the gradual ($p = .003$).

Eye Tracking Results

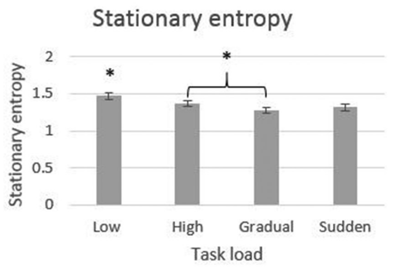
H1: spread metrics. The results of the spread metrics calculations can be seen in Figure 7. There was no significant difference of task load on convex hull area (Figure 7a). There was a significant effect of task load on spatial density (F(2.48,37.24) = 5.24, $p = .006$, $\eta_p^2 = .25$, $\varepsilon = .82$; Figure 7b). There were also significant pairwise



(a)



(b)



(c)

Figure 7. Spread metrics values for each task load condition: (a) convex hull area, (b) spatial density, and (c) stationary entropy.

differences between the low and gradual task load conditions (p = .004) and between low and sudden task load (p = .008). There was also a significant effect of task load on stationary entropy (F(2.21,33.17) = 9.49, p < .001, η_p^2 = .38, ε = .73; Figure 7c). Post hoc tests showed that low task load had a higher entropy than all other task load conditions, with p = .009, p = .001, and p = .007 for the high, gradual, and sudden conditions, respectively. High task load was also significantly higher than gradual (p = .014).

H2: directness metrics. For the directness metrics, which can be seen in Figure 8, there was a significant effect of task load type on mean saccade amplitude (F(2.38,35.76) = 13.81, p < .001, η_p^2 = .47, ε = .79; Figure 8a). Post hoc tests showed that there was a significant difference between low task load and all other conditions (all p < .001). For scanpath length per second, there was also a significant effect of task load (F(2.84,42.61) = 16.32, p < .001; η_p^2 = .52, ε = .94; Figure 8b). Post hoc tests showed that there was a significant difference between low task load and the other conditions (all p < .001). There was no significant effect of task load on backtrack rate (Figure 8c). There was a significant effect of task load on transition rate (F(2.65,39.75) = 4.92, p = .007, η_p^2 = .24, ε = .88; Figure 8d). There was a significant pairwise difference between the low and gradual (p = .029) and low and sudden (p = .005) conditions. For the transition entropy, there was an effect of task load (F(2.16,32.48) = 15.83, p < .001, η_p^2 = .51, ε = .72; Figure 8e). The low task load condition was significantly higher than all other conditions (p < .001, p < .001, and p = .001 for high, gradual, and sudden task loads, respectively). High task load was also significantly higher than the gradual task load condition (p = .007).

H3: duration metric. Finally, for the duration metric, there was no significant effect of task load on mean fixation duration (Figure 9). Table 2 provides a summary of the eye tracking results obtained in this study.

DISCUSSION

The aim of this study was to analyze how performance and attention allocation, as evidenced

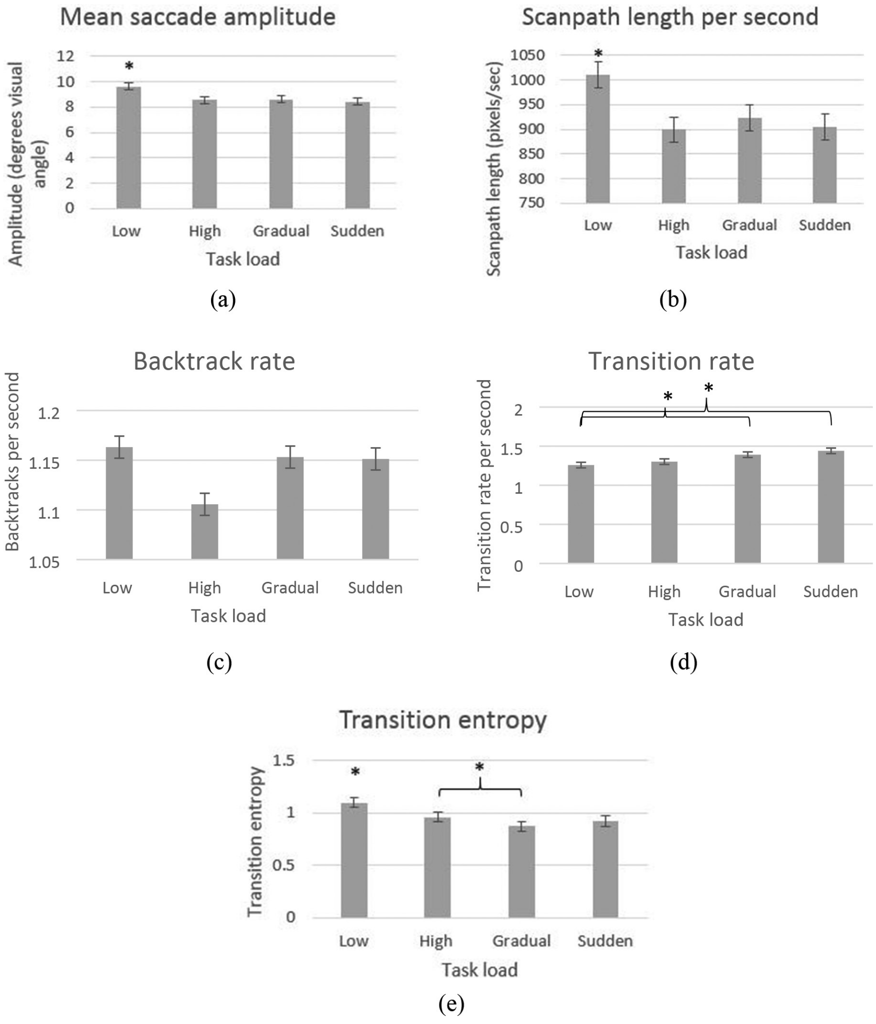


Figure 8. Directness metrics results for each task load condition: (a) scanpath length, (b) mean saccade amplitude, (c) backtrack rate, (d) rate of transitions, and (e) transition entropy.

by eye movements, are affected by gradual and sudden workload transitions when compared to constant low or high workload in a multitasking environment. The performance results obtained in this experiment suggest that there was no significant difference between performance in the gradual and sudden workload conditions. Contrary to our expectations, this suggests that sudden workload transitions are not more detrimental to performance than gradual workload transitions.

The analysis of the primary and secondary tasks provided further insights into the effects of workload transitions. For the primary task, high workload led to the worst primary task performance, which is consistent with the literature on the dangers of high workload (Cain, 2007). The workload transition conditions—both gradual and sudden—led to primary task performance that was in between low and high workload. However, the performance results observed for the secondary task differ from those of the primary task. For the secondary task, low and high workload resulted in the worst performance, whereas the transition conditions resulted in the

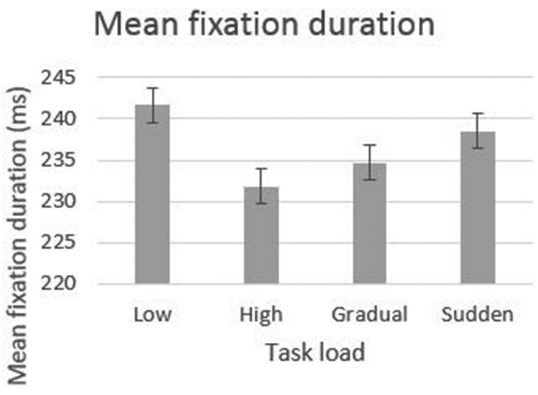


Figure 9. Duration metric results for each task load condition.

best performance. In other words, while task load was fluctuating between low and high, whether suddenly or gradually, secondary task response accuracy was higher compared to when workload was held constant at high or low. In general, the results were contrary to our expectation that, for both the primary and secondary tasks, performance would be increasingly worse in the low, gradual, sudden, and high workload scenarios. Nor do the results support our expectation that there would be a detrimental effect of sudden workload transitions, as found by researchers in other contexts (Cox-Fuenzalida, 2007; Kochan et al., 2004). However, the results are consistent with resource-allocation-based theories of attention, such as Wickens'

TABLE 2: Summary of the Eye Tracking Results

Metric	Result	Summary
Spread Metrics		
Convex hull area (pixels^2)	Not significant	H1: As performance deteriorates, these metrics will exhibit increased spread Findings: Spread of fixations is highest in low task load and lowest in gradual and sudden task load
Spatial density	• Highest value in low task load compared to gradual and sudden task load	
Stationary entropy	• Highest value in low task load • Higher value in high task load than gradual task load	
Directness Metrics		
Mean saccade amplitude (pixels)	• Highest value in low task load	H2: As performance deteriorates, these metrics will exhibit less directness and efficiency Findings: Low task load is the least efficient condition overall, but it has fewer transitions than the sudden and gradual task load conditions
Scanpath length per second (pixels)	• Highest value in low task load	
Backtrack rate (s^{-1})	Not significant	
Transition rate (s^{-1})	• Lowest value in low task load compared to gradual and sudden task load	
Transition entropy	• Highest value in low task load • Higher value in high task load than gradual task load	
Duration Metric		
Mean fixation duration (ms)	Not significant	H3: As performance deteriorates, there will be increased mean fixation duration Findings: Mean fixation duration is not affected by workload or workload transitions

(2002) multiple resource theory and Young and Stanton's (2002) malleable attentional resources theory. These theories posit that very low and very high workload can lead to performance decrements, whereas there exists an optimal middle-ground level of task load that elicits the best performance. Nevertheless, the results are surprising given that participants were asked to prioritize the primary task over the secondary one. The results confirm that workload transitions can affect how people perform their tasks in a way that is different from just high or low workload (Cox-Fuenzalida, 2007; Cumming & Croft, 1973; Goldberg & Stewart, 1980), even in a more complex, multitasking environment. More specifically, the results suggest that the transition conditions of sudden and gradual workload resulted in the participants adjusting their priorities and strategies as part of a multitasking operation.

The discrepancies with the previous literature serve to underline the importance of our main aim in this study, which is to use eye tracking metrics to better understand how these performance effects come about. In general, results were consistent with hypotheses H1 and H2 that there would be increased spread and less directness, respectively, with worse task performance. It is interesting, though, that the task performance that was best reflected in the eye tracking metrics was the secondary task, which relates mainly to transitioning between AOIs. It would appear that the transition between AOIs or the sequence of AOIs is an aspect that should be regularly explored in such studies, contrary to what is currently the case.

However, the eye tracking metrics reflected the poor performance mainly in the low but not high workload condition, although performance was equally poor in both. Under low task load, the *spread* metrics suggest that, in terms of where participants are looking, it appeared that participants were covering wider and more varied areas of the display. The *directness* metrics suggest that, in terms of efficiency, participants were scanning less efficiently under low task load and were more focused and efficient under the other task load conditions, especially in the gradual and sudden conditions. Only with stationary and transition entropy was the high workload condition also significantly different than the gradual workload condition, suggesting that these metrics are particularly sensitive to changes between constant and transitioning workload and would be recommended for studies in this context. This could be because these metrics are based on Markov chains, a stochastic process in which the next state (in this case, the next AOI that is fixated) depends only on the current state (current AOI being fixated). By modeling the scanpath in this way, it appears that one can very well capture differences in transition patterns. Moreover, the only exception to the increased directness under worse performance was observed in terms of the transition rate: sudden and gradual workload transitions exhibited higher transition rates compared to when workload was held constant at low. Finally, the only hypothesis that our findings fully contradicted was H3, which is related to the duration metric; results suggest that there was no effect of workload transitions on how long participants were looking at certain areas. This is surprising given how frequently mean fixation duration is used in the context of workload (e.g., Schulz et al., 2011). It could be that there is no issue of discriminating information in this study, although this would need to be further explored.

In summary, these metrics seem to suggest that low workload results in more spread in terms of where people are scanning, but eye movements become on average less efficient. On the other hand, in the workload transition conditions and under high workload, there is less spread, but higher efficiency. This can be explained by the fact that, under low task loads, participants had more time to scan the entire display without having to be systematic in terms of where they were looking. Although the greater spread may have benefitted the primary task performance, the lack of a systematic approach to scanning adversely affected secondary task performance. This nuance only serves to emphasize the importance of incorporating directness metrics to thoroughly understand attention allocation.

The results can be further interpreted in terms of the two-dimensional framework for the role of attention proposed by Trick et al.

(2004). In the proposed framework, there are two dichotomies: controlled versus automatic and endogenous versus exogenous. In relation to this study, participants in the low workload condition would be operating in the controlled-exogenous mode of attention. This mode promotes more open-ended exploration of the environment/display that is supported by spread metrics observed here under low task load which is the default. However, Trick et al. (2004) note that even though attention in this mode is conscious and voluntary, it can interfere with secondary task performance, which was found to be the case here. Using the same framework, the increased task demands associated with the workload transitions and high workload conditions most closely align with the controlled-endogenous mode, where behavior is deliberate and goal-driven. This is supported by the fact that the increased task demands lead to a need to balance priorities under the dual-task paradigm, which may have come at a cost to primary task performance.

Given that the tasks in this study involved a visual search task (i.e., target detection task), the findings can also be interpreted in light of the considerable visual search literature (see Wolfe and Horowitz (2017) for an overview). Models of visual search generally characterize search as a multistage process, with an early parallel stage obtaining basic information about the wider display area and a subsequent serial, slower stage getting more detailed information, but from limited display areas (Wolfe, 1994). This ability to obtain basic information from the periphery is vital to the visual search process (Rosenholtz, 2016; Wolfe et al., 2017). However, the presence of crowding or clutter, as is the case in this study as the number of active UAVs increases, makes it more difficult to recognize and discriminate objects in the periphery. This made it necessary for participants to saccade to and foveate items detected in the periphery to be able to identify their properties (Wolfe & Whitney, 2014). In the presence of less crowding in low workload, items in the periphery may have been more salient and thus drew participants' visual attention to wide areas of the screen. This saccade to wide areas may have been what led to larger spread in low workload, with the resulting fixations on the targets helping to improve performance in the primary task, which required detailed discrimination. This theory is also consistent with the higher mean saccade amplitude in low workload, which suggests that items in the wider periphery were detected and a (large) saccade was initiated to that item. This increased spread may not have been necessary or helpful for the secondary tasks, though, where peripheral vision may have been enough to understand what needs to be attended to.

Moreover, it has been shown that people can plan saccades in advance of search, although, if they have enough time and an item is salient enough, their planned scanpath can be altered (De Vries et al., 2014). In lower workload, it appears that there was more time for participants to change their planned scanpath when a salient object appeared in the periphery. This would explain the more random and less efficient sequence of eye movements in low workload, with participants' eye movements drawn to different areas at random. In the other workload conditions (high, sudden, and gradual), all of which involved higher workload, it would appear that participants tended to stick to a selected sequence and not deviate from that. This more systematic approach would also explain the increased transition rate in these three conditions, with participants seemingly deciding to move from one area of the screen to another in a planned fashion. The findings here demonstrate the importance of considering multiple types of eye tracking metrics to understand the effects of workload on visual attention allocation as thoroughly as possible.

CONCLUSION

There are three main takeaways from these findings. The first is that, contrary to expectations, sudden workload transitions do not seem to be more detrimental than gradual workload transitions. There does not seem to be any evidence of the "surprise" of a sudden transition leading to any decrements in performance, or any changes in attention allocation, either. The element of surprise and unexpected events may have detrimental effects in certain contexts such as aviation (Wickens, 2001); however, as part of this study, the element of surprise that may have occurred with a sudden workload shift may have been diluted as there was more than one workload transition over the course of the 15 min scenario. If this study were replicated with just one sudden increase, the element of surprise present with the sudden task load condition could be accentuated and lead to decrements in performance.

At the same time, the second conclusion that can be drawn from these findings is that transitioning workload could affect how people prioritize tasks in a multitasking environment. It would seem that switching task load between low and high workload leads to participants changing their strategy. Transitioning workload cannot be treated simply as a "middle" stage between low and high workload, with the expectation that performance will be averaged. Instead, the findings here should be taken into account in the design of tasks within safety-critical, multitasking domains.

The third and final conclusion is the importance of using a combination of eye tracking metrics, together with a knowledge of the context, to understand attention allocation. In particular, the spread and directness metrics, which are much less used as compared to duration measures (e.g., Coyne et al., 2017; Foy & Chapman, 2018), are the ones that provided the more insightful results. Specifically, the combination of spread metrics (namely, spatial density and stationary entropy) and directness metrics (namely, mean saccade amplitude, scanpath length per second, transition rate, and transition entropy) was sensitive to differences and changes in task demands. The directness metrics in general are the ones that helped explain how participants did worse on the secondary task in low workload, contrary to all expectations. Stationary and transition entropies were particularly insightful, suggesting that they should be used more in studies on transitioning workload. Given that stationary entropy reflects just the switching between different areas of the screen, regardless of what the exact task is within each AOI, these metrics could still provide insight into participants' task switching behavior when other tasks are used.

In terms of more long-term future applications, these metrics can potentially be used as the basis for an adaptive display, where information is updated to suit users' needs in real time to support SA (e.g., Monfort et al., 2016). To this end, it is important that the appropriate metrics are used to account for different workload conditions. Once the system detects that the user is struggling with high workload or workload transitions, display adjustments can then be triggered in near real time before significant performance breakdowns occur. This could be extremely valuable in safety- and time-critical domains, such as military operations, aviation, or process control. Using eye tracking allows for the triggering of display adjustments as early as a few seconds into a task (Moacdieh & Sarter, 2017). At the same time, this approach would mean that there would be no need to trigger any unnecessary adjustments that might confuse or distract the user. This approach of *adjusting only when needed* is a cornerstone of adaptive and intelligent displays. However, note that it is critical that these metrics are not considered in isolation and should also take context into account. What works for one task may not work for all tasks in a complex, data-rich domain where operators are tasked with various responsibilities.

Although the results of this empirical work are informative, they are not free of limitations. First, generalizing these results is limited to domains with similar task structure, or ones that consist of monitoring several distinct areas of a screen and where increasing the number of items to attend to is directly proportional to increasing task load. The assumption is also that all the tasks are visual, not auditory. In its simplest form, this experiment involves search and monitoring multiple areas of the screen, and the fact that the eye tracking metrics were able to reflect this transitioning behavior well—namely, with the entropy metrics—suggests that any kind of transitioning between AOIs could also be captured, even in a different context. However, more research is needed to improve the external validity of the study and establish which eye tracking metrics best reflect fluctuations in workload in other multitasking scenarios.

It would also be interesting to look at shorter scenarios in addition to longer time periods, as was done here. Analyzing a single time transition at a time could also provide insight into how operators' attention is affected immediately after transitions. Furthermore, the sampling rate of the eye tracker used in this study (60 Hz) is not ideal for the study of saccade amplitude, although it makes no difference to the detection of fixations (Leube et al., 2017). To further explore the use of that metric, an eye tracker with a higher sampling rate (e.g., 120 Hz) will have to be used in future studies. Also, it would be interesting to further explore how results would change if participants were completely unaware of any changes in workload. In that case, the focus would fully be on the element of surprise.

AUTHOR CONTRIBUTIONS

N. Moacdieh worked on the design of the experiment, the write-up, and the analysis of results. S. Devlin contributed to the design of the experiment, running the experiment, and the write-up. H. Jundi worked on the eye tracking code, the analysis of the metrics, and the write-up of results. S. Riggs worked on the design of the experiment, the write-up, and the analysis of results

ACKNOWLEDGMENTS

This study was supported in part by the National Science Foundation (NSF grant: #1566346; Program Manager: Dr. Ephraim Glinert). The authors would like to thank Aakash Bhagat for his help in developing the simulator for this study.

ORCID iDs

Nadine Marie Moacdieh https://orcid.org/0000-0002-7677-7946

Sara Lu Riggs https://orcid.org/0000-0002-0112-9469

REFERENCES

Bergen, J. R., & Julesz, B. (1983). Parallel versus serial processing in rapid pattern discrimination. *Nature*, *303*(5919), 696–698. https://doi.org/10.1038/303696a0

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., Olmstead, R. E., Tremoulet, P. D., & Craven, P. L. (2007). EEG correlates of task engagement and mental task load in vigilance, learning, and memory tasks. *Aviation, Space, and Environmental Medicine*, *78*(5 Suppl), B231–B244.

Bliss, J. P., & Dunn, M. C. (2000). Behavioural implications of alarm mistrust as a function of task workload. *Ergonomics*, *43*(9), 1283–1300. https://doi.org/10.1080/001401300421743

Boyer, M., Cummings, M. L., Spence, L. B., & Solovey, E. T. (2015). Investigating mental workload changes in a long duration supervisory control task. *Interacting with Computers*, *27*(5), 512–520. https://doi.org/10.1093/iwc/iwv012

Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology*, *42*(3), 361–377. https://doi.org/10.1016/0301-0511(95)05167-8

Cain, B. (2007). *A review of the mental workload literature. Technical report*, Defence Research and Development Toronto.

Coral, M. P. (2016). A*nalyzing cognitive workload through eye-related measurements: A meta-analysis* [Doctoral dissertation]. Retrieved from Wright State University database.

Cox-Fuenzalida, L.-E. (2007). Effect of workload history on task performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *49*(2), 277–291. https://doi.org/10.1518/001872007X312496

Cox-Fuenzalida, L.-E., Beeler, C., & Sohl, L. (2006). Workload history effects: A comparison of sudden increases and decreases on performance. *Current Psychology*, *25*(1), 8–14. https://doi.org/10.1007/s12144-006-1012-6

Coyne, JT., Sibley, C., Sherwood, S., Foroughi, CK., Olson, T., & Vorm, E. (2017, July). *Assessing workload with low cost eye tracking during a supervisory control task* [Conference session]. International Conference on Augmented Cognition, Vancouver, Canada, Springer, 139–147.

Cumming, R. W., & Croft, P. G. (1973). Human information processing under varying task demand. *Ergonomics*, *16*(5), 581–586. https://doi.org/10.1080/00140137308924548

De Rivecourt, M., Kuperus, M. N., Post, W. J., & Mulder, L. J. M. (2008). Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight. *Ergonomics*, *51*(9), 1295–1319. https://doi.org/10.1080/00140130802120267

De Vries, J. P., Hooge, I. T. C., & Verstraten, F. A. J. (2014). Saccades toward the target are planned as sequences rather than as single steps. *Psychological Science*, *25*(1), 215–223. https://doi.org/10.1177/0956797613497020

Di Nocera, F., Camilli, M., & Terenzi, M. (2007). A random glance at the flight deck: Pilots' scanning strategies and the real-time assessment of mental workload. *Journal of Cognitive Engineering and Decision Making*, *1*(3), 271–285. https://doi.org/10.1518/155534307X255627

Di Stasi, L. L., Antolí, A., & Cañas, J. J. (2013). Evaluating mental workload while interacting with computer-generated artificial environments. *Entertainment Computing*, *4*(1), 63–69. https://doi.org/10.1016/j.entcom.2011.03.005

Dixon, S. R., & Wickens, C. D. (2006). Automation reliability in unmanned aerial vehicle control: A reliance-compliance model of automation dependence in high workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *48*(3), 474–486. https://doi.org/10.1518/001872006778606822

Dixon, S. R., Wickens, C. D., & Chang, D. (2005). Mission control of multiple unmanned aerial vehicles: A workload analysis. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *47*(3), 479–487. https://doi.org/10.1518/001872005774860005

Duchowski, A. T. (2007). *Eye tracking methodology: Theory and practice*. Springer.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 32–64. https://doi.org/10.1518/001872095779049543

Endsley, MR. (1988). *Design and evaluation for situation awareness enhancement* [Paper presentation]. Proceedings of the Human Factors Society Annual Meeting, Los Angeles, CA, 97–101.

Eyetracking. (2011). Hardware: Eye tracking systems. http://www.eyetracking.com/Hardware/Eye-Tracker-List

Feitshans, G., Rowe, A., Davis, J., Holland, M., & Berger, L. (2008). *Vigilant spirit control station (VSCS): The face of COUNTER* [Conference session]. AIAA Guidance, Navigation and Control Conference and Exhibit, Honolulu, HI, American Institute of Aeronautics and Astronautics. 6309

Foy, H. J., & Chapman, P. (2018). Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation. *Applied Ergonomics*, *73*, 90–99. https://doi.org/10.1016/j.apergo.2018.06.006

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics*, *24*(6), 631–645. https://doi.org/10.1016/S0169-8141(98)00068-7
Goldberg, R. A., & Stewart, M. R. (1980). Memory overload or expectancy effect? 'Hysteresis' revisited. *Ergonomics*, *23*(12), 1173–1178. https://doi.org/10.1080/00140138008924824
Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*(2), 95–112. https://doi.org/10.1007/BF02289823
Hampson, R. E., Opris, I., & Deadwyler, S. A. (2010). Neural correlates of fast pupil dilation in nonhuman primates: Relation to behavioral performance and cognitive workload. *Behavioural Brain Research*, *212*(1), 1–11. https://doi.org/10.1016/j.bbr.2010.03.011
Hancock, P. A., Williams, G., & Manning, C. M. (1995). Influence of task demand characteristics on workload and performance. *The International Journal of Aviation Psychology*, *5*(1), 63–86. https://doi.org/10.1207/s15327108ijap0501_5
Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, *52*, 139–183.
Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.
Hoover, A., Singh, A., Fishel-Brown, S., & Muth, E. (2012). Real-time detection of workload changes using heart rate variability. *Biomedical Signal Processing and Control*, *7*(4), 333–341. https://doi.org/10.1016/j.bspc.2011.07.004
Huey, B. M., & Wickens, C. D. (1993). *Workload transition: Implications for individual and team performance*. National Academy Press.
Hwang, S. L., Yau, Y. J., Lin, Y. T., Chen, J. H., Huang, T. H., Yenn, T. C., & Hsu, C. C. (2008). Predicting work performance in nuclear power plants. *Safety Science*, *46*(7), 1115–1124. https://doi.org/10.1016/j.ssci.2007.06.005
Kochan, JA., Breiter, EG., & Jentsch, F. (2004, September). *Surprise and unexpectedness in flying: Database reviews and analyses* [Conference session]. Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 48, No. 3, pp. 335–339), Los Angeles, CA.
Krejtz, K., Szmidt, T., Duchowski, A. T., & Krejtz, I. (2014, March). *Entropy-based statistical analysis of eye movement transitions* [Symposium]. Proceedings of the Symposium on Eye Tracking Research and Applications (pp. 159–166), Safety Harbor, FL, ACM.
Leube, A., Rifai, K., & Wahl, S. (2017). Sampling rate influences saccade detection in mobile eye tracking of a reading task. *Journal Eye Movement Ressearch*, *10*(3), 4–14.
Matthews, M. L. (1986). The influence of visual workload history on visual performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *28*(6), 623–632. https://doi.org/10.1177/001872088602800601
Matthews, G., Reinerman-Jones, L. E., Barber, D. J., & Abich IV, J. (2015). The psychometrics of mental workload: Multiple measures are sensitive but divergent. *Human Factors*, *57*(1), 125–143.
Moacdieh, N., & Sarter, N. (2015a). Clutter in electronic medical records: Examining its performance and attentional costs using eye tracking. *Human Factors*, *57*(4), 591–606.
Moacdieh, N., & Sarter, N. (2015b). Display clutter: A review of definitions and measurement techniques. *Human Factors*, *57*(1), 61–100.
Moacdieh, N. M., & Sarter, N. (2017). Using eye tracking to detect the effects of clutter on visual search in real time. *IEEE Transactions on Human-Machine Systems*, *47*(6), 896–902. https://doi.org/10.1109/THMS.2017.2706666
Monfort, SS., Sibley, CM., & Coyne, JT. (2016, May). *Using machine learning and real-time workload assessment in a high-fidelity UAV simulation environment* [Conference session]. In Next-Generation Analyst IV (Vol. 9851, p. 98510B), Baltimore, MD, International Society for Optics and Photonics, San Diego, CA.
Morgan, J. F., & Hancock, P. A. (2011). The effect of prior task loading on mental workload: An example of hysteresis in driving. *Human Factors*, *53*(1), 75–86.
National Transportation Safety Board. (1978). Aircraft Accident Report - United Airlines, Inc., McDonnell-Douglas DC-8-61, N8082U. Portland, Oregon.
Poole, A., & Ball, L. J. (2005). Eye tracking in human-computer interaction and usability research. In C. Ghaoui (Ed.), *Encyclopedia of human computer interaction (pp. 211–219)*. Idea Group.
Prytz, E. G., & Scerbo, M. W. (2015). Changes in stress and subjective task load over time following a task load transition. *Theoretical Issues in Ergonomics Science*, *16*(6), 586–605. https://doi.org/10.1080/1463922X.2015.1084397
Rantanen, E. M., & Goldberg, J. H. (1999). The effect of mental task load on the visual field size and shape. *Ergonomics*, *42*(6), 816–834. https://doi.org/10.1080/001401399185315
Reid, G. B., & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P. A. Hancock & N. Meshkati (Eds.), *Advances in psychology* (Vol. 52, pp. 185–218). North-Holland.
Rosenholtz, R. (2016). Capabilities and limitations of peripheral vision. *Annual Review of Vision Science*, *2*(1), 437–457. https://doi.org/10.1146/annurev-vision-082114-035733
Rouse, W. B., Edwards, S. L., & Hammer, J. M. (1993). Modeling the dynamics of mental workload and human performance in complex systems. *IEEE Transactions on Systems, Man, and Cybernetics*, *23*(6), 1662–1671. https://doi.org/10.1109/21.257761
Rozado, D., & Dunser, A. (2015). Combining EEG with pupillometry to improve cognitive workload detection. *Computer*, *48*(10), 18–25. https://doi.org/10.1109/MC.2015.314
Savage, S. W., Potter, D. D., & Tatler, B. W. (2013). Does preoccupation impair hazard perception? A simultaneous EEG and eye tracking study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *17*, 52–62. https://doi.org/10.1016/j.trf.2012.10.002
Schulz, C. M., Schneider, E., Fritz, L., Vockeroth, J., Hapfelmeier, A., Wasmaier, M., Kochs, E. F., & Schneider, G. (2011). Eye tracking for assessment of workload: A pilot study in an anaesthesia simulator environment. *British Journal of Anaesthesia*, *106*(1), 44–50. https://doi.org/10.1093/bja/aeq307
Solovey, E. T., Zec, M., Garcia Perez, E. A., Reimer, B., & Mehler, B. (2014, April). *Classifying driver workload using physiological and driving performance data: Two field studies* [Conference session]. Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems (pp. 4057–4066), Toronto, Canada, ACM.
Taylor, R. (1990). *Situational Awareness Rating Technique (SART): The development of a tool for aircrew systems design*[Symposium]. AGARD, Situational Awareness in Aerospace Operations 17 P(SEE N 90-28972 23-53), Neuilly-Sur-Seine, France.
Teichner, W. H. (1974). The detection of a simple visual signal as a function of time of watch. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *16*, 339–352. https://doi.org/10.1177/001872087401600402
Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5
Trick, L. M., Enns, J. T., Mills, J., & Vavrik, J. (2004). Paying attention behind the wheel: A framework for studying the role of attention in driving. *Theoretical Issues in Ergonomics Science*, *5*(5), 385–424. https://doi.org/10.1080/14639220412331298938
Van Benthem, K. D., Herdman, C. M., Tolton, R. G., & LeFevre, J. A. (2015). Prospective memory failures in aviation: Effects of cue salience, workload, and individual differences. *Aerospace Medicine and Human Performance*, *86*(4), 366–373. https://doi.org/10.3357/AMHP.3428.2015
Veltman, J. A., & Gaillard, A. W. K. (1996). Physiological indices of workload in a simulated flight task. *Biological Psychology*, *42*(3), 323–342. https://doi.org/10.1016/0301-0511(95)05165-1

Wickens, C. D. (1992). Workload and situation awareness: An analogy of history and implications. *Insight: The Visual Performance Technical Group Newsletter*, *14*(4), 1–3.

Wickens, C. D. (2001, May). *Attention to safety and the psychology of surprise* [Symposium]. Proceedings of the 2001 Symposium on Aviation Psychology, Columbus, OH, The Ohio State University

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 159–177. https://doi.org/10.1080/14639220210123806

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238. https://doi.org/10.3758/BF03200774

Wolfe, B., Dobres, J., Rosenholtz, R., & Reimer, B. (2017). More than the useful field: Considering peripheral vision in driving. *Applied Ergonomics*, *65*, 316–325. https://doi.org/10.1016/j.apergo.2017.07.009

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Publishing Group*, *1*, 1–8. https://doi.org/10.1038/s41562-017-0058

Wolfe, B. A., & Whitney, D. (2014). Facilitating recognition of crowded faces with presaccadic attention. *Frontiers in Human Neuroscience*, *8*, 103. https://doi.org/10.3389/fnhum.2014.00103

Young, M. S., & Stanton, N. A. (2002). Malleable attentional resources theory: A new explanation for the effects of mental underload on performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *44*(3), 365–375. https://doi.org/10.1518/0018720024497709

Nadine Marie Moacdieh is an assistant professor in Industrial Engineering and Management at the American University of Beirut. She obtained her PhD in Industrial and Operations Engineering from the University of Michigan, Ann Arbor in 2015.

Shannon P. Devlin is a PhD candidate in Systems Engineering in the Department of Engineering Systems and Environment at the University of Virginia. She received her MS in Industrial Engineering from Clemson University in 2018.

Hussein Jundi is a Masters student in Data Engineering and Analytics at the Technical University of Munich. He received his BE in Industrial Engineering and Management from the American University of Beirut in 2018.

Sara Lu Riggs is an assistant professor in Systems Engineering in the Department of Engineering Systems and Environment at the University of Virginia. She obtained her PhD in Industrial and Operations Engineering from the University of Michigan, Ann Arbor in 2014.